Received September 21, 2020, accepted October 5, 2020, date of publication October 14, 2020, date of current version October 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030912

Deep Learning-Based Computational Color Constancy With Convoluted Mixture of Deep Experts (CMoDE) Fusion Technique

HO-HYOUNG CHOI[1] **AND BYOUNG-JU YUN**[2]

[1]School of Dentistry, Advanced Dental Device Development Institute, Kyungpook National University, Daegu 41940, South Korea
[2]School of Electronics Engineering, IT College, Kyungpook National University, 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 funded by the Ministry of Education under Grant NRF-2019R1I1A3A01061844, and in part by the NRF of Korea funded by the Ministry of Education under Grant NRF-2018R1D1A1B07040457.

ABSTRACT In the human and computer vision, color constancy is the ability to perceive the true color of objects in spite of changing illumination conditions. Color constancy is remarkably benefitting human and computer vision issues such as human tracking, object and human detection and scene understanding. Traditional color constancy approaches based on the gray world assumption fall short of performing a universal predictor, but recent color constancy methods have greatly progressed with the introduction of convolutional neural networks (CNNs). Yet, shallow CNN-based methods face learning capability limitations. Accordingly, this article proposes a novel color constancy method that uses a multi-stream deep neural network (MSDNN)-based convoluted mixture of deep experts (CMoDE) fusion technique in performing deep learning and estimating local illumination. In the proposed method, the CMoDE fusion technique is used to extract and learn spatial and spectral features in an image space. The proposed method distinctively piles up layers both in series and in parallel, selects and concatenates effective paths in the CMoDE-based DCNN, as opposed to previous works where residual networks stack multiple layers linearly and concatenate multiple paths. As a result, the proposed CMoDE-based DCNN brings significant progress towards efficiency of using computing resources, as well as accuracy of estimating illuminants. In the experiments, Shi's Reprocessed, gray-ball and NUS-8 Camera datasets are used to prove illumination and camera invariants. The experimental results establish that this new method surpasses its conventional counterparts.

INDEX TERMS Color constancy, CMoDE fusion technique, multi-stream deep neural network (MSDNN), illumination estimation, residual networks.

I. INTRODUCTION

In the computer vision, the perceived color of objects is significantly impacted by the color of illumination in the scene [1]. In contrast, the human vision has visual perception that enables them to effortlessly distinguish objects by color. This visual perception is referred to as color constancy [2], [3] in the computer vision. When it comes to a captured image, its recorded color carries a combined effect of the primary peculiarities of the source illuminant, surface reflectance and sensitivity function of the digital imaging device [4]. In particular, the color of the illuminant has a crucial repercussion on the recorded color of the digital image. For this reason, it is significantly meaningful to eliminate undesirable illuminant effects in order to facilitate robust performance of color-based computer vision systems such as human-computer interaction, video analytics, object tracking, color feature extraction and digital photography [5], [6]. For the past two decades, computer vision researchers have suggested a lot of color constancy adjustment techniques to cope with coloration and color cast in digital images. Broadly speaking, the previous works can be categorized as statistics-based, gamut-based, physics-based, learning-based and biologically inspired methods.

The statistics-based approach aims to analyze and represent images not by estimating illumination but by using the

The associate editor coordinating the review of this manuscript and approving it for publication was Chang-Hwan Son.

features that stay constant irrespective of the scene light. This approach has developed into various methods based on a variety of hypotheses. On the achromatic side, the gray world assumption is that the maximum or mean of an image's three RGB color components is achromatic [7], [8]. With this gray world assumption, once the algorithms are given images, they use their color information and perform color constancy adjustment. The algorithms are intended to deliver neural gray when calculating the average of the colors of the color-balanced image. When the approach is first published, it is recognized as groundbreaking color balancing. However, the approach has to take many complicated steps to extract many different features. Its major drawback is that this approach is solely applicable to a uniform illumination condition. In other words, color variations in the given images profoundly affect the performance of the method, which appears in the color-corrected images in a way that colors tend to be skewed towards a dominant color [7].

Based on the Retinex theory, the Max-RGB method assumes that white surface of an image possibly fits into illuminant chromaticity. This method predicts the illuminant by using the chromaticity of the three maximum RGB color values of the image [8]. The Max-RGB method has advanced, as compared with the gray-world method, in the sense that it uses an increasing amount of color information for color balancing. Yet the approach sometimes fails to meet the Retinex theory requirements depending on input image data types. When there is an influence of input data image on the color balancing, its corrected colors appear skewed toward a dominant color of the image.

The next advanced method is the Max-RGB or White Patch method which adopts the subsampling process in ref. [9]. This method uses an arbitrary selection of pixels, not all pixels of an image, to compute the chromaticity of the maximum RGB values. The Shades of Gray (SoG) method assumes that when image data undergoes Minkowski norm, the resultant generalized or power mean is an achromatic color [10]. The authors of this article show that their method produces the best performance when using Minkowski norm with $p = 6$. The Gray Edge (GE) in ref. [11] hypothesizes that the mean absolute deviation of red, green and blue color components is an achromatic color. This method is a combined technique of SoG, Max-RGB and GW (Gray World), and uses the edge information of objects to achieve color constancy. To remove noise, this approach uses an LPF (Low Pass Filter) prior to edge detection. Consequently, the approach delivers increased robustness against noise, but adding one more step comes with a drop in efficiency. The last example of statistic-based methods is the Weighted GE proposed in ref. [12]. This technique extracts the edge information of all different objects across the image based on color correction factors.

Next, the gamut mapping method is introduced by Forsyth based on the hypothesis that an image represents only a few perceived colors due to the illuminant impact [13] and carries color variation because of the changing colors of the light sources. In an image, the recorded colors of objects can differ from their actual colors, depending on given illuminants under which they are captured. The gamut mapping approach captures an image under an unknown illuminant and uses the predetermined canonical gamut to accomplish color constancy. The gamut mapping-based method delivers superior performance to the statistics-based methods in most cases. However, this technique has a drawback of high computational cost. Finlayson and his co-authors extend into 2D gamut mapping approach to reduce the complexity of implementation. The authors also postulate that the gamut mapping approach is enabled to use the chromatic color space [14]. Finlayson and Hordley [15] propose 3D gamut mapping which performs slightly better than the earlier 2D gamut mapping. The authors use convex programming in implementing the 3D gamut mapping method to enhance efficiency, but find that the result is not much different from that of the original method in ref. [16]. Mosley and Funt [17] simplify the gamut mapping approach by replacing the convex hull of the pixel values with a simple cube. Another gamut mapping approach is proposed by Gijsenij *et al.* [18]. The authors come up with the diagonal-offset model to avoid the failure of the diagonal model [19]. They unveil a number of extended gamut mapping versions by combining various n-jet-based gamut mapping methods. In result, they reveal that the gamut mapping achieves the best illuminant estimation with the use of a possible intersection of the gamut maps. Gijsenij and Gevers [20] also propose a gamut mapping-based color constancy approach that displays the best performance with a certain image dataset that they experiment with. The method computes the Weibull parameters such as grain size and contrast, and uses the outcomes in extracting image characteristics. The authors apply a MoG classifier to their algorithms which are then able to learn weighting and correlation between Weibull parameters and diverse image attributes such as textures, edges and SNR. They also come up with a selection of parameters, which contributes to performing the best color constancy. Their method makes a 20% improvement in estimation performance, as compared with its best conventional counterparts.

