

Received August 27, 2021, accepted September 23, 2021, date of publication September 27, 2021, date of current version October 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3115942

Very Deep Learning-Based Illumination Estimation Approach With Cascading Residual Network Architecture (CRNA)

HO-HYOUNG CHOI¹, (Member, IEEE), AND BYOUNG-JU YUN²

¹School of Dentistry, Advanced Dental Device Development Institute, Kyungpook National University, Jung-gu, Daegu 41940, South Korea

²School of Electronics Engineering, IT College, Kyungpook National University, Buk-gu, Daegu 41566, South Korea

Corresponding author: Byoung-Ju Yun (bjisyun@ee.knu.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation (NRF) of Korea by the Ministry of Education under Grant NRF-2019R1I1A3A01061844, and in part by the NRF of Korea by the Ministry of Education under Grant NRF-2018R1D1A1B07040457.

ABSTRACT For the imaging signal processing (ISP) pipeline of digital image devices, it is of high significance to remove undesirable illuminant effects and obtain color invariance, commonly known as ‘computational color constancy’. Achieving the computational color constancy requires going through two phases: the illumination estimation, which will be the primary focus of this work, and the human visual perception-based chromatic adaptation. At the first phase, illumination estimation is to predict RGB triplets, the numeric representations of incident illuminant colors, by calculating the values of image pixels. How much the network can increase its estimation accuracy is a key to realizing computational color constancy. With recent advances in deep learning (DL), a lot of deep learning-based approaches have been suggested, bringing higher accuracy to computer vision applications, but there are still quite a few obstacles to overcome such as instability of learning. In an attempt to address this ill-posed problem in the illumination estimation space, this article presents a novel deep learning-based approach, the Cascading Residual Network Architecture (CRNA), which incorporates the ResNet and cascading mechanism into the deep convolutional neural network (DCNN). The cascading mechanism enables the proposed network to restrain from suddenly varying in size, serves to mitigate learning instability, and accordingly reduces the quality degradation. This is attributed to the ability of the cascading mechanism that fine-tunes the pre-trained DCNN. Considerable amounts of datasets and comparative experiments highlight that the proposed approach delivers more stable and robust results and imply the potential for generalization of the proposed approach across deep learning applications.

INDEX TERMS Image signal processing pipeline, computational color constancy, learning instability, cascading mechanism, illumination estimation.

I. INTRODUCTION

In digital photography, digital images may carry undesired color casts due to an unintended source illuminant in a scene. A great way to understand the unwanted illuminant effect is taking multiple photos of the same scene with the same camera under varying illuminant conditions. The resulting digital images depict the same scene but in varying colors. The imaging model [1] is used to calculate the pixel values based on three factors: the spectrum of the source illuminant, the reflectance of the object surface and the spectral

sensitivity of the camera sensor. The latter two factors: the reflectance of the object surface and the spectral sensitivity of the camera sensor, are kept constant by shooting the same scene with the same camera whereas the spectrum of the source illuminant is varied by taking the photos under varying illuminant conditions. This points to a fact that a digital camera can capture any incident source illuminant but cannot detect the illuminant itself. Therefore, it is of top concern to build the illuminant invariance feature into the ISP pipeline of digital cameras, based on the concept of human visual perception-based (HVP) color constancy [2]. It has been verified that the color constancy is of great use in a wide range of image processing applications such as

