Deep Learning-Based Illumination Estimation Using Light Source Classification

KARLO KOŠČEVIĆ, MARKO SUBAŠIĆ, AND SVEN LONČARIĆ
Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
Corresponding author: Karlo Koščević (karlo.koscevic@fer.hr)
This work was supported by the Croatian Science Foundation under Project IP-06-2016-2092.

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

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

I. INTRODUCTION
One of the first steps in the image formation pipeline of contemporary digital cameras is computational color constancy. Computational color constancy refers to the removal of the influence of illumination color on the colors of objects in the observed image scene. It is motivated by the ability of the human vision system (HVS) to perceive object color invariant to the illumination color, namely color constancy [1]. Computational color constancy is performed in two steps. The first step is the illumination estimation step, where one or multiple illumination vectors are estimated from the target image. Illumination vector is a three-component vector with one value for each color channel \( c \in \{R, G, B\} \).

In the second step, estimated illumination vectors are used to divide out the illumination and object reflectance. This step is called chromatic adaptation and it is achieved by multiplying each image pixel with a diagonal matrix with diagonal values \( d_{11} = 1/e_R, d_{22} = 1/e_G, \) and \( d_{33} = 1/e_B \), where \( \begin{bmatrix} e_R & e_G & e_B \end{bmatrix}^T \) is the illumination vector. After the chromatic adaptation is applied, the colors in the image should appear as if they are captured under the white illumination, i.e., the illumination where \( e_R = e_G = e_B \).

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

\[
  f_c(x) = \int \omega I(\lambda, x)R(\lambda, x)\rho_c(\lambda)d\lambda \tag{1}
\]

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

\[
e(x) = \int_{\omega} I(\lambda, x) \rho(\lambda) d\lambda.
\]

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

\[
e = \int_{\omega} I(\lambda) \rho(\lambda) d\lambda.
\]

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

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

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

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

II. RELATED WORK

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

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

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

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

III. MOTIVATION

When capturing an image with a digital camera, the target scene can be illuminated with many different light sources. Some of these light sources can produce illuminations similar to the white illumination, and they do not affect pixel color significantly. However, for instance, in indoor environments, it is common that the illumination color significantly differs from white, i.e., values between red, green, and blue color channels are different. Such illuminations cause considerable color bias in the image towards the color of the illumination. The difference between an image captured in a light that is close to white and an image captured in yellow light is shown in Figure 1. It can be observed that the yellow illumination in Figure 1b has a great impact on pixel colors. A similar effect can be observed with other artificial light sources, e.g., when taking a picture of an outdoor scene in the night when street lights are turned on.

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

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

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

IV. THE PROPOSED METHOD

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

![Figure 1. The difference between images captured under near-white illumination (a) and yellow illumination (b).](image-url)
is given in Algorithm 1. Both classification and illumination estimation steps are described in more detail in the following sections.

Algorithm 1 Illumination Estimation Using Light Source Classification

**Input:** image I

**Output:** illumination vector e

1. $CN = ClassificationNet()$
2. $IE_j = IlluminationEstNet_i(I), i \in \{0, 1, 2\}$
3. $(p_0, p_1, p_2) = CN.predict(I)$ \quad \triangleright \text{Class probabilities}$
4. $j = \arg \max(p_i)$
5. $e = IE_j.predict(I)$

**A. IMAGE CLASSIFICATION**

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

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

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

**B. ILLUMINATION ESTIMATION**

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

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

**V. EXPERIMENTAL RESULTS**

**A. DATASET PREPARATION**

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

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

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

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

**B. PERFORMANCE METRICS**

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

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

**C. TRAINING SETUP**

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

1) **IMAGE CLASSIFICATION**

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

2) **ILLUMINATION ESTIMATION**

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

\[
L(e, \hat{e}) = \frac{1}{N} \sum_{i=1}^{N} \left( 1 - \frac{e_i \cdot \hat{e}_i}{||e_i|| \cdot ||\hat{e}_i||} \right),
\]  

(4)
where \( i \)th ground-truth illumination vector and estimated illumination vector are denoted as \( \mathbf{e}_i \) and \( \mathbf{\hat{e}}_i \), respectively. \( N \) denotes the number of samples, \( \cdot \cdot \) is the vector dot product, and \( \| \cdot \| \) is the vector \( L_2 \) norm. In the following paragraphs, the parameters specific for each illumination estimation network are given.

