Scene illumination color estimation methods based on convolutional neural networks

Koščević, Karlo

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Karlo Koščević

SCENE ILLUMINATION COLOR ESTIMATION METHODS BASED ON CONVOLUTIONAL NEURAL NETWORKS

DOCTORAL THESIS

Zagreb, 2022

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Karlo Koščević

SCENE ILLUMINATION COLOR ESTIMATION METHODS BASED ON CONVOLUTIONAL NEURAL NETWORKS

DOCTORAL THESIS

Supervisor: Professor Sven Lončarić, PhD

Zagreb, 2022

Sveučilište u Zagrebu

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Karlo Koščević

METODE PROCJENE BOJE OSVJETLJENJA SCENE ZASNOVANE NA KONVOLUCIJSKIM NEURONSKIM MREŽAMA

DOKTORSKI RAD

Mentor: prof. dr. sc. Sven Lončarić

Zagreb, 2022.

This doctoral thesis was completed at the University of Zagreb Faculty of Electrical Engineering and Computing, on Department of Electronic Systems and Information Processing.

Supervisor: Professor Sven Lončarić, PhD

The thesis has 124 pages.

Thesis No.: _____

About the Supervisor

Sven Lončarić was born in Zagreb in 1961. He received Diploma of Engineering and Master of Science degrees in electrical engineering from University of Zagreb Faculty of Electrical Engineering and Computing (FER) in 1985 and 1989, respectively. He received Doctor of Philosophy (Ph.D.) degree in electrical engineering from University of Cincinnati, USA, in 1994. Since 2011, he has been a tenured full professor in electrical engineering and computer science at FER. He was researcher or project leader on a number of research projects in the area of image processing and computer vision. From 2001-2003, he was an assistant professor at New Jersey Institute of Technology, USA. He has been the head of the Image Processing Laboratory at FER. He founded the Center for Computer Vision at University of Zagreb. Prof. Lončarić has been a co-director of the national Center of Research Excellence in Data Science and Cooperative Systems and the director of the Center for Artificial Intelligence at FER. He authored or co-authored more than 200 scientific publications. Prof. Lončarić is a senior member of IEEE and a member of Croatian Academy of Technical Sciences. He was the Chair of the IEEE Croatia Section. He founded and organized several international scientific conferences. He is an editor, a program committee member and reviewer for a number of scientific conferences and journals. He is a recipient of the National Award for Outstanding Scientist, University of Zagreb „Fran Bošnjaković“ Award, Croatian Academy of Engineering "Rikard Podhorsky" Award, FER Science Award, Fulbright stipend, and an IEEE Croatia Section Award.

O mentoru

Sven Lončarić rođen je u Zagrebu 1961. godine. Diplomirao je i magistrirao u polju elektrotehnike na Fakultetu elektrotehnike i računarstva (FER) Sveučilišta u Zagreba, 1985. i 1989. godine. Doktorirao je u polju elektrotehnike na Sveučilištu u Cincinnatiju, SAD, 1994. godine. U zvanje redoviti profesor u trajnom zvanju u polju elektrotehnike i polju računarstva na FER-u izabran je 2011. godine. Bio je suradnik ili voditelj na brojnim istraživačkim i razvojnim projektima u području razvoja metoda za obradu slika i računalnog vida. Od 2001. do 2003. bio je Assistant Professor na Sveučilištu New Jersey Institute of Technology, SAD. Voditelj je istraživačkog laboratorija za obradu slike na FER-u. Osnivač je i voditelj Centra izvrsnosti za računalni vid na Sveučilištu u Zagrebu. Suvoditelj je nacionalnog Znanstvenog centra izvrsnosti za znanost o podatcima i kooperativne sustave i voditelj Centra za umjetnu inteligenciju FER-a. Sa svojim suradnicima publicirao je više od 200 znanstvenih i stručnih radova. Prof. Lončarić član je stručne udruge IEEE i Akademije tehničkih znanosti Hrvatske. Bio je predsjednik Hrvatske sekcije IEEE. Osnivač je i organizator više međunarodnih znanstvenih skupova i ljetnih škola. Bio je urednik, član uredničkih i programskih odbora i recenzent za više međunarodnih znanstvenih skupova i časopisa. Dobitnik je Državne nagrade za znanost, Nagrade „Fran Bošnjaković“ Sveučilišta u Zagrebu, Nagrade "Rikard Podhorky" Akademije tehničkih znanosti Hrvatske, Nagrade za znanost Fakulteta elektrotehnike i računarstva, Fulbrajtove stipendije i Nagrade Hrvatske sekcije IEEE.

Preface

I would like to express my sincere gratitude to my thesis supervisor prof. Sven Lončarić for his guidance and continuous support during my work on this thesis, as well as for his help and contribution to the scientific papers included in this thesis. Many thanks to prof. Marko Subašić for suggesting me the idea to enroll in PhD and for his help and guidance throughout my whole higher education.

A special thanks go to all great doctoral and postdoctoral researchers I met and had the privilege to work with, share tough and joyful times.

I also wish to express my special thanks to my parents, sister, family and friends for their support and encouragement during my doctoral study and years of education.

Most importantly, to my wife Anita and my son Roko - thank you! Anita, thank you for your never-ending support, for your patience, for believing in me, and for pushing me when I needed it the most. Roko, your birth empowered me to write this thesis. You are my greatest motivation, and nothing compares to having you in our lives.

Abstract

Color can be defined as "the property possessed by an object of producing different sensations on the eye as a result of the way the object reflects or emits light"*. Color is a perceptual term that describes the response of the human eye to radiation in the visible range of the electromagnetic spectrum. Due to the varying reflectance properties, different objects in the same scene emit the incident light differently. The human visual system (HVS) perceives these objects as differently colored. The appearance of an object in terms of color varies depending on the light source in the scene because the perceived color of an object is subject to its reflectance properties and the light source's spectrum. However, HVS is very robust to changes in the observed scene and adapts rapidly. HVS can perceive objects' colors invariant of the present light source; therefore, a lemon is yellow in, e.g., sunlight and under the light of an incandescent light bulb. This ability is called color constancy. Image sensors in digital cameras, on the other hand, do not have this ability. Therefore, in digital cameras, a pre-processing step is dedicated to achieving invariance of colors to the scene illumination. That step is referred to as computational color constancy. The impact of the illumination color on colors in digital images is usually removed in two steps. First, a method estimates the color of the light source. Second, chromatic adaptation using the estimate renders an illumination-invariant image. The outcome is an image in which white objects indeed appear white. Thus, it is referred to as white balancing.

The focus of this thesis is on illumination estimation. By definition, it is an ill-posed problem. Given only image pixels, a vector representing scene illumination has to be estimated; however, two (or more) individual reflections in the scene can map to the same pixel value. A straightforward approach toward the solution in such tasks is to relax the problem using assumptions. The aim of this research is the analysis of various approaches to illumination estimation. Each approach is implemented in a method using deep learning methodology. Many illumination estimation methods already exist, but often they become inaccurate when more complex scenes occur. On the other side, deep neural networks achieve state-of-the-art results in many computer vision tasks. They were set apart by their exceptional generalization capability. Deep learning methods based on convolutional neural networks accomplish great accuracy in illumination estimation as well. Deep learning usually involves the question of large datasets that are not typical in illumination estimation. A part of the research in this thesis is dedicated to studying existing datasets and establishing a set of desired features a dataset for illumination estimation should have. Finally, a novel dataset conforming to the defined features is presented.

Keywords: color constancy, convolutional neural networks, deep learning, illumination estimation, white balancing

*Oxford dictionary

Prošireni sažetak

Metode procjene boje osvjetljenja scene zasnovane na konvolucijskim neuronskim mrežama

Boja se može definirati kao svojstvo objekta za stvaranje različitih osjeta u oku kao rezultat načina na koji objekt reflektira ili emitira svjetlost (preuzeto iz Oxfordskog rječnika). Boja je percepcijski pojam koji opisuje odgovor ljudskog oka na zračenje u vidljivom rasponu elektromagnetskog spektra. Zbog različitih svojstava refleksije, različiti objekti u istoj sceni drugačije emitiraju upadnu svjetlost. Ljudski vizualni sustav percipira te objekte kao različito obojene. Izgled objekta u smislu boje mijenja se ovisno o izvoru svjetlosti jer je percipirana boja objekta ovisna o njegovim svojstvima refleksije i spektru izvora svjetlosti. Međutim, ljudski vizualni sustav je vrlo robustan na promjene u promatranoj sceni i brzo se prilagođava. Ljudski vizualni sustav može percipirati boje objekata neovisno o izvoru svjetlosti. Stoga je limun žute boje na sunčevoj svjetlosti i na svjetlu žarulje sa žarnom niti. Ova sposobnost naziva se postojanost boja. Senzori za stvaranje slike u digitalnim fotoaparatima nemaju tu mogućnost. Stoga u digitalnim fotoaparatima na početku procesa formiranja slike postoji korak predobrade namijenjen postizanju invarijantnosti boja na osvjetljenje u sceni. Taj korak naziva se računalna postojanost boja. Utjecaj boje izvora svjetlosti na boje u digitalnim slikama obično se uklanja u dva koraka. U prvom koraku metoda procjenjuje boju izvora svjetlosti. U drugom koraku, kromatskom adaptacijom koristeći dobivenu procjenu dobiva se slika invarijantna na osvjetljenje. Rezultat ovog postupka je slika na kojoj bijeli objekti zaista izgledaju bijelo. Stoga se ovaj proces često naziva i podešavanje bijele.

Fokus ove disertacije je na procjeni osvjetljenja. Po definiciji, to je loše postavljen problem. Koristeći samo piksele slike procjenjuje se vektor koji označava osvjetljenje u sceni. Međutim, dvije ili više različitih refleksija u sceni mogu rezultirati istom vrijednošću piksela. Izravan pristup rješavanju takvih zadataka je pojednostaviti problem korištenjem pretpostavki. Cilj ovog istraživanja je analiza različitih pristupa za procjenu osvjetljenja. Svaki razmatrani pristup implementiran je u metodu koja koristi metodologiju dubokog učenja. Mnoge metode za procjenu osvjetljenja već postoje, ali one su često netočne za vrlo složene scene. S druge strane, duboke neuronske mreže postižu vrhunske rezultate u mnogim zadacima računalnog vida. Odlikuju se iznimnom sposobnošću generalizacije. Metode dubokog učenja zasnovane na konvolucijskim neuronskim mrežama postižu visoku točnost i u procjeni osvjetljenja. Duboko učenje obično povlači i pitanje velikih skupova podataka, koji nisu karakteristični za procjenu osvjetljenja. Dio istraživanja prikazanog u ovoj disertaciji posvećen je proučavanju postojećih skupova podataka kako bi se ustanovio skup značajki poželjnih za skup podataka za procjenu osvjetljenja. Konačno, u posljednjem dijelu disertacije prezentiran je novi skup podataka za procjenu osvjetljenja zasnovan na definiranom skupu značajki.

Izvorni znanstveni doprinos ove disertacije podijeljen je u četiri dijela:

- metoda procjene boje osvjetljenja zasnovana na konvolucijskoj neuronskoj mreži s mehnizmom pažnje;
- metoda procjene boje osvjetljenja korištenjem klasifikacije osvjetljenja scene zasnovana na dubokom učenju;
- metoda procjene boje osvjetljenja zasnovana na povratnoj dubokoj neuronskoj mreži s višestupanjskom funkcijom gubitka;
- skup podataka za procjenu osvjetljenja Cube++.

U prvom dijelu disertacije predložena je metoda za procjenu osvjetljenja zasnovana na konvolucijskoj neuronskoj mreži s mehanizmom pažnje. Osnovna pretpostavka ove metode je da su za procjenu osvjetljenja osim vrijednosti piksela slike bitne i dodatne informacije, npr. vrijeme snimanja i lokacija. Međutim takve informacije su rijetko dostupne i ne može ih se uzeti u obzir, nego se osvjetljenje procjenjuje koristeći samo vrijednosti piksela slike. Drugi pristup je korištenje nekakvog mehanizma koji na osnovu same slike može izvući dodatne informacije o sceni. Jedan takav mehanizam u dubokom učenju naziva se mehanizam pažnje. Mehanizam pažnje kvantitativno određuje uvjerenost neuronske mreže da je određeni dio ulaznih podataka koristan za dani problem. U problemu procjene osvjetljenja mehanizam pažnje određuje koje regije u slici sadrže dovoljno informacija za određivanje vektora osvjetljenja. Tako se neuronska mreža dodatno usmjerava prema regijama koje su korisne za dani zadatak i izbjegava fokusiranje na nebitne i ponekad višeznačne regije. U ovoj disertaciji opisane su dvije metode s mehanizmom pažnje. Princip rada obje metode je isti, a metode se razlikuju u implementaciji mehanizma pažnje. Za određivanje mape značajki ulazne slike koristi se arhitektura iste konvolucijske neuronske mreže. Mapa značajki paralelno se obrađuje u dvije nezavisne grane. Jedna grana računa lokalne procjene osvjetljenja, a druga grana je mehanizam pažnje. Mehanizam pažnje prve metode implementiran je koristeći tri konvolucijska sloja te izračunava zasebnu mapu uvjerenosti za svaki kanal slike. Dodatna karakteristika ovog mehanizma pažnje je činjenica da se za svaku lokalnu regiju koriste zasebni filteri konvolucijskog sloja, tj. za svaku lokalnu regiju određuje se uvjerenost neovisno o sadržaju susjednih regija. Lokalne procjene osvjetljenja iz prve grane množe se s odgovarajućim vrijednostima koje je odredio mehanizam pažnje, umnožak se uprosječuje i normalizira. Rezultat je vektor koji predstavlja globalno osvjetljenje u sceni. Drugi mehanizam pažnje razlikuje se u načinu određivanja mape uvjerenosti i brojem tih mapa. Ovaj mehanizam računa samo jednu mapu značajki koja se koristi za sve kanale slike. Vrijednosti mape se ne računaju zasebno za svaku regiju već se koristi tradicionalni konvolucijski pristup dijeljenja težina. Prvi mehanizam pažnje u skladu je s pretpostavkom neovisnih kanala slike, koja se korist kod korekcije slike dijagonalnom matricom. Drugi mehanizam zasniva se na pretpostavci da ljudski vizualni sustav regije interesa određuje s obzirom na teksture, oblike i sveukupni dojam boje. Eksperimentalni rezultati pokazali su

velika poklapanja mapa uvjerenosti s vrijednostima gradijenta slike. Ovo pokazuje da predložene metode uspijevaju odvojiti regije slike s korisnim informacijama od nebitnih regija. Kod procjene osvjetljenja nebitnima se smatraju višeznačne regije, a to su područja kod kojih bez uvođenja dodatnog konteksta postoji više kombinacija boje površine i boje osvjetljenja koje rezultiraju istim vrijednostima piksela. Jedan primjer takve regije su jednobojni, čisti zidovi.

U drugom dijelu disertacije prezentirana je metoda za procjenu osvjetljenja koja koristi klasifikaciju izvora svjetlosti zasnovanu na dubokom učenju. Osvjetljenja u stvarnome svijetu na različite načine utječu na boje u digitalnim slikama. Izvori svjetlosti bliski bijelom osvjetljenju imaju vrlo mali učinak na boje objekata u digitalnim slikama dok umjetna osvjetljenja vrlo često znatno mijenjaju boje objekata prilikom snimanja digitalnim kamerama. Raznolikost boja izvora svjetlosti najbolje se može prikazati uvidom u referentne vrijednosti osvjetljenja postojećih skupova podataka. Kada se referentne vrijednosti promatraju u domeni kromatičnosti uočava se grupiranje u tri odvojene skupine. Promatrani skup podataka sadržan je od slika pod dnevnom svjetlošću, slika u zatvorenim prostorima i noćnih slika. Podjela u domeni kromatičnosti poklapa se upravo s takvom klasifikacijom slika te se na njoj zasniva i prezentirana metoda za procjenu osvjetljenja. Poznavajući informaciju o skupini kojoj slika pripada smanjuje se prostor mogućih osvjetljenja jer svaka skupina zauzima jednu manju zatvorenu regiju u domeni kromatičnosti. Prezentirana metoda iskorištava tu činjenicu te za svaku skupinu koristi zasebni estimator. Metoda se sastoji od četiri konvolucijske neuronske mreže. Zadatak jedne mreže je klasifikacija ulaznih slika u sljedeće skupine: vanjske scene pod umjetnim osvjetljenjima, vanjske scene pod prirodnim osvjetljenjima i unutrašnje scene pod umjetnim osvjetljenjima. Ovisno o rezultatu klasifikacije, slika se prosljeđuje jednoj od tri konvolucijske neuronske mreže za estimaciju osvjetljenja. Sve tri konvolucijske mreže za estimaciju osvjetljenja zasnivaju se na istoj arhitekturi. Međutim, svaka od te tri mreže trenirana je za estimaciju osvjetljenja isključivo na slikama jedne skupine te se stoga i primjenjuje samo za slike klasificirane u tu skupinu. Uporaba klasifikacije slika i tri estimatora specijalizirana za pojedine kategorije slika pokazala se preciznijom od jednostavne procjene osvjetljenja gdje se ista konvolucijska neuronska mreža koristi za procjenu osvjetljenja svih slika. Postignuto je poboljšanje od 30% s obzirom na najveće pogreške te je eksperimentalnim rezultatima pokazano da je prezentirana klasifikacija bolja od klasifikacije s obzirom samo na vrstu scene (slike unutrašnje ili vanjske scene) ili vrstu osvjetljenja (slike pod prirodnim ili umjetnim osvjetljenjem). Dodatna značajka metode je što klasifikatori ne ovise o raspodjeli slika u skupu podataka, tj. broj slika unutar kategorija ne mora biti isti.

U trećem dijelu disertacije prezentirana je metoda za procjenu osvjetljenja zasnovana na iterativnoj procjeni osvjetljenja uporabom konvolucijske neuronske mreže. Ideja na kojoj se zasniva ova metoda potiče od činjenice da oba koraka za postizanje računalne postojanosti boja imaju svoje mane. Procjena osvjetljenja loše je postavljen problem te se pokušava riješiti pret-

postavkama. S druge strane, korekcija boja aproksimira se jednostavnom dijagonalnom matricom. Stoga je računalnu postojanost boja u nekim slučajevima teško postići. Istraživanje u sklopu ove disertacije pokazuje da i proizvođači digitalnih kamera vrlo vjerojatno uzimaju takve činjenice u obzir. Pokazalo se da automatsko podešavanje bijele u Canon digitalnim kamerama ograničava prostor kromatičnosti na samo mali dio omeđen četverokutom te tako sprječava velike pogreške u korekciji slike, koje mogu biti uzrokovane žarko obojenim osvjetljenjima. Prezentirana metoda procjenjuje osvjetljenje u više koraka. Svaki korak sastoji se od procjene osvjetljenja ograničene na područje u blizini bijelog svjetla i korekcije slike uporabom procijenjenog osvjetljenja pri čemu se korigirana slika koristi kao ulaz sljedećeg koraka. Produkt ograničenih procjena osvjetljenja svih koraka odgovara globalnom osvjetljenju početne slike. Ovakav pristup procjeni osvjetljenja cilja na smanjenje pogrešaka u najgorim slučajevima za koje se pretpostavlja da se mogu pojaviti kada osvjetljenje jako odstupa od bijelog svjetla. Kako bi se ograničio rad metode samo na područje osvjetljenja u blizini bijeloga, osmišljena je specifična funkcija računanja pogreške. Ova funkcija sastoji se od dva dijela. U prvom dijelu računa se kut između procijenjenog globalnog vektora osvjetljenja i stvarnog globalnog vektora osvjetljenja. U drugom dijelu penaliziraju se procjene načinjene u svakom pojedinom koraku na način da se za svaku procjenu računa kut s obzirom na bijelo svjetlo pomnožen s težinskim faktorom kojim se određuje maksimalno dopušteno odstupanje od bijelog svjetla. Težinski faktor povećava se proporcionalnu rednom broju koraka čime se osigurava postepeno smanjenje maksimalno dopuštenog odstupanja od bijelog svjetla. U idealnom slučaju, u posljednjem koraku kut između ograničene procjene osvjetljenja i vektora bijelog svjetla trebao bi iznositi nula, tj. posljednja procjena bi trebala biti neutralna. Metoda je implementirana kao konvolucijska neuronska mreža koja koristi iste težine u svim koracima i množenje s dijagonalnom matricom za korekciju boja slike.

U konačnici, u četvrtom dijelu disertacije prezentiran je novi skup podataka za procjenu osvjetljenja u kojem su sumirana sva znanja stečena istraživačkim radom. Prezentirani skup podataka jedan je od najvećih u području s 4890 slika visoke rezolucije. Međutim, ono što je iznimno bitno je da je taj skup podataka napravljen u skladu s novo-definiranim nizom značajki za skupove podataka za procjenu osvjetljenja, a to su: raznolikost sadržaja i osvjetljenja, veliki broj uzoraka, bogatstvo različitih informacija o slici, jednostavno ažuriranje i praćenje promjena, provjerljivost, jednostavnost pristupa i usklađenost s GDPR-om. Mnoge od ovih značajki motivirane su nedostacima prijašnjih skupova podataka za procjenu osvjetljenja kao što su neusklađenost s pretpostavkom o postojanju samo jednog osvjetljenja, nekoliko loše sinkroniziranih verzija istog skupa podataka te nedostatan broj slika i osvjetljenja. Prezentirani skup podataka korisniku nudi mnoštvo informacija za svaku sliku te tako pruža visoku razinu slobode u razvoju metoda. Velika količina informacija (EXIF podaci, vrijeme nastanka slike, tip osvjetljenja, informacije o lokaciji, sjenama, broju izvora osvjetljenja, raznolikosti scene i kali-

bracijskom objektu) te raznolikost scena i osvjetljenja plodno su tlo za razvoj metoda temeljenih na dubokom učenju, poput metoda opisanih u ovoj disertaciji.

Doktorska disertacija sastoji se od šest radova diseminiranih u časopisima velikog faktora odjeka i na konferencijama. Priloženi radovi predstavljaju izvorni znanstveni doprinos ove disertacije koji se sastoji od prethodno opisana četiri dijela. Disertacija započinje opisom metodologije procjene osvjetljenja i pregledom literature. Zatim je predstavljen izvorni znanstveni doprinos popraćen znanstvenim radovima.

Ključne riječi: postojanost boja, konvolucijske neuronske mreže, duboko učenje, procjena osvjetljenja, podešavanje bijele

Contents

Chapter 1

Introduction

1.1 Computational color constancy

In Figure 1.1, three images of the same scene captured with the same camera are shown. Only the white balance setting was modified for each captured image. Starting from the left image, *Shade*, *Tungsten light*, and *White fluorescent light* were used. These settings tell the camera about the scene light source and which correction to use to remove its influence on image colors. Different auto white balance settings yield differently colored images. However, if a person is placed in the scene while the light source changes, most likely it would perceive colors the same regardless of the light source; due to the property of the human vision system called color constancy. For humans, this operation is innate and subconscious. However, in digital photography, this is an ill-posed problem called computational color constancy.

The computational color constancy objective is to render images invariant of the illumination color. When performed well, a white surface within the scene should appear white in the image capturing the scene, that is, $R = G = B$ in the camera's RGB color space. Therefore, in digital photography, computational color constancy is also called white balancing. For instance, in Figure 1.1, the middle image was captured with the most appropriate white balance setting

Figure 1.1: The impact of different white balance settings on the appearance of digital images. Images of the same scene are captured with three white balance settings. The settings used are the following: left image - *Shade*; middle image - *Tungsten light*; right image - *White fluorescent light*. All images were captured with the same Canon EOS 550D camera using ISO 3200, exposure time 1/15, and aperture f/5.6.

(*Tungsten light*). Computational color constancy is achieved in two steps that are illumination estimation and chromatic adaptation, with the former being more researched.

The straightforward way to illumination estimation is to capture an image that contains at least one achromatic ("white") object. For white light, the pixel values of an achromatic object should coincide with the achromatic line $R = G = B$. For colored illuminations, the pixel values for the same object would deviate from that line. Hence, the color of the scene illumination can be represented by the R, G, and B values of the achromatic object. This approach is possible in two cases; either a calibration object is placed in the image scene or an achromatic surface is a part of the scene, and its location is known. Both cases are equally unlikely; having a calibration object in the image is highly undesirable, and an achromatic object may not appear in all scenes. Moreover, even if an achromatic object occurs, its location must be determined, which is in the scope of a completely different research area. Therefore, illumination estimation algorithms are applied onboard cameras at the beginning of the image formation pipeline.

The result of achieving computational color constancy is a color-corrected image. Being able to achieve this consistency was shown beneficial in many other computer vision tasks. For example, improperly white-balanced images impact the accuracy of classification and segmentation [1, 2].

1.2 Scope of the thesis

From the two steps in computational color constancy, illumination estimation is considered more challenging and critical task. The research community adopted a sufficiently accurate model for chromatic adaptation. However, illumination estimation is still widely researched. In this thesis, the emphasis is on methods for illumination estimation. The thesis aims at developing new methods utilizing convolutional neural networks. Based on the number of light sources, illumination estimation methods are local (multiple light sources) or global (single light source). Typically, in the real world, light is emitted from several sources. For example, in a room, sunlight may pass through the windows, and multiple light bulbs may illuminate the room. Therefore, some parts of the room are illuminated by the mixture of all these light sources. However, local illumination estimation is a difficult task. In addition, it is hard to collect the data and evaluate methods objectively, since ground-truth information is difficult to determine. Global illumination estimation is theoretically an easier problem since a single dominant light source is assumed. Nevertheless, no method meets the task. Therefore, in this thesis, methods for global illumination estimation are researched.

Nowadays, deep learning is state-of-the-art in many image-related fields. Convolutional neural networks are mainly used. They are designed assuming images as inputs. Deep neural networks have a substantial capacity to capture a wide range of events and extract various fea-

tures given a task. Estimating the light source color is an ill-posed problem, i.e., many inputs map to the same output. Therefore, tackling it with deep learning could yield improvements, especially in some border cases where traditional methods fail due to their underlying assumptions. In terms of accuracy, the methods proposed in this thesis aim to improve the illumination estimation in such cases.

Quality datasets are required to tackle deep learning. However, data in computational color constancy is rarely suitable for deep learning-based methods. The most commonly used datasets are small and, moreover, tend to deviate from necessary prerequisites, e.g., they violate the uniform illumination assumption. One part of this thesis is dedicated to illumination estimation datasets.

1.3 Scientific contribution

The scientific contribution of this thesis is contained in six scientific publications, and it pursues the illumination estimation problem regarding two interacting viewpoints: methods for illumination estimation and data acquisition. In this thesis, three illumination estimation methods are presented:

- illumination color estimation method based on a convolutional neural network with an attention mechanism,
- illumination color estimation method using scene lighting classification based on deep learning,
- illumination color estimation method based on recurrent deep neural network with a multistage loss function.

In the scope of data acquisition, a novel illumination estimation dataset is presented, including the acquisition methodology and technical information.

1.4 Organization of the thesis

The organization of the thesis is as follows. Chapter 2 is an overview of the computational color constancy. It provides a reader with the methodology for achieving computational color constancy and an overview of the literature on illumination estimation methods. In Chapter 3, a brief introduction to convolutional neural networks is given, and illumination estimation approaches utilizing convolutional networks are described. The evaluation of illumination estimation methods is described in Chapter 4. In Chapter 5, the scientific contribution, contained in six scientific publications, is presented. Thesis conclusions and plans for future research are given in Chapter 6. The list of scientific publications referring to the thesis contributions and author's contribution in each publication are given in Chapter 7 and Chapter 8, respectively.

Chapter 2

Overview

Computational color constancy is implemented at the beginning of the image formation pipeline and applied to raw-RGB images; an illustration of the typical image processing pipeline is shown in Figure 2.1. Computational color constancy is accomplished in two steps. First, an illumination vector is estimated from the pixel values; this step is called illumination estimation. Second, all pixels are corrected considering the estimated vector; this step is called chromatic adaptation. Whether a single vector is estimated for the whole image or multiple vectors relating to different local image regions are estimated, global and local methods exist, respectively. Computational color constancy is applied on RGB images with minimal processing and without any non-linearity applied. Since only linear operations may have been applied, images are referred to as "linear" or raw images. The required processing includes black level subtraction and the removal of overexposed pixels.

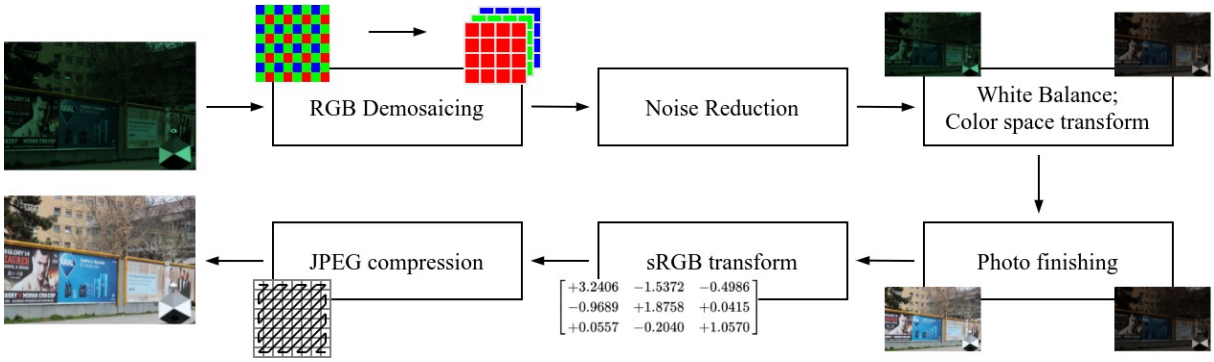


Figure 2.1: An illustration of the typical image formation pipeline in digital cameras; may vary depending on the camera manufacturer and model complexity. White balancing is performed at the beginning, before non-linear operations such as compression, tone mapping, and various proprietary color enhancement operations.

This chapter introduces concepts of computational color constancy and gives an overview of the literature. In this thesis, the taxonomy of illumination estimation methods is based on one defined by Gijsenij et al. [3]. Since the gamut-based methods require learning of the canonical gamut, they are presented within the learning-based category. Methods are divided into

two groups: static methods and learning-based methods. The emphasis of this thesis is on illumination estimation methods based on deep learning. For convenience, they are described separately in Chapter 3 3. Many illumination estimation methods exist, and therefore, only the most influential and well-known ones are described in the scope of this thesis.

2.1 Illumination estimation

Depending on whether the specular reflection is modeled, two image formation models can be considered: the dichromatic reflection model [4] and the Lambertian model [5]. Lambertian model is the simplest image formation model since it disregards specular reflection. This model defines reflection as absolutely diffusive. That means that no matter the viewing angle of an observer, the brightness of a pixel remains the same. The image formation model under the Lambertian assumption is

$$I_c(x,y) = \int_\omega L(x,y,\lambda)R(x,y,\lambda)\rho_c(\lambda)\,\mathrm{d}\lambda \quad c \in \{R,G,B\}. \tag{2.1}$$

Intensity I of a pixel at the position (x,y) for the color channel $c \in \{R,G,B\}$ is obtained by integration over the wavelengths λ in the visible light spectrum* ω. For each pixel, $L(x,y,\lambda)$ and $R(x,y,\lambda)$ denote the spectral distribution of the light source and surface reflectance, respectively. Camera-dependency is modeled with $\rho_c(\lambda)$ that is the sensitivity of the camera sensor for the color channel c.

Global illumination estimation methods assume that the light source spectrum is the same regardless of the position in the image, which simplifies the model to

$$I_c(x,y) = \int_\omega L(\lambda)R(x,y,\lambda)\rho_c(\lambda)\,\mathrm{d}\lambda. \tag{2.2}$$

Then the vector of the observed, assumed global, light source color is

$$e_c = \int_\omega L(\lambda)\rho_c(\lambda)\,\mathrm{d}\lambda, \quad c \in \{R,G,B\}. \tag{2.3}$$

The objective of illumination estimation is to estimate components e_c knowing only pixel values. Both $L(\lambda)$ and $\boldsymbol{\rho}(\lambda)$ are, in general, unknown. That makes the problem of estimating the illumination ill-posed. Common practice is to pose assumptions that re-formulate the task to have a feasible solution. Since illumination can be approximated only up to a scaling factor, occasionally, chromaticities are used instead of RGB values; thus, only the ratio of red, green, and blue is considered while disregarding the intensity. Given R, G, and B components, the

*Visible light spectrum is in the range from 380 to 750 nm.

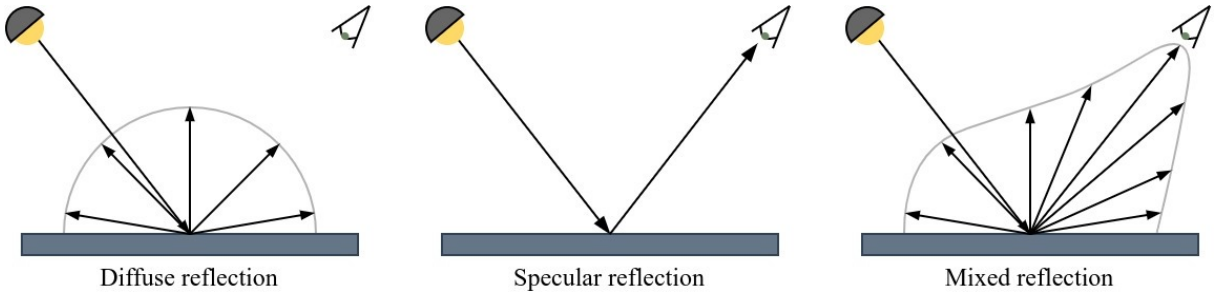


Figure 2.2: Reflection types used in image formation models: Lambertian model [5] assumes only diffuse reflection; dichromatic reflection model [4] assumes both diffuse and specular reflection.

corresponding r, g, and b chromaticities are

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B}. \tag{2.4}$$

Furthermore, some approaches [6, 7] rely on uv log-chrominance components [8, 9]:

$$u = \log \frac{G}{R}, \quad v = \log \frac{G}{B}. \tag{2.5}$$

The dichromatic reflection model accounts for the specular reflection. It is a generalization of the Lambertian model. The dichromatic reflection model is given as

$$I_c(x,y) = m_b(x,y) \int_\omega L(x,y,\lambda) R(x,y,\lambda) \rho_c(\lambda) \, \mathrm{d}\lambda + m_s(x,y) \int_\omega L(x,y,\lambda) \rho_c(\lambda) \, \mathrm{d}\lambda, \tag{2.6}$$

where $m_b(x,y)$ and $m_s(x,y)$ are scaling factors relating to the amount of body and specular reflection, respectively.

Of the two, the Lambertian model is obviously simpler. Nevertheless, it is sufficiently accurate, and it is the basis for most methods. Throughout this thesis, the Lambertian model is used. Figure 2.2 illustrates both image formation models.

2.2 Chromatic adaptation

Chromatic adaptation is the second step in achieving computational color constancy. It corrects the color bias of the image and renders the image under a canonical illumination. Chromatic adaptation is formed as a linear transformation of a color-biased pixel $\tilde{\mathbf{I}} = [\tilde{I}_R, \tilde{I}_G, \tilde{I}_B]$ to its canonical representation $\mathbf{I} = [I_R, I_G, I_B]$ using simple scaling operation as (for simplicity, here, the spatial information (x,y) is omitted from the notation)

$$\mathbf{I} = \mathbf{M}\tilde{\mathbf{I}}, \tag{2.7}$$

Figure 2.3: An example of applying von Kries-based chromatic adaptation model [10, 11]. Left: original color-biased image and the corresponding illumination color; right: corrected image is shown.

where $\mathbf{M}$ is a 3-by-3 correction matrix. A widely used simplification of this model was introduced by von Kries [10, 11]; a multiplicative scaling of each color component independently was proposed. In other words, $\mathbf{M}$ is approximated as a diagonal matrix

$$\mathbf{M} = \operatorname{diag}\left(\left[\frac{e_R}{\tilde{e}_R}, \frac{e_G}{\tilde{e}_G}, \frac{e_B}{\tilde{e}_B}\right]\right), \tag{2.8}$$

where $\mathbf{e} = [e_R, e_G, e_B]$ and $\tilde{\mathbf{e}} = [\tilde{e}_R, \tilde{e}_G, \tilde{e}_B]$ are canonical illumination vector and estimated illumination vector, respectively. [†] Examples of Canon's built-in diagonal white balance matrices that were used for images in Figure 1.1 are:

$$\begin{aligned}
\mathbf{M}^{\text{Shade}} &= \operatorname{diag}([2.4619, 1.000, 1.3125]), \\
\mathbf{M}^{\text{Tungsten light}} &= \operatorname{diag}([1.5488, 1.000, 2.2510]), \\
\mathbf{M}^{\text{White fluorescent light}} &= \operatorname{diag}([1.9180, 1.000, 2.1074]).
\end{aligned}$$

The canonical illumination is most often achromatic, i.e., so-called white light. For white light, $e_R = e_G = e_B$ and, since illumination can be estimated up to a scaling constant only, in practice, $\mathbf{M}$ is mostly

$$\mathbf{M} = \operatorname{diag}\left(\left[\frac{1}{\tilde{e}_R}, \frac{1}{\tilde{e}_G}, \frac{1}{\tilde{e}_B}\right]\right). \tag{2.9}$$

An example of a color-corrected image is shown in Figure 2.3.

[†] $\operatorname{diag}([k,l,m]) = \begin{bmatrix} k & 0 & 0 \\ 0 & l & 0 \\ 0 & 0 & m \end{bmatrix}$

2.3 Static illumination estimation methods

2.3.1 Statistical-based methods

Statistical-based illumination estimation methods are based on different statistical properties of images, e.g., moments and gradients. They do not require any learning and large datasets related to learning techniques. Typically, static methods are pretty simple and effective in terms of low execution time, making them hardware-friendly. Nevertheless, the simplicity of static methods is mostly their major downside. Zakizadeh et al. [12] experimentally showed that specific types of images are hard for statistical methods and that learning-based methods can handle such samples. In terms of accuracy, static methods are inferior compared to learning-based methods.

The simplest and the most famous assumption in illumination estimation is the Gray World assumption [13]. It assumes that the average reflectance in the scene is achromatic, i.e., gray. The deviation of the computed average from achromatic is then due to the impact of the illumination. The color of the illumination equals the mean of the image. Gray World fostered a family of illumination estimation methods that originate from the same base assumption. Finlayson and Trezzi [14] proposed the Shades of Gray method that extends the Gray World by introducing the Minkowski norm. They estimate the illumination vector as

$$\mathbf{e} = k \left(\iint \mathbf{I}(x,y)^p \, \mathrm{d}x \, \mathrm{d}y \right)^{\frac{1}{p}}, \tag{2.10}$$

where p denotes the order of the Minkowski norm and k is the normalizing constant. For $p = 1$, (2.10) equals the Gray World, and for $p \to \infty$, it becomes the Max-RGB method [15]. Max-RGB is also often called White Patch, since it assumes the presence of a white surface in the scene that reflects illumination perfectly. Max-RGB is implemented by computing the maximum intensity for each color channel as

$$\mathbf{e} = k \max_{(x,y)} \mathbf{I}(x,y). \tag{2.11}$$

Joze et al. [16] extend the White Patch by including a gamut of bright pixels in computation. Van de Weijer et al. [17] introduce local smoothing as an additional improvement of the Gray World algorithm. They apply local smoothing by using a Gaussian filter and compute the illumination as

$$\mathbf{e} = k \left(\iint (\mathbf{I}(x,y) \otimes G_\sigma(x,y))^p \, \mathrm{d}x \, \mathrm{d}y \right)^{\frac{1}{p}}, \tag{2.12}$$

where G_σ denotes a Gaussian filter with standard deviation σ. In the same work, Van de Weijer et al. scaled up the Gray World to higher-order statistics. The Gray Edge hypothesis was introduced: "the average of the reflectance differences in a scene is achromatic" [17]. That

leads to computing the illumination vector as

$$\mathbf{e} = k\left(\iint \left|\frac{\partial^n(\mathbf{I}(x,y)\otimes G_\sigma(x,y))^p}{\partial^n x \partial^n y}\right| \mathrm{d}x\,\mathrm{d}y\right)^{\frac{1}{p}}, \tag{2.13}$$

where n denotes the order of the derivative. Assigning weights to the edges based on the edge type further improves the performance [18]. Table 2.1 summarizes statistical-based methods built upon the Gray World assumption that can be derived from (2.13) using different values for n, p, and σ.

Method name	(n, p, σ)	Assumption
Gray World [13]	0, 1, 0	scene average is gray
Max-RGB [15]	0, ∞, 0	scene maximum is gray
Shades of Gray [14]	0, $*$, 0	Minkowski norm of a scene is gray
General Gray World [17]	0, $*$, $*$	Minkowski norm of a smoothed scene is gray
1$^{\text{st}}$ order Gray Edge [17]	1, $*$, $*$	Minkowski norm of scene derivative is gray
2$^{\text{nd}}$ order Gray Edge [17]	2, $*$, $*$	Minkowski norm of scene second derivative is gray

Table 2.1: Statistical-based illumination estimation methods contained in (2.13) that are obtained by varying n, p, and σ parameters. $*$ denotes arbitrary parameter value.

Qian et al. [19] propose finding gray pixels regarding some observed image statistics. The method relies on the dichromatic reflection model characteristic of physics-based approaches described in Section 2.3.2; however, the authors regard the methods as statistical-based due to the underlying statistics used.

2.3.2 Physics-based methods

Physics-based illumination estimation methods examine the physical nature of illumination and object interaction. They commonly rely on a more complex dichromatic reflection model (2.6); hence, highlights and inter-reflections can be modeled. The simplest approach is to find pixels for which only the specular component exists, i.e., pixels (x,y) for which $m^b(x,y) = 0$ in (2.6). Specular reflections are usually brighter than body reflections. Therefore, this approach comes down to Max-RGB [15]. Methods relying on specularities or highlights include [20, 21, 22, 23]. However, in practice, these methods are difficult to apply to real-world images due to ambient light and the lack of specular information. Quite recently, Woo et al. [24] proposed a method that relies on finding a path for which the longest dichromatic line is produced by specular pixels. They assume the Phone reflection model [25] and account for ambient light.

2.4 Learning-based illumination estimation methods

Learning-based methods rely on a substantial amount of data for learning a model that can estimate illumination from unseen image data. Learning-based methods outperform static methods in terms of accuracy but typically at the expense of the increased computational complexity and large memory requirements. In this section, an overview of conventional learning-based methods such as gamut mapping, probabilistic models, and machine learning methods is given. The emerging subgroup of learning-based methods are deep learning methods; for convenience, they are reviewed in Chapter 3.

Gamut mapping [26] is based on the observation that, given an illumination, all observed colors fit inside a convex hull. The set of possible colors in the image under a canonical illumination is called the canonical gamut. Gamut mapping looks for the mapping between the gamut of an image with an unknown light source into the canonical gamut. Usually, many mappings are feasible; thus a single mapping has to be selected regarding some selection criteria, e.g., the mapping resulting in a most colorful scene [26] or averaging and weighted averaging [27].

Gamut mapping can also be applied in 2D chromaticity space [28, 29]. To alleviate the problem of unrealistic illuminations, Finlayson et al. [30] defined a set of plausible illuminations, and for each, they determine the gamut. For an unseen image, they estimate illumination by simply finding the predetermined gamuts that the image data relates to most. Gijsenij et al. [31] improve the Gamut mapping by including image derivatives up to the second-order in gamut calculation, along with pure pixel values.

One line of research focuses on using low-level features to model the posterior probability of different illuminations and reflectances relying on Bayesian learning [32, 33, 34, 35].

Typically, illumination estimation algorithms are assumption-based and, thus, susceptible to error. Therefore, several methods relying on combining outputs of multiple methods were proposed. Cardei and Funt [36] combined methods using weighted average and neural network-based fusion. Schaefer et al. [37] used statistical and physical methods to compute likelihoods for some predefined light sources and combined the likelihoods to estimate the illumination. Gijsenij and Gevers [38, 39] first determine the most suitable illumination estimation method relying on intrinsic properties of natural images and then apply only the selected network to obtain the illumination vector.

Other learning-based approaches also include methods such as limiting the space of possible solutions to illuminations computed by existing illumination estimation methods [40], using linear regression [41, 42, 43], and support vector regression [44, 45].

Chapter 3

Deep learning-based illumination estimation methods

3.1 Convolutional neural networks

Convolutional neural network (CNN) architectures are deep structures designed to tackle the problems with grid-like structured data, such as images. They are named after the mathematical convolution operation although, in practice, CNNs are mainly implemented using cross-correlation (the convolution without flipping the kernel) [46]

$$(F * I)(x,y) = \sum_m \sum_n F(m,n) I(x+m, y+n), \tag{3.1}$$

where F is the convolution filter, I is the input image, (x,y) is the spatial position in the input, and (m,n) denotes indices of individual filter parameters. Convolutional neural networks are based on the following properties: local connectivity [*], weight sharing [†], and equivariance [‡]. The three most common elements in all CNNs are the convolutional layer, nonlinear activation function, and pooling operation.

The convolutional layer contains a set of weights, called kernels or filters, that are learned in the training phase. The spatial extent of a filter is much less compared to the input, often 3×3. The depth of a filter matches the depth of the input, e.g., assuming the input is an RGB image, the filter depth should be three. A single convolutional layer typically contains multiple filters,

[*]CNNs account for the spatial structure in the data using filters with small receptive fields (receptive field is the spatial extent a filter is connected to). Therefore, each filter is connected only to a small spatial portion of the input data but along the entire input depth.

[†]Weight sharing corresponds to using the same set of weights at every position in the input. CNNs do not use a separate filter for each spatial location in the input data. They use a set of filters shared across the whole spatial extent.

[‡]Due to weight sharing, convolutional layers are equivariant to translation. That means that applying convolution to the translated input is the same as applying it to the original input and then translating.

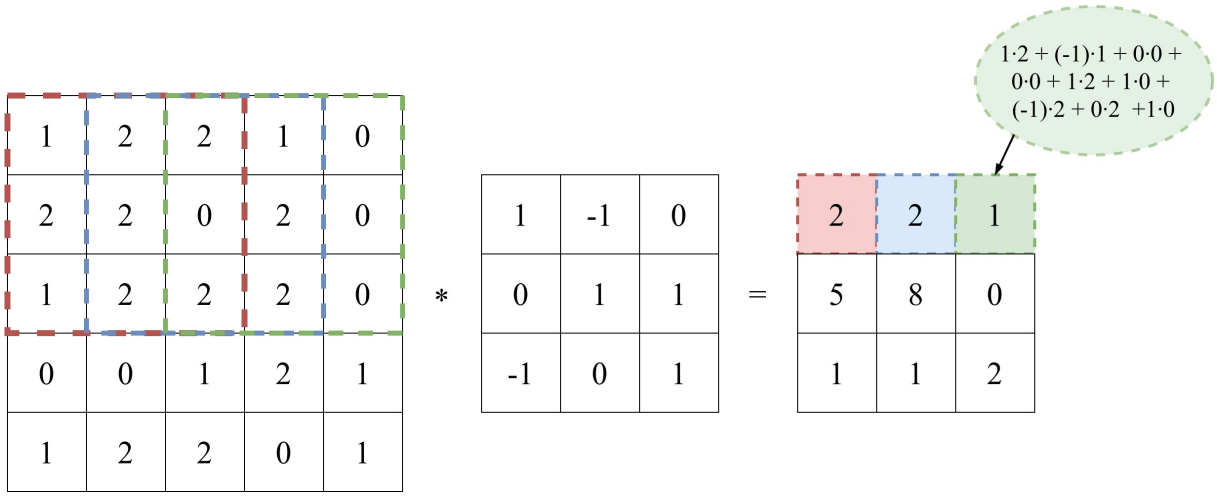


Figure 3.1: Computing the output of a convolutional layer. For illustration purposes, convolution using only one filter is shown. The filter size (width and height) is 3×3, and the stride is one. For values of stride greater than one, the size of the output is spatially reduced, e.g., for stride two, the output is spatially reduced by the factor of two in each direction.

e.g., 128, 256, or 512. The output of the convolutional layer is computed by shifting filters in both spatial directions and then computing the dot product of each filter with the input values for each location. That results in so-called activation maps that are responses of each filter for each local patch. The number of units that filters are shifted by is called stride. The output depth of the convolutional layer, i.e., the number of activation maps, matches the number of filters, and the width and height of the output are computed following

$$d_{out} = \frac{d_{in} + 2p - f}{s} + 1, \tag{3.2}$$

where d_{out} and d_{in} are output and input size, respectively, p denotes padding, f is the filter size, and s is the stride. An illustration of computing the output of the convolutional layer is shown in Figure 3.1. Modern layers types derived from the basic convolutional layer include dilated convolution [47, 48] [§], transposed convolution [49, 50, 51] [¶], depthwise separable convolutions [52, 53] [‖], etc.

The output of the convolutional layer is the linear combination of the input data and layer filters. Stacking multiple such layers yields nothing but a complex linear mapping of the input and output data. Nonlinear activation functions introduce non-linearity in CNN training and

[§]Dilated convolution is a type of convolution that includes pixel skipping, e.g., given the dilation rate of two, every second unit in the input is used for computation.

[¶]Transposed convolution is the opposite of conventional convolution, i.e., it is used to up-sample the input using learnable weights.

[‖]Depthwise separable convolution splits the conventional convolution into two steps. First, convolution is applied to each input channel using filters of depth one, resulting in the same output depth as the input depth. Second, N 1×1 convolutions are applied to expand the depth of the output to N. Such implementation results in the reduced number of trainable weights; hence, computational efficiency is increased, and the chance to overfit is decreased.

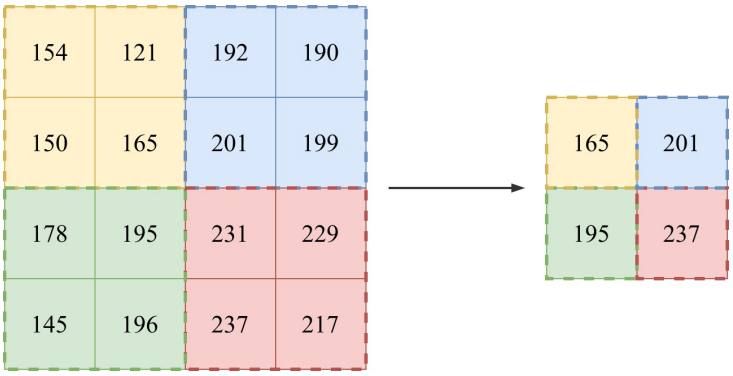


Figure 3.2: The max pooling operator. The width and the height of the pooling region are equal to three, and the stride is equal to two. Stride is the distance between two consecutive pooling locations, i.e. local patches. For demonstration purposes, units affiliated with the same local patch are equally colored.

stimulate CNNs to adapt to the variety of data. Simply, they enable the model to learn diverse features and generalize. Activation functions are applied to the output of the convolutional layer. A frequently used activation function is rectified linear unit (ReLU) $f(x) = \max(0,x)$ [54]. Other activation functions include, e.g., sigmoid function $f(x) = (1+e^{-x})^{-1}$, hyperbolic tangent $f(x) = \tanh(x)$, Leaky ReLU [55], and Scaled Exponential Linear Unit (SELU) [56].

The pooling operator makes the CNN invariant to small translations and distortions in the input [57]. Pooling operates on rectangular non-overlapping patches and outputs summary statistics for each path. A local region is replaced with a single value, and, hence, the spatial extent of the data is reduced. Consequently, pooling enhances the computational efficiency of a CNN. The parameters of the pooling operation are the size of the local patch and the distance between two consecutive pooling locations. The most frequently used is the max pooling operator that computes the maximum of a local patch [46], as shown in Figure 3.2. In addition, average and L2-norm pooling also exist.

Latest CNNs utilize a broad range of additional building blocks such as batch normalization [58], residual blocks [59], skip connections [59, 60, 61], dropout [62], and squeeze and expand blocks [63].

3.2 CNN-based illumination estimation

About two decades ago, a simple multilayer perceptron was trained for illumination estimation [64, 65]. Recently, convolutional neural network (CNN) architectures were applied on raw-RGB images to estimate illumination, motivated by the striking performance that was achieved by employing CNNs on other vision-based tasks [57, 66]. A large portion of CNN-based methods are trained for estimating the global illumination vector [67, 68, 69], i.e., they assume uniform illumination and model illumination estimation as a regression problem. Nevertheless, a number of methods utilize classification [70] and image-to-image translation [71].