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

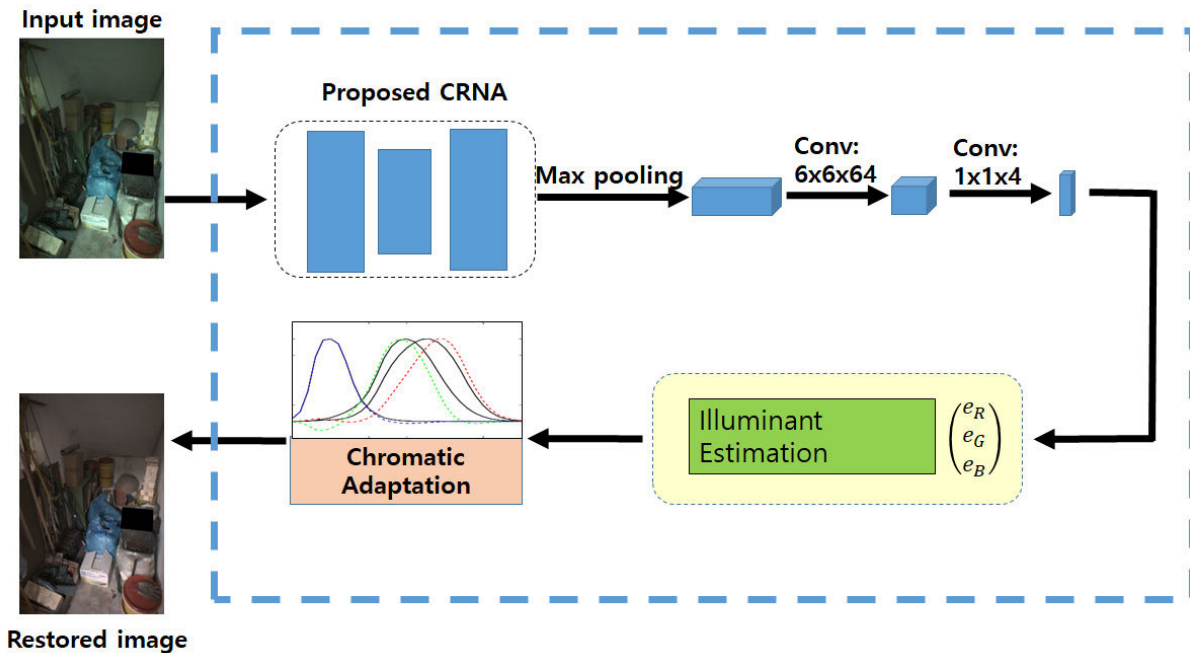


FIGURE 1. The illuminant estimation and chromatic adaptation procedure of the proposed CRNA.

object recognition, scene comprehension and image reproduction [3]. Achieving the computational color constancy requires going through two phases: the illumination estimation, which will be the primary focus of this work, and the HVP-based chromatic adaptation. A network model predicts the colors of the scene illuminant based on the pixel values of the image and removes the estimated undesirable illuminant effect. Over the past several decades, a lot of efforts have gone into achieving more accurate illuminant estimation in the image processing space and also a lot of methods have been proposed, which are largely classified into statistics-based and learning-based categories. In an attempt to obtain accurate illumination estimation, the statistics-based approaches start with the assumption that the illuminant is commonly uniform in the scene, including WP (White-Patch) [4], [5] and its advanced version [6]–[8], and the gray world assumption-based approaches such as GW (Gray-World) [9], SoG (shades of gray) [10], GE (gray edge) [11] and WGE (weight-gray-edge) [12]. These approaches have the benefit of fast computing, but incur huge computation costs, unfortunately. Meanwhile, the learning-based approaches adopt learning models for illuminant color estimation. Recently, more and more learning-based approaches are employing deep learning for illumination estimation, thereby progressing their accuracy. However, a drawback of these approaches is demanding massive learning datasets. In addition, the learning-based approaches require using an increasing amount of computational resources and building more and more complicated architectures. Yet, the latest DL approaches present shallow architectures that consist of only a few convolution layers and

fully connected layers [13], [14]. A content-based CNN is proposed in ref. [15]–[17].

Ref. [18] and [19] navigate the illumination classification issue with the use of DL. Ref. [20] proposes two DCNN architectures and selects the better of the two through comparative studies. Other learning-based approaches include Bayesian learning [21], color moments [22], gamut mapping [23]–[25], spatial localization [26], [27], Choi's illuminant estimation approaches [28]–[30] and others [31], [32]. In short, the learning-based approaches have proven to outperform their statistics-based counterparts in terms of estimation accuracy throughout a lot of literature and studies. There is no denying that deeper network architectures effect higher performance, as evidenced by a lot of proposed studies and literature.

Motivated to address remaining challenges and take opportunities to advance inference accuracy, this article presents the Cascading Residual Network Architecture (CRNA) by building the cascading mechanism and the residual network [33] into the DCNN. The residual network is commonly used in the DCNN. In the proposed CRNA, the key strategy is to increase estimation accuracy to the highest possible level and eliminate undesired illuminant estimates. Estimating and removing the undesired illuminant is referred to as white balance or HVP-based chromatic adaptation. What differentiates the proposed method from conventional DCNN architectures is the cascading mechanism which serves to restrain the network from sudden variations in size, thereby mitigating learning instability and reducing quality degradation. This is attributed to the ability of the cascading mechanism that fine-tunes the pre-trained DCNN.

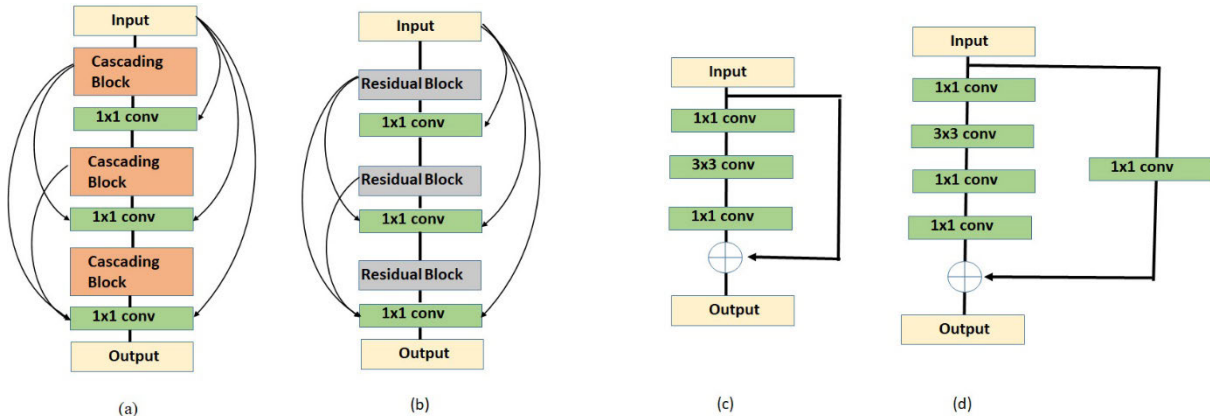


FIGURE 2. The breakdown of the proposed architecture: (a) the proposed CRNA or global cascading connection, (b) local cascading block, (c) residual block and (d) its extended residual block.