The physics-based method is developed based on the dichromatic reflection of the image formation model. Physics-based algorithms factor into physical interactions between the illuminant and an object in the given image. With the hypothesis that all the pixels of a surface align to the RGB color space, the method estimates the illumination color by mapping diverse surfaces onto various planes and causing those planes to interact with one another. However, the physics-based techniques find difficulty in retrieving specular reflections and sometimes cause color clipping, affecting their performance [21]–[23]. Finlayson and Schaefer [24] propose a new method which builds on the dichromatic reflection model to project surface pixels. In this approach, the authors employ the Planckian locus of black-body radiators, which is used to generate possible illuminants.

In the color constancy technology, the learning-based algorithms use diverse machine learning methods to perform the

illuminant estimation [25]–[29]. Baron [25] redefines the color constancy as a localization issue in the log-chromaticity color space. The method is designed to scale the color channels of an image and translate the image into the log chromaticity histogram. Then, the resulting histogram is used as input data in the convolutional neural network (CNN). Bianco *et al.* [26] propose a CNN-based illuminant estimation method with the use of a histogram stretching technique. Their method is made up of a convolutional layer, a fully connected neural network (FCNN) layer and three output nodes. At its last hidden layer, the algorithm uses a histogram stretching technique to adjust the contrast of an image and extract activation values from the adjusted image. Then, the algorithm combines the activation values to predict the illuminant. The authors report convincing, experimental results with a raw image dataset, not with widely benchmarked image datasets. Another CNN-based method in ref. [27] puts Minkowski pooling in the color constancy context to facilitate and promote deep learning of the network. In consequence, the FCNN creates and uses reliable features to perform color balancing. Drew *et al.* [28] report a color constancy approach which uses a log-relative chromaticity space named ''Zeta Image'' and generates better performance than other unsupervised techniques. This supervised technique does not require a process of tuning the parameters of the dataset used in training, in comparison with its unsupervised counterparts which need no training dataset and carry out complex procedures. However, when it comes to estimation accuracy, the two techniques are not much different. Joze and Drew [29] propose an exemplar-based approach which uses the training dataset made up of the weak color constant red, green and blue values and the texture features, and predict the local illuminant by picking the neighboring surfaces in the dataset. Xu *et al.* [30] propose a global-based video enhancement algorithm that significantly increases visual experience of diverse regions of interests (RoI) in images and video frames. Their method explores the features of diverse ROIs to create a global tone mapping curve across the image. They effect immediate and appropriate visual enhancement in various regions of the processed images. Chen *et al.* [31] put forward an intra-and-inter-constraint-based (ACECB) video quality enhancement technique. This method is designed to figure out RoIs in the video frames by applying the AdaBoost-based object detection method. Upon getting the resulting RoIs, the method calculates their mean and standard deviations, analyzes the features of diverse RoIs and create a global piecewise tone mapping curve across the frame. The authors demonstrate that the technique substantially enhances the visual quality of different RoIs within an image to an appropriate level.

The biological color constancy method is intended to mimic and apply the functional characteristics of the human visual system (HVS) to perform the learning-based illuminant estimation, and several models have been presented [32]–[34]. Gao *et al.* [32] propose an HVS-based color rendering approach by taking after the interaction between the single-opponent (SO) cells in the retina, the double-opponent (DO) cells and any possible neurons in the human visual cortexes. The authors display that their framework brings about more advanced outcomes in comparison to their conventional counterparts. In addition, this method does not need to fine-tune a variety of datasets separately. Zhang *et al.* [33] propose a computational color constancy framework which reflects the likeness of the HVS. The proposed framework is designed to resemble the color processing mechanism of the retina. It means that this model eliminates repercussions of the scene illuminant on the visual perception instantly and automatically without going through the illuminant estimation. They report competitive performance relative to their state-of-the-art conventional counterparts. Akbarnia and Parraga [34] suggest a color constancy model which causes multiple asymmetric Gaussian kernels to overlap. The authors in this model intend to strengthen the intensity of a given image by making use of the contrast of the surrounding pixels, i.e. scaling kernels to the changing single-neuron visual receptive field sizes. This model uses the outputs of the most activated receptive fields in order to predict the illuminant. As outlined above, diverse color constancy approaches have been proposed in efforts to achieve the color balancing of images captured under non-white illumination condition. The approaches perform reasonably well at the fundamental illumination conditions. Nevertheless, because of their dataset dependency, their performance has some limitations. For example, when an image has a constant color region, the overall color in the color-corrected image tends to be skewed towards the dominant color.

In this respect, this article presents novel illumination estimation DCNN architecture with the use of the Convoluted Mixture of Deep Expert (CMoDE) which dynamically fuses multiple spectra and modalities. The proposed CMoDE has three parts: the experts which align modalities of spectra domain with outputs of illuminant estimation, the CMoDE which learns probability distributions and accordingly adapts weights to class-specific features of expert networks, and the fusion segment where learning of complementary fused kernels takes place. Traditionally, single DCNN-based conventional models use an easy and straightforward method of stacking multiple layers linearly, but they come with two problems: gradient distortion and overfitting. In contrast, conventional residual networks use identity skip connection and concatenate multiple paths. Accordingly, they go shallow, improve efficiency and solves overfitting, while maintaining performance comparable with that of the previous single DCNN-based models. However, shallow residual networks tend to preserve gradient flow throughout the full depth of the networks, thereby failing to help resolve the vanishing gradient problem. In order to address unresolved problems from earlier, conventional works, the proposed method is meant to stay deeper by adding layers both in series and in parallel, as well as improve efficiency by concatenating a selection of effective paths. In this way, the proposed network promises to bring much progress towards accuracy and efficiency alike. As expected, the experimental results

establish that the proposed network exceeds its state-of-the-art conventional counterparts. With lowered gradient distortion and overfitting, the proposed network is unlikely to make large estimation errors. Color constancy aside, the proposed network is further applicable to other vision problems such as global estimate which responds to aggregated local inference. To the best of our knowledge, this study is new and first in the color constancy approach to build the Deep Experts (CMoDE) fusion technique into scene illuminant estimation. When it comes to the novelty of this study, the CMoDE fusion technique stands in stark contrast to previous works by navigating problems of shallow networks. The proposed CMoDE fusion technique has also a distinct merit of extracting features by region across the image to predict scene illuminant locally, as opposed to its conventional counterparts which extract features from the whole image, perform statistics and predict the illuminant globally. Moreover, as there have been only a few attempts to predict spatially changing illuminants, the proposed architecture, as one of those rare approaches to predicting regionally changing illuminants, is of great importance. The residual network-based CMoDE fusion technique promises to take illumination estimation to a higher level of accuracy performance. The experimental results support that the proposed architecture exceeds other latest approaches in predicting illumination, with a sharp decline in critical estimation errors. Now that the proposed scheme has the distinct merit of aggregating local estimates to produce global estimate, it carries the potential for being applicable to further computer vision matters.