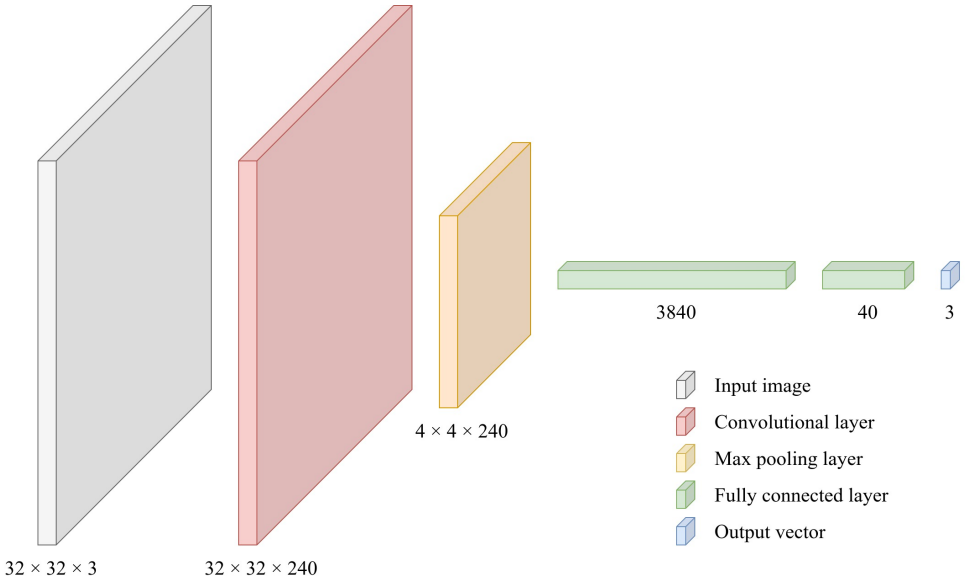


Figure 3.3: The architecture of CNN proposed by Bianco et al. [72]

.

Barron [6] proposed a method that learns a single convolutional filter for illumination estimation. They model the problem as a spatial localization task by translating images to *uv* log-chrominance space defined in (2.5) [8] and learning to localize the corresponding 2D histogram. That was further improved using the Fast Fourier transform to operate in the frequency domain [7].

Among the pioneers in CNN-based illumination estimation are Bianco et al. [72] that use extremely shallow CNN architecture, as shown in Figure 3.3. They used one convolutional layer with 240 filters of the spatial size 1×1, followed by max pooling on 8×8 patches with stride 8 (reducing the feature map eight times in each direction). That was followed by reshaping into a vector and a fully connected layer with 40 weights. The last fully connected layer outputs a three-dimensional vector representing the illumination estimate. The network was trained on pairs of image patches and ground-truth triplets using the euclidean loss function and later fine-tuned using recovery angular error as loss function.

Lou et al. [68] proposed a deep architecture inspired by AlexNet [73]. A total of eight layers was used, including five convolutional and three fully connected layers. They used transfer learning [74] to cope with insufficient dataset sizes. First, they trained their network architecture on the ImageNet dataset [75] for classification. Second, the network was retrained on ImageNet using euclidean loss function and illumination ground-truth generated using Shades of Gray [14] and Gray Edge [17] methods. Last, they fine-tuned the network on the illumination estimation dataset using the euclidean loss function.

Shi et al. [69] proposed Deep Specialized Network (DS-Net). It is a unique CNN architecture consisting of two interacting networks, namely HypNet and SelNet. Hypothesis network (HypNet) aims at estimating hypotheses of the illumination given the input image. The network

outputs two hypotheses of the illumination vector, each computed in a separate output branch forked from the final convolutional layer. A straightforward way to obtain the final illumination vector is to average the hypotheses. However, this was shown inferior to using a CNN as a voter. For that reason, a selection network (SelNet) was proposed. SelNet is trained to select the more plausible illumination vector given the input image and two hypotheses generated by HypNet.

Opposed to regression-based approaches, Oh and Kim [70] approached illumination estimation as illumination classification. They aimed at making the illumination space sparse using the K-means algorithm (and a varying number of cluster centers given the dataset). That means that similar illuminations were grouped and represented by a single central point while pushing central points as far away from each other as possible. They based their network on AlexNet architecture, pre-trained it for classification on ImageNet, and fine-tuned it to output the probability of the input image given the illumination clusters. In the test phase, the illumination vector is approximated as the weighted sum of cluster centers with the network output representing the weights in the summation.

Common to many CNN-based approaches is that they split images into patches [69, 70, 72, 76] due to small datasets and large individual images. They estimate illumination from each patch in the image and then combine estimates regarding some criterion. Typical criteria are averaging and median pooling. However, an attribute of natural images is a high correlation of neighboring pixels, and, thus ambiguous patches occur quite frequently. In the context of illumination estimation, such patches lack semantic information to distinguish reflectance and illumination; the appearance of the ambiguous patch is likely for a wide range of reflectance and illumination combinations. Hu et al. [77] tackled noisy patch-based illumination estimation by proposing a CNN that computes a map of confidence weights; a weight is assigned to each patch to indicate its reliability. They take advantage of the weight-sharing property of CNNs (the same set of weights is applied using the sliding window to all image regions at once) to consider all patches simultaneously. They proposed a fully CNN based on AlexNet and SqueezeNet [63] that computes a four-channel feature map; the first three channels correspond to local illumination estimates, and the fourth channel contains confidence weights. Each point in the four-channel feature map corresponds to a local region in the input image. The final illumination vector is the weighted sum of local estimates using computed confidence weights. They showed that regions containing, e.g., faces, bright pixels, or specular reflections are particularly reliable, which relates to other proposed approaches [16, 22, 67, 78]. Choi et al. [79] similarly approached noisy data but used the residual network and dilated convolution, thus reducing the computational cost.

CNN-based approaches to illumination estimation generally are supervised, i.e., ground-truth value is associated with each image in training data. Laakom et al. [80] proposed an

unsupervised approach. They trained Convolutional Autoencoders (CAE) to reconstruct both labeled and unlabeled data, thus providing the network with samples of images from a wider distribution of scenes and camera models. They proposed two autoencoder-based unsupervised approaches. In the first approach, a CAE was trained to reconstruct images, including labeled and unlabeled samples. Next, that CAE that learned the latent representation of a broad input distribution was fine-tuned for illumination estimation using the recovery angular error as loss function. In the second approach, they extend the binary cross-entropy loss function as

$$\mathrm{L}_{\mathrm{ext}}(D \cup D') = \alpha \frac{1}{|DD'|} \sum_{x \in D \cup D'} \mathrm{L}(x, \hat{x}) + (1-\alpha) \frac{1}{|D|} \sum_{x \in D} \frac{1}{90} \mathrm{RAE}(\mathbf{e}, \tilde{\mathbf{e}}), \tag{3.3}$$

where D and D' are labeled and unlabeled domains, α is the weighting factor, $|\cdot|$ denotes cardinality, $\mathrm{L}(x, \hat{x})$ is binary cross-entropy for input image x and reconstructed image $\hat{x}$. Finally, RAE is the recovery angular loss between the ground-truth illumination vector $\mathbf{e}$ and estimated illumination vector $\tilde{\mathbf{e}}$.Embracing such loss function enabled the training of an autoencoder that simultaneously reconstructs the input data and estimates the illumination vector for labeled samples only. Illumination estimates are obtained from the latent space in which they use only three neurons to match the size of the illumination vector.

Illumination estimation was approached from the perspective of image-to-image translation by Das et al. [71]. Contrasting to the typical approach in which the illumination vector is estimated from the image, they learn to map the input image to its white-balanced variant. They utilized Generative Adversarial Network (GAN) architectures to learn the mapping. Given the input image, the illumination vector is then obtained by inverting (2.7). They analyzed the performance of three well-known GANs: Pix2Pix [81], CycleGAN [82], and StarGAN [83]. Similarly, image-to-image translation using GANs was also used for multi-illuminant computational color constancy [84].

Learning-based methods learn regularities in data and often fail to generalize on unseen examples that do not exhibit the learned regularities. According to (2.1), illumination depends on the camera sensor sensitivity; thus, the same set of illuminations may result in different readings when captured with different cameras. Learning-based methods are sensitive to camera sensor differences and are typically trained using a single camera model. A straightforward solution is to create datasets containing a broad spectrum of camera models. Considering the tremendous number of different cameras, this is impractical. Therefore, methods such as [85, 86, 87] propose the camera-invariant approaches to illumination estimation, e.g., by learning to operate in space independent of the sensor [86].

Chapter 4

Benchmark datasets and evaluation metrics

Illumination estimation methods are usually evaluated by comparing the method's estimate of the scene illumination with some ground-truth value associated with each image. Acquisition of a dataset for illumination estimation is a tedious process since it requires a calibration object to put in the scene for extracting the ground-truth illumination information. Depending on the use-case of the dataset, the calibration object has to be appropriately positioned, e.g., to reflect the most dominant light source or two light sources originating from different directions. Later, the calibration object has to be removed from the image. For illumination estimation, images must be only linearly processed, if necessary at all. That is because illumination estimation is performed at early stages in image processing pipelines in digital cameras before non-linear operations such as tone mapping, gamma correction, or JPEG compression. However, it is necessary to deal with overexposed pixels and subtract the black level, which was neglected in some early datasets but impacts the image appearance [88].

In this chapter, benchmark datasets and the most common evaluation metrics for illumination estimation are described. Since this thesis is on global illumination estimation, the emphasis is on datasets with a single ground-truth illumination. However, for convenience, other existing datasets such as hyper-spectral and multi-illuminant datasets are mentioned as well.

4.1 Benchmark datasets

Ciurea and Funt published GrayBall [89], a large dataset of 11346 images extracted from a two-hour video sequence. It contains images of both outdoor and indoor scenes. To collect illumination information, the gray sphere was mounted to the camera to a fixed position, which ensures the sphere is visible in the camera's field of view in each frame. Since images are consecutive frames from a video sequence, the possibility of a high correlation between some

images should be taken into account. Another downside of the GrayBall is the format of images. Images were published in sRGB image format, which does not comply with the image formation model where it is assumed that no non-linear operations applied to the image. Often used simple fix to this, which only alleviates the problem, was to perform inverse gamma correction with $\gamma = 2.2$. Example images from the GrayBall dataset are shown in Figure 4.1.

Figure 4.1: Example images from the GrayBall dataset [89].

Gehler et al. published the ColorChecker dataset [33] to account for the issues in the GrayBall dataset. Nevertheless, they introduced clipped pixels, nonlinearities, and camera effects by converting RAW images to tiff format with automatic settings [90]. This was later fixed by reprocessing the RAW data appropriately [90]. However, several versions of the dataset and problems such as not subtracting the black level raised the question of the credibility of the evaluation performed on this dataset [88, 91, 92]. With all downsides put aside, the ColorChecker dataset contains 568 images. Two Canon cameras were used, i.e., Canon 1D and Canon 5D. Images captur both indoor and outdoor scenes. For the extraction of the ground-truth information, MacBeth color checker chart was used. Information was extracted from the achromatic patches in the very bottom part of the color checker. Example images from the ColorChecker dataset are shown in Figure 4.2.

Figure 4.2: Example images from the ColorChecker dataset [33]

The same color checker chart was used in the NUS dataset [93]. The novelty of this dataset is that images of the same scene were taken using cameras from various manufacturers, including multiple camera models from some manufacturers. A total of eight cameras were used, and around 200 images per camera were captured. Like in the previous datasets, images are captured in both indoor and outdoor environments. Example images from the NUS dataset are shown in Figure 4.3.

Banić et al. published the Cube and Cube+ datasets [94]. Both share the same acquisition methodology and the majority of features. Cube is entirely contained in the Cube+ and, thus, only the latter will be described. The Cube+ is made of 1707 high-resolution images, which

Figure 4.3: Example images from the NUS dataset [93].

include outdoor and indoor scenes. Moreover, outdoor scenes are captured in both day and night conditions. Outdoor daytime images were originally published in the Cube dataset. This subset contains 1365 images. The rest, which includes outdoor night and indoor scenes, was introduced in Cube+. All images are captured with the same Canon EOS 550D camera. For ground-truth extraction, the SpyderCube was used. Like the GrayBall dataset, it was mounted to the cameras and positioned to appear in the bottom right corner in each image. Example images from the Cube+ dataset are shown in Figure 4.4.

Figure 4.4: Example images from the Cube+ dataset [94].

Recently, the INTEL-TAU dataset [95] was published. The dataset was collected using Canon EOS 5DSR, Nikon D810, and Mobile Sony IMX135 cameras. The acquired number of images per camera was 2910, 2793, and 2120, respectively. INTEL-TAU contains images of outdoor and indoor scenes, and in each scene, an X-Rite ColorChecker Passport was placed for ground-truth extraction. Similar to the NUS dataset, some scenes were captured with all three cameras, enabling cross-camera color constancy. INTEL-TAU is GDPR-compliant, unlike its predecessor dataset Intel-TUT [96]. INTEL-TAU is provided in both raw format and processed format for easier usage. Example images from the INTEL-TAU dataset are shown in Figure 4.5.

Figure 4.5: Example images from the INTEL-TAU dataset [95]. For demonstration purposes, images were tone mapped using Flash tone mapping operator [97].

The described real-world datasets are typically captured in uncontrolled conditions; a scene that complies with some rules can be chosen, but usually, it can not be altered to match the

user's needs. In contrast, there exist hyperspectral datasets and datasets made in laboratory conditions [98, 99, 100, 101, 102]. Such datasets provide a user with the potential to simulate arbitrary RGB values or observe the same object under various illuminations. However, such datasets are scarce in content diversity because the acquisition is complex and time-consuming.

Other research directions in illumination estimation include temporal and multi-illuminant illumination estimation, and the corresponding datasets include [103, 104, 105] and, [106, 107, 108, 109, 110, 111, 112, 113] respectively.

4.2 Evaluation metrics

The most widely used error metric in illumination estimation is the recovery angular error [114, 115]. Due to its widespread usage, it is often referred to only as angular error. If not specified otherwise, in this thesis, recovery angular error is used to measure the method performance. This metric computes the angular distance between the vector of the actual illumination and the vector of the estimated one. It is computed as

$$\varepsilon_{\text{rec}}(\mathbf{e}, \tilde{\mathbf{e}}) = \cos^{-1}\left(\frac{\mathbf{e}\cdot\tilde{\mathbf{e}}}{\|\mathbf{e}\|\|\tilde{\mathbf{e}}\|}\right), \tag{4.1}$$

where $\mathbf{e}$ denotes the actual illumination vector, $\tilde{\mathbf{e}}$ denotes the estimated illumination vector, $\cdot$ is the scalar product, and $\|.\|$ is the Euclidean norm. Finlayson et al. [116] showed that angular error below three is an acceptable error, and below five is not noticeable to humans.

Finlayson et al. [117] proposed reproduction angular error as the angle between the RGB of the actual white surface in the scene and the RGB of the same white obtained after correcting the image with the estimated illumination:

$$\varepsilon_{\text{rep}}(\mathbf{e}, \tilde{\mathbf{e}}) = \cos^{-1}\left(\frac{(\mathbf{e}/\tilde{\mathbf{e}})\cdot\mathbf{U}}{\|(\mathbf{e}/\tilde{\mathbf{e}})\sqrt{3}\|}\right). \tag{4.2}$$

The reproduction error is motivated by observing that recovery angular error can vary, although the color-corrected images appear the same. Finlayson et al. argued that since illumination estimates are used to reproduce image colors as they appear under the white light, the evaluation should follow this use case. Therefore, this metric measures the method's capability of reproducing achromatic surfaces.

Whether recovery of reproduction angular error is used to compute the accuracy of an illumination estimation method, the performance on a benchmark dataset is computed in the same way. Once the error is computed for each image, this set of error values is analyzed with some summary statistics. Very often, the method's performance is indicated by the mean and median measures. It is recommended to opt for the median since the error distribution tends to

be skewed, and the mean performs poorly in such situations. Besides, minimum, maximum, trimean, the mean of 25% lowest errors, and the mean of 25% highest errors are reported. Barron [6] introduced average error as the geometric mean of the mean, median, trimean, and the averages of the values in the lowest and highest 25% errors.

Chapter 5

The main scientific contributions of the thesis

The main scientific contribution of this thesis is contained in the following. First, a method for the estimation of the illumination color from images based on a convolutional neural network guided with an attention mechanism, disseminated in [Pub1] and [Pub2]. Second, illumination color estimation method which uses light source classification based on deep learning, disseminated in [Pub3]. Third, a method for the estimation of the illumination color based on recurrent deep neural network with a multistage loss function, disseminated in [Pub4] and [Pub5]. Finally, a novel benchmark dataset and data acquisition methodology for illumination estimation, disseminated in [Pub6]. In the appendix, an additional research paper that extends the scope of the thesis on data acquisition is included.

5.1 Illumination color estimation method based on a convolutional neural network with an attention mechanism

For some images, it can be ambiguous to determine the color of the illumination considering only pixel intensities since different combinations of light source and surface can result in the same pixel value. An example is an image capturing very few objects and textures, thus lacking the information for determining the light source color. Moreover, if an image is dissected into smaller regions, i.e., patches, it is likely that many patches are only flat regions with no information about the illumination. An illumination estimation method may give a different estimate for each such patch. Since each patch is a part of the same image and it is assumed that illumination is uniform, this conflicts with the assumption. However, illumination estimation is an ill-posed problem, and the color of a pixel does not have to be uniquely defined. Therefore, such inconsistent behavior of an illumination estimation method is indeed expected for

monotone regions. Using additional image information, such as the location, daytime, or information that is contained in the image EXIF data, could help to direct the illumination estimation methods towards the correct illumination color. However, such information is rarely available, and illumination should be estimated from pixel values only. Another solution is to construct a mechanism for determining certainty that a region in the image is informative enough to use the illumination estimate originating from that region. In deep learning, such a mechanism is called an attention mechanism. It helps neural networks to focus more on parts in the image that are relevant for the task at hand rather than using all image data. In [Pub1] and [Pub2], convolutional neural networks with attention mechanisms were used for illumination estimation. In the context of illumination estimation enclosed by the publications, the role of an attention mechanism is to diminish the influence of monotonous regions in the image. The forward pass of the network architecture proposed in [Pub1] is as follows. First, an input image is passed through a feature extractor, which is a set of pre-trained convolutional blocks. Additional convolutional layer processes the resulting feature map and forks into two branches. One branch is the attention mechanism. It is modeled as a compound of three convolutional layers. The attention mechanism produces an individual attention map for each color channel, i.e., three attention maps for RGB images. The second branch consists only of a single convolutional block which computes the map of illumination vectors. Each vector corresponds to one point in the attention map, and they both relate to the same part of the input image. The final global illumination estimate is obtained by multiplying attention maps with the map of illumination vectors then summing and normalizing the product for each color channel. The proposed architecture is trained in five stages. For the first stage, the gradient updates are not applied on the pre-trained feature extractor to prevent initial gradients, which can mislead the training, from altering the pre-trained weights in an undesirable way. In each of the remaining stages, updates of one additional pre-trained convolutional layer were enabled to refine the model for the task of illumination estimation. The attention mechanism of the proposed illumination estimation network separates the color channels, i.e., considers them independent. That complies with the use of the diagonal matrix for the chromatic adaptation step. However, it can be argued that the human vision system does not separate color channels when looking for regions of interest in the observed scene but rather looks for some distinct features such as shape, texture, and overall appearance of the color. Motivated by that assumption in [Pub2], a modified attention mechanism was proposed (denoted as 1D-attention from hereon). 1D-attention computes only a single attention map which is shared across color channels. The computation of the global illumination estimate is performed similarly to the former method: image is passed to a pre-trained feature extractor, intermediate RGB illumination estimates and a single attention map are computed from the obtained feature map, each color channel of intermediate estimates is multiplied with 1D-attention map, the sum of the values in each channel after the multiplication is computed

and normalized. It has been shown that highly confident regions in attention maps produced by 1D-attention to a great extent coincide with the parts of an image where gradients are high. That implies that 1D-attention indeed separates informative and less informative regions in an image. Additionally, it was shown that swapping attention mechanism with image gradients, i.e., using image gradients as attention map, is inferior to using attention maps that the network is trained to compute. This further implies the following two features of the proposed network. First, it can separate content-rich from flat image regions. Second, it further selects content-rich regions based on the amount of information valuable for illumination estimation.

5.2 Illumination color estimation method using scene lighting classification based on deep learning

Colors in digital images occur from the interaction between the object material and the light source illuminating the object. By combining different materials and light sources, differently colored images are obtained, and not all light sources have the same impact on the colors. For instance, illuminations close to white light often have a minor influence on the final colors. On the other hand, the effect of artificial illuminations is often more prominent. Nevertheless, both these light sources are typical in the real world. The diversity of illuminations can nicely be observed in the Cube+ dataset. It is a dataset of real-world scenes captured in various illuminations with the color of the light source known and extracted based on the calibration object that was placed in the scene. The majority of illuminations in images are due to daylight, and a smaller image batch consists of outdoor scenes captured at night and indoor scenes, both illuminated with artificial light sources. When observed in rb-chromaticity space, ground truth illuminations in this dataset occupy few distinct groups, as shown in [Pub3]. These groups correspond to clustering images based on the illumination type (i.e., natural or artificial illumination) and scene (i.e., outdoor or indoor). Following these observations, a new classification in the following groups was proposed in [Pub3]: a) outdoor natural illumination; b) outdoor artificial illumination; c) indoor artificial illuminations. To the human eye, these types of illuminations and scenes form several distinct groups. The illumination estimation method that exploits the classification is then proposed in the same publication. The approach is two-stage and uses multiple convolutional neural networks. For the first stage, one convolutional neural network is trained to classify input images regarding the proposed classification system. For the second stage, three instances of the convolutional neural network from [Pub2] are trained for illumination estimation. Each one is trained on a single class of images in the proposed three-class classification system. Based on the outcome in the classification stage, in the second stage, the network corresponding to the predicted class is used to estimate the illumination. Classification of images into groups with distinct properties enables the training of estimators

specialized for specific image types, which is more accurate than using a single estimator for a broad distribution of scenes and illumination colors. That was manifested in the reduction of the maximum error by over 30% when using the proposed approach compared to the single estimator case. Using multiple estimators avoids the problem of the uneven number of images between classes since each estimator is trained only on a single image class. Although the training depends on the number of available images, the unbalanced dataset can not be the reason for an erroneous estimator. However, the problem remains in the first step, the classification of input images. In the proposed approach, it was alleviated using standard procedures such as oversampling and undersampling. It was shown that the proposed classification is more suitable than the classification concerning only scene type (indoor or outdoor scene classification) and concerning only the illumination type (natural or artificial light source classification). However, it was also shown as the method's weak spot since estimator accuracy highly depends on the classification outcome. When misclassification occurs, the estimator is forced to process an image from the distribution different than the one used in the training process. Nevertheless, estimators still tend to produce outputs as close as possible to the region of illuminations to which the correct classification would point, with accuracy subject to the proximity of the predicted and target class. Therefore, it can be concluded that it is crucial to develop a reliable classifier or a mechanism to detect misclassification and guide the estimation step accordingly for further improvement of the method's accuracy.

5.3 Illumination color estimation method based on recurrent deep neural network with a multistage loss function

Both steps necessary for achieving computational color constancy have their flaws. Illumination estimation is an ill-posed problem. Many methods try to solve it but often fail due to the assumptions they use. For chromatic adaptation, a simple multiplication with the diagonal matrix is used. Although sufficient for satisfactory results, it is still an approximation of a much more complex problem. [Pub4] analyzed the limits of the cameras' built-in white balancing. It was shown that camera manufacturers limit the chromaticity space. For some Canon cameras, it was shown that the camera's illumination estimates get clipped inside the bounds of a regular polygon. This way, camera manufacturers avoid correcting image colors for some undesired illuminations defined by the bounds of the polygon. These illuminations are usually far from the neutral white light and may spoil image colors to the extent that simple multiplication with the diagonal matrix can not remove their influence. Consequently, high errors likely to appear for such illuminations are reduced.

In [Pub5], a novel approach to global illumination estimation is proposed. It is motivated by the limitations to the cameras' chromaticity space and aims to reduce maximum errors.

The premise of the method is that the highest estimation errors occur due to highly colored light sources, and the current color correction model is inadequate for such illuminations types but works well for illuminations closer to white light. It is also assumed that illumination estimation is less prone to errors when near-white light sources are estimated. Therefore, it is proposed to decompose the process of illumination estimation into many multiplicative steps. Instead of estimating the illumination vector directly from the image, it is obtained as a series of intermediate vectors. Each vector is constrained to a subset of illuminations in the vicinity of the white light. In each step, one intermediate vector is estimated from the input image corresponding to that step. Then the input image is corrected using the estimated vector and passed as the input in the next step. In the end, all intermediate vectors are multiplied channel-vise to acquire the final estimation corresponding to the total illumination associated with the original input image. The method was constructed to gradually build the illumination vector and correct the image so that, in the perfect scenario, in the last step, the intermediate vector would be equal to the vector of the white light, and the image would appear as it was captured under the white illumination. That was all embedded in a convolutional neural network. A custom loss function was built to ensure gradual convergence toward the correct scene illumination. In the first part of the loss function, the error of the final estimation is computed. It is the angle between the ground-truth record and the final estimation, which is the product of all intermediate vectors. Intermediate vectors are forced to be near white light with the second part of the loss function. For each intermediate vector, the angle from the vector representing the white light is computed. All computed values are fused using the weighted sum. Each weight controls how far intermediate estimates can be from the white light in terms of the angle. The highest distance is allowed to the first intermediate estimation and the smallest to the last one. Total loss is the sum of both parts. The proposed iterative approach was shown very reliable for illumination estimation. The mean of the worst 25% errors on the test set was 3.20 degree. That is less when compared to other methods that are evaluated on the same dataset. Thereby the goals set to achieve are met. Extensive experimental evaluation was performed to validate that the intermediate estimations form trajectories toward the white light and progress toward the scene illumination when combined.

5.4 The Cube++ Illumination Estimation Dataset

Deep learning methods rely on the availability of large amounts of data to learn important features from diverse perspectives given the task. Large datasets are not usual for illumination estimation due to the tedious and time-consuming acquisition process. However, with the rise of deep learning-based methods, they became a must. In [Pub6], Cube++, a novel dataset designed for illumination estimation, is proposed. It contains 4890 images associated with illumination

information and a variety of additional semantic data. Cube++ was designed to comply with the following set of properties: diversity in content and illuminations, a large number of samples, rich in various image information, easy to update and track changes, verifiable, easily accessible, and GDPR compliant. Many of the listed properties were motivated by the errors reported by the research community for previous illumination estimation datasets. Each image in the Cube++ dataset is supplemented with various information. For each image sample, four vectors that describe the scene illumination were given. By doing so, the dataset is not fixed only for single-illuminant estimation. Nevertheless, a procedure for obtaining a single ground truth is given as well. Other image information include EXIF data, time of the day when an image was captured, type of the illumination, scene information (indoor or outdoor scene, how rich the scene content is, are there any shadows or light sources in the scene), and information about the calibration object. One issue with the previous datasets was the lack of information about the dataset usage. Therefore, all the steps necessary to properly use the Cube++ dataset were listed and described. These include black level subtraction, saturation removal, and color target masking. Aside from the dataset itself, the methodology for collecting such a dataset was given too. The technical setup was described and verified. Ground-truth extraction, data filtration, and peculiarities in data collection were described. Finally, a foundation for an online benchmark for illumination estimation was set. The field of illumination estimation would benefit from such a benchmark as it could remove existing problems and provide a reliable source of information.

Chapter 6

Conclusions and future directions

6.1 Main conclusions of the thesis

Illumination estimation is a fundamental but very challenging task in digital photography. It is an essential part of image processing pipelines in digital cameras, enabling them to capture and display of the actual color of an object in digital images regardless of the color of the light source illuminating the scene. Unfortunately, it is rather hard to perform illumination estimation from image pixels without additional knowledge about the captured scene since distinguishing illumination from surface reflectance is an ill-posed problem. One possible way to overcome this issue is to simplify the problem by posing assumptions about the solution. The main focus of this thesis is on analyzing different assumptions for illumination estimation. Three illumination estimation methods were proposed. Each of the methods was based on an assumption about the illumination estimation task. The proposed methods were implemented utilizing deep learning techniques with an emphasis on convolutional neural networks.

The motivation for the first part of the thesis is that color is a distinguishing characteristic of many real-world objects. This information is convenient for illumination estimation since it introduces prior knowledge about the scene. It can be expected what color object should have once the illumination color cast is corrected. Motivated by this, the author proposed a convolutional neural network architecture with an attention mechanism for illumination estimation. The attention mechanism was used to find the parts in the image where some distinct features about the illumination color exist. The design of the attention mechanism does not restrict the network to finding only objects of some characteristic color but lets the network learn on its own what regions to analyze. Two attention mechanisms were proposed. The first type of mechanism separates color channels and computes an individual attention map for each channel. The second type of attention mechanism produces a single attention map which is shared across all color channels. The latter attention mechanism is more interpretative and comparable. It was shown that such attention mechanism relates to image gradient, which is a well-known term in image

processing. It was shown that the attention mechanism focuses on the parts of the image where the energy of the gradient is high, i.e., it focuses on image regions that are not monotonous.

In the second part of the thesis, the clustering-based illumination estimation method was proposed. The proposed technique relies on clustering input images into three classes, i.e., indoor scenes in artificial illumination, outdoor scenes in artificial illumination, and outdoor scenes in natural illumination. It was shown that such clustering creates three distinct groups of images and reduces the solution space of an illumination estimation method to a well-defined subspace. Therefore, clustering introduces prior knowledge about illumination. Based on this knowledge, it was proposed to select the appropriate illumination estimator, i.e., the one trained for the corresponding image class. Since the class information guides the estimation process, a reduced presence of significantly incorrect estimations is expected. That was shown true in the experimental results, where the worst-case scenarios were reduced by over 30%.

The third part of the thesis elaborates on an iterative illumination estimation procedure. It was shown that illumination estimation could be formulated as a group of sub-tasks where the goal of each sub-task is to estimate a fraction of the global illumination in the scene. When estimates of all sub-tasks are accumulated, the global illumination vector is obtained. The goal of the proposed estimation method was to avoid high estimation errors. Therefore, sub-tasks were designed to operate in the vicinity of white light. Experimentally it was shown that the proposed method indeed gradually estimates the global illumination in the scene. That was possible because chromatic adaptation is performed as multiplication with diagonal matrix, i.e., inter-channel connections are ignored. Therefore, the target diagonal matrix can be modeled as multiplication of many diagonal matrices; thus, enabling gradual illumination estimation from differently colored variants of the input image.

In the last part of the thesis, a new dataset for illumination estimation was presented. The current trend in the overall computer vision research is on deep learning. However, publicly available data for illumination estimation lags behind the explosion of deep learning methods, as was the case in the methods proposed in this thesis. Researchers try to solve the problem of the lack of data by extending the existing dataset using techniques such as splitting images into patches, augmenting data, creating synthetic datasets, and by using various training procedures. Nevertheless, this part of the thesis shows that using existing datasets is limited by their flaws. The guidelines to build a good quality illumination estimation dataset were specified, and a novel dataset that complies with the guidelines is presented. The major drawback of collecting such a dataset is that the process is time-consuming since many requirements need to be satisfied. However, creating such datasets is crucial for the development of learning-based methods. Methods based on learning from data achieve state-of-the-art in many computed vision tasks, including illumination estimation. Therefore, data of high quality and high quantity is a must. Another feature of this dataset, which makes it different from others, is an abundance

of information for each image. Moreover, users of the dataset are not forced to estimate a single illumination vector but are given a set of possible solutions. The dataset was designed in such a way to offer users as much data and freedom as possible for comprehending the ill-posed nature of the illumination estimation task.

6.2 Future directions

The proposed neural network architectures are a proof of concept for the assumptions about the illumination estimation described in this thesis, and each can be further improved. For this research, convolutional neural architectures were mainly based on a feature extractor pre-trained for a tasks such as object classification and object detection. These feature extractors were then supplemented with other existing deep learning structures to suit the corresponding assumption and trained for illumination estimation. However, such architectures are not necessarily optimal for the given task in terms of memory and execution time. Therefore, improvement in reducing the number of architecture parameters and reducing the execution time is a must. Proposed illumination estimation methods in the current form require high-power graphical processing units to run in real-time. They are not suitable for implementation in nowadays very computationally powerful smartphones, let alone in much more modest digital cameras. All proposed methods are complex concerning the number of parameters. For example, the illumination estimation method based on light source classification proposed in the second part of the thesis relies on four convolutional neural networks. One of the networks is used for image classification, and three other networks are instances of the same architecture trained on different image clusters. One may question if three instances of the same complex architecture are necessary. Answering that and similar questions is intended for future work.

Aside from method optimization, the focus of future research is the following. Attention-based illumination estimation methods proposed in the first part of the thesis compute the attention map and intermediate illumination estimates from the whole image. These methods take into account all image data, and that may include information irrelevant for illumination estimation. As opposed to that, an attention mechanism could filter for image regions rich with illumination information first and then perform the illumination estimation only from those regions while completely ignoring other image parts. In other words, a future direction includes inverting the order of operations to first remove all misleading information from the input image and estimate illumination only from a subset of image pixels which should lead to accuracy improvement. The iterative illumination estimation method proposed in the third part of the thesis is trained for the fixed number of iterations. However, the proposed seven iterations are not optimal for all cases. For example, experimental results have shown that highly colored images would benefit from more iterations and the opposite for slightly colored images. Therefore, a

mechanism to determine at which iteration the method should stop processing the input image is to be developed in the future. Such a mechanism should operate on a per-image basis.

Lastly, a significant future direction is regarding the image data. Due to more extensive research of deep learning techniques in the scope of illumination estimation, existing datasets are becoming too simple and of a small scale. Since deep learning-based methods are state-of-the-art in this and many other computer vision tasks, constant improvement of the data is necessary. In the scope of this thesis, a very detailed description of the desirable properties of an illumination estimation dataset and technical methodology for collecting such a dataset are given. In the future, this will be used for collecting additional image data that should set the ground for methods with better generalization properties.

Chapter 7

List of publications

Pub 1 **Koščević, K.**, Subašić, M., Lončarić, S., "Attention-based Convolutional Neural Network for Computer Vision Color Constancy", Proceedings of the 11th International Symposium on Image and Signal Processing and Analysis, Dubrovnik, Croatia, 2019, pp. 372-377.

Pub 2 **Koščević, K.**, Subašić, M., Lončarić, S., "Guiding the Illumination Estimation Using the Attention Mechanism", Proceedings of the 2020 2nd Asia Pacific Information Technology Conference, Bali, Indonesia, 2020, pp. 143-149.

Pub 3 **Koščević, K.**, Subašić, M., Lončarić, S., "Deep Learning-Based Illumination Estimation Using Light Source Classification", IEEE Access, Vol. 8, 2020, pp. 84239-84247.

Pub 4 **Koščević, K.**, Banić, N., Lončarić, S., "Color Beaver: Bounding Illumination Estimations for Higher Accuracy", Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 2019, pp. 183-190.

Pub 5 **Koščević, K.**, Subašić, M., Lončarić, S., "Iterative Convolutional Neural Network-Based Illumination Estimation", IEEE Access, Vol. 9, 2021, pp. 26755-26765.

Pub 6 Ershov, E., Savchik, A., Semenkov, I., Banić, N., Belokopytov, A., Senshina, D., **Koščević, K.**, Subašić, M., Lončarić, S.,"The Cube++ Illumination Estimation Dataset", IEEE Access, Vol. 8, 2020, pp. 227511-227527.

Chapter 8

Author's contribution to the publications

The results presented in this thesis are based on the research carried out during the period of 2018-2021 at the University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, HD-10000 Zagreb, Croatia, as a part of the research projects IP-06-2016-2092 "Research Project - Methods and algorithms for real-time color image enhancement (PefectColor)" and DOK-01-2018 "Young Researchers' Career Development Project - Training of New Doctoral Students - Methods and algorithms for real-time color image enhancement" which were financially supported by the Croatian Science Foundation.

The thesis includes six publications written in collaboration with several coauthors. The author's contribution to each paper consists of text writing, software implementation, performing the required experiments, and the analysis and presentation of the results.

[Pub1] In the paper **"Attention-based Convolutional Neural Network for Computer Vision Color Constancy"**, the author proposed a convolutional neural network architecture with an attention mechanism for illumination estimation. It was motivated by the assumption that some image regions are ambiguous for illumination estimation. An attention mechanism was used to assign weights to estimates from different image regions based on regions' significance for illumination estimation. A separate attention map was computed for each color channel. That follows the use of the diagonal matrix for the chromatic adaptation, i.e., considers color channels independent. The author also proposed a training scheme that does not train the whole network at once but gradually enables the training of more network blocks. The author implemented and experimentally tested the newly proposed illumination estimation approach in the Keras framework.

[Pub2] In the paper **"Guiding the Illumination Estimation Using the Attention Mechanism"**, the author proposed an adaptation of the attention mechanism in [Pub1]. A new version of the attention mechanism computes only a single attention map. This attention map is shared across all color channels. The assumption used for such an attention mechanism was that distinguishing individual color channel values is not how the human visual system looks for regions

of interest in the scene. It is rather beneficial to look for objects, structures, textures, i.e., salient image regions with respect to the scene content. The experimental results confirm that the neural network indeed operates in such a fashion. It was shown that the attention mechanism focuses on regions where image gradients are prominent, i.e., regions where some content exists, and filters such image regions to get the most favorable selection. All of the code and experiments were implemented by the author using the Keras framework.

[Pub3] In the paper **"Deep Learning-Based Illumination Estimation Using Light Source Classification"**, the author proposed to use light source classification to enhance the results of illumination estimation. The following classes were proposed: indoor scene images under artificial illuminations; outdoor scene images under natural illuminations; outdoor scene images under artificial illuminations. The author proposed a method that classifies input images and uses a class-specific illumination estimation network. The same convolutional neural network architecture was used as an estimator for all classes, but a separate instance was trained for each image class. The author performed all analyses and implemented the classification and illumination estimation networks in the Keras framework.

[Pub4] In the paper **"Color Beaver: Bounding Illumination Estimations for Higher Accuracy"**, the author and the second coauthor performed an experimental analysis of illumination estimates obtained by Canon cameras' auto white balancing. The authors observed that cameras estimates are limited to a small region of chromaticity space by a bounding polygon. Based on that finding, it was proposed to look for a more suitable polygon that could be applied to any existing illumination estimation method. The author implemented a genetic algorithm for the search of the most suitable bounding polygon. The author has written the implementation and experimental tests in the Matlab programming language.

[Pub5] In the paper **"Iterative Convolutional Neural Network-Based Illumination Estimation"**, the author proposed an iterative scheme for illumination estimation. The author trained a neural network to sequentially compute multiple illumination estimates from the input image and its chromatically adapted variants. Channel-vise multiplication of estimated illuminations equals the estimated global scene illumination. The goal of the proposed scheme was to enhance the network performance on the worst-performing samples, which was shown to be achieved by the experimental results. The author implemented the proposed iterative technique as a convolutional neural network in the Keras framework and performed experimental validation.

[Pub6] In the paper **"The Cube++ Illumination Estimation Dataset"**, the author participated in the creation of a new dataset for illumination estimation. The goal of publishing the dataset was to alleviate issues with existing datasets in the research field. A detailed methodology, technical details, issues, and best practices for collecting a dataset for illumination estimation were given. The author took part in defining those criteria. The dataset was collected

in many different countries. The author also contributed by collecting images in Croatia, and post-processing and selecting images overall.

Bibliography

[1] Gevers, T., Smeulders, A. W., "Color-based object recognition", Pattern recognition, Vol. 32, No. 3, 1999, str. 453–464.

[2] Afifi, M., Brown, M. S., "What else can fool deep learning? addressing color constancy errors on deep neural network performance", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, str. 243–252.

[3] Gijsenij, A., Gevers, T., Van De Weijer, J., "Computational color constancy: Survey and experiments", IEEE transactions on image processing, Vol. 20, No. 9, 2011, str. 2475–2489.

[4] Shafer, S. A., "Using color to separate reflection components", Color Research & Application, Vol. 10, No. 4, 1985, str. 210–218.

[5] Basri, R., Jacobs, D. W., "Lambertian reflectance and linear subspaces", IEEE transactions on pattern analysis and machine intelligence, Vol. 25, No. 2, 2003, str. 218–233.

[6] Barron, J. T., "Convolutional color constancy", in Proceedings of the IEEE International Conference on Computer Vision, 2015, str. 379–387.

[7] Barron, J. T., Tsai, Y.-T., "Fast fourier color constancy", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, str. 886–894.

[8] Finlayson, G. D., Hordley, S. D., "Color constancy at a pixel", JOSA A, Vol. 18, No. 2, 2001, str. 253–264.

[9] Eibenberger, E., Angelopoulou, E., "The importance of the normalizing channel in log-chromaticity space", in 2012 19th IEEE International Conference on Image Processing. IEEE, 2012, str. 825–828.

[10] Kries, J., "Die gesichtsempfindungen", Nagel's Handbuch der Physiologie des Menschen, Vol. 3, 1905, str. 109.

[11] Fairchild, M. D., Color appearance models. John Wiley & Sons, 2013.

[12] Zakizadeh, R., Brown, M. S., Finlayson, G. D., “A hybrid strategy for illuminant estimation targeting hard images”, in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, str. 16–23.

[13] Buchsbaum, G., “A spatial processor model for object colour perception”, Journal of the Franklin institute, Vol. 310, No. 1, 1980, str. 1–26.

[14] Finlayson, G. D., Trezzi, E., “Shades of gray and colour constancy”, in Color and Imaging Conference, Vol. 2004, No. 1. Society for Imaging Science and Technology, 2004, str. 37–41.

[15] Brainard, D. H., Wandell, B. A., “Analysis of the retinex theory of color vision”, JOSA A, Vol. 3, No. 10, 1986, str. 1651–1661.

[16] Joze, H. R. V., Drew, M. S., Finlayson, G. D., Rey, P. A. T., “The role of bright pixels in illumination estimation”, in Color and Imaging Conference, Vol. 2012, No. 1. Society for Imaging Science and Technology, 2012, str. 41–46.

[17] Van De Weijer, J., Gevers, T., Gijsenij, A., “Edge-based color constancy”, IEEE Transactions on image processing, Vol. 16, No. 9, 2007, str. 2207–2214.

[18] Gijsenij, A., Gevers, T., Van De Weijer, J., “Physics-based edge evaluation for improved color constancy”, in 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, str. 581–588.

[19] Qian, Y., Kamarainen, J.-K., Nikkanen, J., Matas, J., “On finding gray pixels”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, str. 8062–8070.

[20] Lee, H.-C., “Method for computing the scene-illuminant chromaticity from specular highlights”, JOSA A, Vol. 3, No. 10, 1986, str. 1694–1699.

[21] Tominaga, S., Wandell, B. A., “Standard surface-reflectance model and illuminant estimation”, JOSA A, Vol. 6, No. 4, 1989, str. 576–584.

[22] Healey, G., “Estimating spectral reflectance using highlights”, Image and vision computing, Vol. 9, No. 5, 1991, str. 333–337.

[23] Tan, R. T., Nishino, K., Ikeuchi, K., “Color constancy through inverse-intensity chromaticity space”, JOSA A, Vol. 21, No. 3, 2004, str. 321–334.

[24] Woo, S.-M., Lee, S.-H., Yoo, J.-S., Kim, J.-O., “Improving color constancy in an ambient light environment using the phong reflection model”, IEEE Transactions on Image Processing, Vol. 27, No. 4, 2017, str. 1862–1877.

[25] Phong, B. T., “Illumination for computer generated pictures”, Communications of the ACM, Vol. 18, No. 6, 1975, str. 311–317.

[26] Forsyth, D. A., “A novel algorithm for color constancy”, International Journal of Computer Vision, Vol. 5, No. 1, 1990, str. 5–35.

[27] Barnard, K., “Improvements to gamut mapping colour constancy algorithms”, in European conference on computer vision. Springer, 2000, str. 390–403.

[28] Finlayson, G. D., “Color in perspective”, IEEE transactions on Pattern analysis and Machine Intelligence, Vol. 18, No. 10, 1996, str. 1034–1038.

[29] Finlayson, G., Hordley, S., “Improving gamut mapping color constancy”, IEEE Transactions on Image Processing, Vol. 9, No. 10, 2000, str. 1774–1783.

[30] Finlayson, G. D., Hordley, S. D., Tastl, I., “Gamut constrained illuminant estimation”, International journal of computer vision, Vol. 67, No. 1, 2006, str. 93–109.

[31] Gijsenij, A., Gevers, T., Van De Weijer, J., “Generalized gamut mapping using image derivative structures for color constancy”, International Journal of Computer Vision, Vol. 86, No. 2, 2010, str. 127–139.

[32] Brainard, D. H., Freeman, W. T., “Bayesian color constancy”, JOSA A, Vol. 14, No. 7, 1997, str. 1393–1411.

[33] Gehler, P. V., Rother, C., Blake, A., Minka, T., Sharp, T., “Bayesian color constancy revisited”, in 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008, str. 1–8.

[34] Rosenberg, C., Ladsariya, A., Minka, T., “Bayesian color constancy with non-gaussian models”, Advances in neural information processing systems, Vol. 16, 2003, str. 1595–1602.

[35] Sapiro, G., “Color and illuminant voting”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, 1999, str. 1210–1215.

[36] Cardei, V. C., Funt, B., “Committee-based color constancy”, in Color and Imaging Conference, Vol. 1999, No. 1. Society for Imaging Science and Technology, 1999, str. 311–313.

[37] Schaefer, G., Hordley, S., Finlayson, G., “A combined physical and statistical approach to colour constancy”, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), Vol. 1. IEEE, 2005, str. 148–153.

[38] Gijsenij, A., Gevers, T., "Color constancy using natural image statistics", in 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007, str. 1–8.

[39] Gijsenij, A., Gevers, T., "Color constancy using natural image statistics and scene semantics", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, 2010, str. 687–698.

[40] Banić, N., Lončarić, S., "Color Cat: Remembering Colors for Illumination Estimation", Signal Processing Letters, IEEE, Vol. 22, No. 6, 2015, str. 651–655.

[41] Agarwal, V., Gribok, A. V., Koschan, A., Abidi, M. A., "Estimating illumination chromaticity via kernel regression", in 2006 International Conference on Image Processing. IEEE, 2006, str. 981–984.

[42] Agarwal, V., Gribok, A. V., Abidi, M. A., "Machine learning approach to color constancy", Neural Networks, Vol. 20, No. 5, 2007, str. 559–563.

[43] Agarwal, V., Gribok, A., Koschan, A., Abidi, B., Abidi, M., "Illumination chromaticity estimation using linear learning methods", Journal of Pattern Recognition Research, Vol. 4, No. 1, 2009, str. 92–109.

[44] Funt, B., Xiong, W., "Estimating illumination chromaticity via support vector regression", in Color and Imaging Conference, Vol. 2004, No. 1. Society for Imaging Science and Technology, 2004, str. 47–52.

[45] Wang, N., Xu, D., Li, B., "Edge-based color constancy via support vector regression", IEICE transactions on information and systems, Vol. 92, No. 11, 2009, str. 2279–2282.

[46] Goodfellow, I., Bengio, Y., Courville, A., Deep Learning. MIT Press, 2016, http://www.deeplearningbook.org.

[47] Yu, F., Koltun, V., "Multi-scale context aggregation by dilated convolutions", arXiv preprint arXiv:1511.07122, 2015.

[48] Yu, F., Koltun, V., "Multi-scale context aggregation by dilated convolutions", in 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, Bengio, Y., LeCun, Y., (ur.), 2016, dostupno na: http://arxiv.org/abs/1511.07122

[49] Zeiler, M. D., Krishnan, D., Taylor, G. W., Fergus, R., "Deconvolutional networks", in 2010 IEEE Computer Society Conference on computer vision and pattern recognition. IEEE, 2010, str. 2528–2535.

[50] Zeiler, M. D., Fergus, R., "Visualizing and understanding convolutional networks", in European conference on computer vision. Springer, 2014, str. 818–833.

[51] Dumoulin, V., Visin, F., "A guide to convolution arithmetic for deep learning", arXiv preprint arXiv:1603.07285, 2016.

[52] Chollet, F., "Xception: Deep learning with depthwise separable convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, str. 1251–1258.

[53] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., "Mobilenets: Efficient convolutional neural networks for mobile vision applications", arXiv preprint arXiv:1704.04861, 2017.

[54] Nair, V., Hinton, G. E., "Rectified linear units improve restricted boltzmann machines", in Icml, 2010.

[55] Maas, A. L., Hannun, A. Y., Ng, A. Y. *et al.*, "Rectifier nonlinearities improve neural network acoustic models", in Proc. icml, Vol. 30, No. 1. Citeseer, 2013, str. 3.

[56] Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S., "Self-normalizing neural networks", in Proceedings of the 31st international conference on neural information processing systems, 2017, str. 972–981.

[57] LeCun, Y., Bengio, Y., Hinton, G., "Deep learning", nature, Vol. 521, No. 7553, 2015, str. 436–444.

[58] Ioffe, S., Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift", in International conference on machine learning. PMLR, 2015, str. 448–456.

[59] He, K., Zhang, X., Ren, S., Sun, J., "Deep residual learning for image recognition", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, str. 770–778.

[60] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., "Densely connected convolutional networks", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, str. 4700–4708.

[61] Ronneberger, O., Fischer, P., Brox, T., "U-net: Convolutional networks for biomedical image segmentation", in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, str. 234–241.

[62] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., "Dropout: a simple way to prevent neural networks from overfitting", The journal of machine learning research, Vol. 15, No. 1, 2014, str. 1929–1958.

[63] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., Keutzer, K., "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size", arXiv preprint arXiv:1602.07360, 2016.

[64] Cardei, V. C., Funt, B., Barnard, K., "Estimating the scene illumination chromaticity by using a neural network", JOSA a, Vol. 19, No. 12, 2002, str. 2374–2386.

[65] Funt, B., Cardei, V., Barnard, K., "Learning color constancy", in Color and Imaging Conference, Vol. 1996, No. 1. Society for Imaging Science and Technology, 1996, str. 58–60.

[66] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M. S., "Deep learning for visual understanding: A review", Neurocomputing, Vol. 187, 2016, str. 27–48.

[67] Bianco, S., Schettini, R., "Color constancy using faces", in 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, str. 65–72.

[68] Lou, Z., Gevers, T., Hu, N., Lucassen, M. P. *et al.*, "Color constancy by deep learning.", in BMVC, 2015, str. 76–1.

[69] Shi, W., Loy, C. C., Tang, X., "Deep specialized network for illuminant estimation", in European conference on computer vision. Springer, 2016, str. 371–387.

[70] Oh, S. W., Kim, S. J., "Approaching the computational color constancy as a classification problem through deep learning", Pattern Recognition, Vol. 61, 2017, str. 405–416.

[71] Das, P., Baslamisli, A. S., Liu, Y., Karaoglu, S., Gevers, T., "Color constancy by gans: An experimental survey", arXiv preprint arXiv:1812.03085, 2018.

[72] Bianco, S., Cusano, C., Schettini, R., "Color constancy using cnns", in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, str. 81–89.

[73] Krizhevsky, A., Sutskever, I., Hinton, G. E., "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, Vol. 25, 2012, str. 1097–1105.

[74] Pan, S. J., Yang, Q., "A survey on transfer learning", IEEE Transactions on knowledge and data engineering, Vol. 22, No. 10, 2009, str. 1345–1359.

[75] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., "Imagenet: A large-scale hierarchical image database", in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, str. 248–255.

[76] Bianco, S., Cusano, C., Schettini, R., "Single and multiple illuminant estimation using convolutional neural networks", IEEE Transactions on Image Processing, Vol. 26, No. 9, 2017, str. 4347–4362.

[77] Hu, Y., Wang, B., Lin, S., "Fc4: Fully convolutional color constancy with confidence-weighted pooling", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, str. 4085–4094.

[78] Banić, N., Lončarić, S., "Improving the white patch method by subsampling", in 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014, str. 605–609.

[79] Choi, H.-H., Kang, H.-S., Yun, B.-J., "Cnn-based illumination estimation with semantic information", Applied Sciences, Vol. 10, No. 14, 2020, str. 4806.

[80] Laakom, F., Raitoharju, J., Iosifidis, A., Nikkanen, J., Gabbouj, M., "Color constancy convolutional autoencoder", in 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2019, str. 1085–1090.

[81] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., "Image-to-image translation with conditional adversarial networks", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, str. 1125–1134.

[82] Zhu, J.-Y., Park, T., Isola, P., Efros, A. A., "Unpaired image-to-image translation using cycle-consistent adversarial networks", in Proceedings of the IEEE international conference on computer vision, 2017, str. 2223–2232.

[83] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., Choo, J., "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, str. 8789–8797.

[84] Das, P., Liu, Y., Karaoglu, S., Gevers, T., "Generative models for multi-illumination color constancy", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, str. 1194–1203.

[85] Afifi, M., Barron, J. T., LeGendre, C., Tsai, Y.-T., Bleibel, F., "Cross-camera convolutional color constancy", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, str. 1981–1990.

[86] Afifi, M., Brown, M. S., “Sensor-independent illumination estimation for dnn models”, arXiv preprint arXiv:1912.06888, 2019.

[87] Gao, S.-B., Zhang, M., Li, C.-Y., Li, Y.-J., “Improving color constancy by discounting the variation of camera spectral sensitivity”, JOSA A, Vol. 34, No. 8, 2017, str. 1448–1462.

[88] Banić, N., Koščević, K., Subašić, M., Lončarić, S., “The past and the present of the color checker dataset misuse”, in 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE, 2019, str. 366–371.

[89] Ciurea, F., Funt, B., “A large image database for color constancy research”, in Color and Imaging Conference, Vol. 2003, No. 1. Society for Imaging Science and Technology, 2003, str. 160–164.

[90] Shi, L., “Re-processed version of the gehler color constancy dataset of 568 images”, http://www. cs. sfu. ca/˜ color/data/, 2000.

[91] Finlayson, G. D., Hemrit, G., Gijsenij, A., Gehler, P., “A curious problem with using the colour checker dataset for illuminant estimation”, in Color and Imaging Conference, Vol. 2017, No. 25. Society for Imaging Science and Technology, 2017, str. 64–69.

