Robust channel-wise illumination estimation

Firas Laakom[1]
firas.laakom@tuni.fi
Jenni Raitoharju[2]
jenni.raitoharju@syke.fi
Alexandros Iosifidis[3]
ai@ece.au.dk
Jarno Nikkanen[4]
jarnon@xiaomi.com
Moncef Gabbouj[1]
moncef.gabbouj@tuni.fi

[1] Department of Computing Sciences, Tampere University, Tampere, Finland
[2] Programme for Environmental Information, Finnish Environment Institute, Jyväskylä, Finland
[3] Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark
[4] Xiaomi Finland Oy, Tampere, Finland

Abstract

Recently, Convolutional Neural Networks (CNNs) have been widely used to solve the illuminant estimation problem and have often led to state-of-the-art results. Standard approaches operate directly on the input image. In this paper, we argue that this problem can be decomposed into three channel-wise independent and symmetric subproblems and propose a novel CNN-based illumination estimation approach based on this decomposition. The proposed method substantially reduces the number of parameters needed to solve the task while achieving competitive experimental results compared to state-of-the-art methods. Furthermore, the practical application of illumination estimation techniques typically requires identifying the extreme error cases. This can be achieved using an uncertainty estimation technique. In this work, we propose a novel color constancy uncertainty estimation approach that augments the trained model with an auxiliary branch which learns to predict the error based on the feature representation. Intuitively, the model learns which feature combinations are robust and are thus likely to yield low errors and which combinations result in erroneous estimates. We test this approach on the proposed method and show that it can indeed be used to avoid several extreme error cases and, thus, improves the practicality of the proposed technique.

1 Introduction

The human visual system is able to adapt to different lighting conditions to produce invariant representations of the objects [12]. This ability to remove the illumination effect on the colors is known as color constancy. Digital cameras try to mimic this ability in their preprocessing pipelines and try to suppress the light source effect on the colors presented in the scene. The central objective is to recover the true colors of the objects observed as if the light source is a neutral illumination. This task in modern cameras is known as the computational color constancy and several unsupervised [2, 9, 17, 32, 35, 36, 37] and supervised approaches [3, 4, 5, 22, 24, 25, 30, 39] have been proposed to solve it. Achieving an invariant representation

of the objects regardless of the illuminant is critical for many other vision tasks such as classification [18, 33] and scene understanding [13, 31, 38, 38].

Formally, RGB values of an image I at every pixel (x,y) are expressed as a function of the illuminant $\mathbf{e}(x,y,\lambda)$, the surface reflectance $\mathbf{R}(x,y,\lambda)$, and the camera sensitivity $\mathbf{S}(\lambda)$ as follows:

$$\mathbf{I}(x,y) = \int_{\lambda} \mathbf{e}(x,y,\lambda)\mathbf{R}(x,y,\lambda)\mathbf{S}(\lambda)d\lambda, \tag{1}$$

where λ is the wavelength. Computational color constancy approaches simplify the problem by assuming a global uniform illuminant in the whole scene:

$$\mathbf{e} = \mathbf{e}(x,y) = \int_{\lambda} \mathbf{e}(x,y,\lambda)\mathbf{S}(\lambda)d\lambda. \tag{2}$$

This leads to the following final equation:

$$\mathbf{I}(x,y) = \mathbf{R}(x,y) \circ \mathbf{e}, \tag{3}$$

where $\circ$ is element-wise multiplication. Based on this equation, computational color constancy is typically achieved in a two-step process. In the first step, the global illuminant **e** is estimated. Then, in the second step, the original colors **R** are restored by pixel-wise normalization of the raw image **I** by the estimated **e**. As the second step is a straightforward transformation, the computational color constancy problem is reduced to the first step, i.e., illuminant estimation. It can be seen that the latter is an ill-posed problem as it has one known variable **I** and two unknowns, **e** and **R**. For example, given a yellowish pixel, it is impossible to know if it is truly a yellow object under white illuminant or a white object under yellow illuminant.