This article has the following major contributions:

- Proposing novel end-to-end illuminant estimation network architecture based on the residual network framework with dilated convolutions in it
- Comparing the proposed expert network with the benchmarked illumination estimation approaches
- Introducing the CMoDE fusion scheme where the network learns complementary modalities and spectra from robust kernels

II. TECHNICAL APPOACH

A. COLOR CONSTANCY APPROACH

For a Lambertian surface, an image has the intensity value of red, green and blue components $f_c(x) = \{f_R, f_G, f_B\}$ defined by a function of the light source e(λ), the surface reflection s(x, λ) and the camera sensitivity function c(λ):

$$f_c(x) = \int_\omega e(\lambda)\, s(x,\lambda)\, c(\lambda)\, d\lambda; \quad c \in \{R, G, B\}, \tag{1}$$

where ω refers to the visible spectrum of the wavelength λ and x is the spatial coordinate of the captured image. As elaborated in the Introduction section, an image represents the recorded color of a light source with the combined influence of the color of a light source e(λ) and the camera sensitivity function c(λ). The color of the light source is defined by the following equation:

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

The diagonal model is used to indicate changes in light source conditions, especially their colors, as follows [20], [35], [36]:

$$\begin{pmatrix} R_u \\ G_u \\ B_u \end{pmatrix} = \begin{pmatrix} e_R & 0 & 0 \\ 0 & e_G & 0 \\ 0 & 0 & e_B \end{pmatrix} \begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix}, \tag{3}$$

where $\{R_u, G_u, B_u\}$ refers to the recorded color of an unknown light source under which an image is captured. $\{e_R, e_G, e_B\}$ is the outcome of predicting the recorded color of the unknown light source. $\{R_c, G_c, B_c\}$ is the transformed color of a canonical light source. The proposed method uses the diagonal model and the perfect white, $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)^T$, as the canonical illuminant.

B. THE PROPOSED MULTI-STREAM DEEP NEURAL NETWORK

When it comes to the DCNN, a straightforward way of improving performance is merely increasing the size of the network architecture, both in depth and in width. As long as the network is able to cope with a large amount of labeled training dataset, it is an easy and secure way to extend the network with a growing number of layers in parallel and in series in order to develop high-quality models. Of course, it requires much effort, time and cost to collect and provide massive labeled training datasets for the network. When the network expands in size and in the number of parameters, an overfitting problem arises. This is especially true when the network is given a limited amount of labeled training data. It is again costly, challenging and time-consuming for professional evaluators to prepare and classify strong labeled datasets. It is interesting to note that when the network increases in size, its computational resource requirements expands exponentially and enormously. When a multiple-convolutional layer DCNN sees a uniform increase in the number of filters, the network is required to cope with an exponentially expanding amount of computation. If the network fails to make efficient use of the expanded capacity such as near-zero weight, the network could end up with wasting computation efforts. Given the fact that the computational resources are always limited, efficient resource allocation comes ahead of indiscriminate increase in size [37].

To address the aforementioned issues of the fully connected DCNN approaches such as AlexNet and Vgg16 network, the proposed method put forwards a residual network-based color constancy approach. A residual network, as one of DCNNs, has each layer made up of a residual module f_i and a skip connection that bypasses the residual module. The residual network can improve performance with an increasing number of layers [38], [39]. A residual network is also referred to as a residual block in the remainder of this

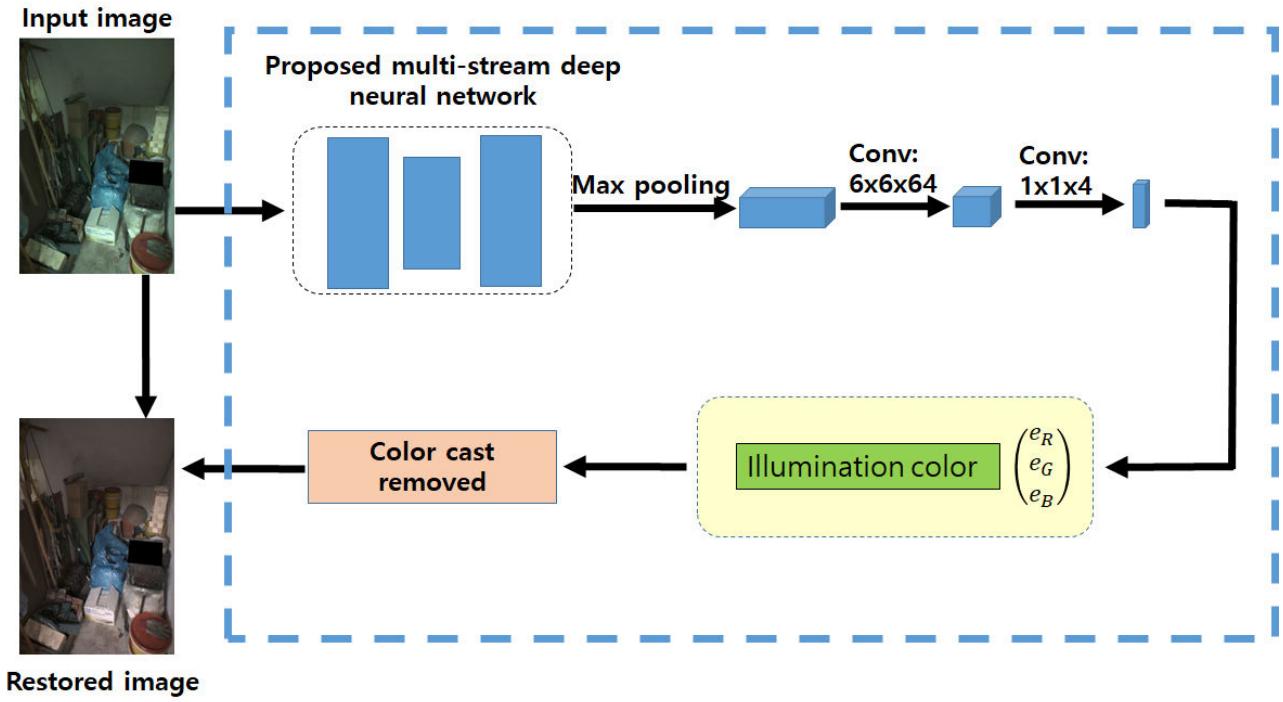


FIGURE 1. The block diagram of the entire stages of the proposed illuminant estimation method.

paper now that the residual network has a large number of convolutional layers. As the input data, y_{i-1}, the recursive output of the i*th* residual block is described below [40]:

$$y_i \equiv f_i(y_{i-1}) + y_{i-1}, \tag{4}$$

where $f_i(x)$ consists of a series of convolutions, batch normalization, $B(x)$ [41], and nonlinear Rectified Linear Units (ReLU), $\sigma(x) \equiv \max(x, 0)$. In the latest formulation of residual block [38], $f_i(x)$ is defined as follows:

$$f_i(x) \equiv W_i \cdot \sigma(B(W_i' \cdot \sigma(B(x)))), \tag{5}$$

where W_i and W_i' are weight matrices and convolution is denoted by a dot, $\cdot$.