[92] Hemrit, G., Finlayson, G. D., Gijsenij, A., Gehler, P., Bianco, S., Funt, B., Drew, M., Shi, L., “Rehabilitating the colorchecker dataset for illuminant estimation”, in Color and Imaging Conference, Vol. 2018, No. 1. Society for Imaging Science and Technology, 2018, str. 350–353.

[93] Cheng, D., Prasad, D. K., Brown, M. S., “Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution”, JOSA A, Vol. 31, No. 5, 2014, str. 1049–1058.

[94] Banić, N., Koščević, K., Lončarić, S., “Unsupervised learning for color constancy”, arXiv preprint arXiv:1712.00436, 2017.

[95] Laakom, F., Raitoharju, J., Nikkanen, J., Iosifidis, A., Gabbouj, M., “Intel-tau: A color constancy dataset”, IEEE Access, Vol. 9, 2021, str. 39 560–39 567.

[96] Aytekin, Ç., Nikkanen, J., Gabbouj, M., “A data set for camera-independent color constancy”, IEEE Transactions on Image Processing, Vol. 27, No. 2, 2017, str. 530–544.

[97] Banić, N., Lončarić, S., “Puma: A high-quality retinex-based tone mapping operator”, in 2016 24th European Signal Processing Conference (EUSIPCO). IEEE, 2016, str. 943–947.

[98] Barnard, K., Martin, L., Funt, B., Coath, A., "A data set for color research", Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, Vol. 27, No. 3, 2002, str. 147–151.

[99] Nascimento, S. M., Ferreira, F. P., Foster, D. H., "Statistics of spatial cone-excitation ratios in natural scenes", JOSA A, Vol. 19, No. 8, 2002, str. 1484–1490.

[100] Foster, D. H., Amano, K., Nascimento, S. M., Foster, M. J., "Frequency of metamerism in natural scenes", Josa a, Vol. 23, No. 10, 2006, str. 2359–2372.

[101] Rizzi, A., Bonanomi, C., Gadia, D., Riopi, G., "Yaccd2: yet another color constancy database updated", in Color Imaging XVIII: Displaying, Processing, Hardcopy, and Applications, Vol. 8652. International Society for Optics and Photonics, 2013, str. 86520A.

[102] Skauli, T., Farrell, J., "A collection of hyperspectral images for imaging systems research", in Digital Photography IX, Vol. 8660. International Society for Optics and Photonics, 2013, str. 86600C.

[103] Qian, Y., Käpylä, J., Kämäräinen, J.-K., Koskinen, S., Matas, J., "A benchmark for temporal color constancy", arXiv preprint arXiv:2003.03763, 2020.

[104] Prinet, V., Lischinski, D., Werman, M., "Illuminant chromaticity from image sequences", in Proceedings of the IEEE International Conference on Computer Vision, 2013, str. 3320–3327.

[105] Yoo, J.-S., Kim, J.-O., "Dichromatic model based temporal color constancy for ac light sources", in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, str. 12 329–12 338.

[106] Afifi, M., Brubaker, M. A., Brown, M. S., "Auto white-balance correction for mixed-illuminant scenes", arXiv preprint arXiv:2109.08750, 2021.

[107] Murmann, L., Gharbi, M., Aittala, M., Durand, F., "A dataset of multi-illumination images in the wild", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, str. 4080–4089.

[108] Bleier, M., Riess, C., Beigpour, S., Eibenberger, E., Angelopoulou, E., Tröger, T., Kaup, A., "Color constancy and non-uniform illumination: Can existing algorithms work?", in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, 2011, str. 774–781.

[109] Beigpour, S., Riess, C., Van De Weijer, J., Angelopoulou, E., "Multi-illuminant estimation with conditional random fields", IEEE Transactions on Image Processing, Vol. 23, No. 1, 2013, str. 83–96.

[110] Beigpour, S., Ha, M. L., Kunz, S., Kolb, A., Blanz, V., "Multi-view multi-illuminant intrinsic dataset.", in BMVC, 2016.

[111] Hao, X., Funt, B., "A multi-illuminant synthetic image test set", Color Research & Application, Vol. 45, No. 6, 2020, str. 1055–1066.

[112] Hao, X., Funt, B., Jiang, H., "Evaluating colour constancy on the new mist dataset of multi-illuminant scenes", in Color and Imaging Conference, Vol. 2019, No. 1. Society for Imaging Science and Technology, 2019, str. 108–113.

[113] Gijsenij, A., Lu, R., Gevers, T., "Color constancy for multiple light sources", IEEE Transactions on Image Processing, Vol. 21, No. 2, 2011, str. 697–707.

[114] Hordley, S. D., Finlayson, G. D., "Reevaluation of color constancy algorithm performance", JOSA A, Vol. 23, No. 5, 2006, str. 1008–1020.

[115] Gijsenij, A., Gevers, T., Lucassen, M. P., "Perceptual analysis of distance measures for color constancy algorithms", JOSA A, Vol. 26, No. 10, 2009, str. 2243–2256.

[116] Finlayson, G. D., Hordley, S. D., Morovic, P., "Colour constancy using the chromagenic constraint", in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1. IEEE, 2005, str. 1079–1086.

[117] Finlayson, G. D., Zakizadeh, R., Gijsenij, A., "The reproduction angular error for evaluating the performance of illuminant estimation algorithms", IEEE transactions on pattern analysis and machine intelligence, Vol. 39, No. 7, 2016, str. 1482–1488.

Publications

Publication 1

Koščević, K., Subašić, M., Lončarić, S., "Attention-based Convolutional Neural Network for Computer Vision Color Constancy", Proceedings of the 11th International Symposium on Image and Signal Processing and Analysis, Dubrovnik, Croatia, 2019, pp. 372-377.

Attention-based Convolutional Neural Network for Computer Vision Color Constancy

Karlo Koščević, Marko Subašić, and Sven Lončarić
Image processing group
Faculty of Electrical Engineering and Computing
University of Zagreb
10000 Zagreb, Croatia
E-mail: {karlo.koscevic, marko.subasic, sven.loncaric}@fer.hr

***Abstract*—Achieving color constancy is an important part of image preprocessing pipeline of contemporary digital cameras. Its goal is to eliminate the influence of the illumination color on the colors of the objects in the image scene. State-of-the-art results have been achieved with learning-based methods, especially when the deep learning approaches have been applied. Several methods that are combining local patches for global illumination estimations exist. However, in this paper, a new convolutional neural network architecture is proposed. It is trained to look for the regions, i.e., patches in the image where the most useful information about the scene illumination is contained. This is achieved with the attention mechanism stacked on top of the pretrained convolutional neural network. Additionally, the common problem of the lack of data in color constancy benchmark datasets is alleviated utilizing the stage-wise training. Experimental results show that the proposed approach achieves competitive results.**

***Index Terms*—Attention mechanism, color constancy, convolutional neural networks, deep learning, illumination estimation, image enhancement**

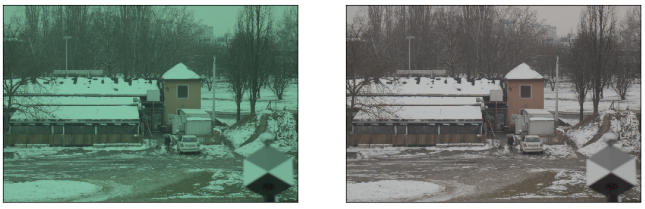
I. Introduction

Colors of objects in an image are determined by three factors, namely intrinsic properties of their surface, the color of the light source, and the camera sensor. The human vision system (HSV) has the ability to constantly perceive colors in a scene notwithstanding the change of the light source. This ability is known as color constancy [1]. As opposed to that, when captured with digital cameras scene colors are affected by the light source color. Therefore the same scene may appear different when the light source changes. This effect is illustrated in Figure 1. Intending to eliminate the color cast, contemporary digital cameras have computer vision color constancy implemented in their image processing pipeline. In the literature, computer vision color constancy is often also referred as computational color constancy, but in [2] it has been shown that there are actually two types of computational color constancy, namely computer vision color constancy and human vision color constancy. The method proposed in this paper relates to the computer vision color constancy. The most important part of the computer vision color constancy is illumination estimation which aims to estimate the light source color by knowing only image pixel values. In computer vision color constancy, the most used image formation model $\mathbf{f}$, which uses Lambertian assumption is

$$f_c(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x}) R(\mathbf{x}, \lambda) \rho_c(\lambda) d\lambda \tag{1}$$

where for each color channel $c \in \{R, G, B\}$, the value at location $\mathbf{x}$ is determined by the spectral distribution of the light source $I(\lambda, \mathbf{x})$, surface reflectance $R(\lambda, \mathbf{x})$, and sensitivity of the camera sensor $\rho_c(\lambda)$. Only wavelengths λ in the visible light spectrum ω are observed. As previously stated, when performing illumination estimation only pixel values $\mathbf{f}$ are known and therefore it is an ill-posed problem. Additional assumptions are necessary to solve it and many methods have been proposed, but the problem still remains open. When the illumination is uniform, which is the most common assumption, the objective of the illumination estimation is to calculate the vector of the light source color $\mathbf{e}$ that is invariant given the position $\mathbf{x}$ in the image scene. i.e.

$$\mathbf{e} = \int_\omega I(\lambda, \mathbf{x}) \boldsymbol{\rho}(\lambda) d\lambda = \int_\omega I(\lambda) \boldsymbol{\rho}(\lambda) d\lambda. \tag{2}$$



(a) (b)

Fig. 1: The same scene (a) before and (b) after the removal of scene illumination. The image has been taken from the Cube+ dataset [3]

Illumination estimation methods can be divided into two main groups, namely statistics-based and learning-based methods. Although statistics-based are much faster and hardware-friendly methods, it is learning-based methods that achieve state-of-the-art accuracy. However, some of the recently proposed learning-based methods have execution times which are comparable with statistics-based methods [4], [5].

Motivated by the success of deep learning approaches in many computer vision tasks, in this paper, a new convolutional

978-1-7281-3140-5/19/$31.00

neural network (CNN) for the task of illumination estimation is proposed. The network has the attention mechanism which can be thought of as a mechanism for a patch-wise illumination estimation where patches are estimated by the network itself and can be of any size and shape. The regions of an image where more informative features for the task of the illumination estimation are located are given more weight and they have a higher impact on the final global estimate. Since the network has the large number of parameters it has been trained in stages, i.e., the updates of layers of pretrained network were enabled progressively.

The rest of the paper is structured as follows: Section II is a brief overview of existing illumination estimation methods. In Section III detailed descriptions of the proposed network architecture and training procedure are given. Section IV presents the experimental setup and obtained results. Section V concludes the paper.

II. Related Work

As illumination estimation is an ill-posed problem many methods with different assumptions have been proposed. One group of methods rely on assumptions based on low-level image statistics and therefore they are hardware-friendly and fast. These are methods like White-patch [6], [7] and its improvements [8]–[10], Gray-world [11], Shades-of-Gray [12], 1st and 2nd order Gray-Edge [13], using gray pixels [14] or bright pixels [15], using bright and dark colors [16], exploiting illumination statistics perception [17].

The second main group are the methods that rely on data and the information they can exploit from it. These methods usually require more parameter tuning, take longer time to train and are more computationally demanding. However, they produce most accurate illumination estimations. Such methods are gamut mapping (pixel, edge, and intersection based) [18], [19], using neural networks [20], using high-level visual information [21], natural image statistics [22], Bayesian learning [23], spatio-spectral learning [24], simplifying the illumination solution space [5], [25], [26], using color or edge moments [27], regression trees with simple features from color distribution statistics [4], performing various kinds of spatial localizations [28], [29], using genetic algorithms [30], modelling colour constancy by using the overlapping asymmetric Gaussian kernels with surround pixel contrast-based sizes [31], finding paths for the longest dichromatic line produces by specular pixels [32], detecting gray pixels with specific illuminant-invariant measures in logarithmic space [33], channel-wise pooling the responses of double-opponency cells in LMS color space [34]. The most recent improvement in accuracy in illumination estimation is due to the convolutional neural networks [28], [29], [35]–[39].

III. Network architecture

The proposed architecture uses VGG16 [40] as the base convolutional neural network. Since the number of images in color constancy datasets is not sufficient to train a deep convolutional neural network from scratch, the VGG16 network has been first pre-trained on the ImageNet dataset [41]. Fully connected layers were replaced with two additional convolutional layers $C1$, $C2$, and an attention mechanism. Both convolutional layers have kernel sizes 3×3 with 512 filters for layer $C1$ and 3 filters for layer $C2$. The reasoning behind only three filters for layer $C2$ is to use each filter to estimate one color channel $c \in \{R, G, B\}$ of the illumination vector. Based on the features obtained from convolutional layer $C1$ the attention mechanism calculates a separate attention map for each color channel $c \in \{R, G, B\}$. The illumination vector is calculated as the normalized sum of the channel-wise product of attention maps and the output of the convolutional layer $C2$. There are no fully connected layers and therefore there are no restrictions on the size of input images. All the layers have ReLU activation function, except for the last layer in the attention mechanism that uses sigmoid. The proposed architecture is illustrated in Figure 2.

A. Training

As described in Section IV-B, the angular error is the most often used metric for evaluation of illumination estimation methods. However, in [42] has been shown that angular error function is not an appropriate loss function for the convolutional neural network training due to the complexity and instability of its derivative. Hence, $1 - cos(\epsilon)$ is proposed as a better loss function choice. Following that, the minibatch loss can be calculated as

$$L = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{\mathbf{e}_i \cdot \mathbf{e}_i^{Est}}{\|\mathbf{e}_i\| \left\|\mathbf{e}_i^{Est}\right\|} \right) \tag{3}$$

where N is the number of training samples in a minibatch, $\mathbf{e}_i^{Est}$ is the estimated illumination vector for the ith training sample, $\mathbf{e}_i$ is the corresponding ground-truth vector, '$\cdot$' is the vector dot product, and $\|.\|$ is vector $L2$ norm.

The proposed network was trained with Equation 3 as a loss function. Adam [43] was used as the optimizer with the learning rates $3{\times}10^{-4}$ and $3{\times}10^{-5}$. Although some research shows that change of learning rate can boost the performance [44], lowering the learning rate for Adam optimizer in this research improved the accuracy only in the last training stage.

The proposed network architecture was trained in stages versus the most common end-to-end manner. There are two reasons for such a training strategy. The first is the insufficient amount of training data to train the whole VGG16 network from scratch. The second is that at the beginning of the training weight updates can be very large. Therefore, if a network would be trained in an end-to-end manner the profit of using the pretrained model would be alleviated or even nullified. The training has been done in 5 stages. In the first stage, only weights of the convolutional layers $C1$ and $C2$ and attention mechanism are updated. In stages 2-5 updates of the layers of the last convolutional blocks of the pretrained VGG16 network were gradually enabled. Layer $block5\_conv3$ is unfrozen in stage 2, $block5\_conv2$ in stage 3, $block5\_conv1$ in stage 4, and finally in stage 5 layer $block4\_conv3$. Number of training

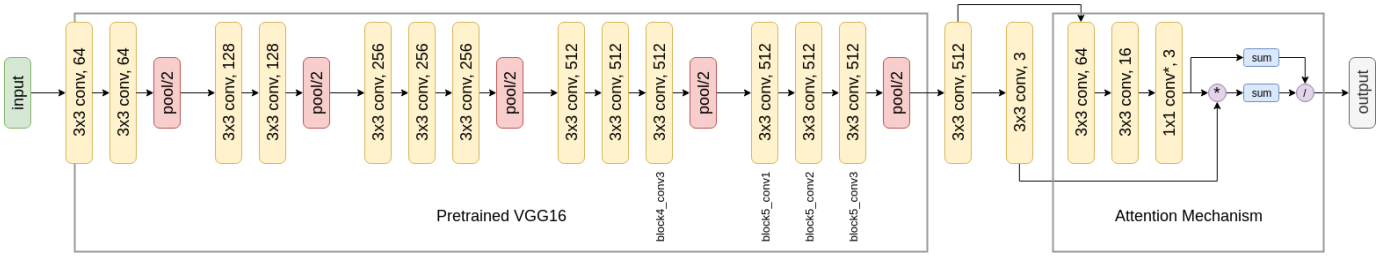


Fig. 2: Illustration of the proposed architecture with the attention mechanism. The *conv** operation uses a different set of weights for each patch of the input, i.e. the weights are unshared. Operators '$*$' and '$/$' resemble the element-wise multiplication and division of two feature maps, respectively. *Sum* is a simple sum pooling operation.

epochs in the stage varied as well: 48 epochs for the stage 1, 56 for the stage 2, 64 for the stage 3, 96 for the stage 4, and only 32 for the 5. In all stages batch size was 32. Learning rate 3×10^{-5} was used for stage 5 whereas 3×10^{-4} was used for all other stages.

IV. Experimental Results

A. Data

Cube+ dataset [3] was used for benchmarking. It is a color constancy dataset of 1707 images with both indoor and outdoor scenes captured during the day and night. Although it is not the largest color constancy dataset, which is a desirable feature when deep neural networks are used, it is the largest publicly available color constancy dataset which does not have drawbacks such as incorrect ground-truth illumination data, a significant amount of noise, violations of some important assumptions, or past misuses [45].

Due to the memory limitations of available hardware the images have been resized to the width and height of 512 pixels. All images were preprocessed as described in [3]. Usually, deep convolutional models have some model specific preprocessing of input images as well and therefore additionally preprocessing specific to convolutional neural network VGG16 has been applied [40].

In order to test the generalization capabilities of the proposed architecture, the data was split into train and test subsets. Train set contains 76.5% of the data and the remaining 23.5% is used as the test data. Both sets have the same ratio of day and night images which is roughly 88% and 22%, respectively. Ground-truth illuminations in both sets occupy the whole domain of the Cube+ dataset. Figure 3 shows the distribution of train and test illuminations in rb-chromaticity plane. During the training 30% of the data in the train set was used for the hyperparameter selection, i.e. as the validation set. The test set was exclusively used to evaluate network performance after the training and hyperparameter optimization have been done.

B. Performance metrics

Since in computer vision color constancy illumination vector is estimated, the most used error measure is angular error between two vectors

$$err = cos^{-1}\left(\frac{\mathbf{e} \cdot \mathbf{e}^{Est}}{\|\mathbf{e}\| \, \|\mathbf{e}^{Est}\|}\right) \quad (4)$$

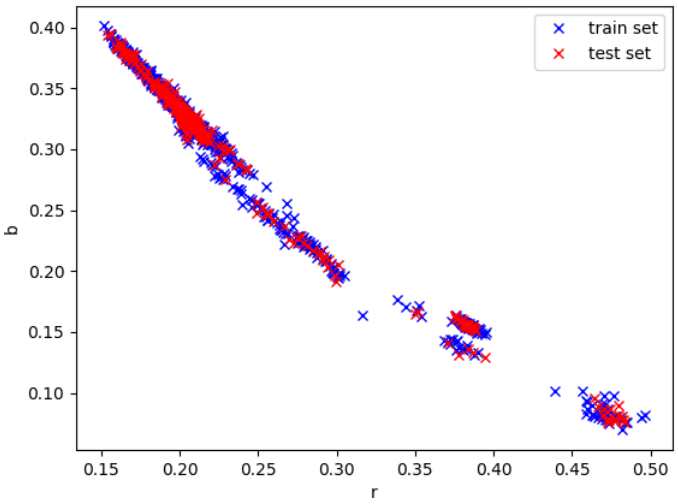


Fig. 3: Distribution of train and test ground-truth illuminations in rb-chromaticity plane.

where $\mathbf{e}$ is the ground-truth illumination, $\mathbf{e}^{Est}$ is the estimated illumination vector, '$\cdot$' is the vector dot product, and $\|.\|$ is the vector $L2$ norm. Angular errors obtained for all images in a dataset are usually combined using summary statistics. The angular error distribution is non-symmetrical and therefore most used measure is median of angular errors [46]. Nonetheless, mean, trimean, best 25%, and worst 25% are also often used for additional comparisons. In [28] a new measure was proposed, i.e. the geometric mean of median, mean, trimean, best 25% and worst 25% summary statistics.

C. Method performance

A few combinations of convolutional layers before the attention have been compared. It has been shown that more than one convolutional layer is not beneficial, but rather makes the architecture unstable and hard to train. One convolutional layer with 512 filters has shown to be a better choice. The accuracy in form of median angular error on validation set is shown in Table I. All the experiments have been done using Adam optimizer, the learning rate of 3×10^{-4} and 100 epochs.

The effect of the stage-wise training strategy can be observed in Figure 5 and Table II. It can be seen how illumination estimations gradually become finer and their distribution in the

TABLE I: Accuracy of different combinations of convolutional layers before the attention mechanism. Median angle is used as the accuracy measure.

Layer combination	Median angle
Conv512 + Conv512	19.6913
Conv512 + Conv256	19.6796
Conv512	**1.8423**
Conv256	1.9152

rb-chromaticity plane tends to match ground-truth distribution. Plotted estimations have been obtained on the test set. The angular error statistics in Table II also show the improvement in the accuracy on the test set. For comparison, the netwok was trained in one stage with the same set of hyperparameter as the proposed stage-wise strategy. Updates of weights of pretrained layers $block5\_conv3$, $block5\_conv2$, and $block5\_conv1$ were enabled from the begining. The network was trained for 264 epochs which is equal to the sum of epochs in first 4 stages of stage-wise training. Adam with learning rate 3×10^{-4} was used. Obtained mean, median and average angular error were 2.05, 1.40, and 1.56 respectively. Stage 5 was not included as different learning rate has been used in stage-wise training and therefore all layers could not be trained at once.

TABLE II: Angular error statistics of the proposed method after each training stage. Results were obtained on the test set. The used format is the same as in [28]. (lower Avg. is better)

Stage	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
Stage 1	2.76	2.00	2.17	0.70	6.11	2.20
Stage 2	2.16	1.48	1.61	0.46	5.08	1.64
Stage 3	2.15	1.30	1.56	0.37	5.39	1.54
Stage 4	1.97	1.17	1.42	0.37	4.86	1.42
Stage 5	1.95	1.13	1.37	0.32	4.92	1.37

In Table III accuracy of the proposed CNN architecture and stage-wise training strategy is compared with other illumination estimation methods. It can be seen that the proposed solution outperforms all of the other methods, except the Color Beaver method. To the best of the authors' knowledge, this is currently the only CNN-based illumination estimation method evaluated on Cube+ dataset. The angular error distribution measured on the test set can be seen in Figure 4. Chromatic adaptation using ground-truth illuminations and estimated illuminations has been performed on images with the lowest, highest and intermediate error. The results are shown in Figure 6. It can be seen that the proposed architecture works better on images captured in more natural environments such as outdoor scenes in the daylight. In contrast, higher error values are obtained on indoor images.

V. Conclusion

In this paper a new convolutional neural network (CNN) architecture for the illumination estimation has been proposed. It uses attention mechanism on top of the pretrained VGG16 model. Attention mechanism guides the network towards the regions in the image that contain the most information about the scene illumination. Additionally, stage-wise training strategy has been proposed to alleviate the lack of the training data. As the stages progress updates of more layers of the pretrained model are enabled. The proposed architecture was trained and evaluated on a newer color constancy benchmark dataset where it outperformed most of the other methods. Future work will include experiments with the attention mechanism architecture and pruning of the base CNN in order to reduce the computational complexity and improve train and inference speed.

TABLE III: Angular error statistics of different color constancy methods on the Cube+ dataset [3]. The used format is the same as in [28]. (lower Avg. is better)

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
White-Patch [7]	9.69	7.48	8.56	1.72	20.49	7.38
Gray-world [11]	7.71	4.29	4.98	1.01	20.19	5.08
Double-opponency (max pooling) [34]	6.76	3.44	4.15	0.79	18.54	4.27
Using gray pixels [33]	6.65	3.26	3.95	0.68	18.75	4.05
Color Tiger [3]	3.91	2.05	2.53	0.98	10.00	2.88
Double-opponency (max pooling) [34]	5.19	1.35	2.10	0.32	16.85	2.40
Color Mule [47]	5.16	1.30	2.03	0.25	16.93	2.25
Shades-of-Gray [12]	2.59	1.73	1.93	0.46	6.19	1.90
2nd-order Gray-Edge [13]	2.50	1.59	1.78	0.48	6.08	1.83
1st-order Gray-Edge [13]	2.41	1.52	1.72	0.45	5.89	1.76
Color Dog [5]	3.32	1.19	1.60	0.22	10.22	1.70
General Gray-World [48]	2.38	1.43	1.66	0.35	6.01	1.64
Proposed method	**1.95**	**1.13**	**1.37**	**0.32**	**4.92**	**1.37**
Color Beaver (using Gray-world) [30]	1.49	0.77	0.98	0.21	3.94	0.99

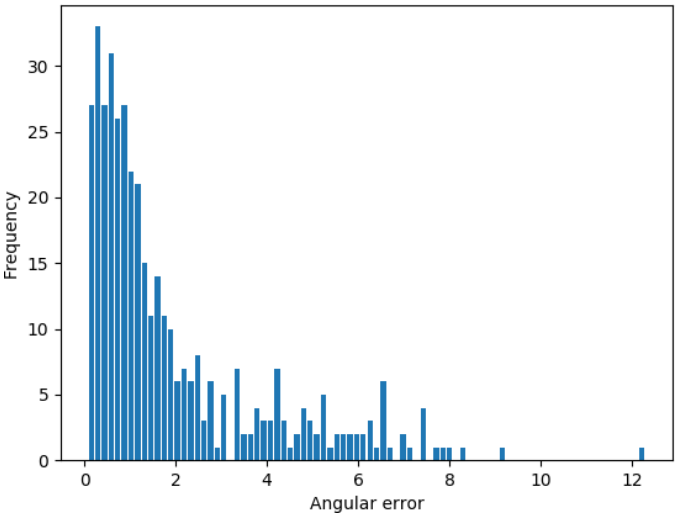


Fig. 4: Angular error distribution on the test set.

Acknowledgment

This work has been supported by the Croatian Science Foundation under Project IP-06-2016-2092. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] M. Ebner, *Color Constancy*, ser. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2007.

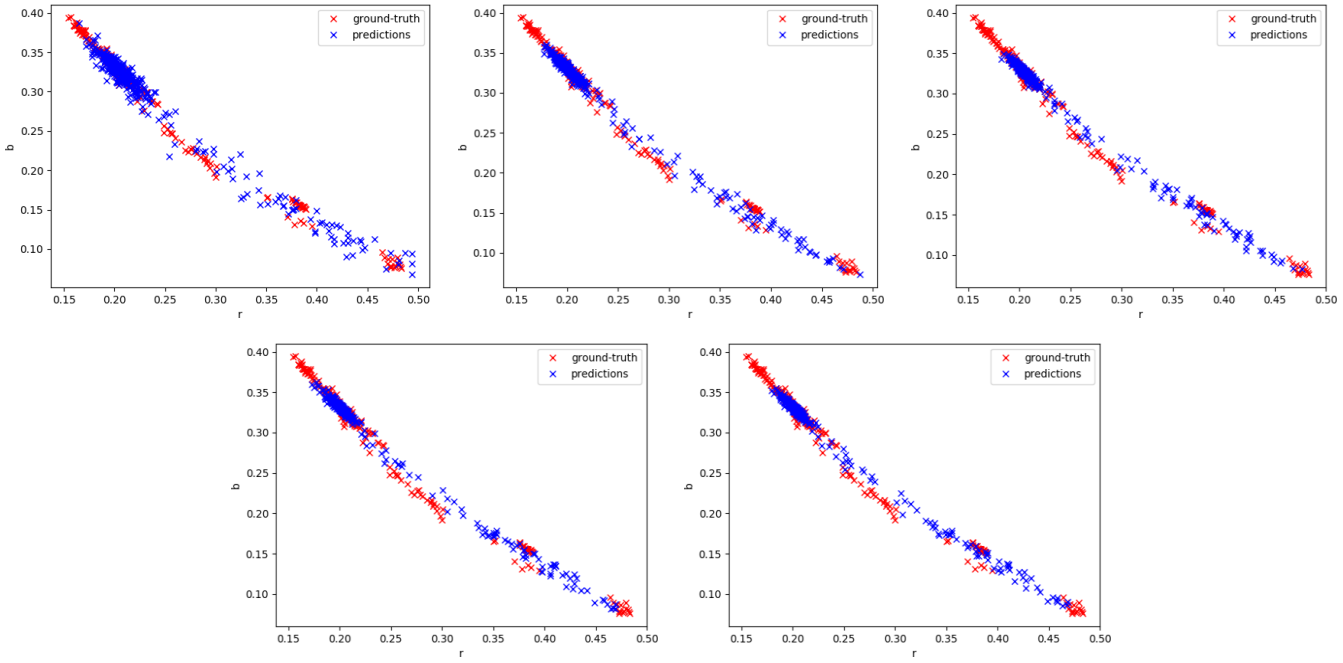


Fig. 5: Illumination estimations refinement with respect to the training stages in chromaticity plane. The red chromaticity is shown on the x axis, whereas on the y axis is the blue chromaticity. In the first row distributions after stages 1, 2, and 3 are shown, whereas in the seconD row distributions after stages 4 and 5.

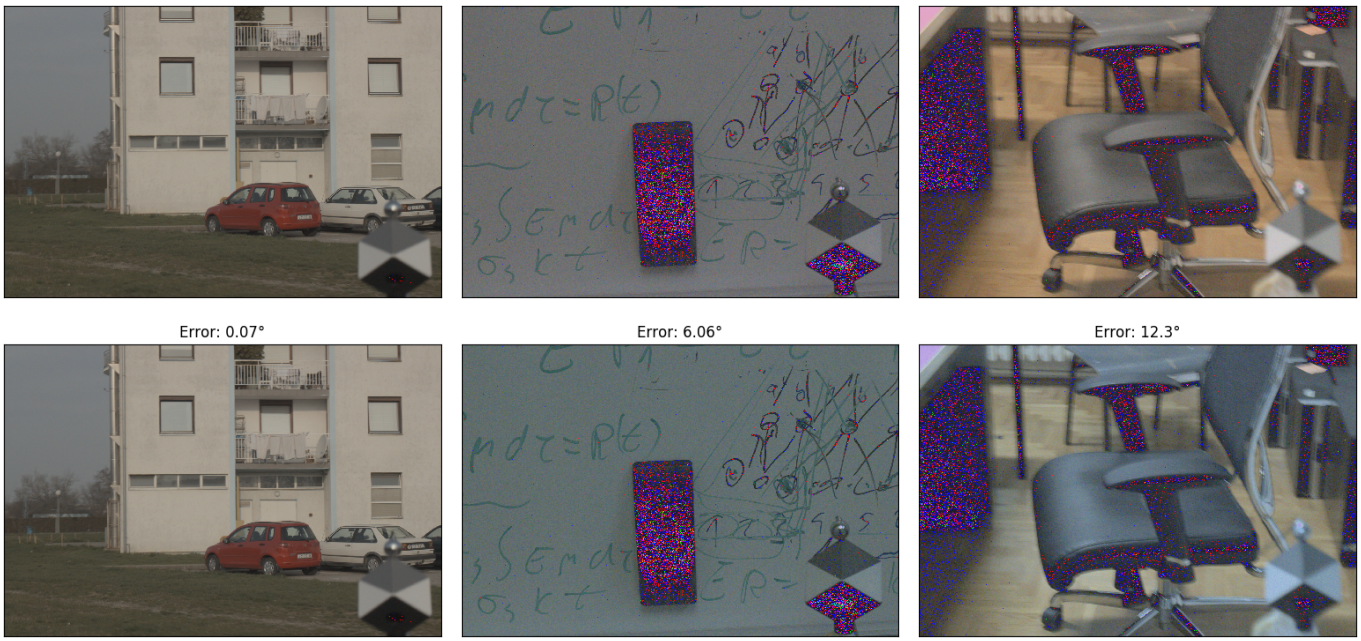


Fig. 6: Examples of images where different angular errors have been obtained. Images after the chromatic adaptation with ground-truth illuminations are shown in the top row and images after the chromatic adaptation with estimated illuminations are shown in the bottom row. For the demonstration purposes images have been tone mapped using Flash tone mapping operator [49].

[2] G. Simone, G. Audino, I. Farup, F. Albregtsen, and A. Rizzi, "Termite retinex: a new implementation based on a colony of intelligent agents," *Journal of electronic imaging*, vol. 23, no. 1, p. 013006, 2014.

[3] N. Banić and S. Lončarić, "Unsupervised Learning for Color Constancy," in *VISAPP*, 2018, pp. 181–188.

[4] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Effective learning-based illuminant estimation using simple features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015,

pp. 1000–1008.
[5] N. Banić and S. Lončarić, "Color Dog: Guiding the Global Illumination Estimation to Better Accuracy," in *VISAPP*, 2015, pp. 129–135.
[6] E. H. Land, *The retinex theory of color vision*. Scientific America., 1977.
[7] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Color and Imaging Conference*, vol. 2010, no. 1. Society for Imaging Science and Technology, 2010, pp. 256–259.
[8] N. Banić and S. Lončarić, "Using the Random Sprays Retinex Algorithm for Global Illumination Estimation," in *Proceedings of The Second Croatian Computer Vision Workshopn (CCVW 2013)*. University of Zagreb Faculty of Electrical Engineering and Computing, 2013, pp. 3–7.
[9] N. Banić and S. Lončarić, "Color Rabbit: Guiding the Distance of Local Maximums in Illumination Estimation," in *Digital Signal Processing (DSP), 2014 19th International Conference on*. IEEE, 2014, pp. 345–350.
[10] N. Banić and S. Lončarić, "Improving the White patch method by subsampling," in *Image Processing (ICIP), 2014 21st IEEE International Conference on*. IEEE, 2014, pp. 605–609.
[11] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of The Franklin Institute*, vol. 310, no. 1, pp. 1–26, 1980.
[12] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color and Imaging Conference*, vol. 2004, no. 1. Society for Imaging Science and Technology, 2004, pp. 37–41.
[13] J. Van De Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *Image Processing, IEEE Transactions on*, vol. 16, no. 9, pp. 2207–2214, 2007.
[14] Y. Qian, S. Pertuz, J. Nikkanen, J.-K. K am ar ainen, and J. Matas, "Revisiting Gray Pixel for Statistical Illumination Estimation," in *VISAPP*, 2019, pp. 36–46.
[15] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey, "The Role of Bright Pixels in Illumination Estimation," in *Color and Imaging Conference*, vol. 2012, no. 1. Society for Imaging Science and Technology, 2012, pp. 41–46.
[16] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," *JOSA A*, vol. 31, no. 5, pp. 1049–1058, 2014.
[17] N. Banić and S. Lončarić, "Blue Shift Assumption: Improving Illumination Estimation Accuracy for Single Image from Unknown Source," in *VISAPP*, 2019, pp. 191–197.
[18] K. Barnard, "Improvements to gamut mapping colour constancy algorithms," in *European conference on computer vision*. Springer, 2000, pp. 390–403.
[19] G. D. Finlayson, S. D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 93–109, 2006.
[20] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *JOSA A*, vol. 19, no. 12, pp. 2374–2386, 2002.
[21] J. Van De Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.
[22] A. Gijsenij and T. Gevers, "Color Constancy using Natural Image Statistics." in *CVPR*, 2007, pp. 1–8.
[23] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
[24] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 8, pp. 1509–1519, 2012.
[25] N. Banić and S. Lončarić, "Color Cat: Remembering Colors for Illumination Estimation," *Signal Processing Letters, IEEE*, vol. 22, no. 6, pp. 651–655, 2015.
[26] N. Banić and S. Lončarić, "Using the red chromaticity for illumination estimation," in *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2015, pp. 131–136.
[27] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1904–1911.
[28] J. T. Barron, "Convolutional Color Constancy," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 379–387.
[29] J. T. Barron and Y.-T. Tsai, "Fast Fourier Color Constancy," in *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Computer Society Conference on*, vol. 1. IEEE, 2017.
[30] K. Koščević, N. Banić, and S. Lončarić, "Color Beaver: Bounding Illumination Estimations for Higher Accuracy," in *VISAPP*, 2019, pp. 183–190.
[31] A. Akbarinia and C. A. Parraga, "Colour constancy beyond the classical receptive field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 9, pp. 2081–2094, Sep. 2018.
[32] S. Woo, S. Lee, J. Yoo, and J. Kim, "Improving color constancy in an ambient light environment using the phong reflection model," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1862–1877, April 2018.
[33] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2254–2263.
[34] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 1973–1985, 2015.
[35] S. Bianco, C. Cusano, and R. Schettini, "Single and multiple illuminant estimation using convolutional neural networks," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4347–4362, 2017.
[36] Y. Hu, B. Wang, and S. Lin, "Fc4: fully convolutional color constancy with confidence-weighted pooling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4085–4094.
[37] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," in *European Conference on Computer Vision*. Springer, 2016, pp. 371–387.
[38] Z. Lou, T. Gevers, N. Hu, M. P. Lucassen *et al.*, "Color constancy by deep learning." in *BMVC*, 2015, pp. 76–1.
[39] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 81–89.
[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
[42] O. Sidorov, "Artificial color constancy via googlenet with angular loss function," *arXiv preprint arXiv:1811.08456*, 2018.
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[44] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 464–472.
[45] N. Banić, M. Subašić *et al.*, "The past and the present of the color checker dataset misuse," *arXiv preprint arXiv:1903.04473*, 2019.
[46] S. D. Hordley and G. D. Finlayson, "Re-evaluating colour constancy algorithms," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 1. IEEE, 2004, pp. 76–79.
[47] N. Banić and S. Lončarić, "A Perceptual Measure of Illumination Estimation Error," in *VISAPP*, 2015, pp. 136–143.
[48] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data," *Image Processing, IEEE Transactions on*, vol. 11, no. 9, pp. 972–984, 2002.
[49] N. Banić and S. Lončarić, "Flash and storm: Fast and highly practical tone mapping based on naka-rushton equation," in *International Conference on Computer Vision Theory and Applications*, 2018.

Publication 2

Koščević, K., Subašić, M., Lončarić, S., "Guiding the Illumination Estimation Using the Attention Mechanism", Proceedings of the 2020 2nd Asia Pacific Information Technology Conference, Bali, Indonesia, 2020, pp. 143-149.

Guiding the Illumination Estimation Using the Attention Mechanism

Karlo Koščević
Faculty of Electrical Engineering and Computing, University of Zagreb
Zagreb, Croatia
karlo.koscevic@fer.hr

Marko Subašić
Faculty of Electrical Engineering and Computing, University of Zagreb
Zagreb, Croatia
marko.subasic@fer.hr

Sven Lončarić
Faculty of Electrical Engineering and Computing, University of Zagreb
Zagreb, Croatia
sven.loncaric@fer.hr

ABSTRACT

Deep learning methods have achieved a large step forward in many computer vision applications. With mechanisms such as attention, deep models can now guide themselves to focus on parts of an image that are more significant for a given task. In computational color constancy, the most important step is to estimate the illumination vector as accurately as possible. Since illumination estimation algorithms can be sensitive to noise, such as ambiguous regions in the image, the ability to have a mechanism to look for specific regions in an image could be helpful. In this paper, a convolutional neural network with an attention mechanism is proposed. The attention mechanism helps the network to focus on regions that contain more content and to avoid regions where ambiguous estimations may occur. In the experimental results, it is shown that the attention mechanism does help the network to obtain more accurate estimations and puts the focus of the network on the regions in an image where gradients are high. The network with the attention mechanism achieves up to 10% increase in accuracy compared to the same network architecture without the attention mechanism.

CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by regression**; **Neural networks**; **Computational photography**; **Image processing**.

KEYWORDS

attention mechanism; computational color constancy; convolution; deep learning; image processing; neural network; regression; white balancing

1 INTRODUCTION

In the image formation pipeline, contemporary digital cameras have a part that removes the influence of light source color on scene colors. Achieving the invariance of scene colors to the color of the light source is called computational color constancy. It is motivated by the ability of the human vision system (HVS) to adapt to changes in illumination in the scene. This ability, named human vision color constancy, enables humans to observe the true color of objects of interest [22]. Unlike HVS, the camera sensor can not distinguish surface reflectance from illumination and, without additional processing, colors of objects in a digital image would be biased to the illumination color. Consequently, with the change of the illumination, the same scene may appear different. In digital cameras, the color constancy is usually achieved in two steps. The first step is to estimate the RGB vector of the illumination color. Therefore it is called the illumination estimation step. In the second step, i.e., the chromatic adaptation step, RGB estimate is divided out from all pixel values of an image. The result should be an image with colors as they would appear when taken under the white illumination, i.e., the illumination whose RGB values are all the same. Hence, computational color constancy is often also referred to as white balancing. To achieve computational color constancy, the image formation model $\mathbf{f}$ with the Lambertian assumption is used. It uses three physical variables to formulate an image: 1) spectral properties of the light source, 2) spectral reflectance of surfaces in the scene, and 3) spectral sensitivity of the camera sensor. It can be expressed as

$$f_c(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x}) R(\lambda, \mathbf{x}) \rho_c(\lambda) d\lambda \tag{1}$$

where $c \in \{R, G, B\}$ is the color channel, $f(\mathbf{x})$ is the image value at pixel location $\mathbf{x}$, $I(\lambda, \mathbf{x})$ is the spectral distribution of the light source, $R(\lambda, \mathbf{x})$ is the surface reflectance, $\rho_c(\lambda)$ is the spectral sensitivity of the camera sensor for color channel c, and λ are the wavelengths in the visible part of the light spectrum ω. From (1) it can be observed that illumination vector $\mathbf{e} = (e_R e_G e_B)$, captured by digital camera, depends on the spectral distribution of the light source and sensitivity of the camera sensor as follows

$$\mathbf{e}(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x}) \boldsymbol{\rho}(\lambda) d\lambda. \tag{2}$$

Unfortunately, in the illumination estimation step, both $I(\lambda, \mathbf{x})$ and $\boldsymbol{\rho}(\lambda)$ are unknown and the illumination vector can only be estimated from image pixel values $\mathbf{f}$. This makes illumination estimation an ill-posed problem. Therefore, many assumptions have been made to overcome this issue. One of the most often used assumptions, which does not make the problem well-posed, but only simplifies it, is that the illumination is uniform in the scene. It is assumed that each position $\mathbf{x}$ in the image is illuminated by the same illumination vector $\mathbf{e}$. Hence, $\mathbf{x}$ can be disregarded from (2) which leads to

$$\mathbf{e} = \int_\omega I(\lambda) \boldsymbol{\rho}(\lambda) d\lambda. \tag{3}$$

During the years, many illumination estimation methods, based on different assumptions, have been proposed. In [37] it is assumed

APIT 2020, January 17–19, 2020, Bali Island, Indonesia

ACM ISBN 978-1-4503-7685-3/20/01…$15.00
DOI: https://doi.org/10.1145/3379310.3379329

that if the content of the scene is taken into account, more accurate estimations can be achieved. A deep neural network with a separate attention mechanism for each color channel $c \in \{R, G, B\}$ was used to achieve that. In this paper, a convolutional neural network with a modified attention mechanism has been proposed. The attention mechanism uses one attention map for all three color channels. Experimental results have shown that such a network can learn to look for the regions in an image where some content exist and ignore ambiguous regions.

The rest of the paper is structured as follows: in Section 2 a short overview of existing illumination estimation methods is given, the motivation for the proposed solution is given in Section 3, the proposed network architecture is described in Section 4, experimental results are presented and commented in Section 5, and Section 6 concludes the paper.

2 RELATED WORK

In the literature, few classifications of illumination estimation methods exist [30, 34, 47]. Most classifications group the methods into statistics-based methods and learning-based methods. Statistics-based are methods such as as White-Patch [26, 38], its improvements [2–4], Gray-World [16], Shades-of-Gray [25], 1^{st} and 2^{nd} order Gray-Edge [45], Weighted Gray-Edge [31], using bright pixels [35], gray pixels [41] or bright and dark colors [19], exploiting illumination statistics perception [11] or expected illumination statistics [9]. Assumptions used in statistics-based methods are based on low-level image statistics. Therefore, they are characterized by very low computational complexity and high execution speed which makes them suitable for hardware-implementation.

On the other hand, most accurate illumination estimations are achieved with learning-based methods, but they require more computational time and parameter tuning, which makes them slower as well. Learning-based methods are methods based on neural networks [17], high-level visual information [46], natural image statistics [29], Bayesian learning [28], spatio-spectral learning [18], methods restricting the illumination solution space [6–8], using color moments [23], regression trees with simple features from color distribution statistics [20], spatial localizations [13, 14], convolutional neural networks [15, 33, 37, 40, 42] and genetic algorithms [36], modelling colour constancy by using the overlapping asymmetric Gaussian kernels with surround pixel contrast based sizes [1], finding paths for the longest dichromatic line produces by specular pixels [48], detecting gray pixels with specific illuminant-invariant measures in logarithmic space [49], channel-wise pooling the responses of double-opponency cells in LMS color space [27]. Although gamut-based methods can be considered learning-based, due to the large impact, in [30] gamut-based methods are presented as a separate group of methods.

3 MOTIVATION

As described in Section 1, illumination estimation is an ill-posed problem. The most simple yet very common situation where illumination estimation methods could produce inaccurate illumination vector is when an image is lacking content. This is because monotone regions in an image are most often ambiguous in illumination estimation. An example is a situation when a flat white wall is illuminated with yellow illumination. An illumination estimation method may not be able to distinguish if the illumination is truly yellow or is it the case that the illumination is yellow and that the wall is white. In contrast, image regions that contain some unambiguous content may lead illumination estimation methods towards more accurate estimations. An illustration of this can be observed in Figure 1.

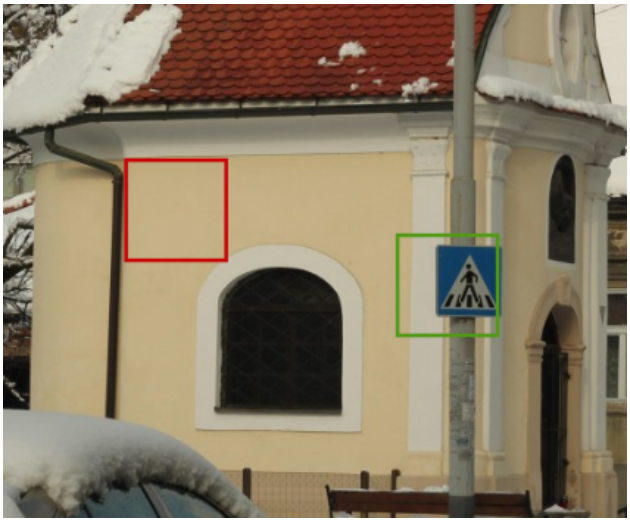


Figure 1. Comparison of the color information in two parts of one image, with red region being ambiguous and green region being informative for illumination estimation.

To guide a convolutional neural network towards the regions more appropriate for a given task, recently, convolutional neural networks have been enriched with an attention mechanism. For a region of an input image, an attention map provides an estimation of how much does that region contribute to the final output [39]. Having an attention map as part of a convolutional neural network for computational color constancy may help it to diminish the influence of ambiguous regions on the final estimate and focus on more informative regions. In [37] a convolutional neural network that uses a separate attention map for each color channel $c \in \{R, G, B\}$ was proposed. The rationale for that is the fact that in chromatic adaptation a diagonal matrix is used [24]. This means that each color channel is considered independent of other channels. However, the solution proposed in this paper is motivated by the assumption that the color of illumination is perceived as a mixture of all color channels and in most cases can not be separated without introducing additional errors.

4 THE PROPOSED NETWORK

The architecture of the proposed deep neural network is shown in Figure 2. Convolutional blocks of VGG16 [44] network are used as a feature extractor, fully connected layers have been replaced with the attention mechanism and one convolutional layer which computes intermediate illumination estimations. Neither feature extractor nor attention mechanism has fully connected layers, but

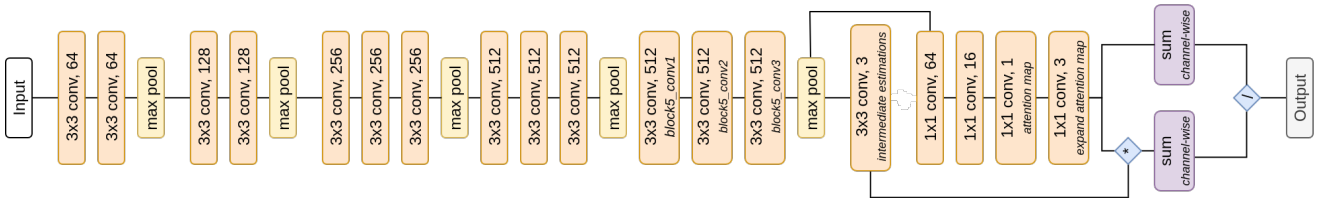


Figure 2. The proposed network architecture. '$*$' is element-wise multiplication, '/' is element-wise division, and *Sum* is the sum operator in spatial dimensions.

only convolutional ones, which makes the proposed architecture fully convolutional. This way, the proposed network can compute illumination estimations invariant to the size of the input image. The attention mechanism is composed of convolutional layers with 1×1 kernels. The first two layers have 64 and 16 filters, respectively. The third convolutional layer has only one output filter which computes the actual attention map. Sigmoid activation function is used to normalize values of the attention map in the range from 0 to 1. The attention map is expanded to match the number of channels of intermediate estimations using a convolutional layer with all weights equal to one and no biases. Expanded attention map is element-wise multiplied with intermediate estimations to produce weighted illumination estimations. Afterward, weighted estimations are summed along spatial dimensions to obtain a three-component vector, i.e. a global illumination vector. To remove the scaling caused by the multiplication with the attention map, the global illumination vector is normalized by division with the spatial sum of the attention map. The normalized vector of global illumination is the output of the proposed network architecture.

5 EXPERIMENTAL RESULTS

5.1 Data preparation

To evaluate the proposed color constancy architecture, the Cube+ dataset [10] was used. At the time of research, it was the largest color constancy dataset which complies with color constancy limitations such as uniform illumination assumption and linear images, i.e., images which have not been processed in any non-linear way. Additionally, the Cube+ dataset has a wide variety of scenes, including outdoor images captured during both day and night and indoor images. Consequently, it has a wide distribution of illuminations. Each image was resized to the size of 512×512 pixels. According to [10], for each image, the black level was subtracted, overexposed pixels were clipped, and the calibration object was masked out. In all experiments, the data was split into train and test sets which contained 80% and 20% of the data, respectively.

5.2 Evaluation metrics

To evaluate the performance of the proposed network architecture, the angular error was used. It is the angle between two vectors which, in this case, are ground-truth illumination vector and the illumination vector estimated by the network. The angular error can be computed as

$$A(\mathbf{e}, \hat{\mathbf{e}}) = cos^{-1}\left(\frac{\mathbf{e} \cdot \hat{\mathbf{e}}}{\|\mathbf{e}\| \, \|\hat{\mathbf{e}}\|}\right), \tag{4}$$

where $\mathbf{e}$ is the ground-truth illumination vector, $\hat{\mathbf{e}}$ is the illumination vector the network estimated, '$\cdot$' is dot product of two vectors, and $\|.\|$ is vector $L2$ norm. Once calculated for each image in a dataset, angular errors are usually summed up using different summary statistics. Most often used statistics are mean, median, trimean, best 25%, worst 25% and the geometric mean of all previous statistics, i.e., so-called average [13]. Since the distribution of angular errors is non-symmetrical it is recommended to use the median instead of the mean as a more significant measure.

5.3 Training setup

VGG16 network was initialized with weights obtained by training on the ImageNet dataset [21] since Cube+ does not contain a sufficient number of images to train the whole VGG16 network. Considering that and the fact that gradients can be large when the training of the network starts, the proposed network was trained by gradually enabling updates of more layers. In the first stage only weights of the attention mechanism were trained. In the following three stages, weights of *block5_conv3*, *block5_conv2*, and *block5_conv1* layers of the pre-trained VGG16 network were updated as well. In each subsequent stage all the layers from the previous stage were kept unfrozen and one additional layer was enabled for updates. This way, in the last training stage the attention mechanism and all convolutional layers in the last convolutional block of the VGG16 network were trained. In the case that weights of all layers were updated from the beginning of the training process, the benefit of pre-trained weights might be lost. At the beginning of the training gradients can be very large since the attention mechanism was not trained yet. These gradients would be propagated through the layers of the pre-trained model and could cause misleading weight updates. Training in stages helps to stabilize an unstable part of the network so it does not disturb other, already stable, parts. Adam optimizer with learning rate 1×10^{-3} was used for all stages. Batch size and the number of training epochs were 32 and 75, respectively. In [43], few loss functions for the training of convolutional neural networks for color constancy have been researched. It has been shown that using a function $1 - cos(A)$ is more suitable then using the angular error described in Subsection 5.2. Therefore in this paper, the following loss function was used

$$L(\mathbf{e}, \hat{\mathbf{e}}) = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{\mathbf{e}_i \cdot \hat{\mathbf{e}}_i}{\|\mathbf{e}_i\| \, \|\hat{\mathbf{e}}_i\|}\right), \tag{5}$$

where $\mathbf{e}_i$ and $\hat{\mathbf{e}}_i$ are ground-truth and estimated illumination vector for nth sample, N is number of samples, '$\cdot$' is the vector dot product, and $\|.\|$ is the vector $L2$ norm.