This article has core contributions as follows:

- Proposing a novel deep learning-based approach, the CRNA, to taking estimation accuracy to a higher level, by building the ResNet and cascading mechanism into the network model.
- Addressing the instability problem and the quality degradation of conventional DCNNs by enabling the architecture to fine-tune the pre-trained DCNN.
- Highlighting that the proposed approach effects higher stability and robustness and implying the potential for generalization of the proposed approach across deep learning applications, as supported by considerable amounts of datasets and comparative experiments.

II. RELATED WORK

In digital imaging, the imaging model is built based on the hypothesis of the Lambertian reflectance. An image f is made up of pixels. Each pixel x , represented by a red, green and blue triplet color $c \in \{R, G, B\}$, equals the sum of the multiplication of every triple of source illuminant spectrum $e(\lambda, x)$, surface reflectance $R(\lambda, x)$ and camera sensitivity $\rho_c(\lambda)$ in the range of wavelengths λ of the visible illuminant spectrum ω , which goes as follows:

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

A. ILLUMINATION ESTIMATION

Figure 1 illustrates how the proposed CRNA network performs illuminant estimation and the chromatic adaptation. The proposed network predicts illuminant colors with the use of image pixels calculated from Eq. (1). The illuminant estimation is determined by the source illuminant spectrum $e(\lambda)$ and camera sensitivity $\rho_c(\lambda)$, which goes as follows:

$$e = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_{\omega} e(\lambda)\rho_c(\lambda)d\lambda \tag{2}$$

Yet the illumination prediction is a big challenge, given that both illuminant spectral distribution $e(\lambda)$ and camera sensitivity $\rho_c(\lambda)$ are out of control.

B. CHROMATIC ADAPTATION TRANSFORMATION

In color constancy, the chromatic adaptation is another critical step to remove the color cast and perform color adjustment. In this white balance process, the proposed approach uses the prevalent diagonal matrix [34], known as the von Kries diagonal matrix model [35]. A triplet pixel $p = (p_R p_G p_B)^T$ multiplied by the diagonal matrix D equals a color rendered pixel, $\hat{p} = (\hat{p}_R \hat{p}_G \hat{p}_B)^T$, which goes as follows:

$$\hat{p} = Dp \tag{3}$$

The diagonal matrix is usually represented as follows:

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

where $e = (e_R e_G e_B)^T$ refers to an undesired illuminant, and $\bar{e} = (\bar{e}_R \bar{e}_G \bar{e}_B)^T$ means the desired illuminant or ground truth illuminant. The perfect white source illuminant $\bar{e} = (1 1 1)^T$ serves as the ground truth illuminant in the proposed approach.

III. THE PROPOSED CRNA APPROACH

The objective of the proposed CRNA approach is to accurately estimate illuminant and efficiently eliminate undesired color casts. This section sets out what the proposed CRNA approach is and how it works. The proposed CRNA approach incorporates the ResNet and cascading mechanism into the DCNN with a view to making the network go deeper and thereby improve performance. The proposed CRNA realizes the cascading mechanism by embedding local and global cascading modules in it. The ResNet architecture is widely used in the DCNN architecture because of its higher learning efficiency and surpassing performance, but does not have both local and global cascading modules. In the proposed

CRNA network, the embedded local and global cascading modules serve to accelerate performance.

To explain how the proposed CRNA works, let f be a convolution function, with an activation function $\tau(z) \equiv \max(0, x)$. In Eq. (5), H_i represents the output of the i -th residual block (RB). W_R^i refers to a set of parameters used in the RB and $W_R^{i,j}$ represents a set of parameters used in the i -th convolution layer inside the i -th RB. In the proposed CRNA, two consecutive convolution layers form a residual block and R_i means the i -th RB, which goes as follows [36]:

$$R_i \left(H^{i-1}; W_R^i \right) = \tau \left(f \left(\tau \left(f \left(H^{i-1}; W_R^{i,1} \right) \right); W_R^{i,2} \right) + H^{i-1} \right) \quad (5)$$

The input image goes through the Eq. (5), gets to the final RB of ResNet and produces the output feature map H^u , which goes as follows:

$$H^u = R_u \left(\dots \left(R_1 \left(f(X; W_R^1) \right) \dots; W_R^u \right) \right) \quad (6)$$