Residual blocks in Eq. (5) are combined to form a network structure, described as follows:

$$y_3 = y_2 + f_3(y_2) \tag{6}$$

$$= [y_1 + f_2(y_1)] + f_3(y_1 + f_2(y_1)) \tag{7}$$

$$= [y_0 + f_1(y_0) + f_2(y_0 + f_1(y_0))] + f_3(y_0 + f_1(y_0) + f_2(y_0 + f_1(y_0))) \tag{8}$$

According to the above equations, the input of each layer is concatenated with the output of the previous layer. However, residual blocks do not exactly follow the above-described pattern due to its inherent nature of the structure. Instead, a residual block provides data for each module $f_i(\cdot)$ by generating a mixture of 2^{i-1} different distributions from every possible configuration of the previous i − 1 residual block.

Fig. 1 is the block diagram of the entire stages of the proposed illuminant estimation method. Estimation accuracy is of vital importance in achieving and improving color constancy performance. For this purpose, the proposed method adopts the CMoDE fusion technique to achieve estimation accuracy of the local illumination. This technique has a distinct merit of selecting and concatenating effective paths, which therefore allows the network to go shallow. This is why the use of CMoDE fusion technique contributes to increasing the estimation accuracy of the proposed network and it is proven in the Experimental Results and Evaluations section. Fig. 2 depicts the proposed residual blocks of Eq. (6). The proposed residual blocks have batch normalization and multiple layers with skip connections. In the proposed residual blocks, the layers are able to select and concatenate effective paths, which allows the networks to go deeper while avoiding gradient degradation. Therefore, the networks can cover large-size receptive fields which often carry highly identifiable features. The key driver behind the adoption of the proposed CMoDE fusion technique is to achieve higher accuracy of predicting the local illumination, which is crucially important for the network to learn to optimize the combination of local estimates and deliver the best performance eventually.

III. EXPERIMENTAL RESULTS AND EVALUATIONS

This section discusses the comparative studies and experimental results between the proposed color constancy method and latest comparable methods. The experiments are conducted with the use of Gray-ball Dataset, and Gehler and Shi's Dataset. For those benchmark illuminant image datasets,

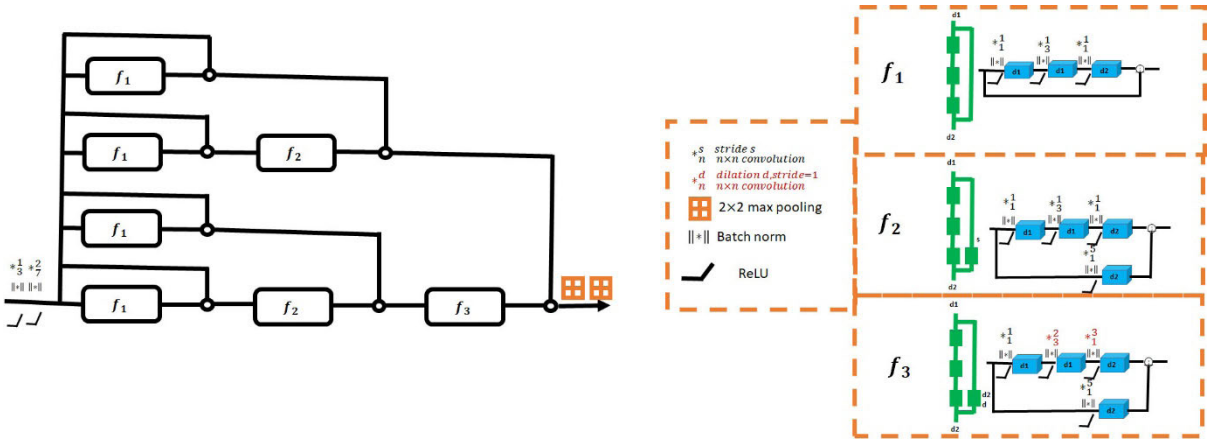
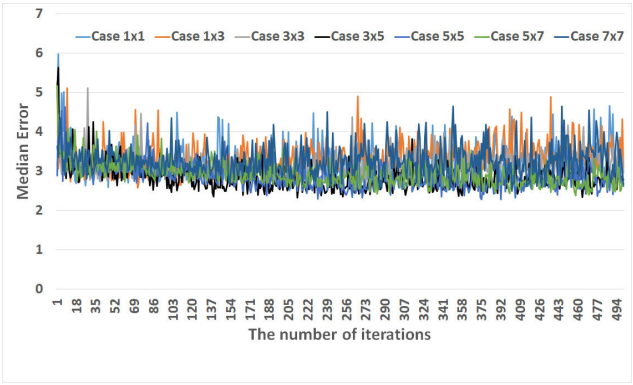
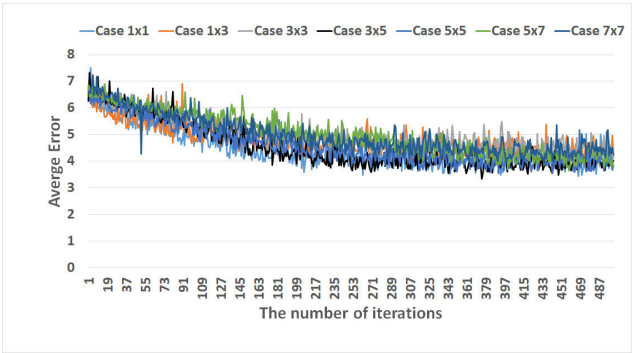


FIGURE 2. The proposed multi-stream deep convolutional neural network with atrous convolution, or called dilated convolution [42]: the atrous convolution indicated in red and the convolution layer in black.



(a)



(b)

FIGURE 3. The experimental results of comparing different filter kernel sizes in terms of (a) median angular error and (b) average angular error.

the images are taken under multiple illumination conditions such as a single illuminant and a combination of several illuminants in order to verify illuminant invariant. For the color reproducibility purpose, Gray-ball Dataset, and Gehler and Shi's Dataset are obtained by placing a gray ball and a Macbeth color checker chart, respectively, in front of video cameras. The Gray-ball Dataset [43] consists of 11,340 different images in 360 × 240 pixels. The Gehler and Shi's Dataset [44], [46] consists of 568 indoor and outdoor scenes

TABLE 1. Cases of different filter kernel sizes of the residual blocks f_1, f_2 and f_3 in Eq. (9).

	f_1	f_2	f_3
Case_1x1	1x1, 1x1, 1x1	1x1, 1x1, 1x1, 1x1	1x1, 1x1, 1x1, 1x1
Case_1x3	1x1, 3x3, 1x1	1x1, 3x3, 1x1, 1x1	1x1, 3x3, 1x1, 1x1
Case_3x3	3x3, 3x3, 3x3	3x3, 3x3, 3x3, 3x3	3x3, 3x3, 3x3, 3x3
Case_3x5	3x3, 5x5, 3x3	3x3, 5x5, 3x3, 3x3	3x3, 5x5, 3x3, 3x3
Case_5x5	5x5, 5x5, 5x5	5x5, 5x5, 5x5, 5x5	5x5, 5x5, 5x5, 5x5
Case_5x7	5x5, 7x7, 5x5	5x5, 7x7, 5x5, 5x5	5x5, 7x7, 5x5, 5x5
Case_7x7	7x7, 7x7, 7x7	7x7, 7x7, 7x7, 7x7	7x7, 7x7, 7x7, 7x7

including people, places and objects. For the experiments, those dataset images are firstly resized in a smaller dimension to 512 × 512 pixels and randomly cropped into 224 × 224 image patches. Those patches are used to train the proposed network model in an end-to-end fashion with back-propagation.