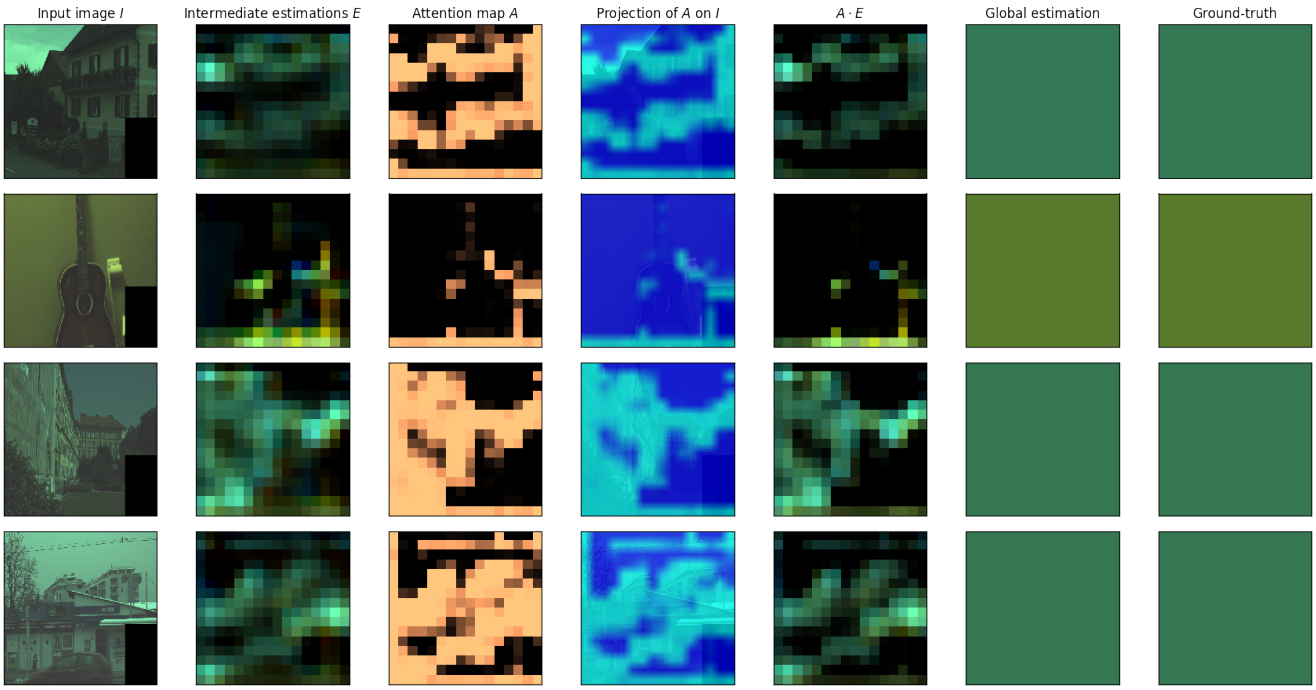


Figure 3. Example outputs of the proposed network architecture.

Table 1. Angular error statistics of different color constancy methods on the Cube+ dataset [10] (lower Avg. is better).

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
White-Patch [26]	9.69	7.48	8.56	1.72	20.49	7.38
Gray-world [16]	7.71	4.29	4.98	1.01	20.19	5.08
Double-opponency (max pooling) [27]	6.76	3.44	4.15	0.79	18.54	4.27
Using gray pixels [49]	6.65	3.26	3.95	0.68	18.75	4.05
Color Tiger [10]	3.91	2.05	2.53	0.98	10.00	2.88
Color Mule [5]	5.16	1.30	2.03	0.25	16.93	2.25
Shades-of-Gray [25]	2.59	1.73	1.93	0.46	6.19	1.90
2nd-order Gray-Edge [45]	2.50	1.59	1.78	0.48	6.08	1.83
1st-order Gray-Edge [45]	2.41	1.52	1.72	0.45	5.89	1.76
Color Dog [7]	3.32	1.19	1.60	0.22	10.22	1.70
General Gray-World [12]	2.38	1.43	1.66	0.35	6.01	1.64
Proposed method	2.05	1.32	1.53	0.42	4.84	1.54
RGB Attention CNN [37]	1.95	1.13	1.37	0.32	4.92	1.37
Color Beaver (using Gray-world) [36]	1.49	0.77	0.98	0.21	3.94	0.99

5.4 Method accuracy

To assess the overall performance of the proposed network architecture and training procedure, a test set that contains 20% of the Cube+ data was used. The images and ground-truth values in the test set were not used during the training of the network. Obtained estimations were compared with corresponding ground-truth values using evaluation metrics described in Subsection 5.2. The summary of obtained angular error values is shown in Table 1. The proposed network architecture achieves comparable results to other illumination estimation methods. The median angular error is much less than 2°, which was suggested as a good enough color constancy performance [32]. However, it is especially important to highlight that the highest estimation errors are much lower than for other methods, except [36] which is in fact constructed to minimize the maximum estimation error. Several network outputs are shown in Figure 3.

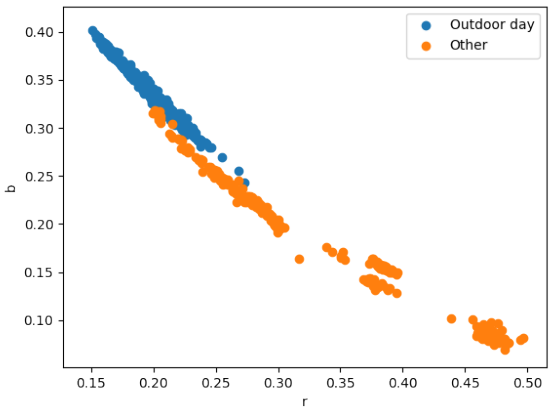


Figure 4. Two clusters of *rb*-chromaticities given the environment in which the image was captured.

In the Cube+ dataset, every fifth image is either an outdoor image captured during the night or an indoor image. The rest of the images are outdoor images captured during the day. According to that, two clusters of illuminations can be observed in the *rb*-chromaticity plane of ground-truth values, as shown in Figure 4. In Figure 5, error

values obtained on the test set for those two clusters are shown. It can be seen that for outdoor images taken during the day the estimation error is much less than for other images. Reasons for that could be the speckle noise present in images captured during the night and sparsity of images in the night and indoor conditions in the Cube+ dataset. In Table 2 the summary of error values on the test for each cluster is shown.

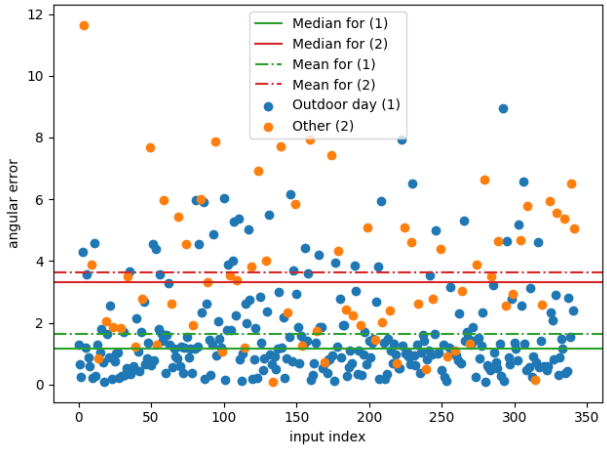


Figure 5. Angular errors obtained on the test set for outdoor images captured during the day ('Outdoor day), and outdoor images captured during the night and indoor images ('Other').

Table 2. Angular error statistics of the proposed network for a cluster with outdoor images taken during the day ('Outdoor day') and cluster with outdoor images taken during the night and indoor images ('Other') (lower Avg. is better).

	Min	Mean	Med.	Tri.	Best 25%	Worst 25%	Max	Avg.
Outdoor day	0.07	1.89	1.39	1.47	0.44	4.25	10.23	1.49
Other	0.62	6.25	4.76	5.15	1.78	13.25	24.43	5.15

5.5 Discussion

One of the ways to interpret the amount of content in an image can be to look into image gradients. In the regions of an image where some changes happen gradient amplitudes are usually high. Whereas, in uniform regions gradient amplitudes tend to be zero. To interpret the effect of attention maps, image gradients were used. For each image in the test set gradients were computed. Experiments confirm that attention maps tend to match regions in an image where some gradient exists. On average, 77% of image energy is located inside regions that attention map focuses on. More detailed correlation of attention maps and image gradients can be observed in Table 3 and Figure 6. In Figure 6 it is shown how estimation error depends on the amount of gradients captured by attention maps. The experimental results show that the proposed network tends to compute inaccurate estimations when the attention map fails to capture image gradients. This confirms the assumption of the model that regions of an image, where more content is, are more significant for illumination estimation. Additionally, the majority of images where the network failed to capture regions with high gradients and estimate accurate illumination vectors are images captured under some unnatural light source (outdoor images captured during the night or indoor images). This again implies that some methods which separate outdoor images in natural lighting from other types of lighting could be useful for datasets such as Cube+. In Figure 7 the comparison of several attention maps and gradients is shown.

Table 3. Ratio of gradient amplitudes captured by attention maps for images in the Cube+ dataset.

	Min	Mean	Med.	Q1	Max
Inside	0.07	0.67	0.77	0.68	1.00
Outside	0.00	0.33	0.23	0.32	0.94

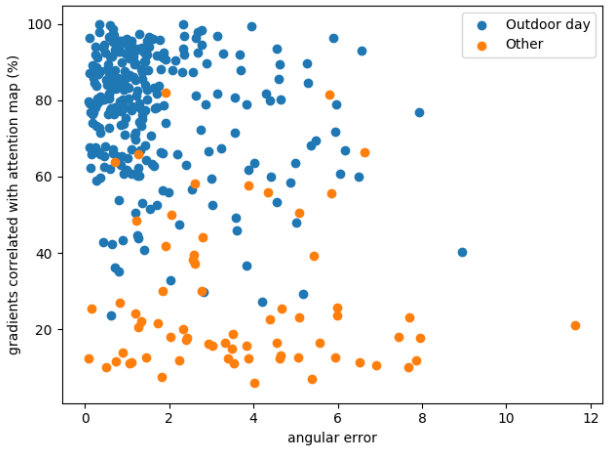


Figure 6. Dependency of estimation error on the amount of gradients captured by attention maps.

To additionally analyze the benefit of attention maps, several experiments were conducted. Since attention maps tend to correlate with image gradient, a network that has a gradient map instead of an attention map was used. Once each image gradients were computed, they were downsampled to the size of the attention map using the max-pooling operator. Max pooling was used to match the pooling operator in the VGG16 network. A new network where the attention mechanism was replaced with a gradient map corresponding to the given input was trained and evaluated. The second experiment was to omit the attention map from the network. In this experiment, the global estimate was obtained as the channel-wise sum of intermediate illumination estimations. Both networks have been trained with the same parameter set as the proposed attention-based network for one stage. The results of the experiments were compared with the proposed attention-based network after the first training stage and given in Table 4. It can be seen that the

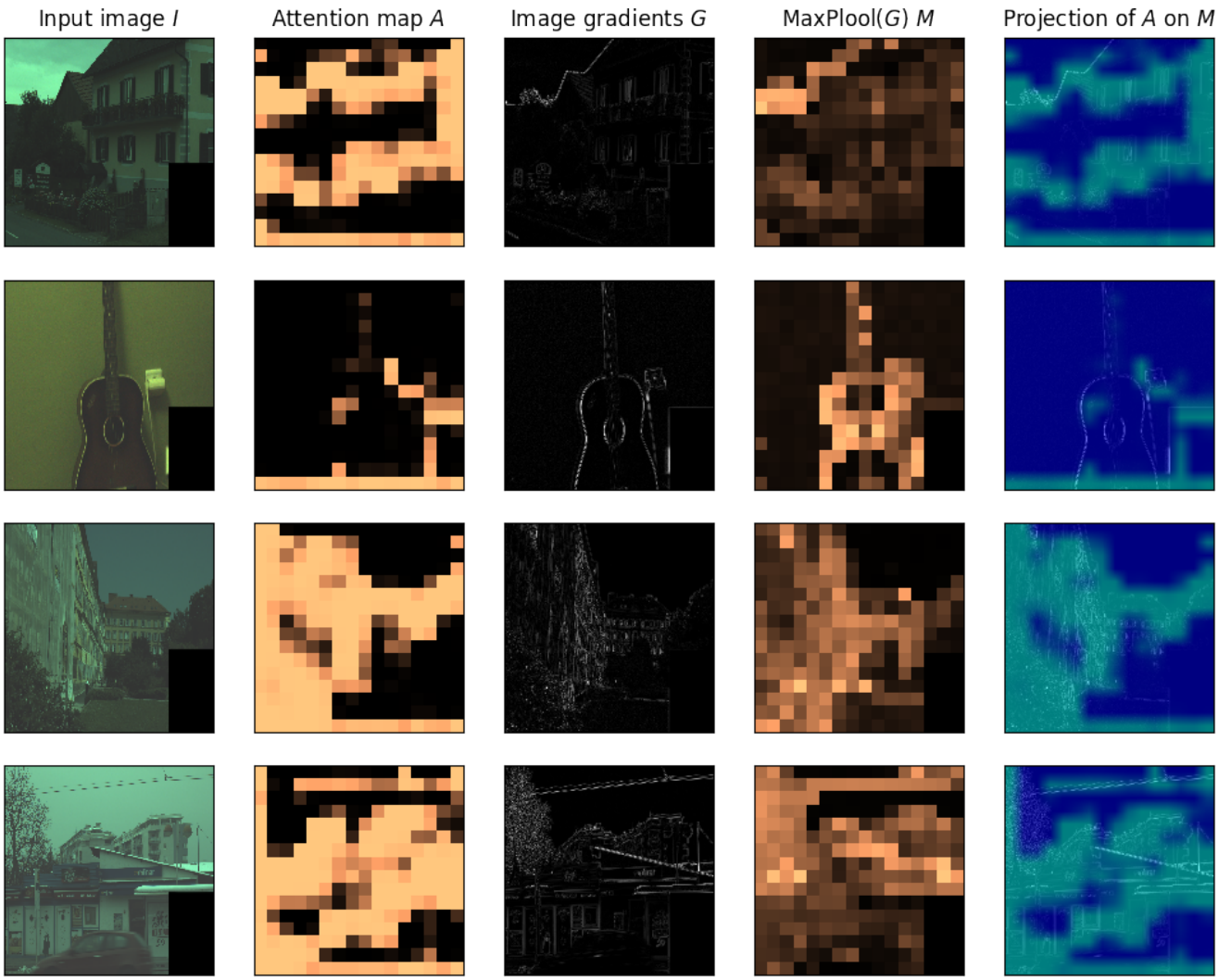


Figure 7. Comparison of the attention map and image gradients.

attention-based network outperforms both the gradient-based network and the network without any attention. The attention-based network can learn to distinguish between image regions with some content and image regions with content significant for illumination estimation.

Table 4. The comparison of angular error statistics obtained on Cube+ dataset for different network architectures.

Network type	Min	Mean	Med.	Tri.	Best 25%	Worst 25%	Max	Avg.
With attention	0.15	3.62	2.69	2.85	0.90	7.93	17.13	2.88
No attention	0.09	3.91	2.97	3.14	1.02	8.48	19.09	3.16
Gradient as attention	0.93	6.53	3.82	3.41	2.52	15.31	32.35	5.32

6 CONCLUSION

In this paper, a convolutional neural network with the attention mechanism for illumination estimation was proposed. It has been shown that the addition of the attention mechanism helps the network to estimate more accurate illumination vectors. Experimental results show that the attention mechanism does indeed guide the network to look for the regions in an image with more content. Even though regions of an image that are rich in content tend to correlate with higher gradient amplitudes, the proposed neural network does not only consider pure content but does a selection of the content. The attention mechanism can learn to select the content which is more important for illumination estimation and to discard regions of an image where content may misguide illumination estimation.

7 ACKNOWLEDGMENTS

This work has been supported by the Croatian Science Foundation under Project IP-06-2016-2092. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

8 REFERENCES

[1] A. Akbarinia and C. A. Parraga. 2018. Colour Constancy Beyond the Classical Receptive Field. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 9 (Sep. 2018), 2081–2094. https://doi.org/10.1109/TPAMI.2017.2753239

[2] Nikola Banić and Sven Lončarić. 2013. Using the Random Sprays Retinex Algorithm for Global Illumination Estimation. In *Proceedings of The Second Croatian Computer Vision Workshopn (CCVW 2013)*. University of Zagreb Faculty of Electrical Engineering and Computing, 3–7.

[3] Nikola Banić and Sven Lončarić. 2014. Color Rabbit: Guiding the Distance of Local Maximums in Illumination Estimation. In *Digital Signal Processing (DSP), 2014 19th International Conference on*. IEEE, 345–350.

[4] Nikola Banić and Sven Lončarić. 2014. Improving the White patch method by subsampling. In *Image Processing (ICIP), 2014 21st IEEE International Conference on*. IEEE, 605–609.

[5] Nikola Banić and Sven Lončarić. 2015. A Perceptual Measure of Illumination Estimation Error. In *VISAPP*. 136–143.

[6] Nikola Banić and Sven Lončarić. 2015. Color Cat: Remembering Colors for Illumination Estimation. *Signal Processing Letters, IEEE* 22, 6 (2015), 651–655.

[7] Nikola Banić and Sven Lončarić. 2015. Color Dog: Guiding the Global Illumination Estimation to Better Accuracy. In *VISAPP*. 129–135.

[8] Nikola Banić and Sven Lončarić. 2015. Using the red chromaticity for illumination estimation. In *Image and Signal Processing and Analysis (ISPA), 2015 9th International Symposium on*. IEEE, 131–136.

[9] Nikola Banić and Sven Lončarić. 2018. Green Stability Assumption: Unsupervised Learning for Statistics-Based Illumination Estimation. *Journal of Imaging* 4, 11 (2018), 127.

[10] Nikola Banić and Sven Lončarić. 2018. Unsupervised Learning for Color Constancy. In *VISAPP*. 181–188.

[11] Nikola Banić and Sven Lončarić. 2019. Blue Shift Assumption: Improving Illumination Estimation Accuracy for Single Image from Unknown Source. In *VISAPP*. 191–197.

[12] Kobus Barnard, Vlad Cardei, and Brian Funt. 2002. A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data. *Image Processing, IEEE Transactions on* 11, 9 (2002), 972–984.

[13] Jonathan T Barron. 2015. Convolutional Color Constancy. In *Proceedings of the IEEE International Conference on Computer Vision*. 379–387.

[14] Jonathan T Barron and Yun-Ta Tsai. 2017. Fast Fourier Color Constancy. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Computer Society Conference on*, Vol. 1. IEEE.

[15] Simone Bianco, Claudio Cusano, and Raimondo Schettini. 2015. Color Constancy Using CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 81–89.

[16] Gershon Buchsbaum. 1980. A spatial processor model for object colour perception. *Journal of The Franklin Institute* 310, 1 (1980), 1–26.

[17] Vlad C Cardei, Brian Funt, and Kobus Barnard. 2002. Estimating the scene illumination chromaticity by using a neural network. *JOSA A* 19, 12 (2002), 2374–2386.

[18] Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler. 2012. Color constancy with spatio-spectral statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34, 8 (2012), 1509–1519.

[19] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. 2014. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A* 31, 5 (2014), 1049–1058.

[20] Dongliang Cheng, Brian Price, Scott Cohen, and Michael S Brown. 2015. Effective learning-based illuminant estimation using simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1000–1008.

[21] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[22] Marc Ebner. 2007. *Color Constancy*. Wiley.

[23] Graham D Finlayson. 2013. Corrected-moment illuminant estimation. In *Proceedings of the IEEE International Conference on Computer Vision*. 1904–1911.

[24] Graham D Finlayson, Mark S Drew, and Brian V Funt. 1994. Color constancy: generalized diagonal transforms suffice. *JOSA A* 11, 11 (1994), 3011–3019.

[25] Graham D Finlayson and Elisabetta Trezzi. 2004. Shades of gray and colour constancy. In *Color and Imaging Conference*, Vol. 2004. Society for Imaging Science and Technology, 37–41.

[26] Brian Funt and Lilong Shi. 2010. The rehabilitation of MaxRGB. In *Color and Imaging Conference*, Vol. 2010. Society for Imaging Science and Technology, 256–259.

[27] Shao-Bing Gao, Kai-Fu Yang, Chao-Yi Li, and Yong-Jie Li. 2015. Color constancy using double-opponency. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 10 (2015), 1973–1985.

[28] Peter V Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. 2008. Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 1–8.

[29] Arjan Gijsenij and Theo Gevers. 2007. Color Constancy using Natural Image Statistics.. In *CVPR*. 1–8.

[30] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. 2011. Computational color constancy: Survey and experiments. *Image Processing, IEEE Transactions on* 20, 9 (2011), 2475–2489.

[31] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. 2012. Improving color constancy by photometric edge weighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34, 5 (2012), 918–929.

[32] Steven D Hordley. 2006. Scene illuminant estimation: past, present, and future. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 31, 4 (2006), 303–314.

[33] Yuanming Hu, Baoyuan Wang, and Stephen Lin. 2017. Fully Convolutional Color Constancy with Confidence-weighted Pooling. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Conference on*. IEEE, 4085–4094.

[34] Hamid Reza Vaezi Joze and Mark S Drew. 2013. Exemplar-based color constancy and multiple illumination. *IEEE transactions on pattern analysis and machine intelligence* 36, 5 (2013), 860–873.

[35] Hamid Reza Vaezi Joze, Mark S Drew, Graham D Finlayson, and Perla Aurora Troncoso Rey. 2012. The Role of Bright Pixels in Illumination Estimation. In *Color and Imaging Conference*, Vol. 2012. Society for Imaging Science and Technology, 41–46.

[36] Karlo Koščević, Nikola Banić, and Sven Lončarić. 2019. Color Beaver: Bounding Illumination Estimations for Higher Accuracy. In *VISAPP*. 183–190.

[37] Karlo Koščević, Marko Subašić, and Sven Lončarić. 2019. Attention-based Convolutional Neural Network for Computer Vision Color Constancy. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 372–377.

[38] Edwin H Land. 1977. *The retinex theory of color vision*. Scientific America.

[39] Kunpeng Li, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, and Yun Fu. 2018. Tell me where to look: Guided attention inference network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 9215–9223.

[40] Seoung Wug Oh and Seon Joo Kim. 2017. Approaching the computational color constancy as a classification problem through deep learning. *Pattern Recognition* 61 (2017), 405–416.

[41] Yanlin Qian, Said Pertuz, Jarno Nikkanen, Joni-Kristian K am ar ainen, and Jiri Matas. 2019. Revisiting Gray Pixel for Statistical Illumination Estimation. In *VISAPP*. 36–46.

[42] Wu Shi, Chen Change Loy, and Xiaoou Tang. 2016. Deep Specialized Network for Illuminant Estimation. In *European Conference on Computer Vision*. Springer, 371–387.

[43] Oleksii Sidorov. 2018. Artificial Color Constancy via GoogLeNet with Angular Loss Function. *arXiv preprint arXiv:1811.08456* (2018).

[44] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[45] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. 2007. Edge-based color constancy. *Image Processing, IEEE Transactions on* 16, 9 (2007), 2207–2214.

[46] Joost Van De Weijer, Cordelia Schmid, and Jakob Verbeek. 2007. Using high-level visual information for color constancy. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 1–8.

[47] Javier Vazquez-Corral, Maria Vanrell, Ramon Baldrich, and Francesc Tous. 2011. Color constancy by category correlation. *IEEE Transactions on image processing* 21, 4 (2011), 1997–2007.

[48] S. Woo, S. Lee, J. Yoo, and J. Kim. 2018. Improving Color Constancy in an Ambient Light Environment Using the Phong Reflection Model. *IEEE Transactions on Image Processing* 27, 4 (April 2018), 1862–1877. https://doi.org/10.1109/TIP.2017.2785290

[49] Kai-Fu Yang, Shao-Bing Gao, and Yong-Jie Li. 2015. Efficient illuminant estimation for color constancy using grey pixels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2254–2263.

Publication 3

Koščević, K., Subašić, M., Lončarić, S., "Deep Learning-Based Illumination Estimation Using Light Source Classification", IEEE Access, Vol. 8, 2020, pp. 84239-84247.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 3, 2020, accepted April 18, 2020, date of publication May 4, 2020, date of current version May 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992121

Deep Learning-Based Illumination Estimation Using Light Source Classification

KARLO KOŠČEVIĆ, MARKO SUBAŠIĆ, AND SVEN LONČARIĆ
Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia

Corresponding author: Karlo Koščević (karlo.koscevic@fer.hr)

This work was supported by the Croatian Science Foundation under Project IP-06-2016-2092.

ABSTRACT Color constancy is one of the key steps in the process of image formation in digital cameras. Its goal is to process the image so that there is no influence of illumination color on the colors of objects and surfaces. To capture the target scene colors as accurately as possible, it is crucial to estimate the illumination vector with high accuracy. Unfortunately, the illumination estimation is an ill-posed problem, and solving it most often relies on assumptions. To date, various assumptions have been proposed, which resulted in a wide variety of illumination estimation methods. Statistics-based methods have shown to be appropriate for hardware implementation, but learning-based methods achieve state-of-the-art results, especially those that use deep neural networks. The large learning capacities and generalization abilities of deep neural networks can be used to develop the illumination estimation methods, which are more general and precise. This approach avoids introducing many new assumptions, which often only work in some specific situations. In this paper, a new method for illumination estimation based on light source classification is proposed. In the first step, the set of possible illuminations is reduced by classifying the input image in one of three classes. The classes include images captured in outdoor scenes under natural illuminations, images captured in outdoor scenes under artificial illuminations, and images captured in indoor scenes under artificial illuminations. In the second step, a deep illumination estimation network, which is trained exclusively on images in the class that was predicted in the first step, is applied to the input image. Dividing the illumination space into smaller regions makes the training of illumination estimation networks simpler because the distribution of image scenes and illuminations is less diverse. The experiments on the Cube+ image dataset have shown the median illumination estimation error of 1.27°, which is an improvement of more than 25% compared to the use of the single network for all illuminations.

INDEX TERMS Color constancy, illumination estimation, classification, deep learning, white balancing, image enhancement.

I. INTRODUCTION

One of the first steps in the image formation pipeline of contemporary digital cameras is computational color constancy. Computational color constancy refers to the removal of the influence of illumination color on the colors of objects in the observed image scene. It is motivated by the ability of the human vision system (HVS) to perceive object color invariant to the illumination color, namely color constancy [1]. Computational color constancy is performed in two steps. The first step is the illumination estimation step, where one or multiple illumination vectors are estimated from the target image. Illumination vector is a three-component vector with one value for each color channel $c \in \{R, G, B\}$. In the second step, estimated illumination vectors are used to divide out the illumination and object reflectance. This step is called chromatic adaptation and it is achieved by multiplying each image pixel with a diagonal matrix with diagonal values $d_{11} = 1/e_R$, $d_{22} = 1/e_G$, and $d_{33} = 1/e_B$, where $\begin{bmatrix} e_R & e_G & e_B \end{bmatrix}^T$ is the illumination vector. After the chromatic adaptation is applied, the colors in the image should appear as if they are captured under the white illumination, i.e., the illumination where $e_R = e_G = e_B$.

More formally, in computational color constancy, the image formation model $\mathbf{f}$ with Lambertian assumption is mostly used and it can be given as [2]:

$$f_c(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x}) R(\lambda, \mathbf{x}) \rho_c(\lambda) d\lambda \tag{1}$$

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

where $c \in \{R, G, B\}$ is color channel, $\mathbf{x}$ is pixel location, $\mathbf{f}$ is pixel value, $I(\lambda, \mathbf{x})$ is the spectral distribution of the light source, $R(\lambda, \mathbf{x})$ is the surface reflectance, $\rho_c(\lambda)$ is the spectral sensitivity of the camera sensor for color channel c, and λ are the wavelengths in the visible light spectrum ω. From the image formation pipeline, it can be seen that colors in the image are a combination of three physical values. These are the spectral distribution of the light source, spectral reflectance properties of surfaces in the image scene, and the sensitivity of the camera sensor. Additionally, it can be seen that the illumination captured by the camera is a function of the spectral distribution of the light source and the sensitivity of the camera sensor for different wavelengths in the visible light spectrum. Therefore, in an ideal case, illumination vector $\mathbf{e}$ can be computed as:

$$\mathbf{e}(\mathbf{x}) = \int_{\omega} I(\lambda, \mathbf{x})\boldsymbol{\rho}(\lambda)d\lambda. \tag{2}$$

And when it is assumed that illumination is the same in the whole image scene, illumination vector $\mathbf{e}$ is invariant of pixel position $\mathbf{x}$. Therefore, for global illumination estimation methods illumination vector is given as:

$$\mathbf{e} = \int_{\omega} I(\lambda)\boldsymbol{\rho}(\lambda)d\lambda. \tag{3}$$

The major drawback of illumination estimation is that it is an ill-posed problem. Because most often both $I(\lambda)$ and $\boldsymbol{\rho}(\lambda)$ are not known, and only image pixel values $\mathbf{f}$ are known, there is an infinite number of possible illumination and surface reflectance combinations for a given image $\mathbf{f}$. To overcome this issue, different assumptions for the illumination estimation have been proposed, yielding a wide variety of illumination estimation methods.

It has been shown in previous research that both illumination estimation techniques and scene classification methods have been applied jointly in many color image processing procedures. They were combined either by using image classification to improve illumination estimation or by using illumination estimation to perform image classification [3], [4]. In this paper, an illumination estimation method that relies on image classification is proposed. Once the input image is classified based on the scene content and illumination type, it is proposed to apply a deep illumination estimation network specialized for the class of images to which the input image was classified. Classification in three classes is performed by combining the classification of image scenes and the classification of illuminations.

The conducted experimental work has shown that separating the possible illumination space into smaller regions and applying a specialized estimator for each region yields more accurate estimations with the median estimation error reduced by more than 25%.

The rest of the paper is structured as follows: Section II gives a short overview of existing illumination estimation methods, in Section III the motivation for the proposed method is given, Sections IV and V describe the proposed method and experimental results, respectively, and a conclusion is provided in Section VI.

II. RELATED WORK

The illumination estimation methods can be divided into three groups [2]. In the first group are the methods which exploit low-level image statistics and features, such as per channel mean and max or *n*th order image derivations. These methods are referred to as statistics-based methods. They usually use a fixed set of parameters and do not require model training. Low computational complexity and high execution speed make them suitable for hardware implementation. Many statistics-based methods are a direct variation of Gray-World assumption that the average of an image is gray, i.e., the mean of all three channels is equal. Such methods include Gray-World [5], Shades-of-Gray [6], 1st and 2nd order Gray-Edge [7], Weighted Gray-Edge [8]. A slightly different subset of methods that can still be derived from the Gray-World assumption is White-Patch method [9], [10] and its improvements [11]–[13]. Statistics-based methods also include methods which use bright pixels [14], gray pixels [15] or bright and dark colors [16], methods which exploit illumination statistics perception [17] or expected illumination statistics [18]. In the second group are the methods which require training of an illumination estimation model. Thus they are referred to as learning-based methods. Once learned, the model is then used to estimate illuminations, which are correlated with the training data distribution. To train a model with good generalization properties, these methods require larger datasets. Due to the training process, larger datasets and more complex structures, learning-based methods are computationally demanding and most often take a longer time to execute. However, in the end, they produce the most accurate illumination estimations. Learning-based methods are methods based on neural networks [19], high-level visual information [20], natural image statistics [21], Bayesian learning [22], spatio-spectral learning [23], methods restricting the illumination solution space [24]–[27], using color moments [28], regression trees with simple features from color distribution statistics [29], spatial localizations [30], [31], channel-wise pooling the responses of double-opponency cells in LMS color space [32], detecting gray pixels with specific illuminant-invariant measures in logarithmic space [33], modelling color constancy by using the overlapping asymmetric Gaussian kernels with surround pixel contrast based sizes [34], finding paths for the longest dichromatic line produces by specular pixels [35]. Following the classification of illumination estimation methods in [2], gamut-based methods [36]–[38] can be considered as a separate group of illumination estimation methods. Even though they are in some way learning-based methods as well, they had a great impact on the field.

An important type of learning-based methods are deep learning methods. Deep learning became the state-of-the-art in many fields, such as natural language processing, computer vision, finances, advertising, and others. Since the publication

of the AlexNet [39], along with image classification, convolutional neural networks have successfully been applied in many fields of computer vision, including object recognition [40], object detection [41], image segmentation [42], etc. One of the first attempts to apply a convolutional neural network for computational color constancy was in [43]. A deeper convolutional neural network with a more complex training procedure for illumination estimation was proposed in [44]. In [45], two convolutional neural networks have been used for illumination estimation with one network computing multiple estimations and the other selecting for the plausible ones. In [46], a convolutional neural network was used to cast the illumination estimation problem into an illumination classification problem, which computes the global illumination based on the results of k-means clustering and classification probabilities. In [47]–[49], convolutional neural networks with weighted local illumination pooling have been proposed. A major drawback of the aforementioned deep-learning methods is that they are sensor-dependent. In contrast, in [50], deep learning was used to map the input images in a sensor-invariant color space, which enables sensor-independent illumination estimation.

In [3], classification-based illumination estimation is proposed. The authors distinguish between indoor and outdoor images based on the fact that different illuminations and scene content are characteristic for each class. The authors have shown that classification-based methods improve the illumination estimation, especially when indoor-outdoor classification with the addition of uncertainty class is used to determine which illumination estimation method to apply for the input image.

In contrast to [3] and this paper, which are classifying the input image based on its features to reduce the illumination space before the illumination estimation step, in [4], the opposite was proposed, i.e., the illumination estimation has been used for indoor-outdoor image classification. Considering the assumption that outdoor images are usually captured in blueish illuminations and indoor images in reddish illuminations, the authors proposed to apply an illumination estimation method to the input image and classify the image as indoor or outdoor considering the position of the estimated illumination in the chromaticity plane.

III. MOTIVATION

When capturing an image with a digital camera, the target scene can be illuminated with many different light sources. Some of these light sources can produce illuminations similar to the white illumination, and they do not affect pixel color significantly. However, for instance, in indoor environments, it is common that the illumination color significantly differs from white, i.e., values between red, green, and blue color channels are different. Such illuminations cause considerable color bias in the image towards the color of the illumination. The difference between an image captured in a light that is close to white and an image captured in yellow light is shown in Figure 1. It can be observed that the yellow illumination

(a) (b)

FIGURE 1. The difference between images captured under near-white illumination (a) and yellow illumination (b).

in Figure 1b has a great impact on pixel colors. A similar effect can be observed with other artificial light sources, e.g., when taking a picture of an outdoor scene in the night when street lights are turned on.

The most obvious division of illuminations can be made by dividing image scenes into outdoor and indoor classes [3]. Illuminations in outdoor scenes are a combination of natural effects, and these illuminations tend to occupy space around the white illumination in the *rb*-chromaticity plane. On the other hand, in indoor scenes, the majority of illuminations are produced by artificial light sources. These illuminations can vary significantly from those close to natural illuminations to the extreme case of illuminations produced by disco bulbs. However, illuminations in outdoor scenes tend to be close to the white illumination only in daytime conditions. When captured during nighttime, it is most likely that the scene was illuminated with some artificial light source, which differs from light sources in outdoor scenes captured during the daytime and most common light sources in indoor scenes. Therefore, an additional class of illuminations can be introduced, leading to a total of three classes of illuminations:

- outdoor natural illuminations
- outdoor artificial illuminations
- indoor artificial illuminations.

Separating illuminations in multiple clusters and applying a different illumination estimator for each cluster can lead to better estimations since each estimator can be specialized to recognize illuminations in its corresponding cluster. Having a less variable distribution of illuminations for each estimator should be beneficial when training each estimator separately than training one estimator on a dataset with a high variability of image scenes and few different clusters of corresponding illuminations. Additionally, the computational cost of the classification of image scenes into three clusters and training three specialized estimators should be compensated by the fact that the maximum estimation error should be lower than in the case of using one general estimator.

IV. THE PROPOSED METHOD

In this paper, before the illumination estimation, it is proposed to classify an input image into one of three classes listed in Section III. Based on the classification result, the illumination is estimated using the estimator specialized for images in the corresponding class. The pseudocode of the proposed method

is given in Algorithm 1. Both classification and illumination estimation steps are described in more detail in the following sections.

Algorithm 1 Illumination Estimation Using Light Source Classification

Input: image **I**
Output: illumination vector **e**

1: $CN = ClassificationNet()$
2: $IE_i = IlluminationEstNet_i(), i \in \{0, 1, 2\}$
3: $(p_0, p_1, p_2) = CN.predict(\mathbf{I})$ ▷ Class probabilities
4: $j = \arg\max_i(p_i)$
5: $\mathbf{e} = IE_j.predict(\mathbf{I})$

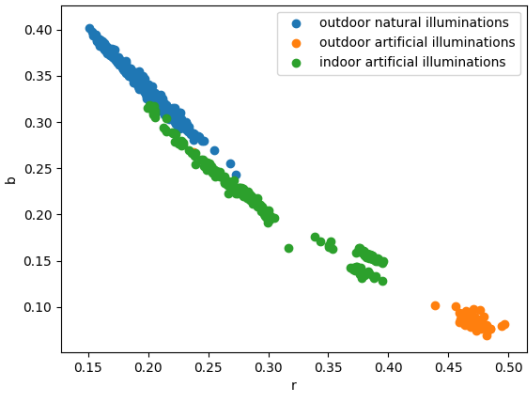


FIGURE 2. Illumination clusters in the Cube+ dataset shown in the form of *rb*-chromaticities.

A. IMAGE CLASSIFICATION

For image classification, a deep neural network is proposed. In the field of illumination estimation, the largest datasets have a few hundred samples, which is, in terms of state-of-the-art image classification, which uses deep neural networks, an insignificant number of samples. The VGG16 network [40] pre-trained for image classification was used to overcome this drawback. Fully connected layers and last convolutional block in the VGG16 network were replaced with a smaller stack of fully connected layers. The newly added stack is structured as follows:

- Flatten Layer
- FC Layer, 256 output neurons
- FC Layer, 128 output neurons
- FC Layer, 64 output neurons
- FC Layer, 3 output neurons,

where the Flatten layer reshapes the feature map produced by the last convolutional layer in the 4^{th} convolutional block of the VGG16 network to match the shape of the following fully connected layer, and FC stands for fully connected. The last fully connected layer has three output neurons, each for one of three target classes. All fully connected layers use the ReLu activation function, except the last fully connected layer, where softmax function is used to compute the probability distribution over target classes. The network structure was experimentally determined and confirmed.

B. ILLUMINATION ESTIMATION

For illumination estimation, the convolutional neural network proposed in [49] was used. It is a fully convolutional neural network. It uses pre-trained VGG16 architecture as a feature extractor on top of which the attention mechanism is placed. The addition of the attention mechanism enables the network to filter the local illumination estimations by considering the usefulness of the information in the corresponding area of an image. Therefore, the network can distinguish between ambiguous and informative regions of an image, where, in the sense of illumination estimation, ambiguous are regions such as flat single-color surfaces.

In this paper, it is proposed to classify images into three classes. Each class has a different set of illuminations. Therefore, three instances of the above-mentioned deep neural network for illumination estimation are trained separately. The first instance is trained to estimate the illuminations on images that are captured in outdoor scenes during the daytime, i.e., under natural illuminations. The second instance is trained to estimate illuminations on images that are captured in outdoor scenes illuminated with artificial light sources. Finally, the third instance is trained on images with indoor scenes where all illuminations are artificial.

V. EXPERIMENTAL RESULTS

A. DATASET PREPARATION

The proposed method was evaluated on the Cube+ dataset [51]. Cube+ is a dataset of 1707 images with a known ground-truth illumination vector for each image, and thus it is appropriate for the evaluation of illumination estimation methods. What makes this dataset significant is not only diverse image scenes but also a very broad distribution of illuminations. Illuminations that occur in the Cube+ dataset can be divided into three clusters, i.e., natural illuminations in outdoor scenes, artificial illuminations in outdoor scenes, and artificial illuminations in indoor scenes. Natural illuminations in outdoor scenes are captured during the daytime, whereas artificial illuminations are captured in the scenes where some artificial light source is present but the corresponding scenes vary between outdoor and indoor scenes. In total, there are 1365 samples with natural outdoor illuminations, 52 samples with artificial outdoor illuminations, and 290 samples with artificial indoor illuminations. In Figure 2, an example distribution of illuminations given as *rb*-chromaticities and split into three clusters is shown. Chromaticities for red and blue channel, i.e., *rb*-chromaticities are calculated as: $r = R/(R+G+B)$, $b = B/(R+G+B)$, where R, G, and B are red, green, and blue pixel intensities, respectively. The difference between images illuminated with a natural outdoor light source, artificial outdoor light source, and artificial indoor light source can be seen in Figure 3.

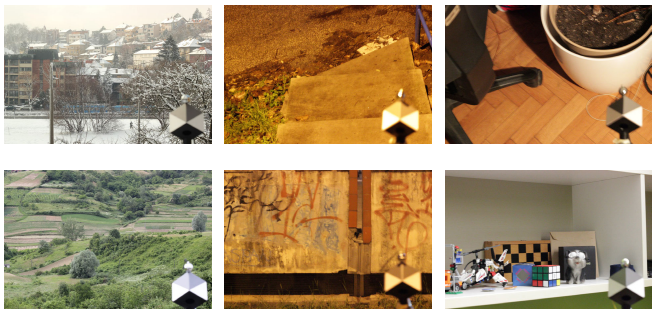


FIGURE 3. Examples of images illuminated with a natural outdoor light source (left column), artificial outdoor light source (middle column), and images illuminated with an artificial indoor light source (right column).

In the following sections, the proposed classes will be referred to as:

- C_0 represents the cluster with outdoor scenes in artificial illuminations
- C_1 represents the cluster with outdoor scenes in natural illuminations
- C_2 represents the cluster with indoor scenes in artificial illuminations.

Cube+ dataset was used to train both illumination estimation and image classification, with a different configuration for each task. For the classification network, the whole Cube+ dataset was used, i.e., all 1707 images. Images were resized to the target size of 224×224 pixels and used in their raw format. Each image was labeled with the corresponding class label, which was then used as ground-truth data. The dataset was split into train and test sets in a ratio of 4 to 1, respectively. The proposed class split in the Cube+ dataset results with imbalanced classes. Namely, when considering both train and test splits together, class C_1 has 1365 samples, whereas classes C_0 and C_2 have only 52 and 290 samples, respectively. Therefore, the data in the train set was balanced. A subset with 20% of samples was first separated from the train set for validation. From the remaining data in the train set, 500 random samples were extracted from class C_1, and the remaining two classes, i.e., class C_0 and class C_2 have been oversampled to match the new number of samples in class C_1. This resulted in a test set with a total of 1500 samples (500 samples per class). The test set was used in its original form.

For regression networks, the Cube+ dataset was split into three parts based on the type of ground-truth illumination. The first part contained only samples in which ground-truth illumination is natural outdoor, the second part contained only samples with accompanied ground-truth illumination from a cluster with artificial outdoor illuminations, and the third part contained only samples which ground-truth is artificial indoor. Accordingly, the first part of the dataset was used to train a network for natural outdoor illumination estimation, the second part was used to train a network for artificial outdoor illumination estimation, and the third part was used to train a network for artificial indoor illumination estimation. The same as for the classification, images were resized to the target size of 224×224 pixels, and each regression dataset was split into train and test sets in a ratio of 4 to 1. Ground-truth data for the regression were ground-truth illumination vectors from the Cube+ dataset.

B. PERFORMANCE METRICS

Illumination estimation method performance for an input image is usually given in the form of the angle between the ground-truth illumination vector and the estimated illumination vector, namely angular error. Different summary statistics are then used to combine individual performances and indicate the overall performance on a dataset. Most often, statistics are min, max, median, mean, best 25%, worst 25%, trimean, and average. Trimean can be calculated as $(Q_1 + 2 \times Q_2 + Q_3)/4$, where Q_1, Q_2, and Q_3 are first, second and third quartile, respectively. The average is the geometric mean of all other mentioned statistics, and it is introduced in [30].

In this paper, the above-mentioned summary statistics were used to evaluate the performance of the proposed illumination estimation method. Emphasis was placed on the median value since the distribution of the angular error is not symmetrical.

C. TRAINING SETUP

The training setup, which includes learning rates, momentum, batch size, and the number of epochs, and which is described in this section, was experimentally determined.

1) IMAGE CLASSIFICATION

VGG16 network was initialized with the weights obtained by training the network for classification on the ImageNet dataset [52]. Fully connected layers in the newly added stack were initialized using the Xavier initialization [53]. During the training, weights in all layers (both VGG16 and the added fully connected stack) have been updated. The network was trained for 20 epochs with 32 samples in the mini-batch. Stochastic gradient optimization with the learning rate of 0.001 and momentum 0.90 was used. Categorical cross-entropy was used as the loss function. The balanced train set described in Section V-A was used to train the classification network.

2) ILLUMINATION ESTIMATION

To obtain the best overall accuracy, the parameters in each illumination estimation network were fine-tuned on the corresponding class of images and illuminations. All networks have been initialized in the same fashion. The initial layer weights were acquired from [49]. All networks have been optimized using the stochastic gradient descent with momentum. The following loss function was used [54]:

$$L(\mathbf{e}, \hat{\mathbf{e}}) = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{\mathbf{e}_i \cdot \hat{\mathbf{e}}_i}{\|\mathbf{e}_i\| \, \|\hat{\mathbf{e}}_i\|} \right), \tag{4}$$

where ith ground-truth illumination vector and estimated illumination vector are denoted as $\mathbf{e}_i$ and $\hat{\mathbf{e}}_i$, respectively, N denotes the number of samples, '$\cdot$' is the vector dot product, and $\|.\|$ is the vector $L2$ norm. In the following paragraphs, the parameters specific for each illumination estimation network are given.

The first illumination estimation network was trained for images with outdoor scenes captured under natural illuminations, i.e., in the daytime. The learning rate and momentum were 0.01 and 0.95, respectively. The network was trained with 10 samples in the mini-batch for 100 epochs. The first four convolutional blocks in the VGG16 network were frozen, i.e., the weights in those convolutional blocks were not updated. Whereas, the weights in the 5$^{\text{th}}$ convolutional block and the weights in the attention mechanism were fine-tuned.

The second illumination estimation network was trained for images capturing outdoor scenes as well, but this time the illuminations were artificial. Stochastic gradient descent was initialized with the learning rate of 0.001 and momentum 0.95. Mini-batch size was 32, and the number of training epochs was 200. For this class of images, the whole network architecture was trained, which includes both the VGG16 network and the attention mechanism.

The final network was trained on images captured in indoor scenes. This class contains only artificial illuminations. Momentum was set to 0.99, and the learning rate to 0.001. The number of training epochs and the mini-batch size was 100 and 10, respectively. Same as for the illumination estimation network for the previous class of images, the weights in all layers have been updated during the training.

Each illumination estimation network was used on a different distribution of input images and ground-truth illuminations. Therefore, when using the same set of parameters for all networks, it is plausible that for a given distribution of input images and ground-truth illuminations, the achieved result if not optimal. In other words, to obtain illumination estimations as accurately as possible, each network was trained using the optimal set of parameters for the corresponding class split.

D. METHOD ACCURACY

1) NATURAL-ARTIFICIAL ILLUMINATION CLASSIFICATION

The major drawback of the Cube+ dataset is that the number of samples between proposed classes varies significantly. However, after balancing the train set, the accuracy given in Table 1 has been achieved. It should be stressed out that the test set contained only 9 samples from class C_0 and 7 samples have been correctly classified.

TABLE 1. The classification accuracy.

Class	Accuracy (%)
C_0	77.78
C_1	98.53
C_2	93.33
total	97.08

TABLE 2. Illumination estimation results for each proposed class of input samples, as well as their combined performance compared to the baseline results.

Class	Min	Mean	Med.	Tri.	Best 25%	Worst 25%	Max	Avg.
C_0	0.65	1.81	1.41	1.47	0.84	3.36	4.25	1.60
C_1	0.07	1.71	1.20	1.33	0.40	3.90	9.86	1.33
C_2	0.18	1.79	1.34	1.45	0.53	3.84	8.07	1.48
combined	0.07	1.73	1.25	1.36	0.42	3.85	9.86	1.36
baseline	0.05	2.34	1.68	1.90	0.50	5.28	15.02	1.82

2) ILLUMINATION ESTIMATION

The baseline for the evaluation of the proposed illumination estimation on data splits, and the parameter setup described in Section V-C.2 is the illumination estimation network trained on the whole Cube+ dataset, i.e., without data splitting. The network has the same architecture as the proposed illumination estimation networks in Section IV-A. The training set contained 80% of the data in the Cube+ dataset. The remaining 20% were used as test data to compute the baseline results. Both train and test sets contained images from classes C_0, C_1, and C_2. The following training configuration was used: stochastic gradient descent with a learning rate of 0.01 and momentum 0.95, mini-batch size of 10 samples, and 100 training epochs. Only the attention mechanism weights and the weights in the 5$^{\text{th}}$ convolutional block of the VGG16 network have been updated. For initialization, the pre-trained weights from [49] were used.

In Table 2, the results of individual illumination estimation networks with the parameter setup from Section V-C.2 are compared with the baseline. In the row labeled *combined*, the combined performance of individual illumination estimation networks is given. The experimental results confirm that the overall illumination estimation accuracy can be improved if the data is carefully split into smaller clusters. Usually, the median is considered the most important statistic in illumination estimation, and indeed, using the proposed approach, its value is improved. However, the most significant improvement is achieved in terms of maximum estimation error. It has been reduced by more than 30%. This confirms that having multiple distinct illumination estimators, which cover different illumination regions, is beneficial over one illumination estimation network that searches the whole illumination space.

The proposed illumination estimation with three clusters has also been compared with the clustering in two classes. Two combinations have been researched. For the first combination, the scene type was considered, which resulted in the following clusters: a cluster with outdoor images and a cluster with indoor images. For this kind of split, in the outdoor cluster, both artificial and natural illuminations exist,

TABLE 3. The comparison of the results obtained by illumination estimation in three clusters (3C) and by illumination estimation in two clusters based on scene type (2C scn) and illumination type (2C ill).

Class	Min	Mean	Med.	Tri.	Best 25%	Worst 25%	Max	Avg.
2C scn	0.01	4.65	1.65	1.98	0.45	14.30	30.91	2.50
2C ill	0.07	1.83	1.29	1.42	0.43	4.04	13.36	1.42
3C	0.07	1.73	1.25	1.36	0.42	3.85	9.86	1.36

while the indoor cluster is the same as it was in the proposed approach. In the second combination, clustering was done based on the illumination type. The first cluster contained only outdoor images captured under natural illuminations, and in the second cluster, images captured under artificial illuminations in both indoor and outdoor scenes have been contained. The proposed clustering approach outperformed both of these clustering combinations. One plausible explanation is that the illumination distributions are more compact when three class split is used instead of any of the two-class splits. For instance, in the second combination, where indoor and outdoor artificial illuminations are combined, the illumination distribution is very diverse. It contains illuminations from near-white in indoor scenes to strong, distinct yellow illuminations in outdoor scenes. The comparison of the proposed approach in its combined form and clustering in two classes is given in Table 3.

3) COMPARISON WITH OTHER ILLUMINATION ESTIMATION METHODS

In Table 4, the overall results of the proposed approach are given and can be compared with other illumination estimation methods evaluated on the Cube+ dataset. The results have been obtained by first classifying the input images and then applying the illumination estimation network trained for the predicted class of images. Due to the classification error, the overall illumination estimation error is slightly higher than in Table 2. However, it has been shown that, even though the input images are misclassified, the illumination estimation networks tend to estimate the illuminations which are close to the actual ground-truth distribution of illuminations for the corresponding images.

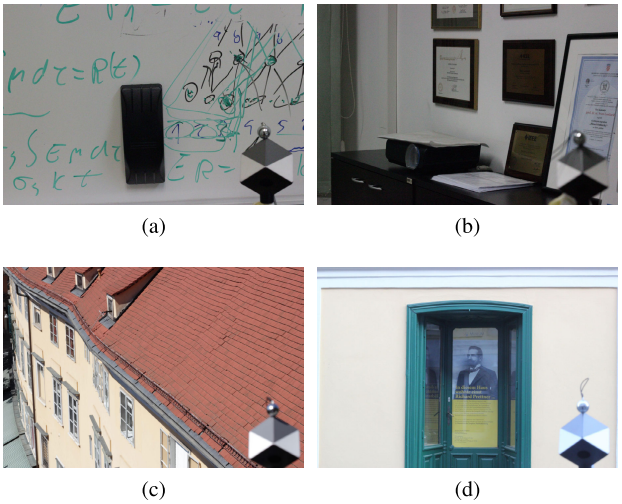
On a test set of 342 images, 10 images were misclassified. In Table 5, the angular error statistics obtained on images that were misclassified are compared with the angular errors that would be obtained if a classifier with 100% accuracy is used, i.e., if all images were classified in their true class. It can be seen that the classification is crucial for good illumination estimation since the error values on misclassified examples are much higher than the overall results. Even though the classification helps to reduce the illumination space, it can be the major limitation of the proposed method. It has been shown that for a given image, the method tends to estimate the illuminations as close as possible to their ground-truth, even when misclassified, but if the ground-truth class and predicted class are not adjacent, the estimation error can be high. Examples of misclassified images are shown in Figure 4. One plausible explanation for misclassification is that these samples have near-white illumination and the classifier is not able to distinguish their class based only on the scene content.