Each RB is followed by a single convolution layer $f(X; W_c)$ which has the parameter, W_c . In general, ResNet does not have the cascading block, but the proposed CRNA has incorporated the local cascading block (CB) into it. B_i is the output of the i -th RB and the input of the i -th local CB, and the i -th local CB has a set of parameters W_c^i . The i -th local CB is described as follows:

$$B_{local}^i \left(H^{i-1}; W_c^i \right) \equiv B^{i,U} \quad (7)$$

where $B^{i,U}$ works in a recursive manner as follows:

$$B^{i,0} = H^{i-1} \quad (8)$$

$$B^{i,u} = f \left(\left[I, B^{i,0}, \dots, B^{i,u-1}, R^u \left(B^{i,u-1}; W_R^u \right) \right]; W_c^{i,u} \right) \quad (9)$$

for $u = 1, \dots, U$

In the end, the final CB produces the output feature map H^b which represents a combination of both local and global cascading mechanisms. H^0 refers to the output of the first convolution layer and the parameters are fixed at $u = b = 3$ throughout the network.

$$H^0 = f(X; W_i) \quad (10)$$

$$H^0 = f \left(\left[H^0, \dots, H^{b-1}, B_{local}^u \left(H^{b-1}; W_B^b \right) \right] \right); \quad (11)$$

for $b = 1, \dots, B$

The key difference between the proposed CRNA and the ResNet lies in whether the cascading mechanism exists or not. As indicated above, the proposed CRNA has both global and local cascading mechanisms. Figure 2 (a) shows how the global cascading connection works and Figure 2 (b) depicts how the local cascading blocks work to form the global cascading connection. ResNet and its extended network in Figure 2 (c) and (d), respectively, are built into the local cascading block. Cascading at the local and global levels has two benefits: 1) it is possible for the architecture to combine

multiple-level features from multiple layers. 2) Multiple-level cascading connections serve as multiple-level shortcut connections that rapidly propagate information from lower to higher layers in the proposed CRNA. In brief, the local and global cascading mechanisms advance the proposed CRNA to a higher level of estimation accuracy and computational efficiency.

IV. EXPERIMENTAL RESULTS AND EVALUATIONS

This section explores comparative experiments conducted to optimize parameters, select better-performing residual network, compare the proposed approach with its latest learning-based competitors, and verify illuminant invariance and camera invariance. The experiments use benchmark image datasets: Gehler and Shi image dataset [21], NUS-8 camera dataset [37] and Gray Ball dataset [38]. To explain the dataset used in the experiments, the NUS-8 camera dataset consists of 8 subsets of images captured by 8 different cameras and each subset contains 210 images. The Gray Ball dataset includes 11,340 images which depict a variety of scenes and each image has a gray ball in sight. In addition, Gehler and Shi dataset is made up of 568 images that represent a lot of different indoor and outdoor images. Importantly, these images are taken under diverse source illuminants and each image has a color checker in sight for the purpose of evaluating the effect of the illuminant condition on the image.

In optimizing parameters, the initial training rate is chosen to be optimized among other parameters because it has the most significant impact on the accuracy of illuminant estimation. To find the optimal initial training rate, several initial training rates are compared in an experiment that uses Gehler and Shi image dataset. Other parameters including the weight decay of 5×10^{-5} and a momentum of 0.9 are kept fixed. In the experiment, the network is designed to resize the images to 512×512 pixels and the batch size is set to 16. Figure 3 shows the comparative result of applying several initial training rates to the proposed CRNA. The proposed CRNA is programmed based on Tensorflow [39] and operates on a single NVIDIA Titan RTX GPU. The training takes place 10K iterations for 1.5 days. Figure 3 (a) plots the median angular error and Figure 3 (b) plots the average angular error at the different initial training rate conditions. As a result, both median and average angular errors are least at the initial training rate of $3.00E-4$ which translates into 3×10^{-4} . In the deep learning space, it is well-known by various studies and experiments that going deeper is a great way to accelerate performance. So, a natural motivation is to test it through an experiment. In the next experiment, it is designed to determine which residual block performs better: Figure 2 (c) or Figure 2 (d). The residual block in Figure 2 (c) is called the bottleneck network and that of Figure 2 (d) is its extended version.

Figure 4 shows the comparative result of implementing the residual block of Figure 2 (c) and its extended version of Figure 2 (d) in the proposed CRNA. The comparison is made in terms of median angular error and average angular error.

TABLE 1. Comparison of the proposed CRNA and its latest learning-based counterparts in terms of angular error with the use of the Gehler and Shi dataset.