In designing network architecture, one of recent trends is using smaller-size filter kernels which takes the network to a higher level of efficiency. On the one hand, small filter kernels give the network the advantage of having a decreasing number of parameters, facilitating faster and more efficient performance, but small filter kernels come with a possible drop in accuracy of the network, on the other hand. That is, the size of the filter kernel carries a trade-off between efficiency and accuracy. In this work, the experiments are meant to find out and verify the optimal size of the filter kernel in terms of efficiency and accuracy alike. As in Table 1, the experiments test seven different filter kernel sizes on the residual blocks f_1, f_2 *and* f_3 in Eq. (9). Figure 3 illustrates the experimental results of comparing 7 different filter kernel sizes from the angular error perspective. The experiments have set the numbers of filter kernels or channels at 32, 64, 128 and 256.

Figure 3 (a) compares the median angular errors of seven different filter kernel sizes. Experiments in each case run 10K epochs and the median angular errors are recorded every 20 epochs. The experiments use Geforce TITEN XP GPU and

TABLE 2. Comparative evaluation between the proposed method and various conventional methods with Shi's Reprocessed Dataset (The lower, the better).

Methods	Mean	Median	Trimean	Best-25%	Worst-25%
Statistics-Based Methods					
White-Patch[8]	7.55	5.68	6.35	1.42	16.12
Gray-world[7]	6.36	6.28	6.28	2.33	10.58
1st-order GE[11]	5.33	4.52	4.73	1.86	10.03
2nd-order GE[11]	5.13	4.44	4.62	2.11	9.26
SoG Method[10]	4.93	4.01	4.23	1.14	10.2
General-Grey world[11]	4.66	3.48	3.81	1.00	10.09
Modifies White-patch[48]	3.87	2.84	3.15	0.92	8.38
Local Surface Reflectance[49]	3.31	2.80	2.87	1.14	6.39
CCATI Method[58]	2.34	1.60	1.91	0.49	5.28
Learning-Based Methods					
SVR Regression Method[50]	8.08	6.73	7.19	3.35	14.89
Edge-based Gamut[18]	6.52	5.04	5.43	1.90	13.58
Bayesian Method[45]	4.82	3.46	3.88	1.26	10.46
Natural Image Statistics[20]	4.19	3.13	3.45	1.00	9.22
Intersection-based Gamut[50]	4.20	2.39	2.93	0.51	10.70
CART-based combination[51]	3.90	2.91	3.21	1.02	8.27
Spatio-spectral[52]	3.59	2.96	3.10	0.95	7.61
EM-based Method[29]	2.89	2.27	2.42	0.82	5.97
19-Edge Corrected-moment[53]	2.86	2.04	2.22	0.70	6.34
CNN based Method[54]	2.75	1.99	2.14	0.74	6.05
ED-based Method[55]	2.42	1.65	1.75	0.38	5.87
H. Zhan et al [59]	2.29	1.90	2.03	0.57	4.72
DS-Net[57]	2.24	1.46	1.68	0.48	6.08
SqueezeNet-FC4[56]	2.23	1.57	1.72	0.47	5.15
AlexNet-FC4[56]	2.12	1.53	1.64	0.48	4.78
Choi's method [60]	2.09	1.42	1.60	0.35	4.78
Proposed Network	**2.05**	**1.06**	**1.42**	**0.29**	**4.50**

take one and a half days. As Figure 3 (a) exhibits, Case_3 $\times$ 5 delivers the best performance, resulting in the lowest median angular error. Figure 3 (b) compares the average angular errors of seven different filter kernel sizes. Experiments in each case also run 10K epochs and the average angular errors are recorded every 20 epochs. As in their comparative results, Case_3 $\times$ 5 also turns out to be the optimal filter kernel size. The experiments use the error metric proposed by Hordley and Finlayson [46]. The error metric represents the angle between the red, green and blue triplets of estimated illuminant (ρ_w) and the measured ground truth illuminant ($\hat{\rho}_w$), described as follows:

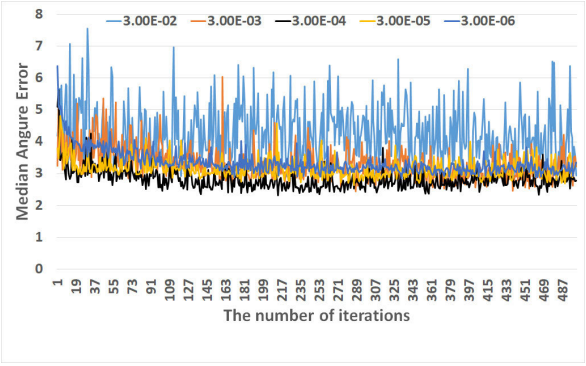
$$e_{ANG} = arccos\left(\frac{\rho_w^T \hat{\rho}_w}{\|\rho_w\| \left\|\hat{\rho}_w\right\|}\right) \tag{9}$$

As Figure 4 shows, experiments continue to test the effects of different parameters on the illumination estimation performance. Figure 4 (a) compares the median angular errors of different initial training rates. The symbol "3.00E-2" translates into 3×10^{-2}. In this experiment, the proposed DCNN makes the lowest median error at the initial learning rate of 3×10^{-4}. Figure 4 (b) is the comparative experiment of the estimation performance with the multi-stream versus without main-stream in Eq. (6). As a result, the proposed color constancy method performs better with the multi-stream. In continuing efforts to search for the optimal condition, the proposed network employs Adam [47] with a batch size of 16. To avoid overfitting, the proposed method sets parameters for all its layers, for instance, a 0.5 dropout probability for f3 convolution layer in Figure 2, a weight decay for 5×10^{-5} and a momentum of 0.9. Figure 5 plots the impact of learning rates on learning behavior or convergence, and the impact of total training loss during learning on the performance of the proposed network. In Figure 5, the experiments use Gehler and Shi's Dataset to test the proposed network. In result, the proposed network makes the lowest total training loss at the learning rate of 3×10^{-4}.

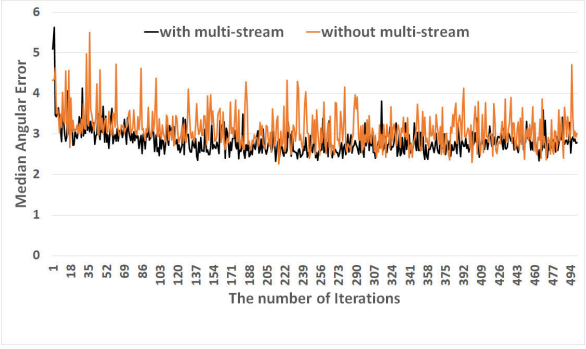
To verify the illumination invariant, the angular errors and their distributions are compared between the proposed network and the latest 28 diverse approaches. In the comparative experiment, the proposed network is set at the optimal parameter conditions based on the experiments above.