TABLE 4. The comparison of angular error statistics of different color constancy methods on the Cube+ dataset [51] (lower Avg. is better).

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
White-Patch [10]	9.69	7.48	8.56	1.72	20.49	7.38
Gray-world [5]	7.71	4.29	4.98	1.01	20.19	5.08
Double-opponency (max pooling) [32]	6.76	3.44	4.15	0.79	18.54	4.27
Using gray pixels [33]	6.65	3.26	3.95	0.68	18.75	4.05
Color Tiger [51]	3.91	2.05	2.53	0.98	10.00	2.88
Color Mule [55]	5.16	1.30	2.03	0.25	16.93	2.25
Shades-of-Gray [6]	2.59	1.73	1.93	0.46	6.19	1.90
2nd-order Gray-Edge [7]	2.50	1.59	1.78	0.48	6.08	1.83
1st-order Gray-Edge [7]	2.41	1.52	1.72	0.45	5.89	1.76
Color Dog [26]	3.32	1.19	1.60	0.22	10.22	1.70
General Gray-World [56]	2.38	1.43	1.66	0.35	6.01	1.64
Attention CNN [49]	2.05	1.32	1.53	0.42	4.84	1.54
Proposed approach	1.86	1.27	1.39	0.42	4.31	1.43
RGB Attention CNN [48]	1.95	1.13	1.37	0.32	4.92	1.37
Color Beaver (Gray-world) [27]	1.49	0.77	0.98	0.21	3.94	0.99

TABLE 5. The comparison of estimation errors obtained for misclassified samples with the estimation errors for the perfectly accurate classification.

	Min	Mean	Med.	Max
misclassified	1.01	7.35	8.84	12.83
perfectly classified	0.62	3.38	2.79	9.86



(a) (b) (c) (d)

FIGURE 4. Examples of misclassified images. True classes of example images are C_2 (a), C_2 (b), C_1 (c), and C_1 (d), and the network classified them as C_1 (a), C_1 (b), C_2 (c), and C_2 (d).

VI. CONCLUSION

In this paper, a new light source classification-based illumination estimation method is proposed. It uses deep neural

networks to classify input images and estimate illumination vectors. Three clusters, i.e., classes, are proposed: cluster with outdoor scenes in natural illuminations, cluster with outdoor scenes in artificial illuminations, and cluster with indoor scenes in artificial illuminations. For each cluster, a separate deep illumination estimation network is trained. With the experimental results, it has been confirmed that training multiple illumination estimation networks using smaller portions of illumination space outperforms a single illumination estimation network. The experiments have shown that the clustering of the illumination space has to be performed carefully and considering not only pure illuminations but features such as scene content as well.

ACKNOWLEDGMENT

The authors would like to thank NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

[1] M. Ebner, *Color Constancy* (The Wiley-IS&T Series in Imaging Science and Technology). Hoboken, NJ, USA: Wiley, 2007.
[2] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.
[3] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Improving color constancy using indoor–outdoor image classification," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2381–2392, Dec. 2008.
[4] N. Banić and S. Lončarić, "Illumination estimation is sufficient for indoor-outdoor image classification," in *Proc. German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2018, pp. 473–486.
[5] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
[6] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imag. Conf.*, 2004, pp. 37–41.
[7] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.
[8] A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 918–929, May 2012.
[9] E. H. Land, *The Retinex Theory of Color Vision*. New York, NY, USA: Scientific America., 1977.
[10] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. Color Imag. Conf.*, 2010, pp. 256–259.
[11] N. Banić and S. Lonäariä, "Using the random sprays retinex algorithm for global illumination estimation," in *Proc. 2nd Croatian Comput. Vis. Workshopn (CCV)*, Sep. 2014, pp. 3–7.
[12] N. Banic and S. Loncaric, "Color rabbit: Guiding the distance of local maximums in illumination estimation," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 345–350.
[13] N. Banic and S. Loncaric, "Improving the white patch method by subsampling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 605–609.
[14] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey, "The Role of Bright Pixels in Illumination Estimation," in *Proc. Color Imag. Conf.*, 2012, pp. 41–46.
[15] Y. Qian, S. Pertuz, J. Nikkanen, J.-K. Kämäräinen, and J. Matas, "Revisiting gray pixel for statistical illumination estimation," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 36–46.
[16] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: Why spatial-domain methods work and the role of the color distribution," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 31, no. 5, pp. 1049–1058, May 2014.
[17] N. Baniá and S. Lonäari, "Blue shift assumption: Improving illumination estimation accuracy for single image from unknown source," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 191–197.
[18] N. Baniá and S. Lonäari, "Green stability assumption: Unsupervised learning for statistics-based illumination estimation," *J. Imag.*, vol. 4, no. 11, p. 127, 2018.
[19] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 19, no. 12, pp. 2374–2386, Dec. 2002.
[20] J. van de Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
[21] A. Gijsenij and T. Gevers, "Color Constancy using Natural Image Statistics," in *Proc. CVPR*, 2007, pp. 1–8.
[22] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[23] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1509–1519, Aug. 2012.
[24] N. Banic and S. Loncaric, "Color cat: Remembering colors for illumination estimation," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 651–655, Jun. 2015.
[25] N. Banic and S. Loncaric, "Using the red chromaticity for illumination estimation," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2015, pp. 131–136.
[26] N. Banic and S. Loncaric, "Color Dog–Guiding the global illumination estimation to better accuracy," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, 2015, pp. 129–135.
[27] K. Koščević, N. Banić, and S. Lončarić, "Color beaver: Bounding illumination estimations for higher accuracy," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 183–190.
[28] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1904–1911.
[29] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Effective learning-based illuminant estimation using simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1000–1008.
[30] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.
[31] J. T. Barron and Y.-T. Tsai, "Fast Fourier color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 886–894.
[32] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1973–1985, Oct. 2015.
[33] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2254–2263.
[34] A. Akbarinia and C. A. Parraga, "Colour constancy beyond the classical receptive field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2081–2094, Sep. 2018.
[35] S.-M. Woo, S.-H. Lee, J.-S. Yoo, and J.-O. Kim, "Improving color constancy in an ambient light environment using the phong reflection model," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1862–1877, Apr. 2018.
[36] D. A. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 5–35, Aug. 1990.
[37] K. Barnard, "Improvements to gamut mapping colour constancy algorithms," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2000, pp. 390–403.
[38] G. D. Finlayson, S. D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 93–109, Apr. 2006.
[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[42] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2015, pp. 3431–3440.
[43] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 81–89.
[44] Z. Lou, T. Gevers, N. Hu, and M. P. Lucassen, "Color constancy by deep learning," in *Proc. Brit. Mach. Vis. Conf.*, Oct. 2015, pp. 1–76.

[45] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for Illuminant estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 371–387.
[46] S. W. Oh and S. J. Kim, "Approaching the computational color constancy as a classification problem through deep learning," *Pattern Recognit.*, vol. 61, pp. 405–416, Jan. 2017.
[47] Y. Hu, B. Wang, and S. Lin, "FC^4: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4085–4094.
[48] K. Koscevic, M. Subasic, and S. Loncaric, "Attention-based convolutional neural network for computer vision color constancy," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 372–377.
[49] K. Košćević, M. Subaăić, and S. Lonćarić, "Guiding the illumination estimation using the attention mechanism," in *Proc. 2nd Asia Pacific Inf. Technol. Conf.*, New York, NY, USA, Jan. 2020, p. 143–149, doi: 10.1145/3379310.3379329.
[50] M. Afifi and M. S. Brown, "Sensor-independent illumination estimation for DNN models," 2019, *arXiv:1912.06888*. [Online]. Available: http://arxiv.org/abs/1912.06888
[51] N. Baniv and S. Lonärić, "Unsupervised learning for color constancy," in *Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2018, pp. 181–188.
[52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[53] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[54] O. Sidorov, "Artificial color constancy via GoogLeNet with angular loss function," 2018, *arXiv:1811.08456*. [Online]. Available: http://arxiv.org/abs/1811.08456
[55] N. Banić and S. Lonärić, "A perceptual measure of illumination estimation error," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, 2015, pp. 136–143.
[56] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 972–984, Sep. 2002.

KARLO KOŠČEVIĆ received the B.Sc. and M.Sc. degrees in computer science, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree in technical sciences in the scientific field of computing with the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. His research interests include image processing, image analysis, and deep learning. His current research interest includes color constancy with a focus on learning-based methods for illumination estimation.

MARKO SUBAŠIĆ received the Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. Since 1999, he has been working at the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently as an Associate Professor. He teaches several courses at the graduate and undergraduate levels. His research interests include image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is a member of the IEEE of Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.

SVEN LONČARIĆ received the B.Sc., M.Sc., and Ph.D. degrees, in 1985, 1989, and 1994, respectively. After earning his doctoral degree, he continued his academic career as an Assistant Professor with the University of Zagreb. He was an Assistant Professor with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ, USA, from 2001 to 2003. His main research interests include medical image analysis and biomedical imaging. Together with his students and collaborators, he has published more than 200 publications in scientific peer-reviewed journals and has presented his work at international conferences. He was a recipient of the 2014 Annual Award for Scientific Achievements of the University of Zagreb Faculty of Electrical Engineering and Computing.

Publication 4

Koščević, K., Banić, N., Lončarić, S., "Color Beaver: Bounding Illumination Estimations for Higher Accuracy", Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 2019, pp. 183-190.

Color Beaver: Bounding Illumination Estimations for Higher Accuracy

Karlo Koščević, Nikola Banić and Sven Lončarić
Image Processing Group, Department of Electronic Systems and Information Processing,
Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia

Keywords: Chromaticity, Color Constancy, Genetic Algorithm, Illumination Estimation, Image Enhancement, White Balancing.

Abstract: The image processing pipeline of most contemporary digital cameras performs illumination estimation in order to remove the influence of illumination on image scene colors. In this paper an experiment is described that examines some of the basic properties of illumination estimation methods for several Canon's camera models. Based on the obtained observations, an extension to any illumination estimation method is proposed that under certain conditions alters the results of the underlying method. It is shown that with statistics-based methods as underlying methods the proposed extension can outperform camera's illumination estimation in terms of accuracy. This effectively demonstrates that statistics-based methods can still be successfully used for illumination estimation in digital cameras. The experimental results are presented and discussed. The source code is available at https://ipg.fer.hr/ipg/resources/color_constancy.

1 INTRODUCTION

Among many abilities human visual system (HVS) can recognize colors of objects regardless of scene illumination. This ability is known as color constancy (Ebner, 2007). Achieving computational color constancy is an important pre-processing step in image processing pipeline as different scene illuminations may cause the image colors to differ as shown in figure 1. In order to remove the influence of illumination color, accurate illumination estimation followed by chromatic adaptation must be preformed. For both tasks the following image **f** formation model, which includes Lambertian assumption, is most often used:

$$f_c(x) = \int_\omega I(\lambda,\mathbf{x})R(\mathbf{x},\lambda)\rho_c(\lambda)d\lambda \qquad (1)$$

where c is a color channel, $\mathbf{x}$ is a given image pixel, λ is the wavelength of the light, ω is the visible spectrum, $I(\lambda,\mathbf{x})$ is the spectral distribution of the light source, $R(\mathbf{x},\lambda)$ is the surface reflectance, and $\rho_c(\lambda)$ is the camera sensitivity of c-th color channel. With the assumption of uniform illumination the problem can be simplified, as now $\mathbf{x}$ is removed from $I(\lambda,\mathbf{x})$ and the observed light source color is given as:

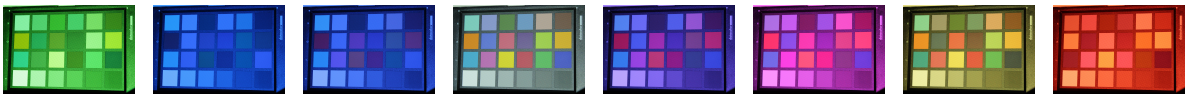
$$\mathbf{e} = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_\omega I(\lambda)\rho(\lambda)d\lambda \qquad (2)$$

(a) (b)

Figure 1: The same scene (a) with and (b) without illumination color cast.

For a successful chromatic adaptation, what is required is only the direction of **e** (Barnard et al., 2002). Since it is very common that only image pixel values **f** are given and both $I(\lambda)$ and $\rho(\lambda)$ remain unknown, calculating **e** is an ill-posed problem. To solve this problem, additional assumptions must be made, which leads to many color constancy methods that are divided into two major groups. First group of methods are low-level statistic-based methods like White-patch (Land, 1977; Funt and Shi, 2010), its improvements (Banić and Lončarić, 2013; Banić and Lončarić, 2014a; Banić and Lončarić, 2014b), Grayworld (Buchsbaum, 1980), Shades-of-Gray (Finlayson and Trezzi, 2004), Gray-Edge (1st and 2nd order) (Van De Weijer et al., 2007a), using bright and dark colors (Cheng et al., 2014). The second group is formed of learning-based methods like gamut mapping (pixel, edge, and intersection based) (Finlayson

Koščević, K., Banić, N. and Lončarić, S.
Color Beaver: Bounding Illumination Estimations for Higher Accuracy.
DOI: 10.5220/0007393701830190
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 183-190
ISBN: 978-989-758-354-4



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2: Color checker cast with projector light of various colors.

et al., 2006), using high-level visual information (Van De Weijer et al., 2007b), natural image statistics (Gijsenij and Gevers, 2007), Bayesian learning (Gehler et al., 2008), spatio-spectral learning (maximum likelihood estimate, and with gen. prior) (Chakrabarti et al., 2012), simplifying the illumination solution space (Banić and Lončarić, 2015a; Banić and Lončarić, 2015b; Banić and Lončarić, 2015b), using color/edge moments (Finlayson, 2013), using regression trees with simple features from color distribution statistics (Cheng et al., 2015), performing various kinds of spatial localizations (Barron, 2015; Barron and Tsai, 2017), using convolutional neural networks (Bianco et al., 2015; Shi et al., 2016; Hu et al., 2017; Qiu et al., 2018).

While learning-based method have a much higher accuracy, it are low-level statistics-based methods that are still being widely used among digital camera manufacturers since they are much faster and often more hardware-friendly than learning-based methods. This is also one of the reasons why statistics-based methods are still important for research. Nevertheless, since cameras are commercial systems, the fact that they still widely use statistics-based methods is not publicly stated. In this paper an experiment is described that examines some of the basic properties of illumination estimation methods for several Canon's camera models. Based on the obtained observations, an extension to any illumination estimation method is proposed that under certain conditions alters the results of the underlying method by bounding them to a previously learned region in the chromaticity plane. The bounding procedure is simple and does not add any significant cost to the overall computation. It is shown that with statistics-based methods as underlying methods the proposed extension can outperform camera's built-in illumination estimation in terms of accuracy. This effectively demonstrates that statistics-based methods can still be successfully used for illumination estimation in digital cameras' pipelines.

The paper is structured as follows: Section 2 lays out the motivation for the paper, in Section 3 the proposed method is described, Section 4 shows the experimental results, and Section 5 concludes the paper.

2 MOTIVATION

2.1 Do statistics-based Methods Matter?

Digital cameras are being used ever more widely, especially with the growing number of smartphones. This definitely means that the results of the research on computational color constancy now also have a higher impact so the importance of this research grows, especially when considering that it is an ill-posed problem. In literature and in the reviews of papers it is sometimes claimed that there is little purpose in researching low-level statistics-based methods since there are now much more accurate learning-based methods that significantly outperform them in accuracy. In contrast to that many experts with experience in the industry claim that many commercial white balancing systems are still based on low-level statistics-based methods. The main reason for that is their simplicity, low cost of implementation, and hardware-friendliness. If this is indeed so, then the research on such methods is definitely still important and should be further conducted and supported.

To check to what degree all these claims are true, it should be enough to examine some of the white balancing systems widely used in commercial cameras. In the world of professional cameras Canon has been the market leader for 15 years (Canon, 2018) and in 2018 it held an estimated 49% of the market share (PhotoRumors, 2018). Since practically every digital camera performs white balancing in its image processing pipeline, it can be claimed that Canon's white balancing system is one of the most widely spread ones. However, since Canon is a commercial company, full details of the white balancing system used in its digital cameras are not publicly known.

2.2 Learning from Existing Systems

One approach to gain more information on Canon's white balancing system is to look at the results of illumination estimation for various images taken under illumination of numerous colors. The following three camera models have been used to perform this experiment: EOS 550D, EOS 6D, and EOS 750D.

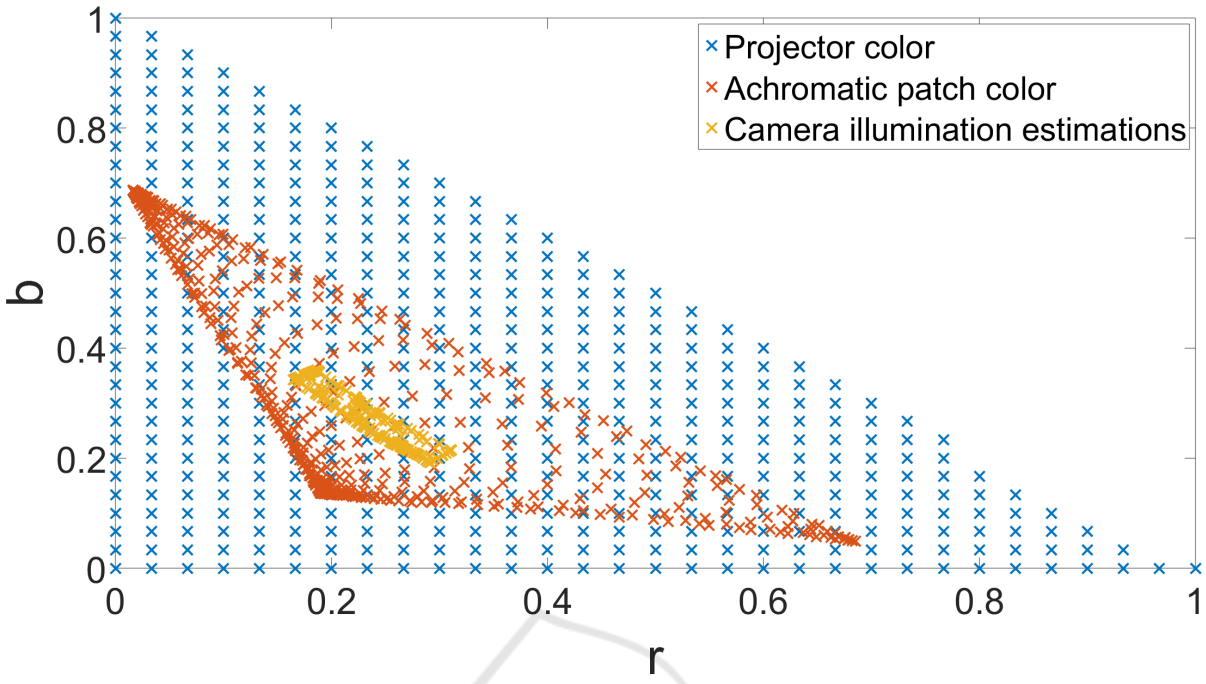


Figure 3: Comparison of chromaticities of projector light color, color of the second achromatic color checker patch, and camera's illumination estimation for Canon EOS 550D in the *rb*-chromaticity plane. The red chromaticity is shown on the *x* axis, while the blue chromaticity is shown on the *y* axis.

The experiment was conducted in a dark room where only a projector has been used as a light source. The projector was used to cast illumination of various colors, with chromaticities evenly spread in the chromaticity plane, on a color checker as shown in Figure 2. These images of the color checker were taken with every of the three previously mentioned cameras.

Although the illumination color was supposed to be computationally determined by projecting specifically created content, due to the projector and camera sensor characteristics the effective illumination color is altered. Its value as perceived by the camera can be read from the achromatic patches in the last row of the color checker and it serves as the ground-truth illumination for the given image. Ideally, it is this color that an illumination estimation method should predict.

Finally, the last step of the experiment was to check the results of illumination estimation performed by each of the cameras. The results of a camera's illumination estimation for a taken image can be reconstructed from the Exif metadata stored in the image file. The fields needed for this are *Red Balance* and *Blue Balance*, which have the values of channel gains i.e. the factors by which the red and blue channels have to be multiplied to perform chromatic adaptation. For practical reasons in cameras the green gain is fixed to 1. The combined inverse values of these gains give the illumination estimation vector. When this vector is normalized, it represents the chromaticity of camera's illumination estimation, which can be directly used to calculate the estimation accuracy by comparing it to the ground-truth illumination.

A comparison between the chromaticities for projected illumination color, achromatic patch color, and camera illumination estimations for Canon EOS 550D camera is given in Figure 3. The values read from achromatic white patches are squeezed with respect to the ones sent by the projector, but a more interesting observation is that none of the camera's illumination estimation are outside of a surface that resembles a parallelogram. As shown in Figure 4, similar results are obtained for other used camera models as well. Although there are some differences between the parallelograms mostly visible on two opposite sides, the parallelograms otherwise mostly cover a similar space in the chromaticity plane.

2.3 Observations

Based on these observations it can be concluded that one of the core properties of Canon's white balancing system is limiting its illumination estimation so that they do not appear outside of a polygon very similar to a parallelogram. Such limitation can be justified by the goal of avoiding unlikely illuminations and thus minimizing the occurrence of too high errors. This can be useful if it can be assumed that the expected illuminations are similar to black body radiation, but

sometimes it can be an disadvantage if artificially colored illumination sources are present like in Figure 2.

On the other hand, there is little that can be said about the white balancing system's behavior inside of the parallelogram. Nevertheless, the limitation observation is already useful because of its potential to limit maximum errors for illumination estimations. As for the behavior of illumination estimation inside the parallelogram, a possible solution is to use some of the already existing methods. Additionally, it can be immediately remarked that a parallelogram is a relatively regular quadrangle and polygon in general.

At least two questions can be raised here: first, is there a better quadrangle i.e. polygon for bounding the illuminations, and second, which method to use as the baseline underlying method that gets bounded?

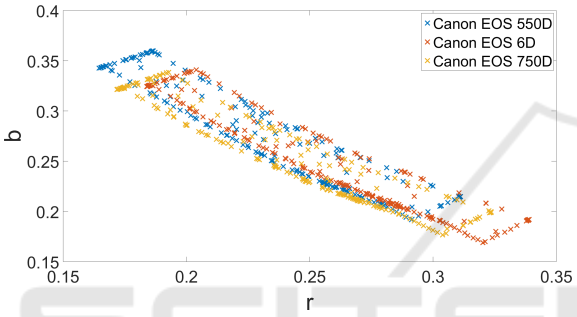


Figure 4: Comparison of cameras' illumination estimation for Canon EOS 550D, Canon EOS 6D, and Canon EOS 750D. The red chromaticity is shown on the x axis, while the blue chromaticity is shown on the y axis.

3 PROPOSED METHOD

Inspired by the bounds used by Canon cameras observed in Figure 4 and in order to give an answer to the two questions from the previous section, in this paper a new method i.e. extension to any chosen underlying illumination estimation method is proposed. The extension learns a bounding polygon with an arbitrary number of vertices that is used to restrict the illumination estimations of the initially chosen underlying method to the chromaticity region specified by the bounding polygon. As explained in the previous section, the motivation for this are the observations of boundaries used by Canon cameras and it can be applied to any illumination estimation method.

A genetic algorithm is used to learn the boundaries. First, the illumination estimations for the initially chosen underlying method are calculated on a given set of images. The boundary polygon population of size s is initialized by taking randomly chosen ground-truth illumination chromaticities as polygon vertices. Empirically, it has been shown that the four-point polygons i.e. quadrangles are generally a good fit for illumination restriction and there is no significant gain when the number of points is increased. The fitness function calculation for a given quadrangle is based on the ground-truth illuminations and the restricted illuminations that are the result of applying the boundary polygon to the underlying method's illumination estimations. Empirically, it has been concluded that the negative sum of the median angular error and a tenth of the maximum angular error is generally a good fitness function; angular error is explained in more detail in Section 4.1. More formally, if $\mathbb{A} = \{a_1, \ldots, a_n\}$ is the set of angular errors on n images, then the chosen fitness function is given as

$$\mathbf{f}(\mathbb{A}) = -\left(\text{med}\,(\mathbb{A}) + \frac{1}{10}\max(\mathbb{A})\right). \quad (3)$$

The maximum error was also included in the fitness function in order to discourage quadrangles that perform very well on the majority of the images, but have poor performance of several outliers. As the selection method the 3-way tournament selection (Mitchell, 1998) with random sampling is used. Averaging crossover function of the two of the best individuals produces a new child which is randomly mutated. The quadrangle with the lowest fitness value among the three ones chosen in the selection procedure is replaced in the current population by the newly created child quadrangle. The mutation is done by translating each vertex of a bounding polygon by the value from the normal distribution with $\mu = 0$ and $\sigma = 0.2$. Mutation rate, which states whether the whole individual should me mutated, is set to 0.3. After all training iterations have finished, the boundary quadrangle with the highest fitness value is chosen as the final result. Figure 5 shows an example of a learned quadrangle.

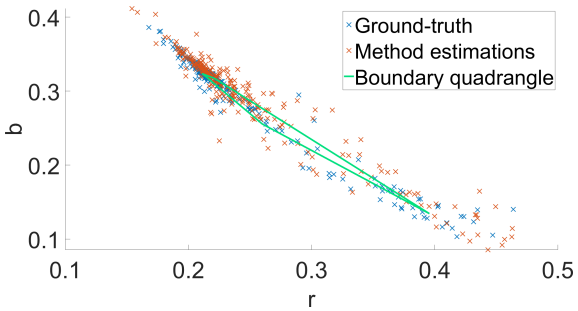


Figure 5: Example of a learned boundary quadrangle for the Canon1 dataset (Cheng et al., 2014) in the chromaticity plane. The red chromaticity is shown on the x axis, while the blue chromaticity is shown on the y axis.

Since the proposed extension bounds illumination estimations and beavers are known to bound water

flows by building dams, the proposed extension was named Color Beaver. In the rest of the paper extending a method **M** by the Color Beaver extension will be denoted as Color Beaver + **M**. The training procedure for Color Beaver is summarized in Algorithm 2.

Algorithm 1: Color Beaver Training.

Input: training images $\mathbb{I}$, ground truth $\mathbb{G}$, method **M**, iterations number N, population size s, fitness function **f**

Output: boundary polygon **P**

$\mathbb{E} = estimateIllumination(\mathbb{I}, \mathbf{M})$
$\mathbb{P} = initializePolygonPopulation(s)$
for i $\in \{1, ..N\}$ **do**
  $\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3} = tournamentSelection(\mathbb{P}, 3, \mathbf{f})$
  $\mathbf{t}' = crossover(\mathbf{t_1}, \mathbf{t_2})$
  $\mathbf{t}'.mutateMaybe(0.3)$d
  $\mathbb{R} = restrictIllumination(\mathbb{E}, \mathbf{t}')$
  $\mathbb{P}.ReplaceExistingWith(\mathbf{t_3}, \mathbf{t}')$
end for
$\mathbf{P} = \mathbb{P}.GetFittest(\mathbf{f})$

Algorithm 2: Color Beaver Application.

Input: image **I**, method **M**, boundary polygon **P**

Output: illumination estimation **e**

$\mathbf{e}_M = estimateIllumination(\mathbf{I}, \mathbf{M})$
$\mathbf{e} = restrictIllumination(\mathbf{e}_M, \mathbf{P})$

4 EXPERIMENTAL RESULTS

4.1 Experimental Setup

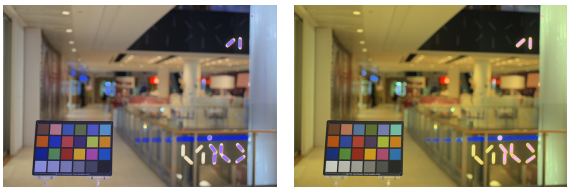
Eight linear NUS datasets (Cheng et al., 2014) and the Cube dataset (Banić and Lončarić, 2017) have been used to test the proposed extension and compare its performance to the one of other methods. All these datasets have linear images, which is also expected by the model described by Eq. (3). The ColorChecker dataset (Gehler et al., 2008; Shi and Funt, 2018) has not been used because of much confusion that is still present in many papers due to of its misuses in the past (Lynch et al., 2013; Finlayson et al., 2017).

The most commonly used accuracy measure among many proposed (Gijsenij et al., 2009; Finlayson and Zakizadeh, 2014; Banić and Lončarić, 2015a) is the angular error. It is the angle between the vectors of illumination estimation and the ground-truth illumination. When the angular errors obtained on each individual image of a given benchmark dataset need to be summarized, one of the most important statistics is the median angular error (Hordley and Finlayson, 2004). Despite that fact, the geometric mean of several statistics including the median angular error has increasingly been gaining popularity in recent publications (Barron and Tsai, 2017) and the same format as there is also used in this paper.

Table 1: Performance of different color constancy methods on the Cube dataset (Banić and Lončarić, 2017) in terms of angular error statistics (lower Avg. is better). The used format is the same as in (Barron and Tsai, 2017).

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
White-Patch (Funt and Shi, 2010)	6.58	4.48	5.27	1.18	15.23	4.88
Gray-world (Buchsbaum, 1980)	3.75	2.91	3.15	0.69	8.18	2.87
Camera built-in	2.96	2.56	2.64	0.82	5.79	2.49
Color Tiger (Banić and Lončarić, 2017)	2.94	2.59	2.66	0.61	5.88	2.35
Shades-of-Gray (Finlayson and Trezzi, 2004)	2.58	1.79	1.95	0.38	6.19	1.84
2nd-order Gray-Edge (Van De Weijer et al., 2007a)	2.49	1.60	1.80	0.49	6.00	1.84
1st-order Gray-Edge (Van De Weijer et al., 2007a)	2.45	1.58	1.81	0.48	5.89	1.81
General Gray-World (Barnard et al., 2002)	2.50	1.61	1.79	0.37	6.23	1.76
Color Beaver Camera + built-in (proposed)	1.70	0.96	1.15	0.31	4.38	1.20
Color Beaver + WP (proposed)	1.59	0.87	1.04	0.25	4.15	1.08
Restricted Color Tiger (Banić and Lončarić, 2017)	1.64	0.82	1.05	0.24	4.37	1.08
Color Dog (Banić and Lončarić, 2015b)	1.50	0.81	0.99	0.27	3.86	1.05
Smart Color Cat (Banić and Lončarić, 2015b)	1.49	0.88	1.06	0.24	3.75	1.04
Color Beaver + SoG (proposed)	1.51	0.81	1.00	0.22	3.97	1.01
Color Beaver + GW (proposed)	1.48	0.76	0.98	0.21	3.90	0.98

For both the NUS datasets and the Cube dataset a three-fold cross-validation with folds of equal size was used like in previous publications. The source code for recreating the results reported later in the paper is publicly available at https://ipg.fer.hr/ipg/resources/color_constancy.



(a) (b)

Figure 6: A failure case for Color Beaver + SoG with chromatic adaptation results based on a) the restricted illumination estimation with angular error of 10.74° and b) the ground-truth illumination.

Table 2: Combined performance of different color constancy methods on eight NUS dataset in terms of angular error statisrics (lower Avg. is better). The used format is the same as in (Barron and Tsai, 2017).

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
White-Patch (Funt and Shi, 2010)	9.91	7.44	8.78	1.44	21.27	7.24
Pixels-based Gamut (Gijsenij et al., 2010)	5.27	4.26	4.45	1.28	11.16	4.27
Grey-world (Buchsbaum, 1980)	4.59	3.46	3.81	1.16	9.85	3.70
Edge-based Gamut (Gijsenij et al., 2010)	4.40	3.30	3.45	0.99	9.83	3.45
Color Beaver + WP (proposed)	5.40	2.12	2.75	0.58	16.08	3.12
Shades-of-Gray (Finlayson and Trezzi, 2004)	3.67	2.94	3.03	0.98	7.75	3.01
Color Beaver + GW (proposed)	3.73	2.65	2.90	0.72	8.55	2.82
Natural Image Statistics (Gijsenij and Gevers, 2011)	3.45	2.88	2.95	0.83	7.18	2.81
Local Surface Reflectance Statistics (Gao et al., 2014)	3.45	2.51	2.70	0.98	7.32	2.79
2nd-order Gray-Edge (Van De Weijer et al., 2007a)	3.36	2.70	2.80	0.89	7.14	2.76
1st-order Gray-Edge (Van De Weijer et al., 2007a)	3.35	2.58	2.76	0.79	7.18	2.67
Bayesian (Gehler et al., 2008)	3.50	2.36	2.57	0.78	8.02	2.66
General Gray-World (Barnard et al., 2002)	3.20	2.56	2.68	0.85	6.68	2.63
Spatio-spectral Statistics (Chakrabarti et al., 2012)	3.06	2.58	2.74	0.87	6.17	2.59
Bright-and-dark Colors PCA (Cheng et al., 2014)	2.93	2.33	2.42	0.78	6.13	2.40
Corrected-Moment (Finlayson, 2013)	2.95	2.05	2.16	0.59	6.89	2.21
Color Beaver + SoG (proposed)	2.86	1.99	2.21	0.59	6.62	2.17
Color Tiger (Banić and Lončarić, 2017)	2.96	1.70	1.97	0.53	7.50	2.09
Color Dog (Banić and Lončarić, 2015b)	2.83	1.77	2.03	0.48	7.04	2.03
Shi et al. 2016 (Shi et al., 2016)	2.24	1.46	1.68	0.48	6.08	1.74
CCC (Barron, 2015)	2.38	1.48	1.69	0.45	5.85	1.74
Cheng 2015 (Cheng et al., 2015)	2.18	1.48	1.64	0.46	5.03	1.65
FFCC (Barron and Tsai, 2017)	1.99	1.31	1.43	0.35	4.75	1.44

4.2 Accuracy

Tables 1 and 2 show the comparisons between the accuracies of methods extended by the proposed extension and other illumination estimation methods. It can be seen that all of the extended methods outperform their initial non-extended versions. As a matter of fact, the extended version of the Shades-of-Gray method outperforms the camera built-in method. Additionally, the extended versions also outperform many learning-based methods. All these results demonstrate the usability of the proposed extension. An example of a failure case for the proposed extension of Shades-of-Gray is shown in Figure 6.

While other methods such as Gray-edge could also have been tested and shown in the Tables, Shades-of-Gray was already good enough to outperform camera's built-in methods. Extending Gray-edge also increases its accuracy, but Gray-edge is slower than Shades-of-Gray (Cheng et al., 2014), more complex, and it requires additional memory. Hence it was left out of the testing procedures since Shades-of-Gray is already sufficient to successfully answer the questions that were raised in this paper.

4.3 Discussion

The fact that statistics-based methods extended by the proposed method outperform camera built-in illumination estimation methods is significant for drawing further conclusions about the nature of camera's illumination estimation methods. Namely, if extended statistics-based methods outperform them, it can be freely stated that statistics-based are good enough to be used in digital cameras. Additionally, it may be that the extended method managed to outperform the camera's built-in methods because that they are also statistics-based, which in turn confirms that cameras do indeed use such method. In any of these two cases it can be concluded that research on statistics-based methods still has a large field of applications and the obtained results only further prove its importance.

5 CONCLUSIONS

An experiment was conducted to examine some of the details of built-in illumination estimation methods for several Canon camera models. Inspired by the observed results, an extension to any underlying illumination estimation method has been proposed. It limits the values of the illumination estimations of the underlying method by forcing it to stay inside a previously learned region in the chromaticity plane without adding any significant computation cost. By limiting some of the best-known statistics-based methods, the obtained accuracy outperforms the one of cameras' built-in methods. This effectively demonstrates that by only using slightly modified statistics-based methods it is possible to be more accurate than contemporary cameras. It also proves the claim that statistics-based methods can and probably are used for illumination estimation in digital cameras. Future research will include looking for new method modifications that result in even higher estimation accuracy.

ACKNOWLEDGEMENTS

This work has been supported by the Croatian Science Foundation under Project IP-06-2016-2092.

REFERENCES

Banić, N. and Lončarić, S. (2015a). Color Cat: Remembering Colors for Illumination Estimation. *Signal Processing Letters, IEEE*, 22(6):651–655.

Banić, N. and Lončarić, S. (2015b). Using the red chromaticity for illumination estimation. In *Image and Signal Processing and Analysis (ISPA), 2015 9th International Symposium on*, pages 131–136. IEEE.

Banić, N. and Lončarić, S. (2017). Unsupervised Learning for Color Constancy. *arXiv preprint arXiv:1712.00436*.

Banić, N. and Lončarić, S. (2013). Using the Random Sprays Retinex Algorithm for Global Illumination Estimation. In *Proceedings of The Second Croatian Computer Vision Workshopn (CCVW 2013)*, pages 3–7. University of Zagreb Faculty of Electrical Engineering and Computing.

Banić, N. and Lončarić, S. (2014a). Color Rabbit: Guiding the Distance of Local Maximums in Illumination Estimation. In *Digital Signal Processing (DSP), 2014 19th International Conference on*, pages 345–350. IEEE.

Banić, N. and Lončarić, S. (2014b). Improving the White patch method by subsampling. In *Image Processing (ICIP), 2014 21st IEEE International Conference on*, pages 605–609. IEEE.

Banić, N. and Lončarić, S. (2015a). A Perceptual Measure of Illumination Estimation Error. In *VISAPP*, pages 136–143.

Banić, N. and Lončarić, S. (2015b). Color Dog: Guiding the Global Illumination Estimation to Better Accuracy. In *VISAPP*, pages 129–135.

Barnard, K., Cardei, V., and Funt, B. (2002). A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *Image Processing, IEEE Transactions on*, 11(9):972–984.

Barron, J. T. (2015). Convolutional Color Constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387.

Barron, J. T. and Tsai, Y.-T. (2017). Fast Fourier Color Constancy. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Computer Society Conference on*, volume 1. IEEE.

Bianco, S., Cusano, C., and Schettini, R. (2015). Color Constancy Using CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 81–89.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of The Franklin Institute*, 310(1):1–26.

Canon (2018). Canon celebrates 15th consecutive year of No.1 share of global interchangeable-lens digital camera market.

Chakrabarti, A., Hirakawa, K., and Zickler, T. (2012). Color constancy with spatio-spectral statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8):1509–1519.

Cheng, D., Prasad, D. K., and Brown, M. S. (2014). Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058.

Cheng, D., Price, B., Cohen, S., and Brown, M. S. (2015). Effective learning-based illuminant estimation using simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1000–1008.

Ebner, M. (2007). *Color Constancy*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley.

Finlayson, G. D. (2013). Corrected-moment illuminant estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1904–1911.

Finlayson, G. D., Hemrit, G., Gijsenij, A., and Gehler, P. (2017). A Curious Problem with Using the Colour Checker Dataset for Illuminant Estimation. In *Color and Imaging Conference*, volume 2017, pages 64–69. Society for Imaging Science and Technology.

Finlayson, G. D., Hordley, S. D., and Tastl, I. (2006). Gamut constrained illuminant estimation. *International Journal of Computer Vision*, 67(1):93–109.

Finlayson, G. D. and Trezzi, E. (2004). Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology.

Finlayson, G. D. and Zakizadeh, R. (2014). Reproduction angular error: An improved performance metric for illuminant estimation. *perception*, 310(1):1–26.

Funt, B. and Shi, L. (2010). The rehabilitation of MaxRGB. In *Color and Imaging Conference*, volume 2010, pages 256–259. Society for Imaging Science and Technology.

Gao, S., Han, W., Yang, K., Li, C., and Li, Y. (2014). Efficient color constancy with local surface reflectance statistics. In *European Conference on Computer Vision*, pages 158–173. Springer.

Gehler, P. V., Rother, C., Blake, A., Minka, T., and Sharp, T. (2008). Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Gijsenij, A. and Gevers, T. (2007). Color Constancy using Natural Image Statistics. In *CVPR*, pages 1–8.

Gijsenij, A. and Gevers, T. (2011). Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):687–698.

Gijsenij, A., Gevers, T., and Lucassen, M. P. (2009). Perceptual analysis of distance measures for color constancy algorithms. *JOSA A*, 26(10):2243–2256.

Gijsenij, A., Gevers, T., and Van De Weijer, J. (2010). Generalized gamut mapping using image derivative

structures for color constancy. *International Journal of Computer Vision*, 86(2):127–139.

Hordley, S. D. and Finlayson, G. D. (2004). Re-evaluating colour constancy algorithms. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 76–79. IEEE.

Hu, Y., Wang, B., and Lin, S. (2017). Fully Convolutional Color Constancy with Confidence-weighted Pooling. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Conference on*, pages 4085–4094. IEEE.

Land, E. H. (1977). *The retinex theory of color vision*. Scientific America.

Lynch, S. E., Drew, M. S., and Finlayson, k. G. D. (2013). Colour Constancy from Both Sides of the Shadow Edge. In *Color and Photometry in Computer Vision Workshop at the International Conference on Computer Vision*. IEEE.

Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT press.

PhotoRumors (2018). 2018 Canon, Nikon and Sony market share (latest Nikkei, BCN and CIPA reports).

Qiu, J., Xu, H., Ma, Y., and Ye, Z. (2018). PILOT: A Pixel Intensity Driven Illuminant Color Estimation Framework for Color Constancy. *arXiv preprint arXiv:1806.09248*.

Shi, L. and Funt, B. (2018). Re-processed Version of the Gehler Color Constancy Dataset of 568 Images.

Shi, W., Loy, C. C., and Tang, X. (2016). Deep Specialized Network for Illuminant Estimation. In *European Conference on Computer Vision*, pages 371–387. Springer.

Van De Weijer, J., Gevers, T., and Gijsenij, A. (2007a). Edge-based color constancy. *Image Processing, IEEE Transactions on*, 16(9):2207–2214.

Van De Weijer, J., Schmid, C., and Verbeek, J. (2007b). Using high-level visual information for color constancy. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE.

Publication 5

Koščević, K., Subašić, M., Lončarić, S., "Iterative Convolutional Neural Network-Based Illumination Estimation", IEEE Access, Vol. 9, 2021, pp. 26755-26765.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 21, 2021, accepted February 1, 2021, date of publication February 4, 2021, date of current version February 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3057072

Iterative Convolutional Neural Network-Based Illumination Estimation

KARLO KOŠČEVIĆ, (Graduate Student Member, IEEE), MARKO SUBAŠIĆ, (Member, IEEE), AND SVEN LONČARIĆ, (Senior Member, IEEE)
Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
Corresponding author: Karlo Koščević (karlo.koscevic@fer.hr)

This work was supported by the Croatian Science Foundation under Project IP-06-2016-2092.

ABSTRACT In the image processing pipelines of digital cameras, one of the first steps is to achieve invariance in terms of scene illumination, namely computational color constancy. Usually, this is done in two successive steps which are illumination estimation and chromatic adaptation. The illumination estimation aims at estimating a three-dimensional vector from image pixels. This vector represents the scene illumination, and it is used in the chromatic adaptation step, which aims at eliminating the bias in image colors caused by the color of the illumination. An accurate illumination estimation is crucial for successful computational color constancy. However, this is an ill-posed problem, and many methods try to comprehend it with different assumptions. In this paper, an iterative method for estimating the scene illumination color is proposed. The method calculates the illumination vector by a series of intermediate illumination estimations and chromatic adaptations of an input image using a convolutional neural network. The network has been trained to iteratively compute intermediate incremental illumination estimates from the original image. Incremental illumination estimates are combined by per element multiplication to obtain the final illumination estimation. The approach is aimed to reduce large estimation errors usually occurring with highly saturated light sources. Experimental results show that the proposed method outperforms the vast majority of illumination estimation methods in terms of median angular error. Moreover, in terms of worst-performing samples, i.e., the samples for which a method errs the most, the proposed method outperforms all other methods by a margin of more than 18% with respect to the mean of estimation errors in the third quartile.

INDEX TERMS Chromatic adaptation, color constancy, convolutional neural networks, illumination estimation, image color analysis.

I. INTRODUCTION

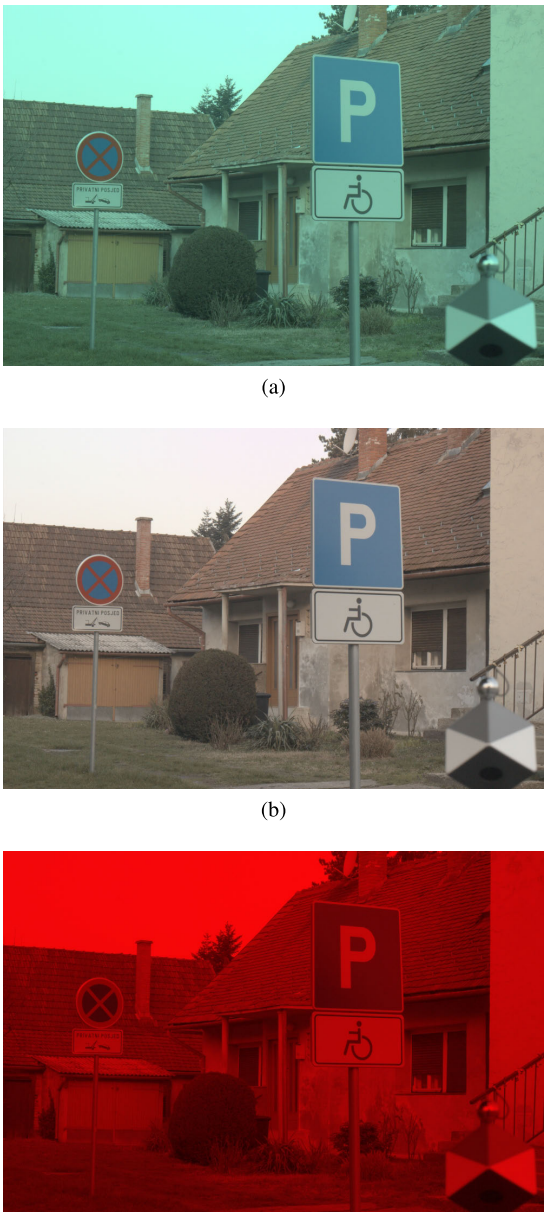
In digital photography, any illumination present in the scene of interest significantly impacts the colors of the objects in digital images. According to the image formation model [1], the value of a pixel in an image is determined by three functions: the spectrum of the light source, the reflectance of the object surface, and the spectral sensitivity of the camera sensor. If the same scene is captured with the same camera (i.e., the reflectance of the object surface and the spectral sensitivity of the camera sensor are constant) whereas the spectrum of the light source changes, the colors in the captured images will most likely differ. The reason for this behavior is that the camera sensor is a device that can only capture the incident light but cannot detect changes in illumination itself. Therefore, for most digital cameras, one of the first steps in the image processing pipeline is dedicated to achieving illumination invariance. This process can be associated with the ability of the human visual system to adapt to changes in scene illumination, namely color constancy [2]. Achieving computational color constancy has proven to be beneficial in many image-related areas such as object recognition, scene comprehension, digital photography, and image reproduction [3]. In order to achieve computational color constancy, two steps are usually required. First, the scene illumination color is estimated based on the image pixel values, and then, in the second step, its influence on the image colors is eliminated. Color constancy is not yet fully understood and

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqi Wang.

modeled, and estimating the scene illumination from the image pixels is an ill-posed problem, which is regularized by various assumptions. During the last few years, many methods for estimating the illumination color have been proposed, with the general assumption that the illumination is uniform in the scene [1]. Since only one illumination vector per image is estimated, a simple diagonal matrix with reciprocal illumination values on the diagonal is usually used to eliminate color distortion.

For a successful computational color constancy, both illumination estimation and chromatic adaptation should be as accurate and similar to the image formation model as possible. However, even though the simple diagonal matrix for chromatic adaptation is computationally efficient and sufficient for a somewhat satisfactory computational color constancy, it is still an approximation. Illumination estimates can be either imprecise or out of the range of illuminations for which color images can be properly corrected using the current chromatic adaptation model. It is expected that the error in computational color constancy is higher for images that are captured in scenes illuminated with highly colored light sources than for scenes affected by near-white illuminations. Such illuminations can corrupt object colors, and if their estimates are imprecise high errors in corrected images can be expected. In [4], it was shown that camera manufacturers bound illuminations to a narrow region in chromaticity space so that chromatic adaptation is never performed with highly colored illuminations. It can be speculated that the cause for this is the inadequacy of the chromatic adaptation model that is unfit for the highly colored illuminations. Therefore, in this paper, a multistage illumination color estimation combined with the current simple chromatic adaptation model is proposed. The individual stages' estimations are restricted from highly colored estimations so that the used chromatic adaptation model is operating in the range of slightly colored illuminations. The final illumination estimation is obtained by combining all of the stage illuminations so that the final illumination estimations can still be highly colored. With this approach, the occurrence of high estimation errors should be alleviated, as shown in experimental results.

For the evaluation of illumination estimation methods, the angular error is used. It is calculated as the angle between the ground-truth illumination vector and the estimated illumination vector. Usually, the RGB color space is used so that both vectors have three components corresponding to the red, green, and blue image channels. The median error value of a test dataset is usually considered the most representative statistic. Nowadays, illumination estimation methods can achieve median error values of less than 2°, which can be regarded as a threshold for a sufficiently accurate illumination estimation [5]. However, even such accurate methods in terms of median or mean error value tend to be flawed in some cases. The maximum error values can be as large as 10° or more. Correcting an image with a highly incorrect illumination color vector can distort the image colors to such an extent that the actual information they carry is effectively lost.



(a)

(b)

(c)

FIGURE 1. Chromatic adaptation example with highly inaccurate illumination vector: (a) original raw image with the influence of illumination; (b) the result of the chromatic adaptation of image (a) with ground-truth illumination vector $(0.1624\ 0.4533\ 0.3843)^T$; (c) the result of the chromatic adaptation of image (a) with inaccurate illumination vector $(0.0001\ 0.6528\ 0.3471)^T$. The angle between the ground-truth vector and inaccurate illumination vector is 19.54°. For display purposes, images were tone mapped by using the Flash tone mapping operator [6].

An example of a chromatic adaptation with a highly incorrect illumination vector is shown in Fig. 1.

In this paper, an illumination estimation method that reduces maximum estimation errors, which can occur when highly colored illuminations are present in the scene, is proposed. The proposed method combines both illumination estimation and chromatic adaptation, which are usually two distinct steps in the image processing pipeline, to obtain more precise illumination estimates. The global illumination vector is estimated through a series of consecutive intermediate

illumination estimations, and chromatic adaptations of an input image. In each step, intermediate illumination estimation is forced to a subset of illuminations that are close to the white light, i.e., a light that does not alter image colors. Chromatic adaptation of the input image with an estimated intermediate illumination vector is performed, and such a corrected image is then passed as a new input. This procedure was embedded in a deep neural network which uses convolutional architecture for the estimation of intermediate illuminations, and simple matrix multiplications for chromatic adaptation of input images and aggregation of intermediate estimates into one final illumination estimate.

The rest of the paper is structured as follows: In Section II, an overview of related methodology is given, Section III describes the proposed illumination estimation method, experimental results are presented and discussed in Section IV, and in Section V, the conclusion is given.

II. RELATED WORK

The image formation model, commonly used in computational color constancy, which assumes Lambertian reflectance can be formulated as

$$f_c(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x}) R(\lambda, \mathbf{x}) \rho_c(\lambda) d\lambda, \tag{1}$$

where each pixel $\mathbf{x}$ in the image $\mathbf{f}$ with three color channels $c \in \{R, G, B\}$ is computed as the integral of the product of light source spectrum $I(\lambda, \mathbf{x})$, surface reflectance $R(\lambda, \mathbf{x})$, and camera sensor sensitivity $\rho_c(\lambda)$ across all wavelengths λ in the visible light spectrum ω.

A. ILLUMINATION ESTIMATION

The first step in computational color constancy is illumination estimation, which aims to estimate the vector of the scene illumination from image pixels. From (1), it can be observed that illumination can be determined by knowing the light source spectrum $I(\lambda, \mathbf{x})$ and camera sensor sensitivity $\boldsymbol{\rho}(\lambda)$. In the case of global illumination estimation methods, i.e., when it is assumed that there is one dominant light source present in the scene, spatial information $\mathbf{x}$ is disregarded, and the illumination vector is defined as

$$\mathbf{e} = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_\omega I(\lambda) \boldsymbol{\rho}(\lambda) d\lambda. \tag{2}$$

The estimation of $\mathbf{e}$ is an ill-posed problem as usually there is no prior knowledge about $I(\lambda)$ and $\boldsymbol{\rho}(\lambda)$ values.

To make the problem of illumination estimation feasible, illumination estimation methods are often based on some assumptions. One group of illumination estimation methods are methods such as White-Patch [7], [8] and its improvements [9]–[11], and gray world assumption-based methods that include Gray-World [12], Shades-of-Gray [13], Gray-Edge [14], Weighted-Gray-Edge [15]. Although simple and do not generalize well, these methods are suitable for hardware implementation since they use simple image features and statistics, which are fast to calculate and have insignificant computational complexity.