Methods	Mean	Median	Trimean	Best-25%	Worst-25%
SVR [40]	8.08	6.73	7.19	3.35	14.89
EBG [41]	6.52	5.04	5.43	1.9	13.58
BS [21]	4.82	3.46	3.88	1.26	10.46
NIS[42]	4.19	3.13	3.45	1	9.22
IBG [40]	4.2	2.39	2.93	0.51	10.7
CART [43]	3.9	2.91	3.21	1.02	8.27
S-spectral[44]	3.59	2.96	3.1	0.95	7.61
EM [45]	2.89	2.27	2.42	0.82	5.97
19-ECM[22]	2.86	2.04	2.22	0.7	6.34
CNN based [46]	2.75	1.99	2.14	0.74	6.05
ED [47]	2.42	1.65	1.75	0.38	5.87
HZ [48]	2.29	1.9	2.03	0.57	4.72
DSNet[20]	2.24	1.46	1.68	0.48	6.08
SNet-FC4[49]	2.23	1.57	1.72	0.47	5.15
ANet-FC4[49]	2.12	1.53	1.64	0.48	4.78
CM [28]	2.09	1.42	1.6	0.35	4.78
CMoDE[29]	2.05	1.06	1.42	0.29	4.5
PMRN [30]	2.01	1.02	1.35	0.24	3.4
Proposed Method	1.99	1.01	1.33	0.22	3.2

TABLE 2. Comparison of the proposed CRNA and its latest learning-based counterparts in terms of angular error with the use of Gray ball dataset.

Method	Mean	Median	Trimean	Best-25%	Worst-25%
SVR	13.17	11.28	11.83	4.42	25.02
BS	6.77	4.7	5	-	-
NIS	5.24	3	4.35	1.21	11.15
EM	4.42	3.48	3.77	1.01	9.36
CNN-based	4.8	3.7	-	-	-
CM	4.03	1.88	2.6	0.61	10.77
CMoDE	3.28	1.7	2.27	0.53	6.69
PMRN	3.02	1.64	1.74	0.49	6.45
Proposed	2.91	1.62	1.71	0.48	6.1

In deep learning, there is a general belief that deeper networks make better performance. However, this experiment challenges the common belief and demonstrates that going deeper does not always equal higher performance.

In the following experiment, the proposed CRNA is compared with its latest learning-based counterparts, using

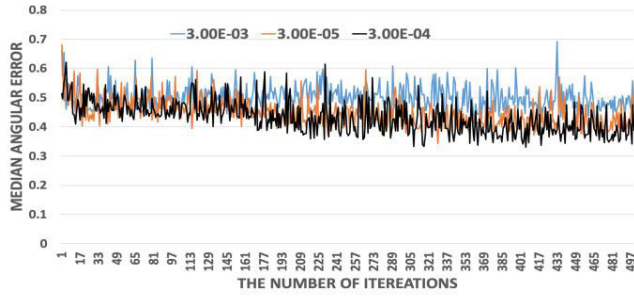
Gehler and Shi dataset. Table 1 is the comparative summary of their respective average, median, trimean, best-25%, and worst-25% angular errors. Noticeably, Hu and his colleague [49] present a segment-wise illuminant estimation approach, known as FC4. Their approach uses the confidence map of each image patch to estimate undesired source

TABLE 3. Comparison of the proposed CRNA and its latest learning-based counterparts in terms of angular error with the use of NUS-8 camera dataset.