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

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

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

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

### D. Method Accuracy

#### 1) Natural-artificial illumination classification

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

#### 2) Illumination estimation

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

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

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

---

**TABLE 1. The classification accuracy.**

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

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

| Class | Min | Mean | Med. | Tri. | Best 25% | Worst 25% | Avg. |
|-------|-----|------|------|-----|----------|-----------|------|
| \( C_0 \) | 0.65 | 1.81 | 1.41 | 1.47 | 0.84 | 3.36 | 4.25 | 1.60 |
| \( C_1 \) | 0.07 | 1.71 | 1.20 | 1.33 | 0.40 | 3.90 | 9.86 | 1.33 |
| \( C_2 \) | 0.18 | 1.79 | 1.34 | 1.45 | 0.53 | 3.84 | 8.07 | 1.48 |
| combined | 0.07 | 1.73 | 1.25 | 1.36 | 0.42 | 3.85 | 9.86 | 1.36 |
| baseline | 0.05 | 2.34 | 1.68 | 1.90 | 0.50 | 5.28 | 15.02 | 1.82 |
TABLE 3. The comparison of the results obtained by illumination estimation in three clusters (3C) and by illumination estimation in two clusters, based on scene type (2C scn) and illumination type (2C ill).

| Class       | Min  | Mean | Med. | Tri. | Best 25% | Worst 25% | Max | Avg  |
|-------------|------|------|------|------|---------|-----------|-----|------|
| 2C scn      | 0.01 | 4.65 | 1.65 | 1.98 | 0.45    | 14.30     | 30.91| 2.30 |
| 2C ill      | 0.07 | 1.83 | 1.29 | 1.42 | 0.43    | 4.04      | 13.36| 1.42 |
| 3C          | 0.07 | 1.73 | 1.25 | 1.36 | 0.42    | 3.85      | 9.86 | 1.36 |

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

3) COMPARISON WITH OTHER ILLUMINATION ESTIMATION METHODS

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

On a test set of 342 images, 10 images were misclassified.

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

VI. CONCLUSION

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

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

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

| Algorithm                   | Mean  | Med. | Tri. | Best 25% | Worst 25% | Avg  |
|-----------------------------|-------|------|------|----------|-----------|------|
| White-Patch [10]            | 9.69  | 7.48 | 8.56 | 1.72     | 20.49     | 7.38 |
| Gray-world [5]              | 7.71  | 4.29 | 4.98 | 1.01     | 20.19     | 5.08 |
| Double-opponency (max pooling) [33] | 6.76  | 3.44 | 4.15 | 0.19     | 18.54     | 4.27 |
| Using gray pixels [33]      | 6.65  | 3.26 | 3.95 | 0.68     | 18.75     | 4.05 |
| Color Tiger [51]            | 3.91  | 2.05 | 2.53 | 0.98     | 10.00     | 2.88 |
| Color Mule [55]             | 5.16  | 1.30 | 2.03 | 0.25     | 16.93     | 2.25 |
| Shades-of-Gray [6]          | 2.59  | 1.73 | 1.93 | 0.46     | 6.19      | 1.90 |
| 2nd-order Gray-Edge [7]     | 2.50  | 1.59 | 1.78 | 0.48     | 6.08      | 1.83 |
| 1st-order Gray-Edge [7]     | 2.41  | 1.52 | 1.72 | 0.45     | 5.89      | 1.76 |
| Color Dog [26]              | 3.32  | 1.19 | 1.60 | 0.22     | 10.22     | 1.70 |
| General Gray-World [56]     | 2.38  | 1.43 | 1.66 | 0.35     | 6.01      | 1.64 |
| Attention CNN [49]          | 2.05  | 1.32 | 1.53 | 0.42     | 4.84      | 1.54 |
| Proposed approach           | 1.86  | 1.27 | 1.39 | 0.42     | 4.31      | 1.43 |
| RGB Attention: CNN [48]     | 1.95  | 1.13 | 1.37 | 0.32     | 4.92      | 1.37 |
| Color Beaver (Gray-world) [27] | 1.49  | 0.77 | 0.98 | 0.21     | 3.94      | 0.99 |