On the other hand, there are machine-learning based illumination estimation methods that require computational models to be trained on data. The most recent examples are methods based on deep learning. These methods achieve the most accurate estimates of scene illumination but are highly dependent on training data distribution. Large and diverse datasets are prerequisites for creating deep learning methods that can generalize well. In comparison with illumination estimation methods in the first group, learning-based methods require more computational resources and have more complex structures. The earliest deep learning architectures for illumination estimation were very shallow, containing only a few convolutional and fully connected layers [16], [17]. Content-based convolutional neural networks that combine weighted local illumination estimations have been proposed in [18]–[20]. In [21], [22], illumination estimation was cast into a deep learning classification problem. In [23], from an image, two illuminations were estimated using one convolutional neural network, and then using another convolutional neural network, a more probable one was chosen. The problem of dependency of illumination estimation methods on the camera sensor was tackled in [24], where two convolutional networks were used for sensor space mapping and illumination estimation, respectively. Other learning-based methods use Bayesian learning [25], color moments [26], gamut mapping [27]–[29], spatial localizations [30], [31], visual information of high level [32], illumination space restrictions [4], [33]–[35], gray pixel detection [36], regression trees with simple color features [37], and others.

B. CHROMATIC ADAPTATION

The second step in computational color constancy is chromatic adaptation, which is used to change the color cast in images due to the illumination color. It was shown that using a diagonal matrix can be sufficient for a successful chromatic adaptation [38]. Namely, following this simplification, which is also known as the *von Kries model* [39], camera sensor responses are considered independent. Then, for an image pixel $\mathbf{p} = \begin{pmatrix} p_R & p_G & p_B \end{pmatrix}^T$, a new color corrected pixel $\hat{\mathbf{p}} = \begin{pmatrix} \hat{p}_R & \hat{p}_G & \hat{p}_B \end{pmatrix}^T$ can be computed as

$$\hat{\mathbf{p}} = \mathbf{C}\mathbf{p}, \tag{3}$$

where $\mathbf{C}$ denotes the correction matrix. In general, the correction matrix $\mathbf{C}$ can be computed as

$$\mathbf{C} = \begin{pmatrix} \bar{e}_R/e_R & 0 & 0 \\ 0 & \bar{e}_G/e_G & 0 \\ 0 & 0 & \bar{e}_B/e_B \end{pmatrix}, \tag{4}$$

where $\mathbf{e} = \begin{pmatrix} e_R & e_G & e_B \end{pmatrix}^T$ denotes the illumination vector that should be removed from an image, and $\bar{\mathbf{e}} = \begin{pmatrix} \bar{e}_R & \bar{e}_G & \bar{e}_B \end{pmatrix}^T$ denotes the vector of the desired illumination. In computational color constancy, the input image should be processed so that it appears as it was captured while illuminated

with a white light source, i.e., the light source for which $e_R = e_G = e_B$. Therefore, $\bar{\mathbf{e}} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T$ is used.

III. PROPOSED METHOD

The proposed method estimates the illumination vector from a raw input image in multiple iterations. In each iteration, a restricted intermediate illumination vector is computed from the input image. The estimated vector is then used for chromatic adaptation of the input image according to (3). In the next iteration, the corrected image is used as input. In the end, intermediate illumination vectors estimated in the iterations are element-wise multiplied to produce the final illumination vector that corresponds to the scene illumination captured in the original raw input image. The pseudocode of the proposed illumination estimation method is given in Algorithm 1.

Algorithm 1 Iterative Illumination Estimation

Input: image $\mathbb{I}$, convolutional neural network *CNN*, iteration number N

Output: illumination vector $\mathbf{e}$

$\mathbf{e} \leftarrow \begin{pmatrix} e_R & e_G & e_B \end{pmatrix} \leftarrow \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$
for $k \leftarrow 1$ to N **do**
  $\mathbf{e}^{(k)} \leftarrow CNN.estimate(\mathbb{I})$
  $\mathbf{e} \leftarrow \mathbf{e} \circ \mathbf{e}^{(k)}$
  $\mathbf{C} \leftarrow \text{diag}(1/e_R^{(k)}, 1/e_G^{(k)}, 1/e_B^{(k)})$ ▷ Eq. (4)
  $\mathbb{I}_{x,y} \leftarrow \mathbf{C}\mathbb{I}_{x,y}\ \forall x, y$ ▷ Eq. (3)
  $\mathbb{I} \leftarrow \frac{1}{\max_{x,y}\mathbb{I}} \cdot \mathbb{I}$
end for
$\mathbf{e} \leftarrow \frac{1}{e_R+e_G+e_B} \cdot \mathbf{e}$

In each iteration, an intermediate illumination vector is estimated using the convolutional neural network. Network parameters are the same in each iteration. Convolutional blocks of the VGG16 [40] network architecture were used as a feature extractor,[1] on top of which one additional convolutional layer was placed. This layer has three filters with a kernel of size 1×1. Each filter corresponds to one of three color channels in the RGB image: red, green, and blue. Output activation was a sigmoid function. Global average pooling, which calculates the average across feature maps, was used to accumulate feature maps computed by the last convolutional layer, thus producing one value for each color channel. Global average pooling yields the intermediate illumination vector. On top of this, chromatic adaptation was implemented, which uses the current network input and illumination estimate to compute the network input in the next iteration.

[1]It was experimentally determined to use the VGG16 network as a feature extractor. The architecture of SqueezeNet [41] convolutional neural network was also considered, which matches the accuracy of AlexNet [42] architecture but with fewer weights. However, the VGG16 network outperformed such simpler architectures.

A. DATA NORMALIZATION

The last convolutional layer in the proposed network architecture uses a sigmoid activation function that ensures that intermediate illumination estimates are all in the first octant in three-dimensional illumination solution space. However, the codomain of a sigmoid function is in the range from zero to one. When such values are used for chromatic adaptation, due to the division, the values in the corrected image may span in a different range than original image values. Therefore, in each iteration, the input image is normalized by dividing every image value by the image maximum. Moreover, input normalization was shown beneficial for efficient backpropagation [43].

Estimated intermediate illumination vectors in each iteration were not normalized using the standard normalization in computational color constancy research, i.e., the division of illumination vector with its sum. The reasoning behind this is that the proposed method combines illumination estimation, chromatic adaptation, and the abovementioned image normalization. Namely, if chromatic adaptation is performed with normalized illumination vector and the resulting image is then normalized as well, the factor which would be used to normalize the illumination would be canceled out. Therefore, normalizing intermediate illumination vectors would not have any effect.

B. NETWORK TRAINING

For the training of the proposed illumination estimation network architecture, a custom loss function was used. It is based on the cosine of the angle[2] between two vectors and consists of two parts. The first part of the custom loss function is dedicated to computing the error between ground-truth illuminations and the end-result of the network. The second part is used to control the behavior of intermediate illumination estimates in each iteration by forcing them to be close to the white light. This is achieved by minimizing the angle between intermediate illumination estimates and the vector of the white light. However, the extent of bounding to the white light is not the same in each iteration. With each subsequent iteration, intermediate illuminations have to be closer to white light. That is achieved by assigning the weight to the loss value in each iteration as

$$w_k = \frac{2^{k-1}}{\sum_{j=0}^{N-1} 2^j}, \tag{5}$$

where $k \in \{1, \ldots, N\}$ denotes the current iteration, and N denotes the number of iterations.

[2]The most direct measure of error in illumination estimation is the angle between the ground-truth illumination value and the estimated illumination value. Taking into account that both the ground-truth and the estimation are vectors, the angle between them, once they are both normalized to unit length, is computed as the inverse cosine ($\cos^{-1}$) of their dot product. According to [44], using $\cos^{-1}$ makes the derivative of the loss function more complex and infinite when the absolute value of the dot product is equal to one, and therefore, using $1 - \cos\theta$ as loss function is more appropriate.

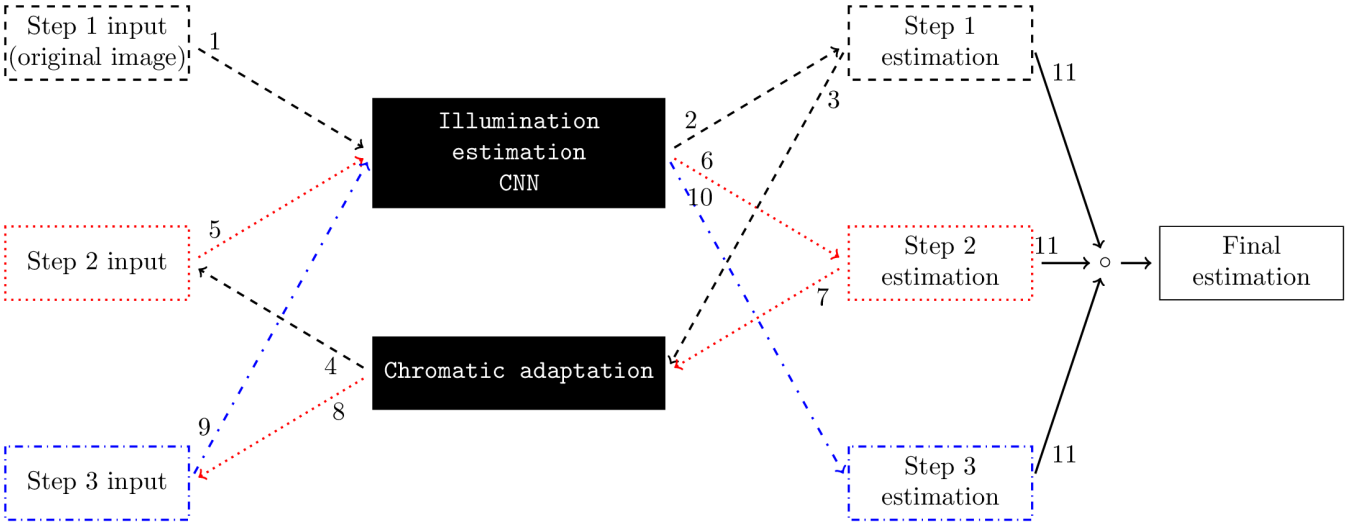


FIGURE 2. The illustration of the forward pass of the proposed method for three iterations. Arrows are enumerated in the order of execution, starting from 1. Different line styles denote different iterations: – – – denotes the first iteration steps, ··· denotes the second iteration steps, and – · – denotes the third iteration steps. The final estimation in step 11 is computed in parallel once the last iteration ends.

In both parts of the loss function, for a mini-batch of M input samples, the loss L was calculated as

$$\mathrm{L}(\mathbb{E}, \hat{\mathbb{E}}) = \frac{1}{M}\sum_{m=1}^{M}\left(1 - \frac{\mathbb{E}^{(m)} \cdot \hat{\mathbb{E}}^{(m)}}{\|\mathbb{E}^{(m)}\|_2 \|\hat{\mathbb{E}}^{(m)}\|_2}\right), \tag{6}$$

where $\mathbb{E}$ and $\hat{\mathbb{E}}$ denote batches of ground-truth and estimated illumination vectors, respectively, m^{th} ground-truth and estimated illumination vectors in the mini-batch are denoted as $\mathbb{E}^{(m)}$ and $\hat{\mathbb{E}}^{(m)}$, respectively, '$\cdot$' is the vector dot product, and $\|.\|_2$ is vector L2 norm.

The total loss for a mini-batch of images is the sum of the end-result loss and weighted intermediate estimation losses as follows

$$\mathrm{L}(\mathbb{E}, \hat{\mathbb{E}}) + \sum_{k=1}^{N} w_k \mathrm{L}(\mathbb{U}, \hat{\mathbb{E}}_k), \tag{7}$$

where $\mathbb{U}$ and $\hat{\mathbb{E}}_k$ denote batches of white illumination vectors and illumination vectors estimated in k^{th} iteration, respectively.

The forward pass in the proposed approach follows the steps in Algorithm 1. It is crucial to emphasize that the forward pass consists of multiple iterations and that the weights of the network are shared across iterations, i.e., the same set of network weights is used in each iteration in the forward pass. This method of the forward pass can be thought of as recurrent since the network is gradually computing the solution from multiple variations of the input image while keeping the set of weights unchanged. Each iteration results in an image with a slight modification of colors obtained by performing the chromatic adaptation of the input in that iteration with illumination estimate, which is also computed in that iteration. The modified image is the input for the succeeding iteration. An illustration of the flow of the proposed method for three iterations is shown in Fig. 2. The only form of supervision during network training is imposed with the loss function, and, in each iteration, in the forward pass, the network estimates intermediate illuminations, which result in a more accurate final estimate.

With the complex form of the forward pass, the backward pass in the proposed approach is complex as well. This is because the final illumination estimate in the forward pass is the product of intermediate estimates, the loss function penalizes each intermediate estimate, and network weights are shared across iterations. Therefore, the gradients propagating through a network layer consist of the gradients induced by the error of the final estimate and by the error of each intermediate estimate with respect to the white light.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

Cube+ dataset [45] was used to train and test the proposed illumination estimation network and the iterative procedure. It is a dataset containing 1707 images labeled for global illumination estimation. It consists of images of outdoor scenes in day and night and images of indoor scenes with artificial illuminations. Raw images in the Cube+ dataset are 2601 pixels wide and 1732 pixels high. For the reduction of the computational cost and to utilize as many resources as possible, all images have been resized to the size of 224×224 pixels. Additionally, by resizing the images to the specified shape, the input shape of the pre-trained VGG16 network was matched. Apart from image resizing, standard pre-processing steps for the Cube+ dataset were applied. Pre-processing steps include calibration object masking, black level subtraction, and overexposed pixel removal.

The angular error was used to evaluate the network accuracy. It is computed as the angle between the ground-truth illumination vector and the estimated illumination vector as

follows

$$A(\mathbf{e}, \hat{\mathbf{e}}) = \cos^{-1}\left(\frac{\mathbf{e}\cdot\hat{\mathbf{e}}}{\|\mathbf{e}\|_2\|\hat{\mathbf{e}}\|_2}\right) \quad (8)$$

For comparison with existing methods, a standard evaluation procedure for the evaluation of illumination estimation methods was followed. Mean, median, trimean, best 25%, worst 25%, and average [30] error statistics were computed on the test set. However, the focus of this paper is on reducing maximum estimation errors which can occur in cases of images with highly colored illuminations. By forcing the intermediate illumination estimates to be as close to the white light as possible, the reduction of maximal errors is expected. Therefore, the worst cases were additionally explored. Since other illumination estimation methods do not have maximal estimation errors reported, comparison with them could only be conducted by using the worst 25% statistic.

The following convolutional neural network parameters were used: learning rate 1×10^{-4}, number of epochs 200, min-batch size 8. The feature extraction part that corresponds to the VGG16 network was initialized with weights from the Keras Applications module [46] which were pre-trained on the ImageNet [47] dataset. The newly added convolutional layer was initialized by using the Xavier initialization [48].

B. DETERMINING THE NUMBER OF ITERATIONS

The optimal number of iterations for the proposed method was experimentally determined. Cube+ dataset was used for this purpose. It was split into three parts: train, test, and validation. The train part of the dataset was used to train the proposed network architecture for a different number of iterations. In each experiment, training parameters were the same, as described in subsection IV-A. The optimal number of iterations was obtained by evaluating the trained models on the validation part of the dataset and looking for the one with the lowest median angular error. Once determined, the model with the optimal number of iterations was evaluated on the test part of the dataset, and these results are reported in subsection IV-C.

An important role in determining the optimal number of iterations is the model complexity, which increases in accordance with the number of iterations. The higher the number of iterations is, the more computational memory is needed. Since the proposed method was trained and tested by using the GPU, the size of the GPU memory was a limiting factor for the conducted experiments.

Taking into account method accuracy and GPU memory limits, models with the number of iterations in the range from one to nine were considered, and, as the optimal one, the model with seven iterations was chosen. Therefore in the proposed method and experimental results the number of iterations and, thus, the number of intermediate illumination estimations is set to seven. For comparison, the model performances for a different number of iterations on the test part of the dataset are shown in Fig. 3.

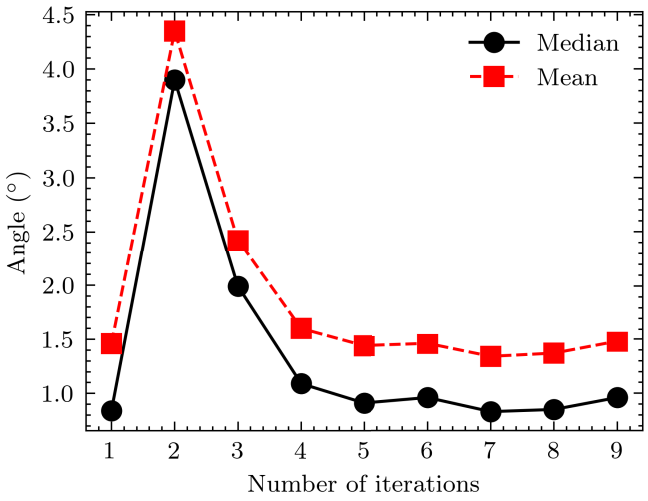


FIGURE 3. Performance of the proposed method for a different number of iterations with respect to the median and mean error statistics.

The proposed multistage approach aimed to achieve the asymptotic convergence of the illumination correction towards no correction. In other words, the preliminary limiting factor was only the amount of the available GPU memory. However, from the experiments, it can be seen that such convergence was not achieved since both mean and median errors start to increase after seven iterations. There are several possible factors for such behavior, with the main one being the imperfection of the simple chromatic adaptation model. Other possible factors include floating-point arithmetic rounding and neural network capacity. Therefore, the proposed search for determining the optimal number of iterations was conducted.

C. METHOD PERFORMANCE

1) COMPARISON WITH EXISTING ILLUMINATION ESTIMATION METHODS

In Table 1, the illumination estimation methods' accuracy on the Cube+ dataset is shown. For evaluation and comparison of the proposed method, final network estimation, i.e., the product of intermediate illumination estimates is used. It can be seen that the proposed method outperforms all other methods on average and in worst-case scenarios. Additionally, both the proposed method and Color Beaver [4] have comparable median and average error statistics that outperform other methods by a notable margin.

The proposed method was tested on a system with Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz central processing unit. The average execution time on the test set using only one core was 2.04 seconds per input image. The proposed model has 14,716,227 weights which is less compared to deep learning-based illumination estimations methods evaluated on the Cube+ dataset in [19], [20], [22], which all use VGG16 network structure for feature extraction, but have more complex additional layer structures, such as attention

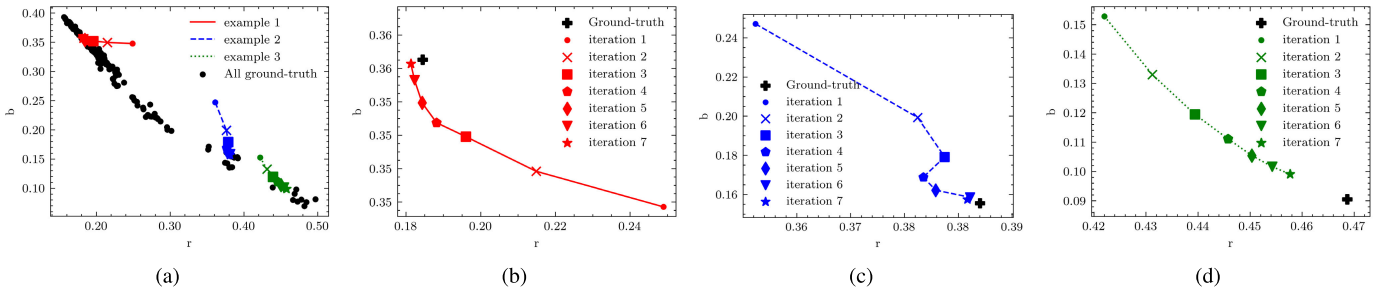


FIGURE 4. Examples of cumulative estimation trajectories with respect to the ground-truth in *rb*-chromaticity space for the proposed approach with seven iterations: (a) cumulative estimation trajectories in comparison to all ground-truth chromaticities; (b), (c), and (d) magnified trajectories for examples 1, 2, and 3 in (a), respectively.

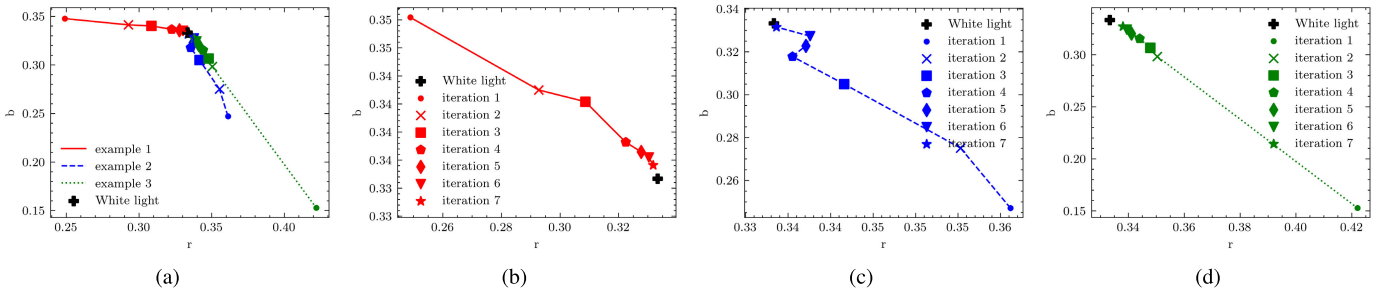


FIGURE 5. Examples of intermediate illumination estimation trajectories with respect to the white light in *rb*-chromaticity space for the proposed approach with seven iterations: (a) estimation trajectories in comparison to white light chromaticities; (b), (c), and (d) magnified trajectories for examples 1, 2, and 3 in (a), respectively. Trajectories correspond to the same examples as in Fig. 4.

TABLE 1. The comparison of angular error statistics of different color constancy methods on the Cube+ dataset [45] (sorted by Avg., lower is better).

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
White-Patch [8]	9.69	7.48	8.56	1.72	20.49	7.38
Gray-world [12]	7.71	4.29	4.98	1.01	20.19	5.08
Double-opponency (max pooling) [49]	6.76	3.44	4.15	0.79	18.54	4.27
Using gray pixels [36]	6.65	3.26	3.95	0.68	18.75	4.05
Color Tiger [45]	3.91	2.05	2.53	0.98	10.00	2.88
Color Mule [50]	5.16	1.30	2.03	0.25	16.93	2.25
Shades-of-Gray [13]	2.59	1.73	1.93	0.46	6.19	1.90
2nd-order Gray-Edge [14]	2.50	1.59	1.78	0.48	6.08	1.83
1st-order Gray-Edge [14]	2.41	1.52	1.72	0.45	5.89	1.76
Color Dog [35]	3.32	1.19	1.60	0.22	10.22	1.70
General Gray-World [3]	2.38	1.43	1.66	0.35	6.01	1.64
Attention CNN [20]	2.05	1.32	1.53	0.42	4.84	1.54
Light Source Classification [22]	1.86	1.27	1.39	0.42	4.31	1.43
RGB Attention CNN [19]	1.95	1.13	1.37	0.32	4.92	1.37
Proposed approach	**1.34**	0.83	**0.97**	0.28	**3.20**	**0.99**
Color Beaver (Gray-world) [4]	1.49	**0.77**	0.98	**0.21**	3.94	**0.99**

blocks, or have multiple instances of the same network structure with different weights.

2) METHOD BEHAVIOR VALIDATION

For the rest of the paper, it is important to define the term cumulative estimate. A cumulative estimate in iteration k is the element-wise product of all intermediate estimates up to and including the iteration k. In other words, cumulative estimate in the iteration k can be thought of as the final output of the network if the total number of iterations is equal to k.

The proposed method introduces iterative illumination estimation which forces intermediate illumination estimates computed in each iteration to be close to the white light and when multiplied element-wise altogether to be equal to the scene illumination. By the construction of the method, it is expected for intermediate estimates to be closer to the white light with each iteration. Also, it is expected for cumulative estimates to be closer to the ground-truth as iterations progress. Neither intermediate estimates nor cumulative estimates should fluctuate in illumination space. Such behavior can be verified in Fig. 4, and Fig. 5 where few examples of estimation trajectories for different input images with respect to the ground-truth and white light are shown. A trajectory represents the path enclosed by either intermediate or cumulative estimates through iterations. In Fig. 4 cumulative estimations with respect to the ground-truth are considered, and in Fig. 5 intermediate estimations with respect to the white light are considered.

Since the proposed method uses estimates from multiple versions of an input image to compute the color of scene illumination, naturally, a question of the benefit of using more estimations compared to a single estimate arises. Therefore, the proposed network architecture was also trained for one iteration only. The same set of parameters was used as described in subsection IV-A: learning rate 1×10^{-4}, epoch 200, and mini-batch size 8. When only one iteration is used, chromatic adaptation is not performed, and the first intermediate estimate is actually the final network estimate.

TABLE 2. The comparison of angular error statistics of the proposed method and the baseline (lower is better).

Algorithm	Min	Max	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
Baseline	0.02	9.47	1.46	0.84	0.98	0.23	3.73	1.01
Proposed approach	0.03	7.36	1.34	0.83	0.97	0.28	3.20	0.99

TABLE 3. The comparison of angular error statistics of the proposed method and the baseline on worst-performing samples for the baseline on the test set (lower is better).

Algorithm	Min	Max	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
Baseline	3.93	9.47	5.71	5.55	5.51	4.24	7.57	5.62
Proposed approach	0.76	7.36	3.57	3.39	3.44	1.36	5.93	3.20

TABLE 4. The comparison of angular error statistics of the proposed method and the baseline on worst-performing samples for the proposed method on the test set (lower is better).

Algorithm	Min	Max	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
Baseline	0.50	9.47	4.11	4.33	4.09	1.18	7.17	3.61
Proposed approach	3.29	7.36	4.55	4.34	4.39	3.42	6.07	4.48

In other words, illumination is estimated from the original image directly. Consequently, calculating the loss during the network training consisted only of the first part of the loss calculation, which is based on the cosine of the angle between the ground-truth and final illumination estimation. In further text, this experiment with one iteration will be referred to as the baseline. In Table 2, the comparison of the angular error statistics of the baseline with the proposed method is shown. It can be seen that the proposed method outperforms the baseline, especially in the case of the mean statistic and worst-performing samples.

To further validate the benefit of the proposed method, additional comparisons were made. In Table 3, estimation error statistics for the proposed method and the baseline method on worst performing samples for the baseline are shown. Worst performing samples are samples with estimation angular error higher than the value of the worst 25% statistic on the whole test set. For the baseline method, that value is 3.73°, and 33 samples have a higher error value. For 90.01% of such samples, the proposed method outperforms the baseline. Considering only the samples for which the proposed method is more accurate, the mean absolute error difference between estimates of the proposed method and estimates of the baseline is 2.43°, and when only the samples for which the baseline is more accurate are considered the difference is 0.73°. The same experiment was repeated with a different set of worst-performing samples. In Table 4, estimation error statistics for the proposed method and the baseline method on worst performing samples for the proposed method are shown. Worst performing samples were sampled using the same criterion as in the previous example. This time the threshold value was 3.20° since that is the value of the worst 25% statistic on the whole test set for the proposed method. Even though these samples were the ones for which the proposed method had the lowest accuracy, for 45.71% of samples the proposed method outperformed the baseline. The mean absolute error difference between proposed method estimates and baseline estimates when considering only the samples for which the proposed method was more accurate was 1.45°, and 2.04° when considering only the samples for which the baseline was more accurate. Finally, estimation error statistics for the proposed method and the baseline method on the intersection of worst-performing samples for both the proposed method and the baseline are given in Table 5. It can be seen that the proposed method outperforms the baseline by a significant margin.

TABLE 5. The comparison of angular error statistics of the proposed method and the baseline on the worst-performing samples for both the proposed method and the baseline on the test set (lower is better).

Algorithm	Min	Max	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
Baseline	4.08	9.47	5.92	5.55	5.56	4.35	8.02	5.77
Proposed approach	3.29	7.36	4.82	4.77	4.68	3.47	6.33	4.73

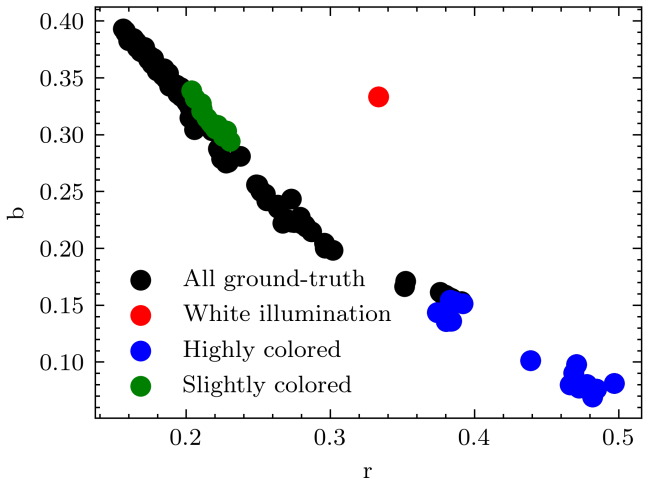


FIGURE 6. The distribution of highly colored ground-truth illuminations and slightly colored ground-truth illuminations in the test set.

Further method validation includes the comparison of method performance on images in two extrema. One extreme is images of scenes in artificial illuminations where scene illumination significantly differs from white illumination (in further text highly colored images). The second extreme contains images in daylight where the illumination was near white, i.e., illumination did not have a significant effect on image colors (in further text slightly colored images). To sample highly and slightly colored images, firstly, the angular distances between the ground-truth illuminations in the test set and a white illumination were computed according to (8). Then, highly colored images were sampled by taking images

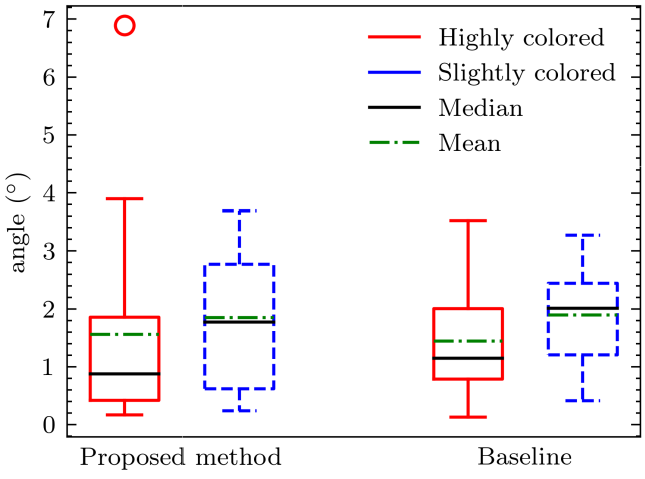


FIGURE 7. Box plot of angular errors of the proposed method and the baseline on highly colored images and slightly colored images.

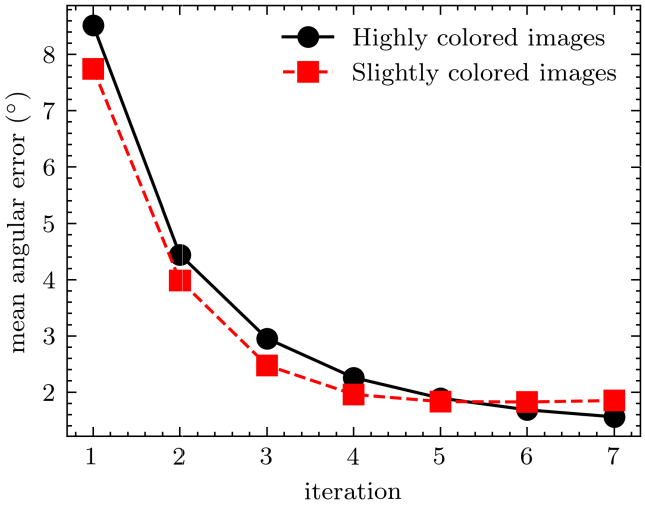


FIGURE 8. Mean angular error between cumulative estimates in each iteration and ground-truth illuminations for highly colored images and slightly colored images.

with corresponding angular distance within the 5% highest values, and slightly colored images were sampled by taking images with corresponding angular distance within the 5% lowest values. In Fig. 6, *rb*-chromaticities of ground-truth illuminations separated based the classification of highly and slightly colored images are shown.

In Fig. 7, the box plot of angular errors for the proposed method and the baseline on highly colored images and slightly colored images is given. For both groups of images, the proposed method outperforms the baseline with median angular errors 0.88° and 1.78° for highly colored images and slightly colored images, respectively. Median angular errors for the baseline were 1.15° for highly colored images and 2.01° for slightly colored images.

Since the proposed method reduces maximal estimation errors by forcing the intermediate illumination estimations to

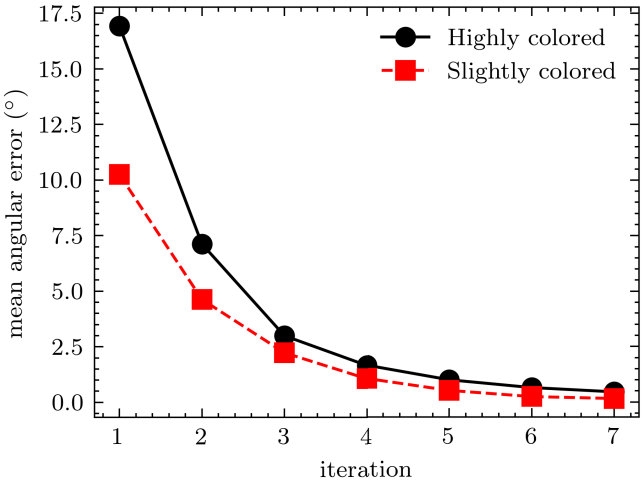


FIGURE 9. Mean angular error between intermediate estimates in each iteration and a white light illumination for highly colored images and slightly colored images.

be close to the white light, it is expected that the convergence to the ground-truth illumination is slower on highly colored images than on slightly colored images. Such behavior is shown in Fig. 8 and Fig. 9. In Fig. 8, it can be seen that for slightly colored images cumulative illumination estimates approach close to ground-truth values much faster than for highly colored images and, what is more important, after the convergence the angular error does not increase in remaining iterations. In Fig. 9, the same trend can be observed with respect to the convergence of intermediate illumination estimates on highly colored images and slightly colored images towards the white light.

V. CONCLUSION

Illumination estimation is an ill-posed problem and as such, it can not be explicitly solved. Moreover, in computational color constancy, it is usually followed by a chromatic adaptation that uses an illumination estimation expressed as a diagonal matrix which assumes independence of image color channels. Both processes are simple and may fail in some cases but when combined together in a controlled manner they could be used for iterative illumination estimation. In this paper, such an illumination estimation method is proposed. It combines illumination estimation and chromatic adaptation in a sequence. The convolutional neural network is used to compute multiple intermediate illumination estimates from an input image, which, when multiplied, correspond to the real scene illumination. By forcing the intermediate illumination estimates to be close to the white light, the proposed method avoids the estimation of highly inaccurate illuminations. The experimental results successfully validate the proposed method and its accuracy, especially in the case of worst-performing samples. Future research will include looking for an early stopping mechanism that should stop the method from entering further iterations if it already converged to the best solution it can calculate.

ACKNOWLEDGMENT

The authors would like to thank the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

[1] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.
[2] M. Ebner, "Color constancy," in *The Wiley-IS&T Series in Imaging Science and Technology*. Hoboken, NJ, USA: Wiley, 2007.
[3] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 972–984, Sep. 2002.
[4] K. Koščević, N. Banić, and S. Lončcarić "Color beaver: Bounding illumination estimations for higher accuracy," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 183–190.
[5] S. D. Hordley, "Scene illuminant estimation: Past, present, and future," *Color Res. Appl.*, vol. 31, no. 4, pp. 303–314, 2006.
[6] N. Banić and S. Loncarić, "Flash and storm: Fast and highly practical tone mapping based on naka-rushton equation," in *Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2018, pp. 47–53.
[7] E. H. Land, *The Retinex Theory of Color Vision*. Boston, MA, USA: Scientific America, 1977.
[8] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. 18th Color Imag. Conf. Final Program Color Imag. Conf.*, vol. 2010, no. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2010, pp. 256–259.
[9] N. Banić and S. Lončarić, "Using the random sprays Retinex algorithm for global illumination estimation," in *Proc. The 2nd Croatian Comput. Vis. Workshopn (CCVW)*. Zagreb, Croatia: Univ. Zagreb Faculty of Electrical Engineering and Computing, 2013, pp. 3–7.
[10] N. Banic and S. Loncaric, "Color rabbit: Guiding the distance of local maximums in illumination estimation," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 345–350.
[11] N. Banic and S. Loncaric, "Improving the white patch method by subsampling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 605–609.
[12] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
[13] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imag. Conf.*, vol. 2004, no. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2004, pp. 37–41.
[14] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.
[15] A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 918–929, May 2012.
[16] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 81–89.
[17] Z. Lou, T. Gevers, N. Hu, and M. P. Lucassen, "Color constancy by deep learning," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–76.
[18] Y. Hu, B. Wang, and S. Lin, "Fully Convolutional Color Constancy with Confidence-weighted Pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4085–4094.
[19] K. Koscevic, M. Subasic, and S. Loncaric, "Attention-based convolutional neural network for computer vision color constancy," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 372–377.
[20] K. Koščević, M. Subašić, and S. Lončarić, "Guiding the illumination estimation using the attention mechanism," in *Proc. 2nd Asia Pacific Inf. Technol. Conf.*, Jan. 2020, pp. 143–149, doi: 10.1145/3379310.3379329.
[21] S. W. Oh and S. J. Kim, "Approaching the computational color constancy as a classification problem through deep learning," *Pattern Recognit.*, vol. 61, pp. 405–416, Jan. 2017.
[22] K. Koscevic, M. Subasic, and S. Loncaric, "Deep learning-based illumination estimation using light source classification," *IEEE Access*, vol. 8, pp. 84239–84247, 2020.
[23] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 371–387.
[24] M. Afifi and M. S. Brown, "Sensor-independent illumination estimation for DNN models," 2019, *arXiv:1912.06888*. [Online]. Available: http://arxiv.org/abs/1912.06888
[25] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[26] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1904–1911.
[27] D. A. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 5–35, Aug. 1990.
[28] K. Barnard, "Improvements to gamut mapping colour constancy algorithms," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2000, pp. 390–403.
[29] G. D. Finlayson, S. D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 93–109, Apr. 2006.
[30] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.
[31] J. T. Barron and Y.-T. Tsai, "Fast Fourier color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 886–894.
[32] J. van de Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[33] N. Banic and S. Loncaric, "Color cat: Remembering colors for illumination estimation," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 651–655, Jun. 2015.
[34] N. Banic and S. Loncaric, "Using the red chromaticity for illumination estimation," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2015, pp. 131–136.
[35] N. Banic and S. Loncaric, "Color dog–guiding the global illumination estimation to better accuracy," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, Mar. 2015, pp. 129–135.
[36] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2254–2263.
[37] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Effective learning-based illuminant estimation using simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1000–1008.
[38] G. D. Finlayson, M. S. Drew, and B. V. Funt, "Diagonal transforms suffice for color constancy," in *Proc. 4th Int. Conf. Comput. Vis.*, May 1993, pp. 164–171.
[39] G. West and M. H. Brill, "Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy," *J. Math. Biol.*, vol. 15, no. 2, pp. 249–258, Oct. 1982.
[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[41] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360
[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[43] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks Trade*. Berlin, Germany: Springer, 2012, pp. 9–48.
[44] O. Sidorov, "Artificial color constancy via GoogleNet with angular loss function," *Appl. Artif. Intell.*, vol. 34, no. 9, pp. 643–655, 2020.
[45] N. Banić, K. Koščević, and S. Lončarić, "Unsupervised learning for color constancy," 2017, *arXiv:1712.00436*. [Online]. Available: http://arxiv.org/abs/1712.00436
[46] F. Chollet. (2015). *Keras*. [Online]. Available: https://keras.io
[47] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[48] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[49] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1973–1985, Oct. 2015.
[50] N. Banić and S. Lončarić, "A perceptual measure of illumination estimation error," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, Mar. 2015, pp. 136–143.

KARLO KOŠČEVIĆ (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science in 2016 and 2018, respectively. He is currently in his second year of the technical sciences in the scientific field of computing Ph.D. program with the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. His research interests include image processing, image analysis, and deep learning. His current research is in the area of color constancy with a focus on learning-based methods for illumination estimation.

MARKO SUBAŠIĆ (Member, IEEE) received the Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. Since 1999, he has been working with the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently an Associate Professor. He teaches several courses at the graduate and undergraduate levels. His research interests include image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is also a member of the IEEE Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.

SVEN LONČARIĆ (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Cincinnati, Cincinnati, OH, USA, in 1994, as a Fulbright Scholar. He was an Assistant Professor with the New Jersey Institute of Technology, Newark, NJ, USA, from 2001 to 2003. He is currently a Professor of Electrical Engineering and Computer Science at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. He was the principal investigator on a number of R&D projects. He is the Director of the Center for Computer Vision, University of Zagreb and the Head of the Image Processing Group. He is a Co-Director of the Center of Excellence in Data Science and Cooperative Systems. He has coauthored more than 250 publications in scientific journals and conferences. His research interests include image processing and computer vision. He was the Chair of the IEEE Croatia Section. He is a member of the Croatian Academy of Technical Sciences. He received several awards for his scientific and professional work.

Publication 6

Ershov, E., Savchik, A., Semenkov, I., Banić, N., Belokopytov, A., Senshina, D., **Koščević, K.**, Subašić, M., Lončarić, S.,"The Cube++ Illumination Estimation Dataset", IEEE Access, Vol. 8, 2020, pp. 227511-227527.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 23, 2020, accepted December 7, 2020, date of publication December 16, 2020,
date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045066

The Cube++ Illumination Estimation Dataset

EGOR ERSHOV[1], ALEXEY SAVCHIK[1], ILLYA SEMENKOV[1], NIKOLA BANIĆ[2], (Member, IEEE), ALEXANDER BELOKOPYTOV[1], DARIA SENSHINA[1], KARLO KOŠČEVIĆ[3], (Student Member, IEEE), MARKO SUBAŠIĆ[3], (Member, IEEE), AND SVEN LONČARIĆ[3]

[1]Institute for Information Transmission Problems, Russian Academy of Sciences, 119991 Moscow, Russia
[2]Gideon Brothers, 10000 Zagreb, Croatia
[3]Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia

Corresponding author: Egor Ershov (ershov@iitp.ru)

This work was supported in part by the Croatian Science Foundation (under data collection) under Project IP-06-2016-2092, and in part by the Russian Science Foundation (under data processing) under Grant 20-61-47089.

ABSTRACT Computational color constancy has the important task of reducing the influence of the scene illumination on the object colors. As such, it is an essential part of the image processing pipelines of most digital cameras. One of the important parts of the computational color constancy is illumination estimation, i.e. estimating the illumination color. When an illumination estimation method is proposed, its accuracy is usually reported by providing the values of error metrics obtained on the images of publicly available datasets. However, over time it has been shown that many of these datasets have problems such as too few images, inappropriate image quality, lack of scene diversity, absence of version tracking, violation of various assumptions, GDPR regulation violation, lack of additional shooting procedure info, etc. In this paper a new illumination estimation dataset is proposed that aims to alleviate many of the mentioned problems and to help the illumination estimation research. It consists of 4890 images with known illumination colors as well as with additional semantic data that can further make the learning process more accurate. Due to the usage of the SpyderCube color target, for every image there are two ground-truth illumination records covering different directions. Because of that, the dataset can be used for training and testing of methods that perform single or two-illuminant estimation. This makes it superior to many similar existing datasets. The datasets, it's smaller version SimpleCube++, and the accompanying code are available at https://github.com/Visillect/CubePlusPlus/.

INDEX TERMS Color constancy, dataset, illumination estimation, white balancing, multiple illumination, mixed illumination.

I. INTRODUCTION

The human visual system is able, in some conditions, to recognize colors despite the influence of the illumination on their appearance through the ability known as color constancy [1]. It is not yet fully understood how this ability functions and therefore it is not possible to directly model it. Nevertheless, various computational color constancy methods are used in the pipelines of digital cameras. They are usually designed to first identify the chromaticity of the light source and then to remove its influence on the scene. The last one is described in details here [2]–[5]. For both of these tasks, the commonly used image formation model that also includes the Lambertian assumption is usually given as

$$f_c(\mathbf{x}) = \int_\omega I(\lambda, \mathbf{x}) R(\lambda, \mathbf{x}) \rho_c(\lambda) d\lambda \tag{1}$$

where $\mathbf{x}$ is a pixel in the image $\mathbf{f}$, $c \in \{R, G, B\}$ is the color channel, λ is a wavelength in the visible light spectrum ω, $I(\lambda, \mathbf{x})$ is the spectral distribution of the light source, $R(\lambda, \mathbf{x})$ is the surface reflectance, and $\rho_c(\lambda)$ is the camera sensitivity for the color channel c. It is often assumed that the scene illumination is uniform. This means that the spatial information is not required in the illumination estimation equations and so the color of the observed light source $\mathbf{e}$ is

$$\mathbf{e} = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_\omega I(\lambda) \boldsymbol{\rho}(\lambda) d\lambda. \tag{2}$$

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif.

For a somewhat satisfying color correction, it is already enough to know the direction of **e** [6], which means that **e** can be described by chromaticities instead of colors. For example, r, g, and b chromaticity components are calculated as R, G, and B color components divided by their sum so that $r + g + b = 1$. Thus, knowing only two of them is enough.

Since there are more unknowns than equations, illumination estimation is an ill-posed problem and additional assumptions have to be made in order to tackle it. Because of that, numerous illumination estimation methods with various assumptions have been proposed and they are often divided into two groups: the low-level statistics-based methods and the learning-based methods.

The low level statistics-based methods include White-Patch [7], [8] and its improvements [9]–[11], Gray-World [12], Shades-of-Gray [13], 1st and 2nd order Gray-Edge [14], Weighted Gray-Edge [15], using bright pixels [16], gray pixels [17] or bright and dark colors [18], exploiting illumination perception [19] and expectation [20], etc. Interesting to note that ≪gray balancing≫ occurs also in scope of printer calibration [21].

Learning-based methods include neural networks [22], high-level visual information [23], natural image statistics [24], Bayesian learning [25], [26], spatio-spectral learning [27], methods restricting the illumination solution space [28]–[30], color moments [31], regression trees with simple features from color distribution statistics [32], spatial localizations [33], [34], convolutional neural networks [35]–[38] and genetic algorithms [39], modelling color constancy by using the overlapping asymmetric Gaussian kernels with surround pixel contrast based sizes [40], finding paths for the longest dichromatic line produced by specular pixels [41], detecting gray pixels with specific illuminant-invariant measures in logarithmic space [42], channel-wise pooling the responses of double-opponency cells in LMS color space [43], sensor-independent learning [44], [45], and numerous others. Learning-based methods have much higher accuracy than statistics-based ones, but they are usually slower [46].

While the number of the proposed illumination estimation methods is ever-growing, there are not too many illumination estimation datasets and even the existing ones have various problems. These include too few images, inappropriate image quality, lack of scene diversity, multiple poorly synchronized versions of the same dataset, violation of various assumptions, etc. A high-quality illumination estimation dataset should be:

- *Diverse*. The more content and illumination cases are covered, the higher is the testing quality.
- *Large*. It is important that the datasets are not only diverse but that they also contain many images for each particular case. This makes it possible to notice quality improvement even for rare cases [47].
- *Informative*. Dataset should contain as much information about each captured image as possible. Precisely the information available during shooting procedure, meta-information about scene properties, information about light sources from different angles, etc.
- *Updatable*. Every illumination estimation dataset usually contains ground-truth illumination errors. Because of that, the dataset infrastructure should provide simple and reliable way for dataset debugging and tracking of its versions.
- *Verifiable*. From the previous point, it follows that the dataset should be available for verification, namely all provided markup and ground-truth can be collected and, if necessary, recreated by anyone who just downloads the source images.
- *Accessible*. The value of a dataset is decreasing when the downloading process is too complicated or time-consuming.
- *GDPR compliant*. Even a very good dataset can be of limited use for European researches if it is not compliant with GDPR, because it may prevent the researchers from publishing some of their results without breaking the regulations.

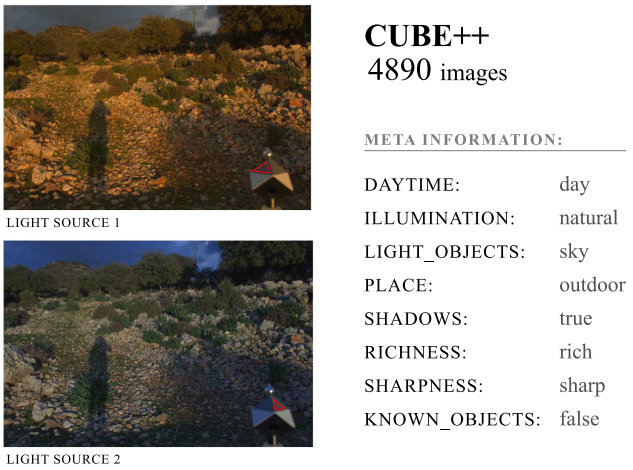


FIGURE 1. Examples of chromatic adaptation based on two captured ground-truth illumination colors for an image from Cube++, a new large dataset where each image is accompanied by ground-truth illumination from several directions and semantic information about the scene content. This enables the research of single and multiple illumination scenarios as well as selection of images by various criteria.

In this paper a new illumination estimation dataset named Cube++ with all of these properties is described. It contains 4890 images (see Fig. 2) carefully calibrated so as to get highly accurate ground-truth illumination. The images were collected in numerous countries, places, and illumination conditions. The countries in question include Austria, Croatia, Czechia, Georgia, Germany, Romania, Russia, Slovenia, Turkey, and Ukraine. In order to enable easy selection of images with specific properties, each image is accompanied by additional semantic information such as whether there are shadows in the image, whether it is an indoor or an outdoor image, whether the scene contains objects with known coloration, etc. An example of an image from Cube++ is shown in Fig. 1. The dataset is appropriate for different light source

FIGURE 2. Example images from the newly created Cube++ dataset.

estimation use cases such as: single light source estimation, two light sources estimation, or estimation of at least one significant light source in the scene. Finally, some of the collected images were not included in the dataset and they are kept aside to be released later as part of a future illumination estimation benchmark somewhat similar to [48].

The paper is structured as follows: Section II describes the most important existing illumination estimation datasets and the problems associated with them, Section III gives the motivation for creating a new dataset, Section IV describes the methodology used to collect the dataset, Section V describes the newly proposed Cube++ dataset, Section VI presents a discussion about the scientific usefulness of contemporary datasets' form and about a potential improvement, and finally, Section VII concludes the paper.

II. INFLUENTIAL EXISTING DATASETS

One of the first illumination estimation datasets with a large number of images was the GreyBall dataset [49]. A gray ball was placed in the scene of each of the 11346 images to extract the ground-truth illumination. The main problem with this dataset is that the images are non-linearly processed and as such, they do not comply with the image formation model given in Eq. (1). Furthermore, the images in the GreyBall dataset are relatively small with a size of 240×360. Finally, the images were extracted from a video that was captured at several locations, which means that many of them have highly correlated illuminations and content. To cope with this problem of high redundancy, it has been proposed to use only a subset of 1135 images from GreyBall [50], [51].

In 2008, the ColorChecker dataset [25] with its 568 images was published and the ground-truth illumination was extracted by means of putting a color checker instead of a gray ball in the image scenes. This dataset was created by two different cameras and its images, which are individually bigger than the ones in the GreyBall dataset, also underwent non-linear processing, which means that similarly as with the GreyBall dataset they are given as 8-bit per-channel JPEG images.

In 2011, the reprocessed version of the ColorChecker dataset that contains only linearly processed images was published [52]. However, as observed already in 2013 [53], it was not mentioned clearly enough that the black level was supposed to be subtracted before using the images. Despite this observation, a lot of papers continued publishing results of methods obtained on the technically unprepared images with the black level included. This effectively led to the circulation of at least three versions of the ColorChecker datasets and the problem was formally addressed in [54] by also bringing into question the alleged advances in the illumination estimation research. In 2018, there was an attempt to rehabilitate the ColorChecker dataset by publishing the recalculated accuracy of various methods by using the allegedly correct ground-truth [55]. However, this attempt was marred by serious technical faults and wrong calculations that included comparing the estimations obtained on older versions to the new ground-truth, which only introduced further confusion [56]. This effectively opens the possibility of more future versions of the results on the ColorChecker dataset. In short, using the ColorChecker dataset can be very confusing and

problematic due to many circulating versions of the alleged results and consequent inappropriate comparisons, and therefore, to avoid problems, it should probably be omitted as the primary dataset choice.

In 2014, nine new NUS datasets with each of them taken by one of nine different cameras were published [18]. The images were only linearly processed, the black level subtraction was performed from the start in the initial paper, and the number of images was sufficiently high. The calibration object used to extract the ground-truth illumination was again a color checker. However, the problems with the NUS datasets include violations of uniform illumination assumption when having only a single ground-truth illumination, a relatively small number of images per camera with 268 being the maximum, having the same scenes repeated in images, and not being GDPR compliant as well as neither of the previous datasets.

In 2017, the Cube dataset [44] was published with 1365 images taken with a single camera and with a SpyderCube [1] calibration tool used for calibration. Due to its geometry that is superior to the one of a color checker, SpyderCube allows for easier detection of the presence of two illuminations and their extraction. This was extensively used to carefully calibrate each of the images of the Cube dataset and to obtain an accurate ground-truth. Special care was also taken to avoid the violation of the uniform illumination assumption as much as possible. The main drawback of the Cube dataset is that it contains only outdoor images, which also negatively affects the ground-truth illumination distribution. This drawback was alleviated in the Cube dataset's extension named the Cube+ dataset [44]. It contains 342 additional indoor images for a total of 1707 images and a wider span of ground-truth illumination distribution similar to the one in other datasets.