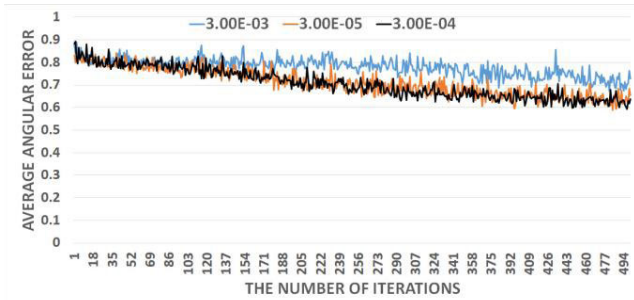
Method	PG	BF	SS	NIS	CM	CMo	PMR	ProM
Camera	Mean-Angular Error							
Canon1Ds	6.13	3.58	3.21	4.18	3.18	3.05	2.93	2.88
Canon600D	14.51	3.29	2.67	3.43	2.35	2.21	2.12	2.09
FujiXM1	8.59	3.98	2.99	4.05	3.1	2.95	2.76	2.73
NikonD5200	10.14	3.97	3.15	4.1	2.35	2.23	2.21	2.18
OlympEPL6	6.52	3.75	2.86	3.22	2.47	2.32	2.26	2.21
LumixGX1	6	3.41	2.85	3.7	2.46	2.34	2.25	2.23
SamNX2000	7.74	3.98	2.94	3.66	2.32	2.18	2.13	2.11
SonyA57	5.27	3.5	3.06	3.45	2.33	2.21	2.12	2.09
Camera	Median-Angular Error							
Canon1Ds	4.3	2.8	2.67	3.04	2.71	2.43	1.57	1.55
Canon600D	14.83	2.35	2.03	2.46	2.19	2.05	1.27	1.22
FujiXM1	8.87	3.2	2.45	2.96	2.82	2.55	2.01	1.97
NikonD5200	10.32	3.1	2.26	2.4	1.92	1.66	1.35	1.32
OlympEPL6	4.39	2.81	2.24	2.17	2.12	1.81	1.78	1.74
LumixGX1	4.74	2.41	2.22	2.28	1.42	1.25	1.22	1.21
SamNX2000	7.91	3	2.29	2.77	1.32	1.15	1.12	1.11
SonyA57	4.26	2.36	2.58	2.88	1.65	1.53	1.5	1.48
Camera	Trimean-Angular Error							
Canon1Ds	4.81	2.97	2.79	3.3	2.69	2.52	2.04	2.02
Canon600D	14.78	2.4	2.18	2.72	2.33	2.21	1.49	1.48
FujiXM1	8.64	3.33	2.55	3.06	2.88	2.74	2.12	2.07
NikonD5200	10.25	3.36	2.49	2.77	1.95	1.9	1.63	1.59
OlympEPL6	4.79	3	2.28	2.42	2.18	1.95	1.9	1.86
LumixGX1	4.98	2.58	2.37	2.67	1.81	1.64	1.62	1.59
SamNX2000	7.7	3.27	2.44	2.94	1.65	1.51	1.48	1.47
SonyA57	4.45	2.57	2.74	2.95	1.91	1.75	1.74	1.73
Camera	Mean of Best-25%							
Canon1Ds	1.05	0.76	0.88	0.78	0.65	0.52	0.49	0.46
Canon600D	9.98	0.69	0.68	0.78	0.73	0.59	0.47	0.46
FujiXM1	3.44	0.93	0.81	0.86	0.75	0.81	0.72	0.69
NikonD5200	4.35	0.92	0.86	0.74	0.57	0.43	0.42	0.38
OlympEPL6	1.42	0.91	0.78	0.76	0.8	0.75	0.73	0.71
LumixGX1	2.06	0.68	0.82	0.79	0.65	0.53	0.51	0.49
SamNX2000	2.65	0.93	0.75	0.75	0.53	0.39	0.33	0.32
SonyA57	1.28	0.78	0.87	0.83	0.57	0.42	0.4	0.39
Camera	Mean of Worst-25%							
Canon1Ds	14.16	7.95	6.43	9.51	6.67	6.53	6.35	6.32
Canon600D	18.45	7.93	5.77	5.76	5.29	5.19	5.15	5.12
FujiXM1	13.4	8.82	5.99	9.37	5.64	5.53	5.4	5.38
NikonD5200	15.93	8.18	6.9	10.01	4.86	4.72	4.66	4.64
OlympEPL6	15.42	8.19	6.14	7.46	4.62	4.49	4.4	4.35
LumixGX1	12.19	8	5.9	8.74	5.74	5.55	5.21	5.18
SamNX2000	13.01	8.62	6.22	8.16	5.55	5.39	4.58	4.56
SonyA57	11.16	8.02	6.17	7.18	5.12	4.95	4.6	4.59

illuminant colors. In their approach, the network architecture is meant to learn semantic information and build the confidence map. However, it comes with smoothing of the local

region of the image due to partially inaccurate local estimation. The inaccuracy occurs because the confidence map, which is supposed to mask and remove noise, mistakenly



(a)

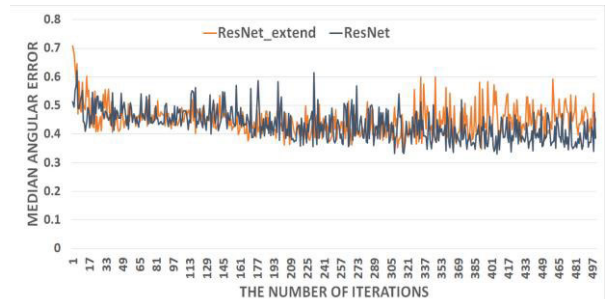


(b)

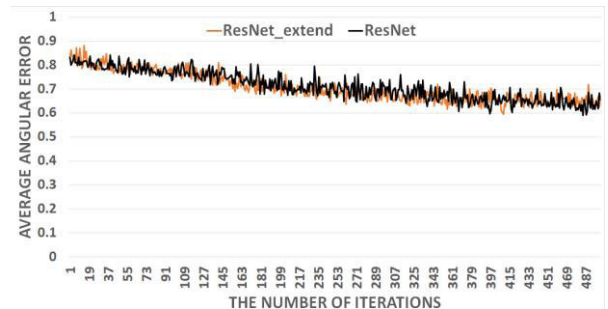
FIGURE 3. Comparison of (a) median angular error and (b) average angular error of several initial training rates on the logarithmic domain with the use of the Gehler and Shi image dataset.

perceives small objects as noise and thus masks them. This leads the architecture to falsely recognize small objects as noise and eliminate them. To solve the inaccuracy problem, Choi and his colleague [28], [29] set forth novel approaches by bringing the ResNet and the dilated convolution to the architecture. The two new approaches surpass their latest competitors from the estimation accuracy perspective. Unfortunately, these approaches have a long-term dependency problem due to the nature of linearity that the input of the current layer is the output of the previous layer. To address the long-term dependency challenge, Choi and Yun [30] come up with another novel approach called PMRN. In continuing efforts to advance the network to the next level, they present the proposed CRNA in this work by building the cascading mechanism and ResNet into the DCNN. The key strategy behind the proposed CRNA is to adopt the cascading mechanism and thereby improve learning stability. The cascading mechanism serves to restrain the network model from sudden variations in size. This cascading mechanism differentiates the proposed approach from conventional DCNN architectures by mitigating learning instability and accordingly reducing the quality degradation. To be concrete, it is attributed to the ability of the cascading mechanism that fine-tunes the pre-trained DCNN. Resulting, the proposed CRNA delivers the most up-to-date performance in the field of color constancy as in Table 1.