TABLE 3. Comparative evaluation between the proposed method and various conventional methods with Gray-Ball Dataset (The lower, the better).

Method	Mean	Median	Trimean	Best-25%	Worst-25%
Statistics-Based Methods					
White-Patch	6.80	5.30	5.77	1.18	14.73
Grey-world	7.87	6.97	7.14	2.16	15.25
1st-order GE	5.91	4.72	5.10	-	-
2nd-order GE	6.12	4.93	5.30	-	-
SoG Method	6.11	5.36	5.55	-	-
General Grey world	6.14	5.36	5.57	-	-
Local Surface Reflectance	6.00	5.13	-	-	-
CCATI Method	3.80	3.80	-	-	-
Learning-Based Method					
SVR-Regression Method	13.17	11.28	11.83	4.42	25.02
Bayesian Method	6.77	4.70	5.00	-	-
Natural Image Statistics	5.24	3.00	4.35	1.21	11.15
EM-based Method	4.42	3.48	3.77	1.01	9.36
CNN-based Method	4.80	3.70	-	-	-
Choi's method	4.03	1.88	2.60	0.61	10.77
Proposed Network	**3.28**	**1.70**	**2.27**	**0.53**	**6.69**



(a)



(b)

FIGURE 4. The experimental results of (a) comparing different initial training rates from the median angular error perspective and (b) comparing median errors with multi-stream versus without multi-stream in Eq. (6).

The 28 different methods consist of both unitary and combinational methods: Gray-world, WP, SoG, 1st- and 2nd-order Gray Edge and CNN-based methods. Their source codes are conducted by matlab and tensorflow [47] and their parameter conditions are set as suggested in their respective articles. Table 2 compares the 28 methods and the proposed method from the angular error perspective: mean, median, trimean, best-25% and worst-25% and the comparative study uses Gehler and Shi's Dataset of real-world scene images mostly. Among the 28 methods, the CNN-based method [54] proposed by S. Bianco et al uses a specially designed CNN for estimating multiple local illuminates. This method uses a multi-illuminant detector to decide whether or not it is necessary to aggregate local outputs to produce a single estimate. More recently, Choi *et al.* [60] propose another color constancy approach based on the residual network architecture to overcome several problems with Alexnet-FC4 [56] such as overfitting, gradient degradation and gradient vanishing and the method delivers advanced performance. Yet this method is designed to stack the residual blocks with skip connection in depth only, not in width. For this reason, the approach still has room to improve computational efficiency, as well as take estimation accuracy to a higher level, even though Choi's method performs higher accuracy than Alexnet-FC4. In order to improve both efficiency and accuracy, this article proposes novel CMoDE fusion technique-based CNNs. The proposed CMoDE fusion technique has the primary advantage of taking local illumination estimation to a higher level of accuracy, which is of vital importance in training the network to achieve the optimal combination of the local estimates and contributing to the final performance. In detail, the multiple paths in the proposed architecture are trained all at once, but they function individually and independently from one another, which gives the network the flexibility to select and concatenate effective paths alone. The select valid paths engage

TABLE 4. Comparative evaluation between the proposed network (PN) and Grey World (GW), White Patch (WP) 4, Shades of Grey (SoG), General Grey World (GGW), 1st-order Grey Edge (GE1), 2nd-order Grey Edge (GE2), Local Surface Reflectance Statistics (LSR), Pixels-based Gamut (PG), Bayesian framework (BF), Spatio-spectral Statistics (SS), Natural Image Statistics (NIS) and Choi'method (CM) with NUS-8 Camera Dataset.