A relatively recent dataset is the INTEL-TAU dataset [57], a successor to the INTEL-TUT dataset [58], with 7022 images taken by three different cameras. While the number of images is sufficiently high, its main drawback is the fact that most of its images do not contain a calibration object in their scenes. Namely, it was removed after the initial calibration. Although this removes the requirement for masking it out, it also makes it impossible to reliably check and verify the ground-truth calibration and it is known that such errors occur [59]. Additionally, since the original raw image files are not provided, the EXIF data with the meta-information that may be important to some methods is also not available. The INTEL-TAU dataset is also completely GDPR compliant. Instead of avoiding problematic scenes, GDPR compliance was achieved by having ''privacy masking applied on all sensitive information'' such as ''recognizable faces, license plates, and other privacy-sensitive information''. The masking was performed so that ''color component values inside the privacy masking area were averaged''. However, this effectively changes the original content and it may be undesirable in some cases.

A relatively recent dataset is the one for temporal color constancy [60], which contains 600 sequences of varying length between 3 and 17 frames. The dataset has not yet been made publicly available at the moment of writing this paper.

It is also important to mention that in contrast to all the described datasets that contain real-world images taken in mostly uncontrolled conditions, there are a lot of datasets made in fully controlled or even laboratory conditions, such as [6], [61]–[66].

The main advantage of the laboratory dataset is that it allows to research particular problem in fully-controlled conditions, but the variability of such datasets is often too low.

While other illumination estimation benchmark datasets also exist, it can be argued that the ones mentioned here are the most influential ones. They also share many problems with other existing datasets and thus their descriptions also cover most of the problems of other datasets. Some characteristics of the datasets mentioned here are summarized in Table 1.

III. MOTIVATION

After laying out the brief descriptions of some of the best-known illumination estimation benchmark datasets, it is possible to identify some of their main problems already recognized by the wider interested research community. Therefore, the motivation for creating a new illumination estimation is to try to reduce or entirely eliminate some of the mentioned problems of the existing datasets.

A. SIMPLE TECHNICAL FAULTS

Probably the most serious and most detrimental problem is the one connected to the technical shortfalls that can happen when creating and publishing a dataset. Some of the main such shortfalls are using non-linearly processed images and providing confusing information on black level subtraction.

As for the non-linearly processed images, the solution is to simply avoid performing non-linear processing and this can be simply carried out.

In the case of the black level subtraction, with the earlier datasets, this problem occurred due to a lack of explicit mentioning of the black level value in the papers that originally described these datasets. Additionally, in some cases, even a script that demonstrates the proper handling of the black level was either missing or put to a somewhat obscure location. In the case of the ColorChecker, such problems have led to multiple circulating versions of the ground-truth data and experimental results. Therefore, in the case of publishing a new dataset, such and similar problems motivate to clearly provide all necessary details on the required data for the black level subtraction and also to provide an example of how to do it.

[1] https://www.datacolor.com/photography-design/product-overview/spydercube/

TABLE 1. Characteristics of different illumination estimation datasets; the Cube++ dataset is described later in Section V.

Characteristic	*GreyBall [49]*	*ColorChecker [52]*	*NUS [18]*	*Cube+ [44]*	*INTEL-TAU [57]*	*TCC [60]*	***Cube++***
Number of images	11346 [2]	568	1736	1707	7022	approx. 10000 [2]	4890
Number of sensor types	1	2	8	1	3	1	1
GDPR compliance	-	-	-	-	✓	✓	✓
Undistored content	✓	✓	✓	✓	-	✓	✓
Color target in the scene	✓	✓	✓	✓	-	✓	✓
Color target type	gray ball	color checker (CC)	CC	SpyderCube	CC	SpyderCube	SpyderCube
Meta-info about the scene	-	-	-	-	-	-	✓
Night images	-	-	-	-	-	-	✓
Winter season	-	-	-	-	-	✓	✓
Year	2003	2020 [3]	2014	2019	2020	2020 [4]	2020
License	-	-	-	-	CC BY-SA 4.0	MIT	CC BY 4.0

B. RELIABLE GROUND-TRUTH

One of the probably least detectable technical faults with serious consequences is erroneous calibration and ground-truth illumination extraction. Based on the experience with existing datasets, it usually happens that there are multiple illuminations in the scene and the calibration object is under the influence of only one of them, which may not even be the dominant one. In that case, even if a method estimates the dominant illumination, it will be penalized because the ground-truth is based on another illumination. As mentioned earlier, this was already reported for the ColorChecker dataset.

To make the ground-truth reliable, one should use such a calibration objects that can detect the presence of multiple illuminations. Examples of such calibration objects include a gray ball such as in the GreyBall dataset or a SpyderCube instance such as in the Cube+ dataset, because they make it possible to simultaneously observe illuminations coming from different angles, and these can then be checked for any significant difference. An example of capturing two significantly different illuminations with a SpyderCube instance and showing the difference in how they affect color correction is given in [44]. If a significant difference is present, additional steps can be taken to either correctly determine which of the illuminations is the dominant one or to discard the image to prevent any future problems, which finally results in a correctly extracted and reliable ground-truth illumination.

C. VERIFIABLE GROUND-TRUTH

While the ground-truth should primarily be reliable, it should also be verifiable in order to add an additional layer of reliability. The simplest way of making the ground-truth of a dataset verifiable is to have all the dataset images contain a calibration objects in their scenes. In that way the ground-truth can easily be extracted by other researchers and then compared to the originally provided one to look for potential errors. Additionally, the very visual information can help identify cases such as e.g. having the calibration object in a shadow while the majority of the scene is outside of that shadow.

D. CONTENT VARIETY

A new illumination estimation dataset should have a high content variety. While this seems rather obvious, it is not always put into practice to the full extent. For example, while the GrayBall dataset contains over 11k images, they are highly correlated and thus effectively not as rich in content as it may seem at first. In the case of datasets such as the ColorChecker dataset or the NUS datasets, all images were taken at the same geographical location and during the same season. None of the images there were taken e.g. during winter or at night. Such content choice restriction results in failure to cover many interesting and challenging environments that illumination estimation methods encounter in real-world applications and that should also be included in the research.

E. ILLUMINATION VARIETY

Having an appropriate ground-truth illumination variety in an illumination estimation dataset is important for several reasons. The most important one is to closely cover as much as possible of the illuminations that are encountered in the real-world applications because in that way the illumination estimation methods can be properly trained and tested.

An additional reason to have a sufficient ground-truth illumination variety is to prevent abuses of some often used error statistics that are possible if the ground-truth illumination are too clustered [67]. Such abuses can lead to false conclusions about the performance of the tested methods and consequently be detrimental for the research community and practitioners.

F. CHECKING FOR MULTIPLE ILLUMINATIONS

The majority of the illumination estimation datasets provide only a single ground-truth illumination per image. This effectively means that in terms of evaluation these datasets implicitly assume an uniform illumination. However, it is know that in illumination estimation datasets this is generally not the case [56]. As a matter of fact, any image with shadows has already effectively at least two illuminations that may differ significantly and this can also have a significant outcome on the later color correction step [44]. Additionally, even if there are no shadows, it is still possible for an image to be under the influence of multiple illuminations. In that case having a calibration object that is designed to successfully capture the illumination from only one direction at a time will fail to capture all the illuminations in the scene, let alone to detect their presence. Capturing only a single illumination when there are more present also leads to a problem during the evaluation. Namely, if a method correctly estimates one of the illuminations, but the other one is marked as the ground-truth, it may be argued that in this case the method is being unfairly judged. Because of that, an illumination estimation dataset should preferably use calibration objects that can simultaneously capture the illumination color from multiple directions. This would solve at least two problems. First, it would detect whether there are multiple illuminations in the first place, and second, if there really are multiple illuminations in the scene, then such a calibration object will capture more information on them. An example of such a calibration object is the SpyderCube object that has been described earlier.

G. NUMBER OF IMAGES

While some of the previous datasets with non-linearly processed images are obviously disadvantageous, some of them like the GreyBall dataset have the advantage of having thousands of images, which still makes them attractive to many researchers. Therefore, besides having a technically correct dataset, it is also important to make it have a sufficiently large number of images. This can result in both making the dataset desirable by offering a lot of useful data as well as simultaneously discouraging researches from using the inferior older datasets just because of their size. As for how large exactly a dataset should be, it should contain several thousands of images to outsize the existing datasets of lower quality and also to enable new breakthroughs. Finally, having a dataset with a large number of images is a prerequisite for achieving the previously mentioned content and illumination variety.

H. SEMANTIC DATA

In numerous cases, additional semantic information can be useful for research of specific illumination estimation methods. For example, some of the methods may be interested in being explicitly trained only on indoor or outdoor images. Others may be interested in training images that contain no shadows whatsoever since they introduce additional illuminations. More generally, it may be useful to know whether there is a violation of the uniform illumination assumption on a given image. In such cases, it can be highly practical to be able to efficiently filter out images from a dataset based on some given criteria.

Because of that, a useful addition to a new illumination estimation benchmark dataset would be semantic information for each image. In that way, the research could be speeded up by not requiring researchers to label the images from scratch. Additionally, if such semantic information were given in advance, a lot of potential label mismatches between the labels created by different researchers could also be prevented.

I. PRIVACY CONCERNS

With the recent arrival of regulations such as the General Data Protection Regulation (GDPR), it becomes ever more important to respect privacy in publicly available images. This also means that using images from previous datasets with e.g. recognizable people or registration plates may nowadays be potentially seen as problematic. With respect to this, for the sake of respecting privacy, any new illumination estimation dataset should also take care of avoiding images that would contain any content that could compromise someones privacy.

On the other hand, if a public dataset is also supposed to be useful for development of methods that rely on e.g. faces [68], [69] or sclera [70], then it should obviously also contain images with faces. However, in that case it would be appropriate to obtain the consent for public use from the persons present in the image scenes. That would enable the researchers to use and show these images publicly in papers.

J. MULTIPLE INSTANCES OF THE SAME SENSOR CLASS

There can be significant differences between spectral characteristics of different sensors used by various cameras. This effectively means that a learning-based method that has been successfully trained on the images created by a camera of one model will not necessarily perform well on images created by a camera of another model without some adjustments. As a result, the problem of inter-camera color constancy has recently started to gain ever more attention [44], [45], [71]. Since almost every dataset was taken with another camera sensor, there is no shortage of training and testing images.

On the other hand, it is known and it can be shown that even for the instances of the same sensor class there are measurable differences in the spectral characteristics [72]. Hence, to check the significance of the impact of these differences on the accuracy of illumination estimation methods, an interesting feature of an illumination estimation dataset would be to have images created by several instances of the same sensor class. In addition to ground-truth illumination,

[2]Images are highly correlated. The GreyBall images are taken from 2 hours of video. The TCC images are taken from 600 video sequences.

[3]The cited version of ground-truth values were published in 2020, the original dataset was published in 2008.

[4]Expected in 2020, the data is not published at the time of writing.

such a dataset would also have to provide the sensor instance labels for each image.

IV. ACQUISITION METHODOLOGY

By identifying the problems with the existing datasets and describing some desired properties of the future datasets, the guidelines for creating a new illumination estimation dataset have been laid out. One of the main goals of this paper is not just to provide a theoretical framework, but also to create and propose a dataset by following these guidelines. The first step in doing so is to describe the used acquisition methodology.

A. TECHNICAL SETUP

1) COLOR TARGET (SpyderCube) CHARACTERISATION

As the calibration tool in the newly proposed dataset, the SpyderCube instances were used. SpyderCube is a color target for photographers whose main purpose is to help them to adjust the white balance manually. The general look of the SpyderCube is given in Fig. 3. A chrome ball is used to analyze specular highlights, the white on two faces is used to estimate true highlight value, the gray on two faces represents the midtone of the image and its color temperature, and the bottom black face is used to evaluate shadow values in the scene in relation to the black trap i.e. the hole, which represents absolute black.

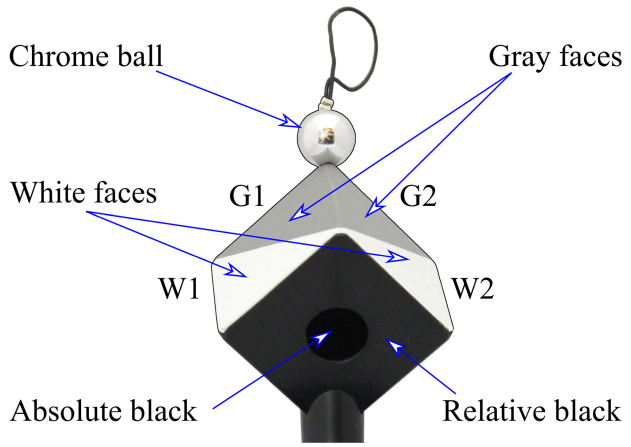


FIGURE 3. SpyderCube general look.

According to the manufacturer company Datacolor, the gray cube faces are neutral gray with a reflection coefficient of 18%.

Since SpyderCube is a relatively low-cost tool, some doubts about its declared optical properties could arise. To validate its properties, two SpyderCube instances, labeled SC1 and SC2, were compared. Individual faces of these SpyderCube instances were named G1, G2, and W1, W2, as shown in Fig. 3.

Reflection spectra of SpyderCube parts were measured using a Eye-One Pro spectrophotometer by X-Rite in the high-resolution mode of 3.3 nm with the help of the spotread

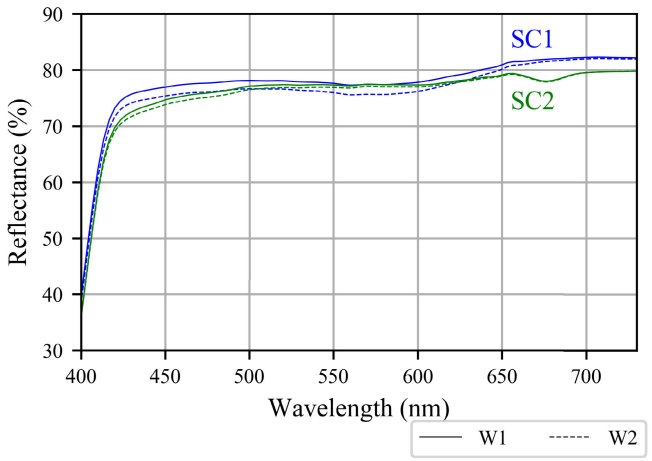


FIGURE 4. SC1 and SC2 white parts reflectance spectra.

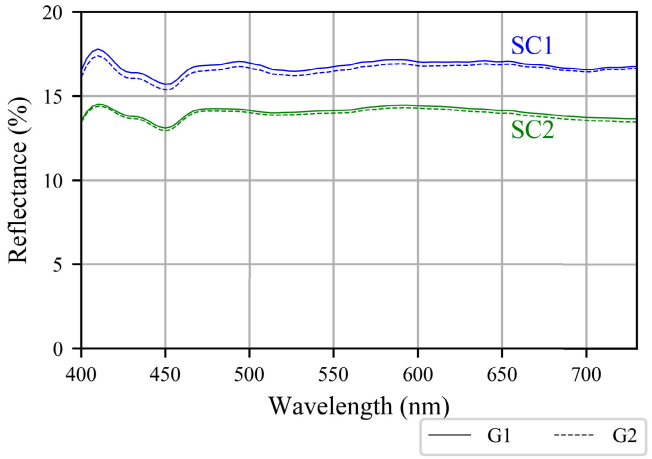


FIGURE 5. SC1 and SC2 gray parts reflectance spectra.

utility from Argyll CMS.[5] For each part, three measurements were made and the results were averaged. Figures 4 and 5 show the spectral reflection coefficients of the white and gray parts of the SC1 and SC2, respectively.

These measurements lead to the following observations:

- Gray parts of both SpyderCube instances are not "ideal" gray, i.e. the reflection spectra slightly depend on the wavelength. The sensitivity of the blue sensor in many cameras has a maximum at around 450 nm wavelength, and the reflection coefficients of gray parts G1 and G2 have a noticeable drop in the blue band.
- Each SpyderCube instance has small differences between reflection coefficients of its own gray parts G1 and G2.
- There are rather big differences between the gray parts reflection coefficients of the two measured SpyderCube instances.

[5]http://www.argyllcms.com/

- White parts of both SpyderCube instances are also not "ideal" white, i.e. the reflection spectra are not horizontal lines.
- Differences of the white parts W1 and W2 reflection coefficients of the both SpyderCube instances are small.

The idea behind SpyderCube as a calibration tool is that it does not distort the color of the illumination source, i.e. it is assumed to be "color neutral". From this point of view, what is important is the similarity between the shapes of the curves of reflection coefficients for the two SpyderCube instances and not the differences between the curves' values. From the measurements, it can be concluded that the curve shapes are indeed very similar. Therefore, the SpyderCube "color neutrality" assumption generally holds with one exception being the blue region of the spectrum for the white faces.

The degree of SpyderCube's color neutrality is one of the most important factors for accurate ground-truth extraction. The height of the grey reflection coefficient curve does not significantly influence the ground-truth extraction. Compared to other uncertainties, the measured deviations from color neutrality have only rather a small impact on the ground-truth.

Nevertheless, to measure the amount of this impact in terms of practical use, several images of two SpyderCube instances that were simultaneously in the same scene were captured with a Canon 600D camera under a D50-like illumination. The average difference between the ground-truth illuminations extracted from the faces of each SpyderCube instance and measured in terms of the angular reproduction error [73] was about 0.15°, which is in terms of color reproduction insignificant and invisible [74].

Performed measurements effectively demonstrate that using grey faces of different SpyderCube instances in different images has no significant effect on the overall ground-truth extraction quality. Still, SpyderCube quality should be studied in details also for all types of complicated artificial light sources (such as gas discharge lamp, etc.).

2) HANDHELD SETUP

To collect the dataset in natural conditions, the following equipment shown in Fig. 6 was used:

- Canon 550D camera or Canon 600D,
- SpyderCube calibration tool, and
- special attachment of the cube to the camera.

Special cube fasteners were built that allows the cube to be positioned so that it appears near the lower right corner of the image. The fasteners can also be rotated both in horizontal and vertical planes. The distance of the cube from the camera can be adjusted using a telescopic monopod and during the dataset, images capturing it was set to 50 cm. The experience gained while collecting the dataset images has led to the conclusion that the custom-built handheld setup is convenient to use.

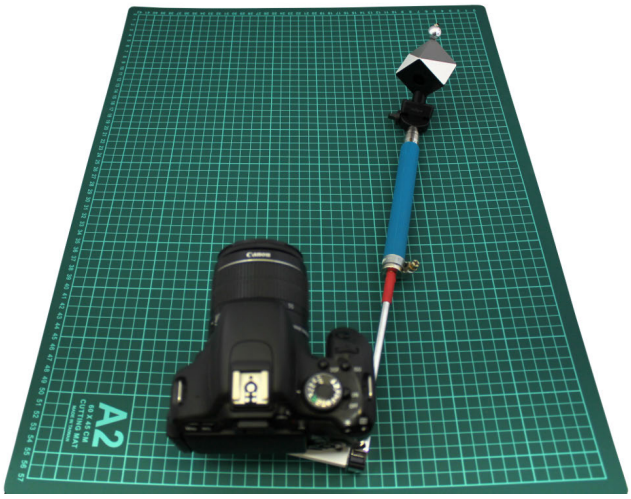


FIGURE 6. The general look of the handheld setup with Canon 600D camera.

FIGURE 7. Examples of images that should be excluded from the dataset: a) the color target is illuminated by the local lantern from the near shop, the color is different from the lighting of the most of scene; b) the color target is illuminated by sources that have almost no effect on the lighting of the observed scene; c) overexposed color target; d) the overly dark image.

B. DATA COLLECTION AND FILTRATION

The main thing to pay attention to during the image capturing was to assure that the used target cube and the majority of the observed scene are under the same illumination or illuminations. Examples of images with scenes where this requirement was not met are shown in Fig. 7.

Another significant factor that prevents accurate ground-truth extraction is the occurrence of glare on the color target. Images with this issue are usually characterized by clipping of the values in one of the color channels on the gray or white faces of the color target. The overexposure can be avoided in at least two ways: either by using manual camera settings or by specifying relative exposure compensation.

Dimming by one step usually turned out to be enough during the image collecting. Manual camera settings and one step lighting can also help to properly deal with the overly dark images. Examples of an overexposed and a too dark image are shown in Fig. 7c and Fig. 7d, respectively.

It should again be mentioned that there may often be several different illumination sources in one scene, commonly two, e.g., sun and sky or sky and streetlight. In this case, especially if the areas of the scene parts illuminated by different light sources are comparable, it is preferable to place the cube so that both illumination colors are captured by different cube faces. By doing so, it is later possible to simultaneously extract the illumination color of both influential scene light sources.

One of the main problems during image acquisition was to find the right position for the photographer to avoid differences between illuminations influencing the target cube and most of the observed scene. A lot of interesting scenes are available only in urban areas where there are a lot of different artificial illuminations. However, scenes in urban areas are usually full of different personal data like faces or plate numbers, which means that there are some difficulties related to GDPR. To make Cube++ GDPR compliant, the images with humans in the scene were filtered out and removed. This was done both automatically by using YOLOv3 [75] and manually by additionally checking each image.

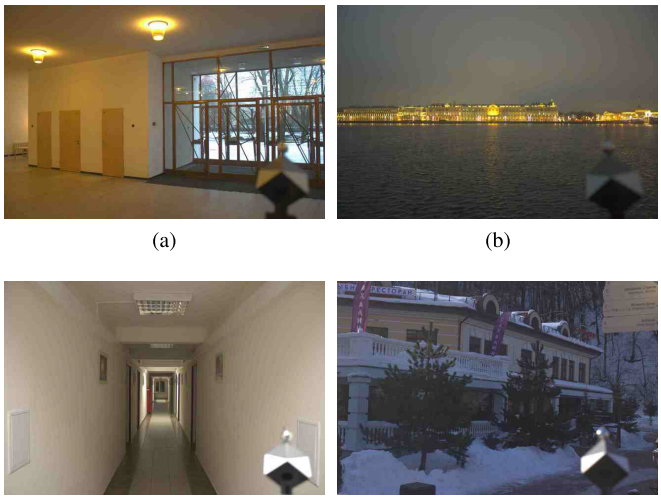


FIGURE 8. Examples of images with partial illumination estimation: a) and b) some scene parts are illuminated by the light source not captured by the cube; c) illumination significantly varies in the scene due to the interreflections d) all of the scene is in the shadow, while one of the cube faces is illuminated by the sun.

During the final quality filtration, all images were divided into three categories: a) images with **full light source estimation**, where the cube was illuminated by all the main light sources in the scene, b) **incorrect** images where the cube does not allow to determine the illumination rating consistent with the scene, and c) the rest i.e. images with **partial light source estimation**, see Fig. 8.

As a result, about 400 images were marked as incorrect and removed from the dataset, about 524 images were marked as difficult images with a partial light source estimation, and the rest of the images were marked as good.

Additionally, the fiber on the top of the cube may fall on a cube or remain on the image after cropping out the color target. To prevent it, the fibers were glued to the cube or just cut off on most images. All of the images are captured horizontally without the use of a camera flash.

C. GROUND-TRUTH EXTRACTION

The ground-truth extraction was performed on raw images. First, a simple debayering has been performed by transforming each RGGB Bayer pattern square into a single pixel. The red and blue channels of the pixel color were obtained directly from the R and B components of the pattern, while the green was obtained by averaging the two G values. No interpolation was performed and therefore the number of image rows and columns was halved. Next, the oversaturated pixels were masked out, and then the black level of 2^{11} was subtracted from all pixels. Finally, the ground-truth illumination values were extracted by calculating the average chromaticity of the manually annotated areas of the SpyderCube triangles.

Four chromaticities were calculated for every image. They correspond to white and gray triangles on the left and right cube faces. Note that on the brightly illuminated cubes, the white triangles may have oversaturated areas that cannot be properly used. On the contrary, the gray triangles chromaticities on the darker images may not be stable due to the black level noise. It is important to note no image contains saturated grey edges, while some of the images contain saturated white edges and in such cases, a corresponding mark is provided.

The illuminations for a triangle were calculated as the mean illumination of its area after 50% downscale to the barycenter. The value of 50% is selected as a simple empirical trade-off. Namely, a full-size triangle may contain non-triangle pixels because of unfocused cube or markup inaccuracies, while a tiny triangle would contain too few pixels and would be affected by noise.

D. SEMANTIC MARKUP

When developing and testing an algorithm for illumination estimation in a scene, it is useful to be able to analyze the structure of errors. The average error over the entire dataset will often not help to reveal whether e.g. the accuracy of the method for indoor images is much less accurate than for outdoor images. To enable performing such and similar checks faster and easier, additional information about the scene and shooting conditions were added to each image in the dataset. In addition to the information available during the shooting, this also includes the following manual annotation:

- Time of day (field **daytime**, with values day/night/unknown).

- The presence of objects with known coloration (**has_known_objects** field with values true/false).
- Scene illumination type (**illumination** field with values artificial/natural/unknown). It is worth noting that there are no flash photos in the dataset.
- Image sharpness (**is_sharp** field with values true/false).
- The presence of light sources in the scene (the **light_objects** multiple choice field with the values lamp/sky/sun/none).
- The place where the image was captured (the **place** field with the values outdoor/indoor/unknown).
- Scene richness (field **richness** with values rich/simple).
- The presence of shadows in the scene (**shadows** field with values yes/no/unknown)
- The cube illumination by all the main light sources in the scene (**estimation** field with values full/partial)

Finally, it is important to note that none of the fields had a preset default value. In that way, the value of every field had to be explicitly set by an annotating person. Namely, if some default field values were to be set in advance, it could increase the annotation bias.

V. THE PROPOSED DATASET

Having in mind all of the concerns and motivation from the previous section, a new dataset named Cube++ is proposed that continues on the previous Cube+ dataset. The dataset download link, the accompanying code, and the technical file description are available at https://github.com/Visillect/CubePlusPlus/.

The Cube++ dataset contains 4890 images. It includes only 1359 of the 1707 images from the Cube+ dataset and only 330 of the 363 images from the 1st Illumination Estimation Challenge (IEC#1) test set [76]. Other images were excluded because they may go against respecting privacy by containing personal data such as faces and license plates or they may be problematic for ground-truth extraction. The remaining 3201 images are brand new.

Cube++ has diverse scene illumination cases as demonstrated by Fig 9. There it can be seen that the chromaticity coverage area is wider than in e.g. Cube+. In other words, the illumination variability has been significantly improved.

The ground-truth illumination distribution for Cube++ and its parts can also be seen from another point of view by taking a look at Fig. 10.

This figure shows that Cube+, which includes Cube, and Cube++ have somewhat similar distributions, which in turn means that a lot of images were taken under outdoor daylight illumination.

One of the important features of the proposed dataset is the fact that it contains two ground-truth illumination records per image, one for each side of the SpyderCube instance. Even though in many of the images there is effectively only one dominant illumination in the scene, Fig. 11 helps to better understand the relation between the two recorded illuminations over the dataset images. Currently, the average angular error of state-of-the-art illumination estimation methods is arguably somewhere between 1° and 2°. With that in mind, all images with larger angular difference between their illumination records can be treated as two-illumination cases.

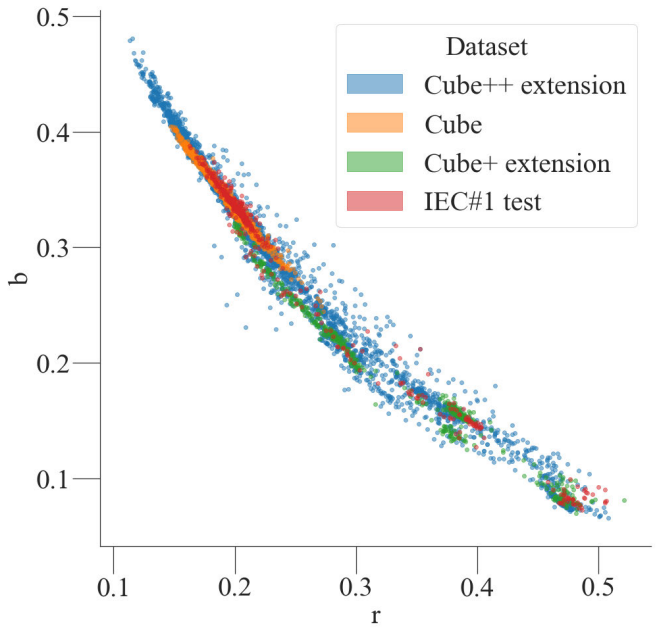


FIGURE 9. Scatter plot of ground truth illumination chromaticities captured by the SpyderCube gray faces.

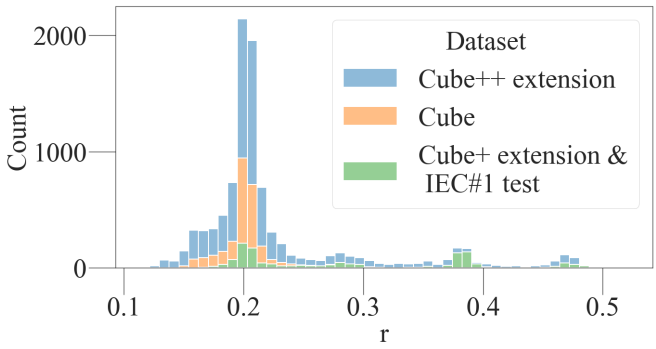


FIGURE 10. Stacked histogram of the red chromaticity values $r = R/(R + G + B)$ of Cube++ ground-truth illuminations.

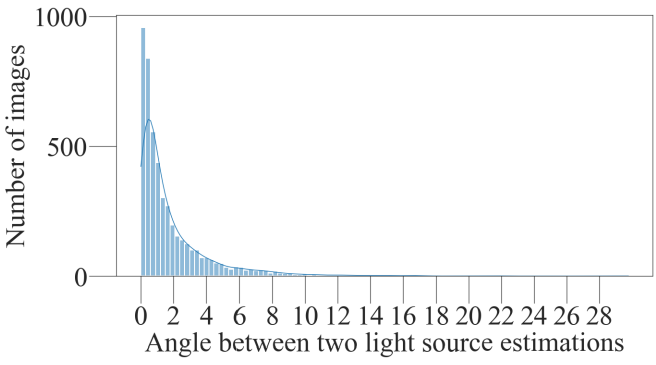


FIGURE 11. Histogram of angular differences between SpyderCube's left and right gray faces for Cube++.

TABLE 2. Feature statistics for various Cube++ subsets.

Feature		Cube++ parts					Cube++ subsets	
		Cube[8]	Cube+[8] extension	IEC#1[8] test	Cube++ extension	**Total Cube++**	Full estimation	**Simple Cube++**
estimation	full	916	306	301	2843	**4366**	4366	**2234**
	partial	115	22	29	358	**524**	0	**0**
daytime	day	1007	24	176	2302	**3509**	3155	**1615**
	night	1	1	1	71	**74**	36	**5**
	unknown	23	303	153	828	**1307**	1175	**614**
place	indoor	12	281	27	454	**774**	653	**379**
	outdoor	1017	46	217	2536	**3816**	3432	**1738**
	unknown	2	1	86	211	**300**	281	**117**
illumination	artificial	2	9	16	306	**333**	233	**109**
	mixed	0	0	1	27	**28**	13	**7**
	natural	995	41	164	2311	**3511**	3171	**1611**
	unknown	34	278	149	557	**1018**	949	**507**
light_objects	none	560	326	287	1505	**2678**	2434	**1246**
	flash	0	0	0	3	**3**	3	**2**
	lamp	1	0	1	125	**127**	58	**29**
	lamp&sky	0	0	0	21	**21**	7	**5**
	sky	470	2	42	1543	**2057**	1861	**952**
	sky&sun	0	0	0	3	**3**	3	**0**
	sun	0	0	0	1	**1**	0	**0**
shadows	yes	264	203	61	1282	**1810**	1573	**696**
	no	767	125	259	1879	**3030**	2751	**1519**
	unknown	0	0	10	40	**50**	42	**19**
richness	simple	3	8	94	337	**442**	400	**198**
	rich	932	152	233	2840	**4157**	3704	**1860**
	unknown	96	168	3	24	**291**	262	**176**
has_known_objects	True	767	49	76	1555	**2447**	2185	**1138**
	False	264	279	254	1646	**2443**	2181	**1096**
is_sharp	True	1002	259	309	3006	**4576**	4088	**2088**
	False	29	69	21	195	**314**	278	**146**
Total		1031	328	330	3201	**4890**	4366	**2234**

Another important feature of the proposed dataset that has to be stressed additionally is that it contains semantic data for each image. All semantic information features are shown in Table 2. Different features provided in the semantic data can be helpful for algorithm tuning and profiling as they give potentially useful information about each image individually.

A. TECHNICAL DESCRIPTION

The dataset consists of several parts. First, there are the raw images with only simple linear debayering performed that are stored in 16-bit PNG format. Next, there are CSV files with ground-truth illumination values and CSV files with additional related properties. Furthermore, there are JPEG images generated by using the `dcraw` open-source tool.[7] Finally, there are also additional files for storing auxiliary information. All these files are automatically built from sources included in the dataset by running a script that is also provided. The sources contain the original CR2 images from the camera and JSON files with the manual annotation data.

The original camera JPEG images are not included as their generation depends on cameras' settings, which means that they cannot be recreated simply or even accurately [77].

1) PNG AND JPEG IMAGES

The main 16-bit PNG images are generated from the original CR2 files in three steps. First, the CR2 files are decoded by using the `dcraw` tool with the options `-D -4 -T`. This generates a 16-bit 1-channel TIFF image. Second, the $[10, 10 + 5184] \times [4, 4 + 3456]$ rectangle was cropped, to have the same area as the default camera JPEG, which comes with certain advantages. Finally, a naive debayering is applied so that every $\begin{smallmatrix} R, & G_1, \\ G_2, & B \end{smallmatrix}$ pattern is converted to a pixel of color $\left(R, \frac{G_1+G_2}{2}, B\right)$. After that the size of the generated PNG images is $2592 \times 1728 = 2^5 3^4 \times 2^6 3^3$. Even though the color channel values have 16 bits of storage, in practice their

[7] https://www.dechifro.org/dcraw/

maximal value is always below $2^{14} - 1$. The black level that can be used for every dataset image is $2048 = 2^{11}$.

For visualization purposes, the modified versions of JPEG images generated by the `dcraw` tool are included as well. The modification includes cropping and downscaling in order for the JPEG images to have the same size as the PNG images. Downscaling is required because JPEG images generated by `dcraw` are not downscaled like the PNG images. On the other hand, JPEG images generated by the camera were not included because they depend on camera settings and the camera's white balancing algorithm, which is proprietary, not fully documented, and it may differ for Canon 550D and 600D cameras that have been used for image capturing. Because of that, they can not be recreated reliably.

2) THE GROUND-TRUTH

The ground-truth illumination records are stored in the `gt.csv` file. Ground-truth illuminations are calculated as described in Section IV-C. The columns are: *image* and for each of the 4 triangles (left, right, left bottom, right bottom) it contains three columns *r, g, b* with the corresponding RGB illumination chromaticities so that $r+g+b = 1$. The triangles brightness values are given in the `properties.csv` file.

Usually, computational color constancy datasets contain only a single ground-truth illumination vector, which represents the dominant illumination in the scene. In the Cube++ dataset such illumination is not given, because the precise single illuminant estimation may require specialist annotation. Moreover, some images have two significantly different illuminations, which makes it harder to select the dominant one. If only a single ground-truth illumination is required and the possible errors that it leads to are acceptable, then one of the following methods can be used to obtain it:

- sample images with relatively similar left and right ground-truth illuminations (the sugested answers for such images are denoted in *properties.csv*);
- select from the left and right sides the brighter one; then select the white triangle for a dark image, and grey triangle for a bright image.

Note that the difference between the sensitivities of the white faces is greater then the difference between the sensitivities of the gray ones (see Section IV-A). Additionally, since the white faces are more often overexposed than the gray ones, using the gray faces should be preferred. On the contrary, using white faces may be better on dark images as mentioned in Section IV-C.

We also estimated if the ground truth values are distorted by the pixels with the clipped values. The images with overexposed grey triangles were removed from the dataset. The images with the clipped values on white triangles are present in the dataset, but the overexposed triangles are marked in *properties.csv*.

3) RELEVANT META-INFORMATION

The `properties.csv` file contains the most relevant meta-information about the Cube++ images. It includes the average triangle brightness $\frac{R+B+G}{3}$, manual annotation data, information about overexposed triangles, and a carefully selected subset of EXIF data fields.

The EXIF data was extracted from CR2 files using the `PyExifTool` library.[8] All the extracted values can be found in the corresponding JSON files. The properties table contains only a few selected ones. The EXIF data format slightly differs between the Canon EOS 550D and 600D cameras: there are 312 common fields, 2 in 550D only, and 21 in 600D only. All the selected EXIF fields are common.

The `cam_estimation.csv` file contains the EXIF fields of the camera that contains the camera's light source estimation

B. IMAGE PREPARATION

Finally, it is important to clearly specify how to properly prepare the provided Cube++ images before handing them over to illumination estimation methods that are to be tested.

There are three main steps that have to be taken.

1) BLACK LEVEL SUBSTRACTION

The first step is to subtract the approximate black level of 2048 from all image pixel color components. In some cases this can result in negative values, but such values should then be set to 0.

2) SATURATION DETECTION

The second step is to calculate the maximum value m for all pixels across all color channels. After that all pixels that have a value greater than or equal to $m - 50$ in any of their channels should have all their channel values set to 0 This would remove most of the incorrect pixels with clipped values. Nevertheless, it would leave some rare overexposed pixels, because demosaicing procedure may mix them with the normal ones. To get precise information about saturated pixels it is recommended to analyse images before demosaicing (the last one can be extracted from CR2 files).

3) COLOR TARGET MASKING

The last step is to mask out the lower right rectangle of the image that contains the color target to remove any potential bias and thus to have a relatively fair testing. The size of this rectangle is 700×1000 for all images. The rectangle is masked out by setting all channel values of all its pixels to 0, i.e. by making it black.

C. INTENDED DATASET USAGE

With all its features, especially the two ground-truth illumination records, Cube++ is appropriate for several illumination estimation use cases. All datasets mentioned in Section II, except for maybe TCC dataset, are designed for the most widely used classical illumination estimation problem: estimation of the single light source in the scene. Therefore, each image is provided with only single light source ground-truth,

[8] https://pypi.org/project/PyExifTool/

even in cases when the scene is obviously under the influence of multiple illuminations. In contrast to these datasets, Cube++ allows to work on following problem statements:

1) Estimate one and only dominant lighting in a scene;
2) Estimate two dominant light sources in a scene;
3) Estimate at least one dominant light source in the scene.

For each of the listed problem statements we propose the following rules to filter the Cube++ images that are suitable for it. To form the dataset subset for the first problem, one needs to select all images where the angular differences between its two extracted ground-truth illuminations is below 1° (except partially light source estimation part, see section IV-B). For the two light source estimation problem, one needs to do the opposite, i.e. to select all images that are not selected for the first problem (except partially light source estimation part). Finally, to work on the third problem, all images can be chosen.

(a) (b)

FIGURE 12. Example of chromatic adaptation based on illumination extracted from a) left and b) right gray face of the SpyderCube calibration object placed in the scene.

Here it is important to mention that the proposed rules are arbitrary, that they will result in some of the images being inappropriately selected, and that they may be improved. One example where these rules fail is shown in Fig. 12. There the extracted ground-truth illuminations differ significantly, but practically the whole scene is mostly under the illumination captured by the right cube face. This means that even though the scene is effectively under uniform illumination, the mentioned rules will result in the opposite conclusion based on the difference between the extracted illuminations on the two cube faces.

Also it is important to mention what at the moment partially light source estimation part of the dataset is not provided with subjective single illumination estimation choice. The plan is to solve this in future work.

D. SimpleCube++ DATASET

In addition to the main 200GB Cube++ dataset, a 2GB-small and simpler version of it is prepared. The small dataset contains 4x downscaled images that have less than 1° difference between ground-truth illuminations of SpyderCube's left and right grey faces. It includes only images with a single illumination source, and consequently, the ground-truth file contains only one ground-truth per image. This ground-truth was extracted in the following manner: firstly, average values for both gray faces were calculated as in the main Cube++ dataset; secondly, they were normalized by using l_1-norm ($r+g+b=1$ for both gray faces); finally, obtained ground-truth values were averaged and normalized again by using l_1-norm. This dataset has two main advantages: small weight (around 2GB) and a single answer per image.

SimpleCube++ contains PNG and JPG files, `gt.csv` with ground-truth data, and `properties.csv` with manual annotation data. In addition, this dataset was divided into train and test parts. Each image was independently assigned to the test set with a probability of 20%.

VI. DISCUSSION

Having another high-quality illumination estimation dataset such as the one proposed in this paper is certainly beneficial to the interested research community as well as the industrial sector and there should probably be no discussion about that. However, proposing a new dataset is still only an incremental move in terms of the overall paradigm of illumination estimation research since this has been done on numerous occasions while the dataset usage has remained relatively unchanged.

A much more constructive and necessary discussion that is rarely taken forward should be about the direction of how to better use or not use the datasets to achieve better progress in illumination estimation research. In terms of that, one of the burning issues is that the results in most illumination estimation papers are unverifiable and thus questionable. Therefore, for the sake of improving the state of the illumination estimation research, it would be quite useful to further discuss this problem as well as the potential solutions to it in more detail.

A. QUESTIONABLE RESEARCH PROGRESS

Obtaining low illumination estimation errors on a benchmark dataset is a regularly used approach when trying to demonstrate the superiority of a proposed illumination estimation method. For all well-known datasets the ground-truth illumination used during the test phase is publicly available and the actual error statistics calculation is usually performed by the authors themselves and published in their papers. However, this introduces several problems with the most serious being data dredging, i.e. *p*-hacking and erroneous reporting.

The problem with data dredging in illumination estimation is that in cases when a model selection is required, the final results that are reported were not always obtained through nested cross-validation [78]. Instead, the reported results are the ones that were used to select the model in the first place. By using these results, a method's true performance on new unknown data may be masked and unfairly shown to be better than it actually is. This can prevent or slow down the progress in illumination estimation research by giving misleading clues about the validity of the method's underlying assumptions.

In the area of visual odometry similar problems with e.g. the KITTI dataset [48] have been prevented by simply

[8]Not including Cube, Cube+, IEC#1 test part images, removed from the Cube++ dataset because of GDPR restrictions or possible problems with ground-truth extraction.

keeping the ground-truth for the test secret. By having the evaluation of the results on the test set carried out by the dataset administrators, any serious attempts of *p*-hacking have been prevented.

Another problem that can be prevented if the evaluation is carried out by a third party is erroneous reporting. For example, in [18] the results of the proposed illumination estimation method on several datasets were allegedly all obtained by using the same value of a hyperparameter. However, trying to re-implement the method fails to produce the same results and only after checking the associated webpage [79] it becomes clear that the hyperparameter value has to be changed for each dataset to fully reproduce the published results.

A somewhat similar example is the 2007 paper by van de Weijer *et al.* [23]. In an erratum published in 2008 [80] it was explained how testing was inadequately performed, which consequently resulted in reporting of erroneous error statistics.

Finally, any doubts in the validity of some reported error statistics could be reduced or fully eliminated if they were calculated not by the authors themselves, but by a reliable third party. This would also help the overall research progress.

B. ILLUMINATION ESTIMATION CHALLENGES

Inspired by the ideas mentioned in the previous subsection, two international illumination estimation challenges have already taken place [76], [81]. The challenges provided the participants with thousands of training images and their respective ground-truth illuminations, while for the test set only the images were provided and the ground-truth remained secret until the end of the challenge. Because of that the error statistics for the illumination estimations sent over by the authors were calculated by the challenge organizers, which prevented a lot of problems described in the previous section. The results were thus more trustworthy and they have shown e.g. high errors for some methods that were previously reported to be highly accurate. Additionally, the challenges helped to recognize additional problems such as training a method to obtain excellent values for a given error metric [82], which results in issues related to the so called Goodhart's law [83].

C. BENCHMARK

While the described international illumination estimation challenges have shown the advantages of having a reliable third-party calculate the error statistics, they were fixed in time and they cannot be repeated on the same images anymore. Therefore, the next step would be to create a benchmark dataset similar to the KITTI dataset with an online user interface for submitting the results at any given time. This would surely represent a significant contribution to the illumination estimation research since it would simultaneously provide the researchers with trustworthy results and also eliminate many of the serious problems that were described earlier in this paper.

For the above reasons, creating such a benchmark is already underway at the time of writing this paper. At present time we are working on the question of benchmark creation. Possible benchmark will be based on the images that were taken during the same time as the rest of the Cube++ images, but that were excluded from its final version. Because of that, in this paper there are purposely no error statistics obtained on the Cube++ dataset by any of the illumination estimation methods. The error statistics will be published online and they will be based on the first version of the benchmark test set. This aims to avoid providing any results obtained on the Cube++ images with known ground-truth illumination. Namely, the idea is to separate the testing and the associated problems from the dataset and to relegate it to the benchmark. Therefore, the overall goal of this paper is to provide high quality training data without any testing. The role of testing data is to be assumed by the future benchmark.

VII. CONCLUSION

A new illumination estimation dataset named Cube++ has been proposed. Unlike similar existing illumination estimation datasets, it provides rich, reliable, and verifiable data on scene illumination with specific care being given to precise calibration. For every one of its 4890 images, there are two ground-truth illumination records as well as a multitude of semantic information and it is GDPR-compliant. Furthermore, a wide variety of scene content is covered, and numerous illuminations are captured. Cube++ contains images taken with several instances of the same model of the camera sensor. In addition to that, a centralized versioning control system for Cube++ has been established to simplify and document possible future changes in the dataset and error handling. By having these properties and novelties, Cube++ is technically superior to most similar illumination estimation datasets. One of the future steps that should also be significant progress in the overall illumination estimation research is to create an online illumination estimation benchmark based on the infrastructure that was used to create the Cube++ dataset.

ACKNOWLEDGMENT

The authors would like to thank Prof. Dmitry Nikolaev, for his help in reading the article and his contribution to the discussions.

The authors would also like to thank Maria Yarykina and Ekaterina Panfilova for the images they added to the dataset and Viacheslav Vasilyev, Vasily Tesalin, Sergey Emelyanov, Oleg Emelyanov, Tatyana Postnikova, Sergey Pavlov, Evgenija Sidorchuk, and Olga Vlasova for their contribution to the annotation.

REFERENCES

[1] M. Ebner, *Color Constancy* (IS&T Series in Imaging Science and Technology). Hoboken, NJ, USA: Wiley, 2007.

[2] G. D. Finlayson, M. S. Drew, and B. V. Funt, "Color constancy: Enhancing von kries adaption via sensor transformations," in *Proc. 4th Hum. Vis., Vis. Process., Digit. Display, Int. Soc. Opt. Photon.*, vol. 1913, Sep. 1993, pp. 473–484.

[3] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Beyond white: Ground truth colors for color constancy correction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 298–306.

[4] D. P. Nikolaev and P. P. Nikolayev, "On spectral models and colour constancy clues," in *Proc. 21st Eur. Conf. Modeling Simulation (ECMS)*, Prague, Czech Republic: Citeseer, 2007, pp. 318–323.

[5] P. P. Nikolaev, "Gybridnaya spectralnaya model v zadache constantnosty tsveta. 1. Zonalniye okraski pri gaussovskom osveshenii," *Sensory Syst.*, vol. 22, no. 4, pp. 287–308, 2008.

[6] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 972–984, Sep. 2002.

[7] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.

[8] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 1, 2010, pp. 256–259.

[9] N. Banic and S. Loncaric, "Using the random sprays Retinex algorithm for global illumination estimation," in *Proc. 2nd Croatian Comput. Vis. Workshopn (CCVW)*. Zagreb, Croatia: Univ. of Zagreb, Faculty of Electrical Engineering and Computing, Sep. 2014, pp. 3–7.

[10] N. Banic and S. Loncaric, "Color rabbit: Guiding the distance of local maximums in illumination estimation," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 345–350.

[11] N. Banic and S. Loncaric, "Improving the white patch method by subsampling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 605–609.

[12] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.

[13] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 1, 2004, pp. 37–41.

[14] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.

[15] A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 918–929, May 2012.

[16] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey, "The role of bright pixels in illumination estimation," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 1, 2012, pp. 41–46.

[17] Y. Qian, S. Pertuz, J. Nikkanen, J.-K. Kämäräinen, and J. Matas, "Revisiting gray pixel for statistical illumination estimation," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2019, pp. 36–46.

[18] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: Why spatial-domain methods work and the role of the color distribution," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 31, no. 5, pp. 1049–1058, 2014.

[19] N. Banic and S. Loncaric, "Blue shift assumption: Improving illumination estimation accuracy for single image from unknown source," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2019, pp. 191–197.

[20] N. Banic and S. Loncaric, "Green stability assumption: Unsupervised learning for statistics-based illumination estimation," *J. Imag.*, vol. 4, no. 11, p. 127, Oct. 2018.

[21] D. A. Tarasov and O. B. Milder, "Mathematics and practice of color space invariants by the example of determining the gray balance for a digital printing system," *Comput. Opt.*, vol. 44, no. 1, pp. 117–126, Feb. 2020.

[22] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *JOSA A*, vol. 19, no. 12, pp. 2374–2386, 2002.

[23] J. van de Weijer, C. Schmid, and J. Verbeek, "Using High-Level Visual Information for Color Constancy," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.

[24] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[25] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[26] D. Hernandez-Juarez, S. Parisot, B. Busam, A. Leonardis, G. Slabaugh, and S. McDonagh, "A multi-hypothesis approach to color constancy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2270–2280.

[27] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1509–1519, Aug. 2012.

[28] N. Banic and S. Loncaric, "Color cat: Remembering colors for illumination estimation," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 651–655, Jun. 2015.

[29] N. Banic and S. Loncaric, "Using the red chromaticity for illumination estimation," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2015, pp. 131–136.

[30] N. Banic and S. Loncaric, "Color Dog–Guiding the global illumination estimation to better accuracy," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, 2015, pp. 129–135.

[31] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1904–1911.

[32] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Effective learning-based illuminant estimation using simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1000–1008.

[33] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.

[34] J. T. Barron and Y.-T. Tsai, "Fast Fourier color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2017, pp. 886–894.

[35] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 81–89.

[36] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 371–387.

[37] Y. Hu, B. Wang, and S. Lin, "FC^4: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4085–4094.

[38] S. W. Oh and S. J. Kim, "Approaching the computational color constancy as a classification problem through deep learning," *Pattern Recognit.*, vol. 61, pp. 405–416, Jan. 2017.

[39] K. Koščevic, N. Banic, and S. Loncaric, "Color beaver: Bounding illumination estimations for higher accuracy," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 183–190.

[40] A. Akbarinia and C. A. Parraga, "Colour constancy beyond the classical receptive field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2081–2094, Sep. 2018.

[41] S.-M. Woo, S.-H. Lee, J.-S. Yoo, and J.-O. Kim, "Improving color constancy in an ambient light environment using the phong reflection model," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1862–1877, Apr. 2018.

[42] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2254–2263.

[43] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1973–1985, Oct. 2015.

[44] N. Banic and S. Loncaric, "Unsupervised learning for color constancy," in *Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2018, pp. 181–188.

[45] M. Afifi and M. S. Brown, "Sensor-independent illumination estimation for DNN models," 2019, *arXiv:1912.06888*. [Online]. Available: http://arxiv.org/abs/1912.06888

[46] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.

[47] J. T. Barron, *Personcal Correspondence*. 2019.

[48] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

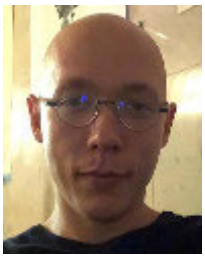
[49] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. Soc. Imag. Sci. Technol., Color Imag. Conf.*, no. 1, 2003, pp. 160–164.

[50] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Improving color constancy using indoor–outdoor image classification," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2381–2392, Dec. 2008.

[51] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Automatic color constancy algorithm selection and combination," *Pattern Recognit.*, vol. 43, no. 3, pp. 695–705, Mar. 2010.