Figure 5 displays another experiment which compares the proposed CRNA approach and its latest learning-based counterparts in terms of total training loss and the resultant



(a)



(b)

FIGURE 4. Comparison of (a) median angular error and (b) average angular error of the residual block and its extended version on the logarithmic domain with the use of the Gehler and Shi image dataset.

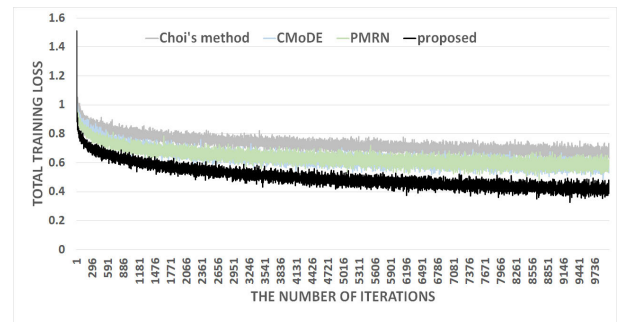


FIGURE 5. Comparison of the proposed CRNA and its latest learning-based counterparts in terms of total training loss and the resultant convergence behavior trend on the logarithmic space with the use of standard stochastic gradient descent (SGD).

convergence behavior trend. It uses the standard stochastic gradient descent (SGD). As for the outcome, the proposed CRNA approach tops its latest learning-based approaches by tending towards the lowest total training loss. Figure 6 plots the angular error distribution, comparing the proposed CRNA approach and some high performers from Table 1: SNet-FC4 (or SqueezeNet-FC4), CMoDE and PMRN. Consequently, the proposed CRNA approach shows an overall lower angular error trend. In order to demonstrate the illuminant invariance, another experiment follows. Using the Gray-ball dataset, the experiment compares the proposed CRNA approach and its learning-based counterparts in terms of mean, median, trimean, best-25%, and worst-25%. Table 2 encapsulates the

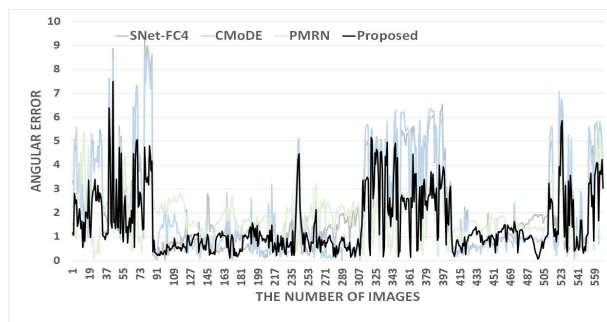


FIGURE 6. Comparison of the proposed CRNA and its latest select learning-based counterparts in terms of angular error distribution with the use of 568 images.

experimental result where the proposed CRNA approach reports the lowest angular error in contrast with its conventional counterparts. In order to verify the camera invariant, the final experiment compares the proposed CRNA approach and its latest conventional counterparts with the use of the NUS-9 camera dataset. Table 3 encapsulates the experimental result where the proposed CRNA approach excels its counterparts regardless of the camera sensitivity.

V. CONCLUSION

This article presents a novel approach to more accurate illuminant estimation by embedding the ResNet and cascading mechanism into the DCNN. The cascading mechanism of the novel approach differentiates the proposed CRNA from conventional DCNN architectures, which restrains the network model from sudden variations in size and thereby addresses the learning instability. It also serves to fine-tune the pre-trained DCNN and thereby contributes to reducing the quality degradation problem. From two aspects, the proposed architecture benefits from the cascading at both local and global levels. First, it is possible for the architecture to combine multiple-level features from multiple layers. Second, multiple-level cascading connections serve as multiple-level shortcut connections that rapidly propagate information from lower to higher layers. In summary, the local and global cascading mechanisms advance the proposed CRNA to a higher level of accuracy and computational efficiency. Prevalent benchmark image datasets and various experiments highlight unprecedented superiority of the proposed CRNA approach over its state-of-the-art conventional competitors. Notwithstanding, it is imperative to continue to work towards the optimization of the DCNN architecture and take the computer vision to the next level.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their insightful comments.

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] E. H. Land, *The Retinex Theory of Color Vision*. Boston, MA, USA: Scientific America, 1977.