	Statistics-Based							Learning-Based					
Method	**GW**	**WP**	**SoG**	**GGW**	**GE1**	**GE2**	**LSR**	**PG**	**BF**	**SS**	**NIS**	**CM**	**PN**
Camera	Mean Angular Error												
Canon1Ds	5.16	7.99	3.81	3.16	3.45	3.47	3.43	6.13	3.58	3.21	4.18	3.18	**3.05**
Canon600D	3.89	10.96	3.23	3.24	3.22	3.21	3.59	14.51	3.29	2.67	3.43	2.35	**2.21**
FujiXM1	4.16	10.2	3.56	3.42	3.13	3.12	3.31	8.59	3.98	2.99	4.05	3.10	**2.95**
NikonD5200	4.38	11.64	3.45	3.26	3.37	3.47	3.68	10.14	3.97	3.15	4.10	2.35	**2.23**
OlympEPL6	3.44	9.78	3.16	3.08	3.02	2.84	3.22	6.52	3.75	2.86	3.22	2.47	**2.32**
LumixGX1	3.82	13.41	3.22	3.12	2.99	2.99	3.36	6.00	3.41	2.85	3.70	2.46	**2.34**
SamNX2000	3.90	11.97	3.17	3.22	3.09	3.18	3.84	7.74	3.98	2.94	3.66	2.32	**2.18**
SonyA57	4.59	9.91	3.67	3.20	3.35	3.36	3.45	5.27	3.50	3.06	3.45	2.33	**2.21**
Camera	Median Angular Error												
Canon1Ds	4.15	6.19	2.73	2.35	2.48	2.44	2.51	4.30	2.80	2.67	3.04	2.71	**2.43**
Canon600D	2.88	12.44	2.58	2.28	2.07	2.29	2.72	14.83	2.35	2.03	2.46	2.19	**2.05**
FujiXM1	3.30	10.59	2.81	2.60	1.99	2.00	2.48	8.87	3.20	2.45	2.96	2.82	**2.55**
NikonD5200	3.39	11.67	2.56	2.31	2.22	2.19	2.83	10.32	3.10	2.26	2.40	1.92	**1.66**
OlympEPL6	2.58	9.50	2.42	2.18	2.11	2.18	2.49	4.39	2.81	2.24	2.17	2.12	**1.81**
LumixGX1	3.06	18.00	2.30	2.23	2.16	2.04	2.48	4.74	2.41	2.22	2.28	1.42	**1.25**
SamNX2000	3.00	12.99	2.33	2.57	2.23	2.32	2.90	7.91	3.00	2.29	2.77	1.32	**1.15**
SonyA57	3.46	7.44	2.94	2.56	2.58	2.70	2.51	4.26	2.36	2.58	2.88	1.65	**1.53**
Camera	Tri-mean Error												
Canon1Ds	4.46	6.98	3.06	2.50	2.74	2.70	2.81	4.81	2.97	2.79	3.30	2.69	**2.52**
Canon600D	3.07	11.40	2.63	2.41	2.36	2.37	2.95	14.78	2.40	2.18	2.72	2.33	**2.21**
FujiXM1	3.40	10.25	2.93	2.72	2.26	2.27	2.65	8.64	3.33	2.55	3.06	2.88	**2.74**
NikonD5200	3.59	11.53	2.74	2.49	2.52	2.58	3.03	10.25	3.36	2.49	2.77	1.95	**1.90**
OlympEPL6	2.73	9.54	2.59	2.35	2.26	2.20	2.59	4.79	3.00	2.28	2.42	2.18	**1.95**
LumixGX1	3.15	14.98	2.48	2.45	2.25	2.26	2.78	4.98	2.58	2.37	2.67	1.81	**1.64**
SamNX2000	3.15	12.45	2.45	2.66	2.32	2.41	3.24	7.70	3.27	2.44	2.94	1.65	**1.51**
SonyA57	3.81	8.78	3.03	2.68	2.76	2.80	2.70	4.45	2.57	2.74	2.95	1.91	**1.75**
Camera	Mean of Best 25%												
Canon1Ds	0.95	1.56	0.66	0.64	0.81	0.86	1.06	1.05	0.76	0.88	0.78	0.65	**0.52**
Canon600D	0.83	2.03	0.64	0.63	0.73	0.80	1.17	9.98	0.69	0.68	0.78	0.73	**0.59**
FujiXM1	0.91	1.82	0.87	0.73	0.72	0.70	0.99	3.44	0.93	0.81	0.86	0.75	**0.81**
NikonD5200	0.92	1.77	0.72	0.63	0.79	0.73	1.16	4.35	0.92	0.86	0.74	0.57	**0.43**
OlympEPL6	0.85	1.65	0.76	0.72	0.65	0.71	1.15	1.42	0.91	0.78	0.76	0.80	**0.75**
LumixGX1	0.82	2.25	0.78	0.70	0.56	0.61	0.82	2.06	0.68	0.82	0.79	0.65	**0.53**
SamNX2000	0.81	2.59	0.78	0.77	0.71	0.74	1.26	2.65	0.93	0.75	0.75	0.53	**0.39**
SonyA57	1.16	1.44	0.98	0.85	0.79	0.89	0.98	1.28	0.78	0.87	0.83	0.57	**0.42**
Camera	Mean of Worst 25%												
Canon1Ds	11.00	16.75	8.52	7.08	7.69	7.76	7.3	14.16	7.95	6.43	9.51	6.67	**6.53**
Canon600D	8.53	18.75	7.06	7.58	7.48	7.41	7.4	18.45	7.93	5.77	5.76	5.29	**5.19**
FujiXM1	9.04	18.26	7.55	7.62	7.32	7.23	7.06	13.4	8.82	5.99	9.37	5.64	**5.53**
NikonD5200	9.69	21.89	7.69	7.53	8.42	8.21	7.57	15.93	8.18	6.90	10.01	4.86	**4.72**
OlympEPL6	7.41	18.58	6.78	6.69	6.88	6.47	6.55	15.42	8.19	6.14	7.46	4.62	**4.49**
LumixGX1	8.45	20.4	7.12	6.86	7.03	6.86	7.42	12.19	8.00	5.90	8.74	5.74	**5.55**
SamNX2000	8.51	20.23	6.92	6.85	7.00	7.23	7.98	13.01	8.62	6.22	8.16	5.55	**5.39**
SonyA57	9.85	21.27	7.75	6.68	7.18	7.14	7.32	11.16	8.02	6.17	7.18	5.12	**4.95**

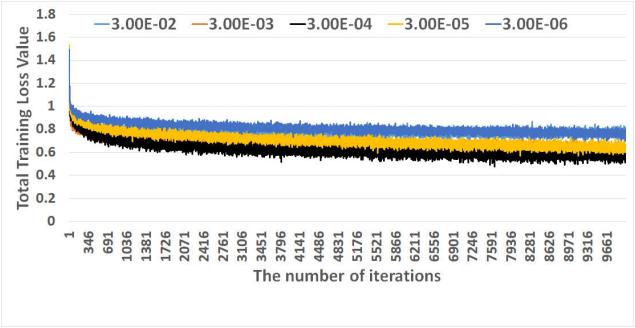


FIGURE 5. Comparative evaluation of the total training loss in the logarithm space with four different initial learning rates to decide on the optimal learning rate.

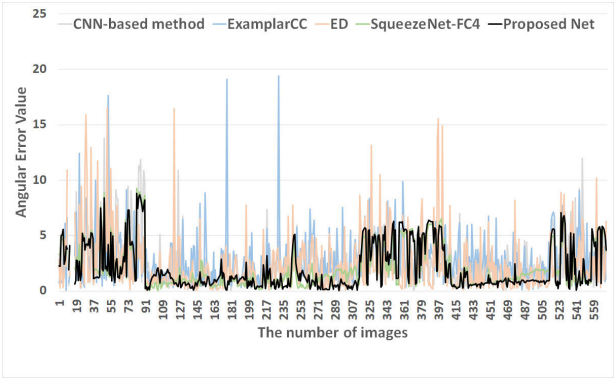


FIGURE 6. The angular error distribution of 'real-world scene' Gehler and Shi's Dataset images.

in implementing the network smoothly. So, despite a large number of layers in the proposed network, they go shallow through the selective use of effective paths only. Therefore, the proposed architecture is designed to improve accuracy and computational efficiency alike. In result, the proposed network makes the lowest angular errors. Next, in order to compare angular error distributions, several high-performing methods such as CNN, ExamplarCC, ED and SqueezeNet-FC4 are picked from Table 2 and compared with the proposed network. Figure 6 exhibits the comparative results.

To verify illumination invariant, another comparative experiment is conducted between the proposed network and 14 different methods, with the use of Grey-ball Dataset and in the aspects of the mean, median and trimean angular error. Table 3 summarizes the experimental outcomes. As a result, the proposed method produces lower angular errors than the 14 different methods.

To verify camera invariant, the proposed network is assessed with the use of the latest and well-known color constancy dataset, the NUS-8 camera image dataset [61]. The dataset includes total 1,736 images resulting from taking pictures of 210 individual scenes with eight cameras. In the experiment with the NUS-8 camera dataset, the proposed network is compared to 13 conventional methods. Table 4 illustrates the comparison results of camera-invariant performance between the 13 methods and the proposed network. The proposed network proves to exhibit robust performance regardless of the camera conditions by exceeding the 13 comparison methods.

IV. CONCLUSION

This article presents a deep learning-based computational color constancy method which trains the CMoDE fusion technique-based CNNs to estimate real illuminant colors. To highlight the distinctive merits of the proposed method, first, it organizes the CMoDE fusion technique-based CNNs in series and in parallel alike. This architecture gives the network to enhance accuracy and efficiency at the same time. The multiple paths in the CMoDE fusion technique function individually and independently one another which enables the network to choose effective paths. Consequentially, the residual blocks of the proposed method help avoid overfitting and make efficient computation, as opposed to a single ultra-deep network whose structure is vulnerable to overfitting and computational efficiency deterioration. Second, the select effective paths are simultaneously trained to estimate illumination, which also contributes to efficiency enhancement and smooth implementation of the network. Finally, the proposed network has multiple layers, but the bottom line is that it goes shallow by using a selection of effective paths only in the end. While a large number of paths are the potential source of the gradient, the proposed network uses shorter-than-expected paths. Accordingly, the proposed method can steeply reduce the training time cost. Further, to verify illumination and camera invariants, the proposed network is compared with various advanced methods, using multiple datasets such as Gehler and Shi's, Gray-ball and NUS-8 camera datasets. The experimental results demonstrate that our proposed network surpasses its comparable counterparts. Notwithstanding, there is still much to be done and this study will continue towards optimizing the structure of CNNs for color constancy.