[52] B. F. L. Shi. (May 2020) *Re-Processed Version of the Gehler Color Constancy Dataset of 568 Images*. [Online]. Available: https://www2.cs.sfu.ca/~colour/data/shi_gehler/

[53] S. E. Lynch, M. S. Drew, and G. D. Finlayson, "Colour constancy from both sides of the shadow edge," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 899–906.
[54] G. D. Finlayson, G. Hemrit, A. Gijsenij, and P. Gehler, "A Curious Problem with Using the Colour Checker Dataset for Illuminant Estimation," in *Proc. Color Imag. Conf., Soc. Imag. Sci. Technol.*, no. 25, 2017, pp. 64–69.
[55] G. Hemrit, G. D. Finlayson, A. Gijsenij, P. Gehler, S. Bianco, B. Funt, M. Drew, and L. Shi, "Rehabilitating the colorchecker dataset for illuminant estimation," in *Proc. Soc. Imag. Sci. Technol. Color Imag. Conf.*, no. 1, 2018, pp. 350–353.
[56] N. Banic, K. Koscevic, M. Subasic, and S. Loncaric, "The past and the present of the color checker dataset misuse," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 366–371.
[57] F. Laakom, J. Raitoharju, A. Iosifidis, J. Nikkanen, and M. Gabbouj, "INTEL-TAU: A color constancy dataset," 2019, *arXiv:1910.10404*. [Online]. Available: http://arxiv.org/abs/1910.10404
[58] C. Aytekin, J. Nikkanen, and M. Gabbouj, "A data set for camera-independent color constancy," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 530–544, Feb. 2018.
[59] R. Zakizadeh, M. S. Brown, and G. D. Finlayson, "A hybrid strategy for illuminant estimation targeting hard images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 16–23.
[60] Y. Qian, J. Käpylä, J.-K. Kämäräinen, S. Koskinen, and J. Matas, "A benchmark for temporal color constancy," 2020, *arXiv:2003.03763*. [Online]. Available: http://arxiv.org/abs/2003.03763
[61] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 103–112, Jan. 2005.
[62] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troger, and A. Kaup, "Color constancy and non-uniform illumination: Can existing algorithms work?" in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 774–781.
[63] A. Rizzi, C. Bonanomi, D. Gadia, and G. Riopi, "YACCD2: Yet another color constancy database updated," in *Proc. 18th Color Imag. Displaying, Process., Hardcopy, Appl.*, vol. 8652, Feb. 2013, p. 86520.
[64] A. Smagina, E. Ershov, and A. Grigoryev, "Multiple light source dataset for colour research," in *Proc. 12th Int. Conf. Mach. Vis. (ICMV)*, Jan. 2020, Art. no. 114332.
[65] L. Murmann, M. Gharbi, M. Aittala, and F. Durand, "A dataset of multi-illumination images in the wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4080–4089.
[66] D. Shepelev, A. Tereshin, P. P. Nikolaev, and E. Ershov, "Stereo correspondence problems in terms of linear theory of spectral stimulus formation," *Sensory Syst.*, vol. 32, no. 2, pp. 150–160, 2017.
[67] N. Banic and S. Loncaric, "A perceptual measure of illumination estimation error," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, 2015, pp. 136–143.
[68] S. Bianco and R. Schettini, "Face-based illuminant estimation," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2012, pp. 623–626.
[69] S. B. Knorr and D. Kurz, "Real-time illumination estimation from faces for coherent rendering," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2014, pp. 113–122.
[70] M. Males, A. Hedi, and M. Grgic, "Colour balancing using sclera colour," *IET Image Process.*, vol. 12, no. 3, pp. 416–421, Mar. 2018.
[71] S. Bianco and C. Cusano, "Quasi-unsupervised color constancy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12212–12221.
[72] S. Koskinen, D. Yang, and J.-K. Kämäräinen, "Cross-dataset color constancy revisited using sensor-to-sensor transfer," in *Proc. Brit. Mach. Vis. Conf.*, 2020.
[73] G. D. Finlayson, R. Zakizadeh, and A. Gijsenij, "The reproduction angular error for evaluating the performance of illuminant estimation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1482–1488, Jul. 2017.
[74] G. D. Finlayson, S. D. Hordley, and P. Morovic, "Colour constancy using the chromagenic constraint," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 1079–1086.
[75] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767
[76] (Sep. 2020) *Illumination Estimation Challenge*. [Online]. Available: https://www.isispa.org/illumination-estimation-challenge
[77] S. Joo Kim, H. Ting Lin, Z. Lu, S. Süsstrunk, S. Lin, and M. S. Brown, "A new in-camera imaging model for color computer vision and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2289–2302, Dec. 2012.
[78] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for Machine Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2020.
[79] (Aug. 2020) *On Illuminant Detection*. [Online]. Available: http://cvil.eecs.yorku.ca/projects/public_html/illuminant/illuminant.html
[80] B. F. L. Shi. (Aug. 2020) *Using High-Level Visual Information for Color Constancy*. [Online]. Available: http://lear.inrialpes.fr/people/vandeweijer/papers/vandeweijer_iccv07.pdf
[81] (Sep. 2020). *Illumination Estimation Challenge*. [Online]. Available: http://chromaticity.iitp.ru/
[82] A. Savchik, E. Ershov, and S. Karpenko, "Color cerberus," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 355–359.
[83] M. Strathern, "'Improving ratings': Audit in the British University system," *Eur. Rev.*, vol. 5, no. 3, pp. 305–321, 1997.

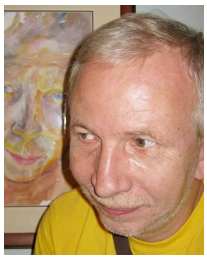


EGOR ERSHOV received the Ph.D. degree from the Faculty of Radio Engineering and Computer Technology, in 2019. He is currently a Senior Researcher with the Vision System Laboratory, Institute of Information Transmission Problems (IITP) (Kharkevic Institute), Russian Academy of Science (RAS). He is also Senior Lecturer with the Computer Science Faculty, Higher School of Economy. He is an Assistant Manager of the Cathedra at Computer Science Faculty, Higher School of Economy. His main areas of research interests include image processing and analysis, particularly color computer vision, color reproduction technologies, colorimetry, human vision systems, Hough transform. He was a recipient of several awards for his scientific and professional work.

ALEXEY (ALEX) SAVCHIK graduated from the Department of Mechanics and Mathematics, Moscow State University, in 2014, and from the Yandex School of Data Analysis, in 2012. He is currently with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. His research interests include deep learning and computer vision.

ILLYA (ILYA) SEMENKOV received the bachelor's degree from the National Research University Higher School of Economics (NRU HSE), in 2019, where he is currently pursuing the master's degree with the Faculty of Computer Science. He is also with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. His research interests include deep learning, computer vision, statistical learning, and optimal transport.

NIKOLA BANIĆ (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer science in 2011, 2013, and 2016, respectively. He is currently working as a Senior Computer Vision Engineer at Gideon Brothers, Croatia. He has worked in real-time image enhancement for embedded systems, digital signature recognition, people tracking and counting, and image processing for stereo vision. His research interests include image enhancement, color constancy, image processing for stereo vision, and tone mapping.



ALEXANDER BELOKOPYTOV graduated from the Moscow Institute of Physics and Technology (MIPT), with specialization in microwave engineering, in 1982.

He is currently with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. While at IITP, he participated in research in peripheral and color human vision, underwater photography. His research interests include human vision (color and peripheral), optoelectronics, and scientific data visualization.

DARIA SENSHINA received the bachelor's degree from the Moscow Institute of Physics and Technology (MIPT), in 2020, where she is currently pursuing the master's degree with Phystech School of Applied Mathematics and Informatics. Simultaneously, she is pursuing the M.Sc. degree in industrial and applied maths with the Université Grenoble Alpes, Grenoble, France. She is also with the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. Her research interests include computer vision and image processing.

KARLO KOŠČEVIĆ (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree in the technical sciences in the scientific field of computing Ph.D. Program at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. His research interests include image processing, image analysis, and deep learning. His current research interest includes the area of color constancy with a focus on learning-based methods for illumination estimation.

MARKO SUBAŠIĆ (Member, IEEE) received the Ph.D. degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 2007. Since 1999, he has been working with the Department for Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently as an Associate Professor. He teaches several courses at graduate and undergraduate levels. His research interests include image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is a member of the IEEE Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.

SVEN LONČARIĆ received the B.Sc., M.Sc., and Ph.D. degrees in 1985, 1989, and 1994, respectively. After earning his Ph.D. degree, he continued his academic career at the Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently a Full Professor. He was an Assistant Professor with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ, USA, from 2001 to 2003. His main areas of research include image processing and analysis. He is the Director of the Center for Artificial Intelligence, and the Co-Director of the National Center of Research Excellence for Data Science and Advanced Cooperative Systems. Together with his students and collaborators, he has published more than 200 publications in scientific peer-reviewed journals and has presented his work at international conferences. He was a recipient of several awards for his scientific and professional work.

Appendix

Appendix 1

Banić, N., **Koščević, K.**, Subašić, M., Lončarić, S., "CroP: Color Constancy Benchmark Dataset Generator", Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing (ICVISP 2020), Association for Computing Machinery, New York, NY, USA, Article 4, 202, pp. 1–9

CroP: Color Constancy Benchmark Dataset Generator

Nikola Banić
Gideon Brothers Radnička 177
10000 Zagreb, Croatia
nbanic@gmail.com

Karlo Koščević
Faculty of Electrical Engineering and Computing
University of Zagreb Unska 3
10000 Zagreb, Croatia
karlo.koscevic@fer.hr

Marko Subašić
Faculty of Electrical Engineering and Computing
University of Zagreb Unska 3
10000 Zagreb, Croatia
marko.subasic@fer.hr

Sven Lončarić
Faculty of Electrical Engineering and Computing
University of Zagreb Unska 3
10000 Zagreb, Croatia
sven.loncaric@fer.hr

ABSTRACT
Implementing color constancy as a pre-processing step in contemporary digital cameras is of significant importance as it removes the influence of scene illumination on object colors. Several benchmark color constancy datasets have been created for the purpose of developing and testing new color constancy methods. However, they all have numerous drawbacks including a small number of images, erroneously extracted ground-truth illuminations, long histories of misuses, violations of their stated assumptions, etc. To overcome such and similar problems, in this paper a color constancy benchmark dataset generator is proposed. For a given camera sensor it enables generation of any number of realistic raw images taken in a subset of the real world, namely images of printed photographs. Datasets with such images share many positive features with other existing real-world datasets, while some of the negative features are completely eliminated. The generated images can be successfully used to train methods that afterward achieve high accuracy on real-world datasets. This opens the way for creating large enough datasets for advanced deep learning techniques. Experimental results are presented and discussed. The source code is available at http://www.fer.unizg.hr/ipg/resources/color constancy/.

CCS Concepts
Computing methodologies → Computational photography; Image representations; Image Processing

Keywords
Color Constancy, Data Augmentation, Illumination Estimation, Image Dataset, White Balancing

ACM Reference format:
Nikola Banić, Karlo Koščević, Marko Subašić and Sven Lončarić. 2020. CroP: Color Constancy Benchmark Dataset Generator. In Proceedings of *2020 4th International Conference on Vision, Image and Signal Processing(ICVISP 2020), December 9-11, 2020, Bangkok, Thailand*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3448823.3448829

ICVISP 2020, December 9-11, 2020, Bangkok, Thailand

ACM ISBN 978-1-4503-8953-2/20/12…$15.00
https://doi.org/10.1145/3448823.3448829

1. INTRODUCTION
Color constancy is the ability of the human vision system (HVS) to perceive the colors of the objects in the scene largely invariant to the color of the light source [25]. Most of the contemporary digital cameras have this ability implemented into their image pre-processing pipeline. The task of computational color constancy is to estimate the scene illumination and then perform the chromatic adaptation in order to remove the influence of the illumination color on the colors of the objects in the scene. Three physical variables can describe the perceived color of objects in the image: 1) spectral properties of the light source, 2) spectral reflectance properties of the object surface, and 3) spectral sensitivity of the camera sensor. Under the Lambertian assumption, the resulting image **f** formation model is

$$f_c(\mathrm{x}) = \int_\omega I(\mathrm{x},\lambda) R(\mathrm{x},\lambda) \rho_c(\lambda) d\lambda \qquad (1)$$

where $f_c(\mathrm{x})$ is the value at the pixel location x for the c-th color channel, $I(\mathrm{x},\lambda)$ is the spectral distribution of light source, $R(\mathrm{x},\lambda)$ is the surface reflectance, and $\rho_c(\lambda)$ is the camera sensor sensitivity for the c-th color channel. The value at pixel location x is obtained by integrating across the all wavelengths λ in the visible light spectrum ω. When estimating the illumination it is often assumed that it is uniform across the whole scene. With this, x can be disregarded and the observed light source e is calculated as

$$\mathrm{e} = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_\omega I(\lambda) \rho(\lambda) d\lambda \qquad (2)$$

Since only pixel values f are known and both $I(\lambda)$ and $\rho(\lambda)$ remain unknown, it is an ill-posed problem to calculate the illumination vector e. Illumination estimation methods try solve this problem by introducing new assumptions. On one side, there are methods that rely on low-level image statistics such as White-patch [44, 32] and its improvements [10,11,12], Gray-world [20], Shades-of-Gray [29], Gray-Edge (1st and 2nd order) [49], using bright and dark colors [22], exploiting the illumination color statistics perception [14], exploiting the

expected illumination statistics [9], using gray pixels [46]. Appropriately, these methods can be found in the literature as statistics-based methods. They are fast, hardware-friendly, and easy to implement. On the other hand, there are learning-based methods, which use data to learn their parameter values and compute more precise estimations, but they also require significantly more computational power and parameter tuning. Learning-based method include gamut mapping (pixel, edge, and intersection based) [28], using high-level visual information [50], natural image statistics [34], Bayesian learning [33], spatiospectral learning (maximum likelihood estimate, and with gen. prior) [21], simplifying the illumination solution space [4, 5, 13], using color/edge moments [26], using regression trees with simple features from color distribution statistics [23], performing various spatial localizations [17, 18], genetic algorithms and illumination restriction [40], convolutional neural networks [19, 48, 38].

To compare the accuracy of these methods, several publicly available color constancy datasets have been created. While they significantly contributed to the advance of the illumination estimation, they have several drawbacks. The main one is that they contain relatively few images due to the significant amount of time required for determining the ground-truth illumination. This was shown to have an impact on the applicability of the deep learning techniques. Other common drawbacks include cases of incorrect groundtruth illumination data, significant noise amounts, violations of some important assumptions, etc. In the worst cases the whole datasets are being used completely wrong in the pure technical sense [2], which may have led to many erroneous conclusions in the field of illumination estimation [27]. In order to try to simultaneously deal with most of these problems, in this paper a color constancy dataset generator is proposed. It is confined only to simulation of taking images of printed photographs under projector illumination of specified colors, but in terms of illumination estimation the properties of the resulting images are shown to resemble many properties of real-world images. The experimental results additionally demonstrate the usability of the generated dataset in real-world applications.

This paper is structured as follows: Section 2 gives an overview of the main existing color constancy benchmark datasets, in Section 3 the proposed dataset generator is described, in Section 4 its properties and capabilities are experimentally validated, and Section 5 concludes the paper.

2. RELATED WORK

2.1 Image Calibration

The main idea of color constancy benchmark datasets is for them to have images for which the color of the illumination that influences their scenes is known. That means that along images every such dataset also has the ground-truth illumination for each of these images. For a given image the ground-truth is usually determined by putting a calibration object in the scene and later reading the value of its achromatic surfaces. Calibration objects include gray ball, color checker chart, SpyderCube, etc. Due to the illposedness of the illumination estimation problem, determining the ground-truth illumination for a given image without calibration objects can often not be carried out accurately enough. While in such images some of the scene surfaces with known color under the white light could be used, this could lead to inaccuracies due to the metamerism.

2.2 Existing Datasets

The first large color constancy benchmark dataset with real-world images and ground-truth illumination provided for each image was the GreyBall dataset [24]. It consists of 11346 images and in the scene of each image a gray ball is placed and used to determine the ground-truth illumination for this image. However, the images in this dataset are non-linear i.e. they have been processed by applying non-linear operations to them and therefore they do not comply with the image formation model assumed in Eq. (1). Additionally, the images are small with only the of size 240×360.

In 2008 the Color Checker dataset has been proposed [33]. It consists of 568 images with each of them having a color checker chart in the scene. Several version of the dataset and its ground-truth illumination found their way into the literature over time with most of them being plagued by several serious problems [27, 36, 2].

Cheng et al. created the NUS dataset in 2014 [22]. It is a color constancy dataset composed of natural images captured with 8 different cameras with both indoor and outdoor scenes under various common illuminations. With the same scene taken using multiple cameras, the novelty of this dataset is that the performance of illumination estimation algorithms can be compared across different camera sensors.

In [7] a dataset with 1365 images was published, namely the Cube dataset. It consists of exclusively outdoor images with the SpyderCube calibration object placed in the lower right corner of each image to obtain the ground-truth illumination. All images were taken with the Canon EOS 550D camera. The main disadvantage of the Cube dataset i.e. restriction to only outdoor illuminations was alleviated in the Cube+ dataset [7]. It is a combination of the original Cube dataset and additional 342 images of both indoor scenes and outdoor scenes taken during the night. Consequently, besides the larger number of images, a more diverse distribution of illuminations was achieved which is the desirable property of the color constancy benchmark datasets.

A dataset for camera-independent color constancy was published in [1]. The images in that dataset were captured with three different cameras with one of them being a mobile phone camera and the other two high-resolution DSLR cameras. The dataset is composed of images in both laboratory and fields scenes taken with all three camera sensors.

Recently a new benchmark dataset with 13k images was introduced [45]. It contains both indoor and outdoor scenes with the addition of some challenging images. Unfortunately, at the time, this dataset is not publicly available. Another relatively large dataset with challenging images which is not publicly available was used in [47]. Although the authors report the performance of their illumination estimation methods on these datasets, comparison with other methods is hard since they are not publicly available.

During the years of research in the field of color constancy numerous other benchmark datasets such as [15, 16] were created, but they are not commonly used for the performance evaluation of illumination estimation methods.

2.3 Problems

The main problem with the previous datasets is the limited number of their images, which is due to the tedious process of the ground-truth illumination extraction. This effectively limits the full-scale application of deep learning methods like for some other problems and various data augmentation techniques have to be used with variable success.

Another problem that can occur during image acquisition is to choose scenes for which the uniform illumination estimation does not hold. This is especially problematic if the less dominant illumination is affecting the calibration object because the extracted ground-truth is then erroneous and results in allegedly hard to estimate image cases [51].

Even if all of the ground-truth illumination data was correctly collected, it often consists of only the most commonly observed illuminations. This lack of variety makes some of the datasets susceptible to abuse cases of methods that aim to fool some of the error metrics [3]. It also prevents the illumination estimation methods from being tested on images formed by the presence of extreme illuminations.

In some of the worst cases, some datasets were used technically inappropriately [2], which made the obtained experimental results to be technically incorrect and put in question some of the allegedly achieved progress [27].

3. THE PROPOSED DATASET GENERATOR

A solution to many problems mentioned in the previous section would be the possibility to generate real-world images whose scenes are influenced by an arbitrary chosen known illumination and exactly such a solution is proposed in this section. When taking into account everything that has been mentioned here, several conditions have to be met:

- there has to be a big number of available illuminations,
- the colors of any material present in the scene that are known for the canonical white illumination have also to be known for every other possible illumination,
- and the influence of a chosen camera sensor on the color of illuminated material has also to be known.

All this can be accomplished by recording enough real-world data and then use it to simulate real-world images. Knowing the behavior of colors of various materials under different illuminations would require too much data both to collect and to control during the image generation process. Because of this and motivated by existence of images like the one in figure 1, the proposed dataset generator is restricted only to the colors printed by the same single printer on the same single sheet of paper. To assure uniform illumination and some control over its color, all scenes are illuminated by a projector that projects single color frames. In short, the proposed dataset generator is able to simulate taking of raw camera images of printed images illuminated by a projector. More details are given in the following subsections.

3.1 Used Illuminations

To assure a big variability of available illuminations, 707 of them were used. They are composed of colors whose chromaticities are uniformly spread and of colors of a black body at various temperatures. The latter colors are important because they occur very often in real-world scenes. The relation between all these colors is shown in figure 2a. Due to the projector and camera characteristics, the final appearance of these colors is changed. For example, if the achromatic surfaces of the SpyderCube calibration object are photographed under all these illuminations, their appearances in the RGB colorspaces of two different cameras described in Section 3.3 are as shown in figure. 2b.

Figure 1. Example of an image from the Cube+ Dataset [7] whose scene consists only of another printed image.

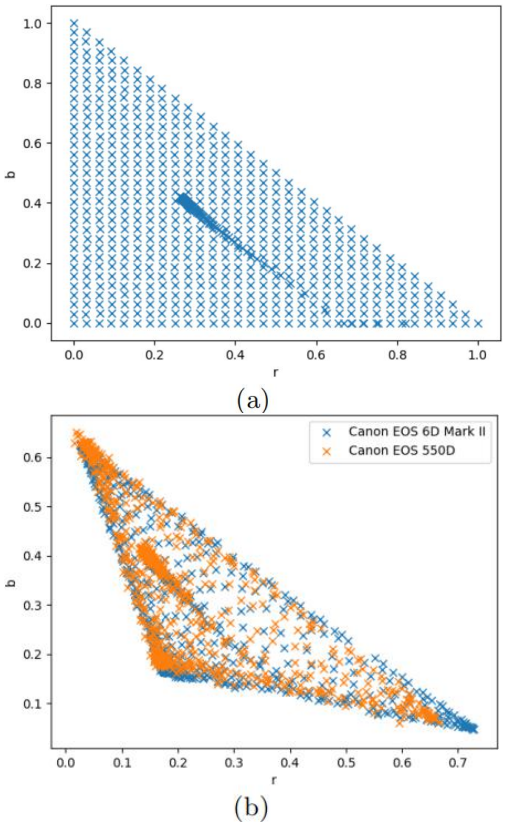


Figure 2. (a) rb-chromaticities of the illuminations used to illuminate the printed color pattern; (b) rb-chromaticities of the achromatic surfaces of the SpyderCube calibration object colors in the camera RGB after It is illuminated by illuminations with colors from (a) and its image taken with a camera.

3.2 Printed Colors

In order to simulate the real-world images, lots of material types would have to be analyzed as the spectral reflectance properties are varying between materials. This is because the material properties determine how a color will change under different illuminations, which is important information for simulating real-world behavior. As handling so much data is hardly feasible in terms of both the data acquisition stage and the image generation stage, the proposed dataset generator uses only one material, namely paper. When printing on paper, RGB colors with 8 bits per channel are used, which leads to a total of 256^3 i.e. more than 16 million different possible colors. For each of these RBG

colors, its behavior when printed on paper has to be known for every illumination chosen in Section 3.1. Such behavior for a given illumination can be recorded by photographing the printed colors under the projector cast. For the illumination to really be the same for all colors, all of them have to be photographed on the paper simultaneously. Namely, if they were taken partially over several shots, there is the possibility of slight projector cast color changing due to e.g. projector lamp heating. If all 256^3 colors were used, they could hardly be printed on one paper and later photographed in a high enough resolution. For this reason, instead of using 256^3 color values, for the proposed generator only 32^3 were used. They were generated by putting the three least significant bits in the red, green, and blue channel to zero. This number of colors was shown to be appropriate for printing on a single paper sheet of size A0, which can be photographed in one shot while still having a high enough resolution. The colors were arranged in the grid shape as shown in figure 3. Each square represents one RBG color under the canonical white illumination. To reflectance properties are constant for each color since they were all printed on the same paper by using the same printer and photographed under the same illumination. Once the printed paper was photographed under all of the 707 chosen illuminations, a 5×5 pixel area was taken from each of the squares to represent a single color under some illumination. This means that for each of 32^3 colors there are 25 realistic representations under for of the 707 chosen illuminations that can be used to simulate the effects of randomness as well as noise.

3.3 Generator Cameras

The printed color pattern was photographed under different illuminations with two Canon cameras, namely Canon EOS 550D and Canon EOS 6D Mark II. In order to obtain the linear PNG images that comply with the model in Eq. (1) from raw images, the dcraw tool with options -D -4 -T was used followed by simple subsampling and debayering. The sensor field resolution for the former Canon camera is 5202×3465, whereas the latter camera model has the sensor field resolution of 6384×4224. Higher camera resolution enables higher precision when extracting the color values from the squares of the photographed color pattern as the boundaries of squares tend to get blurred when using lower resolution images. By investigating figure 2b, which show the rb-chromaticities of the illuminations captured with two cameras, the difference in rb-chromaticities of the illuminations can be noticed. This clearly shows how camera sensor characteristics differ, with the Canon EOS 6D Mark II producing smoother illumination estimations.

Figure 3. Squares in all simplified colors arranged in the pattern that was printed on a single big paper, illuminated by 707 different illuminations, and photographed.

3.4 Image Generation

Generating a new image includes choosing the source image, the desired illumination, and the camera sensor. The source image is first simplified following the same procedure as for the creation of the color pattern described in Section 3.2, i.e. the three least significant bits in the red, green, and blue channel are put to zero. That way, the colors in the source image are constrained to the ones in the color pattern shown in figure 3, whose behavior on paper under the previously selected illumination is known. Then, the color of every pixel in the simplified image is changed to a color observed on the pattern square of the same color when it was photographed under the desired illumination. As mentioned earlier, there are 25 possible choices for this change. Doing this for all pixels gives a raw linear image as if the initially chosen image is printed, illuminated by the projector using the initially chosen illumination, and then photographed. Figure 4 illustrates the described steps for the whole image generation process. Repeating this procedure by having a fixed camera sensor results in a new dataset.

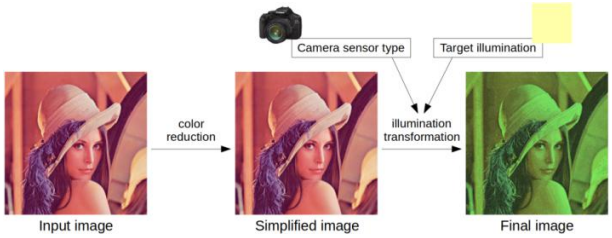


Figure 4. The diagram of the image generation process; the flash tone mapping operator [6, 8] was used for the final image.

4. EXPERIMENTAL VALIDATION

4.1 Error Metrics

The angular error is the most commonly used among many error metrics that have been proposed to measure the performance of illumination estimations methods [35, 3]. There are two kinds of angular error, namely the recovery angular error and the reproduction angular error. When neither of these two is explicitly mentioned, it is commonly understood that the recovery angular error is used. The recovery angular error is defined as the angle between the illumination estimation and the ground-truth illumination

$$err_{recovery} = \cos^{-1}\left(\frac{\rho^{E} \cdot \rho^{E_{st}}}{\| \rho^{E} \| \, \| \rho^{E_{st}} \|}\right) \tag{3}$$

where the $\rho^{E_{st}}$ is the illumination estimation, ρ^{E} is the ground-truth illumination, and '$\cdot$' is the vector dot product. The reproduction angular error [30, 31] has been defined as

$$err_{recovery} = \cos^{-1}\left(\frac{(\rho^{E,W} / \rho^{E_{st}}) \cdot U}{\| \rho^{E,W} / \rho^{E_{st}} \| \sqrt{3}}\right) \tag{4}$$

where $\rho^{E,W}$ is the vector of the white surface color in the image RGB color space under the scene illumination, U is the vector of the ideally corrected white color, i.e. $[1, 1, 1]^T$. Although the recovery angular error has been and still is extensively used, it has been shown in [31] how the change in the illumination of the same scene can cause significant fluctuations of the recovery angular error, while the reproduction angular error has been shown to be stable.

Figure 5. Influence of color reduction: (a) Without color reduction; (b) to (h) with color reduction, starting with only one bit i the red, green, and blue channel put to zero for (b) up to seven bits for (h).

To evaluate the illumination estimation method performance on a whole dataset, the error values calculated for all dataset images are summed up using various summary statistics. As the distribution of the angular errors is non-symmetrical, it is much better to use the median instead of the mean angular error [37]. However, other measures such as mean, trimean, and best and worst *p*% are also used for additional comparisons of methods. In [17] the measure often called as the average was introduced. It is the geometric mean of the mean, median, trimean, best 25%, and worst 25% of the obtained angular errors. In the following experiments, the median angular error of the reproduction angular error has been used as the reference summary statistic.

4.2 Influence of Color Reduction

As described in Sections 3.2 and 3.4, the number of colors in both the printed pattern and the input image are reduced to the total of 32^3 different colors by setting the three least significant bits in the red, green, and blue channel to zero. Figure 5 shows how this type of color reduction influences the quality of sRGB images for different number of bits being set to zero. To test the effect of bits removal on the performance of illumination estimation methods, linear images of the Canon 1Ds Mk III dataset from the NUS datasets [22] were used. Since the dataset generator manages bits on sRGB images, for the sake of simulating bits removal the linear images were first tone mapped and converted to sRGB images with 8 bits per channel by applying the Flash tone mapping operator [6, 8]. Next, the three least significant bits were set to zero, and then the image was returned to its linear form by applying the reversed formula of the Flash tone mapping operator. Finally, illumination estimation methods were applied to such changed images. The results for Gray-world [20], Shades-of-Gray [29], and 1st order Gray-Edge [49] applied on raw images with reduced colors are shown in figure 6. In some cases of bits clearing the median angular error for Gray-World and Shades-of-Gray methods is better than when the original linear images are used. Since bits clearing can eliminate darker pixels, this reminds of [39] where using only bright pixels for illumination estimation resulted in improved accuracy. As opposed to that, the 1st order Gray-Edge method did not improve when removing the bits. This method relies on the edge information to estimate the illuminations and in that case the color reduction can be detrimental since it can reduce edges.

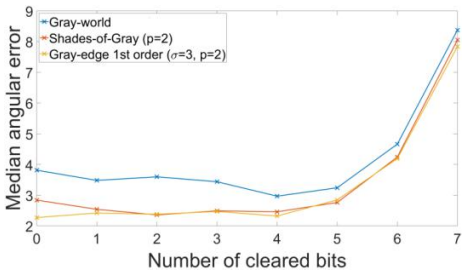


Figure 6. The effect of color reduction on the performance of illumination estimation methods.

4.3 Comparison to the Usual Image Augmentation

One of the techniques of data augmentation used for computational color constancy methods' training is to multiply the image color channels in order to simulate another illumination in rough accordance with Eq. (1). Let $\mathrm{f}^{(\mathrm{e})}$ be an image taken under the observed light source e. If $\hat{\mathrm{f}}^{(\mathrm{e}')}$ is the simulation of $\mathrm{f}^{(\mathrm{e})}$ being taken under the observed light source e', the channel c value of a pixel at location x is then

$$\hat{f}^{(\mathrm{e}')}(\mathrm{x}) = \frac{e'_c}{e_c} f_c^{(\mathrm{e})}(\mathrm{x}) \tag{5}$$

For example this can done by multiplying the color channel values of images and their corresponding ground-truth illuminations by random factors so that $\frac{e'_c}{e_c} \in [0.8, 1.2]$ for every channel c [43, 42] in accordance with the von Kries diagonal model [41]. Since Eq. (5) is a vast oversimplification of Eq. (1) that does not include inter-channel connections, it should have no effect on the error of moment-based methods such as Gray-world if the effects of intensity rounding are ignored. The illumination estimation of Gray-world is

$$\frac{\int \mathrm{f}(x)dx}{\int dx} = \mathrm{e}. \tag{6}$$

The error of Eq. (6) obtained on the augmented images should by definition remain the same except for the rounding errors, which is demonstrated by figure 7. The reproduction angular errors obtained by Gray-world for the images there are 7.0°, 7.06°, 4.63° and 4.62°, respectively, which means that despite a change in the appearance Eq. (5) had little effect on the Gray-world method, while the proposed dataset Generator's result had a significant impact on it.

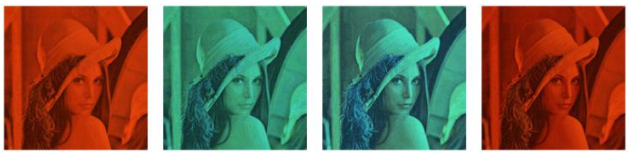


Figure 7. Examples of data augmentation by illumination simulation on images generated by the proposed dataset generator: (a) the linear raw image of the printed photography taken under the red light, (b) the simulation of taking the image under the white light by applying Eq. (5) to the previous image, (c) the linear raw image of the printed photography taken under the white light, and (d) the

simulation of taking the image under the red light by applying Eq. (5) to the previous image.

4.4 Method Performance

Several dataset were created to evaluate the behavior of some simpler illumination estimation methods on generated images and compare it to the behavior on real-world datasets. To create the test datasets, two options were used for the scenes whose printing was to be simulated, two options were used for the camera sensors, and two options were used for the illuminations. When these options were combined through Cartesian product, they resulted in 8 triplets of inputs for the proposed dataset generator and consequently in 8 datasets. Two options for the scenes were the sRGB images of the Canon 1Ds Mk III dataset, which is one of the NUS datasets [22], and synthetic images where all pixel values were randomly drawn from uniform distribution. The camera options included Canon EOS 550D and Canon 6D Mark II. As for the illuminations, the mentioned two options were a subset of illuminations from Section 3.1 that are closest to the ground-truth illuminations of Canon 1Ds Mk III dataset and a subset of randomly chosen illuminations described in Section 3.1. The results for White-Patch [32], Gray-world [20], and Shades-of-Gray [29] on the 8 generated datasets are reported in table 1. The obtained angular error statistics and their relations for different methods are very similar to the ones obtained on other well known real-world datasets [22, 7]. Particularly interesting are the results of the White-patch method. Namely, for the datasets where the Canon EOS 6D Mk II camera was used, the Whitepatch method performed surprisingly well when compared to the datasets where the Canon EOS 550D camera was used. This can be attributed to higher resolution of the former Canon camera as well as of its higher sensor quality due to its being of a significantly newer production date. In other words, the datasets where the Canon EOS 550D camera was used contain more noise then the ones where for the Canon EOS 6D Mk II camera.

4.5 Real-world performance

To check to what degree the datasets generated by the proposed dataset generator resemble the real-world and help coping with it, an experiment with the Cube+ dataset [7] was carried out. This dataset happens to consist of images taken by the very same Canon EOS 550D camera the was used during the creation of the proposed dataset generator. Therefore, the proposed dataset generator was used to simulate the use of the Canon EOS 550D camera to take photos of printed sRGB Cube+ images illuminated by the illuminations similar to Cube+ ground-truth illuminations.

Several learning-based methods were then first trained on the artificially generated dataset and tested on the realworld Cube+ dataset. The obtained results are shown in table 2. Training on real-world images is obviously better, but for methods like Color Beaver the difference in performance with respect to the used training data is not too big and statistics like the median and the trimean angular error are even better. For the Smart Color Cat method the number of bins was restricted due to the colors themselves being restricted. As for the regression trees, their performance was affected the most, but they still obtained relatively accurate results. Some of the performance degrading may be attributed to the Canon EOS 550D data having more noise as previously mentioned, while for Canon EOS 6D Mk II a similar experiment could not have been conducted since it was not used to create any real-world public dataset. The obtained results can be said to serve as a proof-of-concept that learning from realistically generated artificial images can lead to high accuracy on the real-world images.

Table 1. Performance of white-patch [32], gray-world [20], and shades-of-gray [29] on 8 generated datasets (Lower Avg. Is Better). The used format is the same as in [17]. "C1" is the abbreviation for canon 1Ds Mk III dataset, which is one of NUS datasets [22], "550D" represents canon EOS 550D camera, "6D" represents canon 6D mark II camera, and 'RND' is the abbreviation for random.

Scenes, Sensor, Illuminations	Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
C1 6D C1	White-Patch [32]	2.61	2.59	2.50	1.03	4.38	2.38
	Gray-world [20]	6.27	5.32	5.58	3.34	10.75	5.82
	Shades-of-Gray (p=2) [29]	2.79	2.36	2.40	1.28	5.12	2.53
C1 6D RND	White-Patch [32]	2.17	2.05	2.08	0.88	3.73	1.98
	Gray-world [20]	5.79	5.20	5.38	2.64	9.82	5.30
	Shades-of-Gray (p=2) [29]	2.34	1.93	1.96	0.98	4.43	2.08
C1 550D C1	White-Patch [32]	9.41	5.38	5.56	2.60	23.56	7.04
	Gray-world [20]	5.75	5.25	5.39	2.75	9.45	5.31
	Shades-of-Gray (p=2) [29]	2.61	2.07	2.14	0.97	5.20	2.25
C1 550D RND	White-Patch [32]	10.90	6.75	6.81	2.90	27.18	8.31
	Gray-world [20]	5.25	5.04	5.07	2.65	8.32	4.94
	Shades-of-Gray (p=2) [29]	2.15	1.73	1.83	0.67	4.35	1.82
RND 6D C1	White-Patch [32]	2.59	2.23	2.37	1.31	4.31	2.38
	Gray-world [20]	3.84	4.06	3.96	3.06	4.34	3.82
	Shades-of-Gray (p=2) [29]	2.73	2.78	2.78	1.95	3.22	2.66
RND 6D RND	White-Patch [32]	2.46	2.15	2.33	0.88	4.34	2.16
	Gray-world [20]	4.09	4.16	4.20	2.53	5.38	3.96
	Shades-of-Gray (p=2) [29]	2.47	2.64	2.57	1.54	3.17	2.42
RND 550D C1	White-Patch [32]	22.79	10.35	19.52	6.43	51.43	17.24
	Gray-world [20]	3.99	4.28	4.14	2.16	5.65	3.86
	Shades-of-Gray (p=2) [29]	2.36	2.43	2.31	1.13	3.68	2.23
RND 550D RND	White-Patch [32]	25.81	12.30	21.33	7.45	59.80	19.77
	Gray-world [20]	4.25	4.23	4.20	2.44	6.15	4.08
	Shades-of-Gray (p=2) [29]	4.01	2.80	2.81	0.85	9.69	3.04

Table 2. The performance of some learning-based methods on the cube+ dataset [7] with respect to the training (Lower Avg. Is Better). The used format is the same as in [17].

Algorithm	Mean	Med.	Tri.	Best 25%	Worst 25%	Avg.
Trained and tested Cube+ dataset (through cross-validation)						
Smart Color Cat [5]	2.27	1.35	1.61	0.34	5.72	1.58
Regression trees (simple features) [23]	1.57	0.89	1.04	0.20	4.15	1.04
Color Beaver (using Gray-world) [40]	1.49	0.77	0.98	0.21	3.94	0.99
Trained on the generated dataset and tested on the Cube+ dataset						
Regression trees (simple features) [23]	2.54	1.66	1.89	0.45	6.07	1.85
Smart Color Cat [5]	2.47	1.43	1.76	0.40	6.21	1.73
Color Beaver (using Gray-world) [40]	1.73	0.74	0.97	0.37	4.75	1.17

4.6 Influence of More Data on Deep Learning Models

To check whether the proposed dataset generator can help deep learning methods to achieve better accuracy by merely providing an abundance of training data, an experiment with the method described in [19] was performed. Tens of thousands of publicly available real-world images were downloaded from the English Wikipedia and transformed by using the proposed dataset generator and the illuminations close to the ones in the Canon1 dataset [22]. From these images several train sets of various sizes ranging from 100 up to 32000 and used for separate trainings, but always with the same fixed validation set. Additionally, any kind of hyper-parameter tuning was intentionally avoided in order to strictly check only the influence of the train set size on the illumination estimation error. Part of the obtained results shown in figure 8 shows that with the abundance of data even simpler architectures with purposely non-optimal hyper-parameters can achieve state-of-the-art accuracy.

4.7 Comparison to Datasets with Real-world Images

Some of the advantages of using the proposed CroP are:

- there is a large variety of possible illuminations that can be used when images are being created and the illumination distribution can easily be controlled
- the images contain no calibration objects that would have to be masked out to prevent any unfair bias,
- there is no black level and there are no clipped pixels,
- the generated images can be influenced by arbitrary many illuminations with clearly defined ground-truth,
- the number of dataset images can be arbitrarly high. Some of the disadvantages of the proposed CroP include:
- only one material i.e. paper is used in all images,
- the spectral characteristics of the illuminations are limited by the ones of the lamps in the used projector.

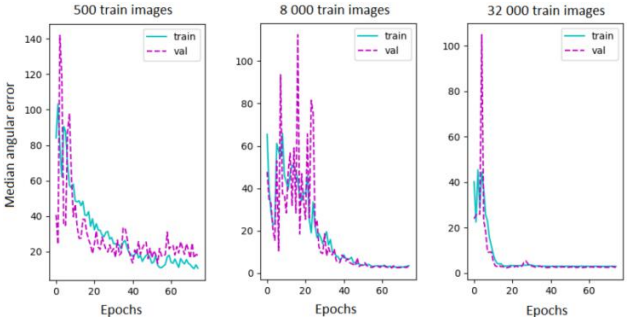


Figure 8. The effect of color reduction on the performance of illumination estimation methods.

5. CONCLUSION

In this paper, a color constancy dataset generator that enables generating realistic linear raw images has been proposed. While image generation is constrained to a smaller subset of possible realistic images, these have been shown to share many properties with the real-world images when statistics-based methods are applied to them. Additionally, it has been demonstrated that these images can be used to train learning-based methods, which then achieve relatively accurate results on the real-world datasets. This potentially means that the proposed dataset generator could be used to create large amounts of images required for some more advanced deep learning techniques. Future work will include experiments with generating images with multiple illuminations and adding new camera models and illuminations.

ACKNOWLEDGMENT

This work has been supported by the Croatian Science Foundation under Project IP-06-2016-2092.

REFERENCES

[1] C. Aytekin, J. Nikkanen, and M. Gabbouj. A data set for camera-independent color constancy. *IEEE Transactions on Image Processing*, 27(2):530–544, 2018.

[2] N. Banić, K. Koščević, M. Subašić and S. Lončarić. The Past and the Present of the Color Checker Dataset Misuse. *arXiv preprint arXiv:1903.04473*, 2019.

[3] N. Banić and S. Lončarić. A Perceptual Measure of Illumination Estimation Error. In *VISAPP*, pages 136 – 143, 2015.

[4] N. Banić and S. Lončarić. Color Cat: Remembering Colors for Illumination Estimation. *Signal Processing Letters, IEEE*, 22(6):651 – 655, 2015.

[5] N. Banić and S. Lončarić. Using the red chromaticity for illumination estimation. In Image and Signal Processing and Analysis (ISPA), *2015 9th International Symposium on*, pages 131 – 136. IEEE, 2015.

[6] N. Banić and S. Lončarić. Puma: A high-quality retinex-based tone mapping operator. In *Signal Processing Conference (EUSIPCO), 2016 24th European*, pages 943 – 947. IEEE, 2016.

[7] N. Banić and S. Lončarić. Unsupervised Learning for Color Constancy. *arXiv preprint arXiv:1712.00436*, 2017.

[8] N. Banić and S. Lončarić. Flash and Storm: Fast and Highly Practical Tone Mapping based on Naka-Rushton Equation. In *International Conference on Computer Vision Theory and Applications*, pages 47 – 53, 2018.

[9] N. Banić and S. Lončarić. Green stability assumption: Unsupervised learning for statistics-based illumination estimation. *Journal of Imaging*, 4(11):127, 2018.

[10] N. Banić and S. Lončarić. Using the Random Sprays Retinex Algorithm for Global Illumination Estimation. In Proceedings of *The Second Croatian Computer Vision Workshopn (CCVW 2013)*, pages 3-7. University of Zagreb Faculty of Electrical Engineering and Computing, 2013.

[11] N. Banić and S. Lončarić. Color Rabbit: Guiding the Distance of Local Maximums in Illumination Estimation. In *Digital Signal Processing (DSP), 2014 19th International Conference on*, pages 345-350. IEEE, 2014.

[12] N. Banić and S. Lončarić. Improving the White patch method by subsampling. In *Image Processing (ICIP), 2014 21st IEEE International Conference on*, pages 605 – 609. IEEE, 2014.

[13] N. Banić and S. Lončarić. Color Dog: Guiding the Global Illumination Estimation to Better Accuracy. In *VISAPP*, pages 129 – 135, 2015.

[14] N. Banić and S. Lončarić. Blue Shift Assumption: Improving Illumination Estimation Accuracy for Single Image from Unknown Source. In *VISAPP*, pages 191 – 197, 2019.

[15] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *IEEE transactions on Image Processing*, 11(9):972–984, 2002.

[16] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms-part ii: Experiments with image data. *IEEE transactions on Image Processing*, 11(9):985–996, 2002.

[17] J. T. Barron. Convolutional Color Constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2015.

[18] J. T. Barron and Y.-T. Tsai. Fast Fourier Color Constancy. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Computer Society Conference on*, volume 1. IEEE, 2017.

[19] S. Bianco, C. Cusano, and R. Schettini. Color Constancy Using CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,* pages 81 – 89, 2015.

[20] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of The Franklin Institute*, 310(1):1–26, 1980.

[21] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8):1509–1519, 2012.

[22] D. Cheng, D. K. Prasad, and M. S. Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA* A, 31(5):1049–1058, 2014.

[23] D. Cheng, B. Price, S. Cohen, and M. S. Brown. Effective learning-based illuminant estimation using simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1000–1008, 2015.

[24] F. Ciurea and B. Funt. A large image database for color constancy research. In *Color and Imaging Conference*, volume 2003, pages 160–164. Society for Imaging Science and Technology, 2003.

[25] M. Ebner. Color Constancy. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2007.

[26] G. D. Finlayson. Corrected-moment illuminant estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1904–1911, 2013.

[27] G. D. Finlayson, G. Hemrit, A. Gijsenij, and P. Gehler. A Curious Problem with Using the Colour Checker Dataset for Illuminant Estimation. In *Color and Imaging Conference*, volume 2017, pages 64–69. Society for Imaging Science and Technology, 2017.

[28] G. D. Finlayson, S. D. Hordley, and I. Tastl. Gamut constrained illuminant estimation. *International Journal of Computer Vision*, 67(1):93–109, 2006.

[29] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[30] G. D. Finlayson and R. Zakizadeh. Reproduction angular error: An improved performance metric for illuminant estimation. *perception*, 310(1):1–26, 2014.

[31] G. D. Finlayson, R. Zakizadeh, and A. Gijsenij. The reproduction angular error for evaluating the performance of illuminant estimation algorithms. *IEEE transactions on pattern analysis and machine intelligence*, 39(7):1482–1488, 2017.

[32] B. Funt and L. Shi. The rehabilitation of MaxRGB. In *Color and Imaging Conference*, volume 2010, pages 256–259. Society for Imaging Science and Technology, 2010.

[33] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition*, 2008. *CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[34] A. Gijsenij and T. Gevers. Color Constancy using Natural Image Statistics. In *CVPR*, pages 1–8, 2007.

[35] A. Gijsenij, T. Gevers, and M. P. Lucassen. Perceptual analysis of distance measures for color constancy algorithms. *JOSA* A, 26(10):2243–2256, 2009.

[36] G. Hemrit, G. D. Finlayson, A. Gijsenij, P. V. Gehler, S. Bianco, and M. S. Drew. Rehabilitating the color checker dataset for illuminant estimation. *CoRR*, abs/1805.12262, 2018.

[37] S. D. Hordley and G. D. Finlayson. Re-evaluating colour constancy algorithms. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 76–79. IEEE, 2004.

[38] Y. Hu, B. Wang, and S. Lin. Fully Convolutional Color Constancy with Confidence-weighted Pooling. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Conference on*, pages 4085–4094. IEEE, 2017.

[39] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey. The role of bright pixels in illumination estimation. In *Color and Imaging Conference*, volume 2012, pages 41–46. Society for Imaging Science and Technology, 2012.

[40] K. Koščević, N. Banić and S. Lončarić. Color Beaver: Bounding Illumination Estimations for Higher Accuracy. In *VISAPP*, pages 183–190, 2019.

[41] J. v. Kries. Theoretische Studien u¨ber die Umstimmung des Sehorgans. *Festschrift der Albrecht-Ludwigs-Universit¨at in Freiburg*, pages 145–148, 1902.

[42] F. Laakom, N. Passalis, J. Raitoharju, J. Nikkanen, A. Tefas, A. Iosifidis, and M. Gabbouj. Bag of color features for color constancy. *arXiv preprint arXiv:1906.04445*, 2019.

[43] F. Laakom, J. Raitoharju, A. Iosifidis, J. Nikkanen, and M. Gabbouj. Color constancy convolutional autoencoder. *arXiv preprint arXiv:1906.01340*, 2019.

[44] E. H. Land. *The retinex theory of color vision*. Scientific America., 1977.

[45] Y. Liu and S. Shen. Self-adaptive Single and Multi-illuminant Estimation Framework based on Deep Learning. *arXiv preprint arXiv:1902.04705*, 2019.

[46] Y. Qian, S. Pertuz, J. Nikkanen, J.-K. K am ar ainen, and J. Matas. Revisiting Gray Pixel for Statistical Illumination Estimation. In *VISAPP*, pages 36–46, 2019.

[47] J. Qiu, H. Xu, Y. Ma, and Z. Ye. PILOT: A Pixel Intensity Driven Illuminant Color Estimation Framework for Color Constancy. *arXiv preprint arXiv:1806.09248*, 2018.

[48] W. Shi, C. C. Loy, and X. Tang. Deep Specialized Network for Illuminant Estimation. In *European Conference on Computer Vision*, pages 371–387. Springer, 2016.

[49] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *Image Processing, IEEE Transactions on*, 16(9):2207–2214, 2007.

[50] J. Van De Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[51] R. Zakizadeh, M. S. Brown, and G. D. Finlayson. A Hybrid Strategy For Illuminant Estimation Targeting Hard Images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 16 – 23, 2015.

Biography

Karlo Koščević was born on February 20, 1995 in Vinkovci, Croatia, where he attended primary school and Natural Sciences-Mathematics High School. After finishing secondary school, he continued his education at the University of Zagreb Faculty of Electrical Engineering and Computing, where he graduated in 2018.

Since October 2018, he has been employed as a research assistant at the Department of Electronic Systems and Information Processing, University of Zagreb Faculty of Electrical Engineering and Computing, where he participated in the scientific project of the Croatian Science Foundation IP-2016-06-2092 "PerfectColor - Methods and algorithms for real-time color image enhancement".

His research interests include computational color constancy, deep learning, and computer vision. He participated in the organization of several international conferences and two illumination estimation challenges. He is an author or co-author of four journal papers and seven conference papers.

Publications

Journal publications

1. Ershov, E., Savchik, A., Semenkov, I., Banić, N., Koščević, K., Subašić, M., Belokopytov, A., Terekhin, A., Senshina, D., Nikonorov, A., Li, Z., Qian, Y., Buzzelli, M., Riva, R., Bianco, S., Schettini, R., Barron, J. T., Lončarić, S., Nikolaev, D. "Illumination Estimation Challenge: The experience from the first 2 years", Color research and application, Vol. 46, No. 4, 2021, pp. 705-718
2. Koščević, K., Subašić, M., Lončarić, S., "Iterative Convolutional Neural Network-Based Illumination Estimation", IEEE Access, Vol. 9, 2021, pp. 26755-26765
3. Ershov, E., Savchik, A., Semenkov, I., Banić, N., Belokopytov, A., Senshina, D., Koščević, K., Subašić, M., Lončarić, S., "The Cube++ Illumination Estimation Dataset", IEEE Access, Vol. 8, 2020, pp. 227511-227527
4. Koščević, K., Subašić, M., Lončarić, S., "Deep Learning-Based Illumination Estimation Using Light Source Classification", IEEE Access, Vol. 8, 2020, pp. 84239-84247

Conference publications

1. Koščević, K., Stipetić V., Provenzi E., Banić N., Subašić, M., Lončarić, S., "HD-RACE: Spray-based Local Tone Mapping Operator", Proceedings of the 12th International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia, 2021, pp. 264-269
2. Koščević, K., Subašić, M., Lončarić, S., "Guiding the Illumination Estimation Using the Attention Mechanism", Proceedings of the 2020 2nd Asia Pacific Information Technology Conference, Bali, Indonesia, 2020, pp. 143-149
3. Koščević, K., Subašić, M., Lončarić, S., "Attention-based Convolutional Neural Network for Computer Vision Color Constancy", Proceedings of the 11th International Symposium on Image and Signal Processing and Analysis, Dubrovnik, Croatia, 2019, pp. 372-377
4. Koščević, K., Banić, N., Lončarić, S., "Color Beaver: Bounding Illumination Estimations for Higher Accuracy", Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 2019, pp. 183-190
5. Banić, N., Koščević, K., Lončarić, S., "The Past and the Present of the Color Checker Dataset Misuse", Proceedings of the 11th International Symposium on Image and Signal Processing and Analysis, Dubrovnik, Croatia, 2019. pp. 366-371
6. Banić, N., Koščević, K., Subašić, M., Lončarić, S., "CroP: Color Constancy Benchmark Dataset Generator", Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing, Bangkok, Thailand, 2020, pp. 1-9
7. Banić, N., Koščević, K., Subašić, M., Lončarić, S., "On Some Desired Properties of Data Augmentation by Illumination Simulation for Color Constancy", Proceedings of 9th International Conference on Signal, Image Processing and Pattern Recognition, Sydney, Australia, 2020. pp. 27-38

Životopis

Karlo Koščević rođen je 20. veljače 1995. godine u Vinkovcima, u Hrvatskoj, gdje je pohađao osnovnu školu te Prirodoslovno-matemetičku gimnaziju. Po završetku srednje škole, upisuje Fakultet Elektroteknike i računarstava Sveučilišta u Zagrebu, gdje je 2018. godine stekao titulu magistra inženjera računarstva.

Od listopada 2018. godine zaposlen je kao asistent na Zavodu za elektroničke sustave i obradbu informacija na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu gdje je sudjelovao na znanstvenom projektu Hrvatske zaklade za znanost IP-2016-06-2092 "Perfect-Color - Metode i algoritmi za poboljšanje slika u boji u stvarnom vremenu".

Njegovi istraživački interesi obuhvaćaju računalnu postojanost boja, duboko učenje i računalni vid. Sudjelovao je u organizaciji nekoliko međunarodnih konferencija i dva natjecanja iz procjene osvjetljenja. Autor je ili suautor četiri rada u znanstvenim časopisima te sedam konferencijskih radova.