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

|                | Min  | Mean  | Med. | Max  |
|----------------|------|-------|------|------|
| misclassified  | 1.01 | 7.35  | 8.84 | 12.83|
| perfectly classified | 0.62 | 3.38  | 2.79 | 9.86 |
networks to classify input images and estimate illumination vectors. Three clusters, i.e., classes, are proposed: cluster with outdoor scenes in natural illuminations, cluster with outdoor scenes in artificial illuminations, and cluster with indoor scenes in artificial illuminations. For each cluster, a separate deep illumination estimation network is trained. With the experimental results, it has been confirmed that training multiple illumination estimation networks using smaller portions of illumination space outperforms a single illumination estimation network. The experiments have shown that the clustering of the illumination space has to be performed carefully and considering not only pure illuminations but features such as scene content as well.

ACKNOWLEDGMENT

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

REFERENCES

[1] M. Ebner, Color Constancy (The Wiley-IS&T Series in Imaging Science and Technology). Hoboken, NJ, USA: Wiley, 2007.
[2] A. Gijsenij, T. Gevers, and J. van de Weijer, “Computational color constancy: Survey and experiments,” IEEE Trans. Image Process., vol. 20, no. 9, pp. 2475–2489, Sep. 2011.
[3] S. Bianci, G. Ciocca, C. Cusano, and R. Schettini, “Improving color constancy using indoor-outdoor image classification,” IEEE Trans. Image Process., vol. 17, no. 12, pp. 2381–2392, Dec. 2008.
[4] N. Banić and S. Lončaric, “Improving the red chromaticity for illumination estimation,” in Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA), Sep. 2015, pp. 131–136.
[5] N. Banić and S. Lončaric, “Color dog—Guiding the global illumination estimation to better accuracy,” in Proc. 10th Int. Conf. Comput. Vis. Theory Appl., 2015, pp. 129–135.
[6] K. Koščević et al., Deep Learning-Based Illumination Estimation Using Light Source Classification.
[45] W. Shi, C. C. Loy, and X. Tang, “Deep specialized network for Illuminant estimation,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 371–387.

[46] S. W. Oh and S. J. Kim, “Approaching the computational color constancy as a classification problem through deep learning,” Pattern Recognit., vol. 61, pp. 405–416, Jan. 2017.

[47] Y. Hu, B. Wang, and S. Lin, “FC**: Fully convolutional color constancy with confidence-weighted pooling,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4085–4094.

[48] K. Koščević, M. Subašić, and S. Lončaric, “Attention-based convolutional neural network for computer vision color constancy,” in Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA), Sep. 2019, pp. 372–377.

[49] K. Koščević, M. Subašić, and S. Lončaric, “Guiding the illumination estimation using the attention mechanism,” in Proc. 2nd Asia Pacific Inf. Technol. Conf., New York, NY, USA, Jan. 2020, p. 143–149, doi: 10.1145/3379310.3379329.

[50] M. Afifi and M. S. Brown, “Sensor-independent illumination estimation for DNN models,” 2019, arXiv:1912.06888. [Online]. Available: http://arxiv.org/abs/1912.06888

[51] N. Baniv and S. Lonarić, “Unsupervised learning for color constancy,” in Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl., 2018, pp. 181–188.

[52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jan. 2009, pp. 248–255.

[53] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in Proc. 13th Int. Conf. Artif. Intell. Statist., 2010, pp. 249–256.

[54] O. Sidorov, “Artificial color constancy via GoogLeNet with angular loss function,” 2018, arXiv:1811.08456. [Online]. Available: http://arxiv.org/abs/1811.08456

[55] N. Banić and S. Lonarić, “A perceptual measure of illumination estimation error,” in Proc. 10th Int. Conf. Comput. Vis. Theory Appl., 2015, pp. 136–143.

[56] K. Barnard, V. Cardei, and B. Funt, “A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data,” IEEE Trans. Image Process., vol. 11, no. 9, pp. 972–984, Sep. 2002.

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

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

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

* * *