REFERENCES

[1] M. Ebner, *Color Constancy*. Hoboken, NJ, USA: Wiley, 2007.
[2] Y. Qian, K. Chen, J.-K. Kamarainen, J. Nikkanen, and J. Matas, "Deep structured-output regression learning for computational color constancy," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Cancún, Mexico, Dec. 2016, pp. 1899–1904.
[3] J. Simao, H. J. A. Schneebeli, and R. F. Vassallo, "An iterative approach for color constancy," in *Proc. Joint Conf. Robot., SBR-LARS Robot. Symp. Robocontrol*, São Carlos, Brazil, Oct. 2014, pp. 130–135.
[4] 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.
[5] G. Shacter, S. Hondley, and G. Finlayson, "A combined physical and statistical approach to color constancy," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 148–155.
[6] T. Jiang, D. Nguyen, and K. Kuhnert, "Auto white balance using the coincidence of chromaticity histograms," in *Proc. 8th Int. Conf. Signal Image Technol. Internet Based Syst.*, Naples, Italy, Nov. 2012, pp. 201–208.
[7] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
[8] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.
[9] N. Banić and S. Lončarić, "Improving the white patch method by subsampling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 605–609.

[10] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. IS&T/SID Color Imag. Conf.*, 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] D. A. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 5–36, 1990.
[14] G. D. Finlayson, "Color in perspective," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 1034–1038, Oct. 1996.
[15] G. Finlayson and S. Hordley, "Improving gamut mapping color constancy," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1774–1783, Oct. 2000.
[16] G. Finlayson and R. Xu, "Convex programming color constancy," in *Proc. IEEE Workshop Color Photometric Methods Comput. Vis.*, Oct. 2003, pp. 1–8.
[17] M. Mosny and B. Funt, "Cubical gamut mapping colour constancy," in *Proc. IS&T's Eur. Conf. Color Graph., Imag. Vis.*, 2010, pp. 466–470.
[18] 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, nos. 2–3, pp. 127–139, Jan. 2010.
[19] G. D. Finlayson, S. D. Hordley, and R. Xu, "Convex programming colour constancy with a diagonal-offset model," in *Proc. IEEE Int. Conf. Image Process.*, Genova, Italy, 2005, pp. 948–951.
[20] 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.
[21] S. Tominaga and B. A. Wandell, "Standard surface-reflectance model and illuminant estimation," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 6, no. 4, pp. 576–584, Apr. 1989.
[22] G. Healey, "Estimating spectral reflectance using highlights," *Image Vis. Comput.*, vol. 9, no. 5, pp. 333–337, Oct. 1991.
[23] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse intensity chromaticity space," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 21, no. 3, pp. 321–334, 2004.
[24] G. D. Finlayson and G. Schaefer, "Solving for colour constancy using a constrained dichromatic reection model," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 127–144, 2001.
[25] J. T. Barron, "Convolutional color constancy," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 379–387.
[26] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Boston, MA, USA, Jun. 2015, pp. 81–89.
[27] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Trémeau, and C. Wolf, "Mixed pooling neural networks for color constancy," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 3997–4001.
[28] M. S. Drew, H. R. V. Joze, and G. D. Finlayson, "Specularity, the zeta image, and information-theoretic illuminant estimation," in *Proc. ECCV*, Florence, Italy, 2012, pp. 411–420.
[29] 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.
[30] N. Xu, W. Lin, Y. Zhou, Y. Chen, Z. Chen, and H. Li, "A new global-based video enhancement algorithm by fusing features of multiple region-of-interests," in *Proc. Vis. Commun. Image Process. (VCIP)*, Tainan, Taiwan, Nov. 2011, pp. 1–4.
[31] Y. Chen, W. Lin, C. Zhang, Z. Chen, N. Xu, and J. Xie, "Intra-and-inter-constraint-based video enhancement based on piecewise tone mapping," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 74–82, Jan. 2013.
[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] X.-S. Zhang, S.-B. Gao, R.-X. Li, X.-Y. Du, C.-Y. Li, and Y.-J. Li, "A retinal mechanism inspired color constancy model," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1219–1232, Mar. 2016.
[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] G. D. Finlayson, M. S. Drew, and B. V. Funt, "Spectral sharpening: Sensor transformations for improved color constancy," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 11, no. 5, pp. 1552–1563, 1994.
[36] B. V. Funt and B. C. Lewis, "Diagonal versus affine transformations for color correction," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 17, no. 11, pp. 2108–2112, 2000.
[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014, pp. 1–12, *arXiv:1409.4842*. [Online]. Available: http://arxiv.org/abs/1409.4842
[38] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," 2016, *arXiv:1603.05027*. [Online]. Available: http://arxiv.org/abs/1603.05027
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*. [Online]. Available: http://arxiv.org/abs/1512.03385
[40] A. Veit, M. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," 2016, *arXiv:1605.06431*. [Online]. Available: http://arxiv.org/abs/1605.06431
[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1–11.
[42] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolution," in *Proc. ICLR*, 2016, pp. 1–13.
[43] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. 11th Color Imag. Conf. Final Program*, 2003, pp. 160–164.
[44] L. Shi and B. Funtes. (2010). Re-processed version of the Gehler color constancy datasets of 568 images. Simon Fraser University. [Online]. Available: http://www.cs.sfu.ca/~colour/data/
[45] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
[46] S. Hordley and G. Finlayson, "Re-evaluating color constancy algorithms," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 76–79.
[47] 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
[48] G. K. Kloss, "Colour constancy using von Kries transformations: Colour constancy, 'Goes to lab,'" *Res. Lett. Inf. Math. Sci.*, vol. 13, nos. 19–33, pp. 19–33, 2009.
[49] S. Gao, W. Han, K. Yang, C. Li, and Y. Li, "Efficient color constancy with local surface reflectance statistics," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 158–173.
[50] W. Xiong and B. Funt, "Estimating illumination chromaticity via support vector regression," *J. Imag. Sci. Technol.*, vol. 50, no. 4, pp. 341–348, 2006.
[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] 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.
[53] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1904–1911.
[54] 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.
[55] 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.
[56] 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.
[57] 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.
[58] M. A. Hussain, A. S. Akbari, and E. Abbott-Halpin, "Color constancy for uniform and non-uniform illuminant using image texture," *IEEE Access*, vol. 7, pp. 7294–72978, 2019.
[59] 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, Jun. 2019.
[60] 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.
[61] D. Cheng, D. K. Prasad, and M. S. Brown, "Illumination 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.

HO-HYOUNG CHOI 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 June 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 the 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.