[5] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. 18th Color Imag. Conf. Final Program Color Imag. Conf.*, vol. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2010, pp. 256–259.

[6] N. Banić and S. Lončarić, "Using the random sprays Retinex algorithm for global illumination estimation," in *Proc. 2nd Croatian Comput. Vis. Workshop (CCVW)*. Zagreb, Croatia: Univ. Zagreb Fac. Electr. Eng. Comput., 2013, pp. 3–7.

[7] N. Banić and S. Lončarić, "Color rabbit: Guiding the distance of local maximums in illumination estimation," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 345–350.

[8] N. Banić and S. Lončarić, "Improving the white patch method by sub-sampling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 605–609.

[9] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.

[10] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color Imag. Conf.*, vol. 1. Springfield, VA, USA: Society for Imaging Science and Technology, 2004, pp. 37–41.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] K. Koscevic, M. Subasic, and S. Lončarić, "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.

[17] K. Košćević, 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.

[18] 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.

[19] K. Koscevic, M. Subasic, and S. Loncaric, "Deep learning-based illumination estimation using light source classification," *IEEE Access*, vol. 8, pp. 84239–84247, 2020.

[20] 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.

[21] 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.

[22] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1904–1911.

[23] D. A. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 5–35, Aug. 1990.

[24] K. Barnard, "Improvements to gamut mapping colour constancy algorithms," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2000, pp. 390–403.

[25] 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.

[26] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 379–387.

[27] 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.

[28] H. H. Choi, H. S. Kang, and B. J. Yun, "CNN-based illumination estimation with semantic information," *Appl. Sci.*, vol. 10, no. 14, pp. 1–17, Jul. 2020.

- [29] H.-H. Choi and B.-J. Yun, "Deep learning-based computational color constancy with convoluted mixture of deep experts (CMoDE) fusion technique," *IEEE Access*, vol. 8, pp. 188309–188320, 2020.
- [30] H.-H. Choi and B.-J. Yun, "Learning-based illuminant estimation model with a persistent memory residual network (PMRN) architecture," *IEEE Access*, vol. 9, pp. 29960–29969, 2021.
- [31] N. Banić and S. Lončarić, "Color cat: Remembering colors for illumination estimation," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 651–655, Jun. 2013.
- [32] N. Banić and S. Lončarić, "Using the red chromaticity for illumination estimation," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2015, pp. 131–136.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [34] 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.
- [35] 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.
- [36] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," 2018, *arXiv:1803.08664*. [Online]. Available: <http://arxiv.org/abs/1803.08664>
- [37] 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.
- [38] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. 11th Color Imag. Conf. Imag. Sci. Technol.*, 2003, pp. 160–164.
- [39] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: <http://arxiv.org/abs/1603.04467>
- [40] W. Xiong and B. Funt, "Estimating illumination chromaticity via support vector regression," *J. Imag. Sci. Technol.*, vol. 50, no. 4, pp. 341–348, Jul. 2006.
- [41] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *Int. J. Comput. Vis.*, vol. 86, no. 2, pp. 127–139, Jan. 2010.
- [42] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 687–698, Apr. 2011.
- [43] 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.
- [44] 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.
- [45] H. R. V. Joze and M. S. Drew, "Exemplar-based color constancy and multiple illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 860–873, May 2014.
- [46] S. Bianco, C. Cusano, and R. Schettini, "Single and multiple illuminant estimation using convolutional neural networks," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4347–4362, Sep. 2017.
- [47] 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.
- [48] H. Zhan, S. Shi, and Y. Huo, "Computational colour constancy based on convolutional neural networks with a cross-level architecture," *IET Image Process.*, vol. 13, no. 8, pp. 1304–1313, Feb. 2019.
- [49] Y. Hu, B. Wang, and S. Lin, "FC⁴: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. CVPR*, Jul. 2017, pp. 4085–4094.



HO-HYOUNG CHOI (Member, IEEE) received the Ph.D. degree in mobile communication engineering from Kyungpook National University, Daegu, South Korea, in 2012. From 2014 to 2019, he was with Chungbuk National University, where he was a Postdoctoral Researcher and an Assistant Professor. Since Jun 2019, he has been with the School of Dentistry, Advanced Dental Device Development Institute, Kyungpook National University, where he is currently a Research Invited Professor. His current research interests include image processing, computer vision, machine vision, color constancy, tone mapping for HDR image, machine learning, and convolutional neural networks.



BYOUNG-JU YUN received the Ph.D. degree in electrical engineering and computer science from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2002. From 1996 to 2003, he was with SK Hynix Semiconductor Inc., where he was a Senior Engineer. From 2003 to 2005, he was with the Center for Next Generation Information Technology, Kyungpook National University, where he was an Assistant Professor. Since 2005, he has been with the School of Electronics Engineering, Kyungpook National University, where he is currently an Invited Professor. His current research interests include image processing, color consistency, multimedia communication systems, HDR color image enhancement, biomedical image processing, and HCI.

• • •