Recently, Convolutional Neural Network (CNN)-based approaches have been extensively used to solve this problem [5, 20, 22, 27], given their strong abilities to generalize and to regress directly from the input raw image to the desired target variable without needing feature extraction or preprocessing. The main problem in using the CNN-based techniques follows from the lack large datasets as even the largest publicly available datasets contain only about 7000 images [28]. The models can be categorized either as patch-wise or single-pass methods. Patch-wise approaches solve the data scarcity problem by training on small patches of the original image [5, 6, 7, 8, 24]. In the test phase, the patch estimates are aggregated directly using the average [5] or using a weighted combination [8] to obtain the final estimate. Various other patch-based approaches using different combination techniques have been proposed [6, 7]. Due to the limited amount of labeled data, an unsupervised pretraining phase of an autoencoder using auxiliary data was proposed in [24]. The single-pass CNN models are trained using the full image as input to estimate the illuminant [11, 19, 22]. Different methods have been proposed in the literature also in this category. Some of them rely on pretrained classic CNN architectures, such as VGG16, to overcome the limited number of training samples [22, 30]. A GAN-based approach was proposed in [11] and a Bag of Color Features (BoCF) approach that discards the spatial information as it is not important in the color constancy context was proposed in [25].

Training large CNNs requires a large amount of data which is not available in the current illuminant estimation datasets. Moreover, CNNs are typically over-parameterized and, thus, expensive computationally and in terms of energy and time which restricts their usage in low computational power devices such as mobile phones. Therefore, reducing the number of parameters is critical for a deployable color constancy model. Most CNN-based illuminant

estimation approaches [5, 20, 22] operate directly on the input image $\mathbf{I}$ without exploiting the specificities and characteristics of Eq. (3) defining the problem. We argue that this is not optimal and show that the problem presented in Eq. (3) can be decomposed channel-wise into three symmetric independent sub-problems. Based on this decomposition, we propose a novel CNN-based computational color constancy approach, named Channel-Wise Color Constancy (CWCC), which leverages the problem characteristics. The introduced dynamics are not only in full corroboration with the color constancy theory, i.e., Eq. (3), but they also enable us to substantially reduce the number of the required parameters by up to 90% while achieving comparable experimental results to previous state-of-the-art approaches. This makes our approach energy and time-efficient and, thus, suitable for low-cost devices.

Furthermore, illuminant estimation approaches typically fail for some samples and yield very high errors. For the practical use of these techniques, it would be important to be able to identify these extreme error cases to know when the model prediction for a given scene is not reliable and a different algorithm should be preferred. This can be seen as an uncertainty estimation problem. Recently, Monte Carlo dropout was proposed in the illuminant estimation context to predict the model uncertainty [27]. However, this requires multiple forward passes of the same image to produce an uncertainty estimate which is problematic as it further increases the energy and time costs of the process. In this work, we propose a novel computational color constancy uncertainty estimation approach that augments the trained model with an auxiliary branch that learns to predict the error based on the feature representation. Intuitively, the model learns which feature combinations are robust and, thus, likely to yield low errors and which combinations result in erroneous estimates. Our approach is efficient as it requires only a single forward pass of the input.

To summarize, our main contributions are as follows:

- We propose a channel-wise decomposition of the illuminant estimation problem into three independent sub-problems.
- We propose a novel CNN-based illuminant estimation approach, called Channel-Wise Color Constancy (CWCC), which leverages the decomposition enabling us to reduce the number of parameters up to 90%
- We propose a novel efficient uncertainty estimation approach that augments the trained model with an auxiliary branch that learns to predict the error based on the feature representation.

2 Channel-wise color constancy

We note that the problem defined in Eq. (3) can be divided into three problems using the color channels (r,g,b):

$$\mathbf{I}_r = \mathbf{R}_r \mathbf{e}_r, \quad \mathbf{I}_g = \mathbf{R}_g \mathbf{e}_g, \quad \mathbf{I}_b = \mathbf{R}_b \mathbf{e}_b. \tag{4}$$

As it can be seen, the system composed of the sub-equations of Eq. (4) is equivalent to the problem defined in Eq. (3). Moreover, we note that the sub-equations in Eq. (4) are symmetric, i.e., the problem defined in each equation is similar. Typically, CNN-based methods operate directly on the input image and optimize the filters of the first layers jointly without

exploiting this symmetry in Eq. (3). In this paper, we argue that this might not be optimal, because it can lead to learning some non-intuitive cross-channel correlations from the scarce training data. Thus, exploiting the symmetry can improve the performance of the CNN-based approaches for illuminant estimation. To this end, we propose a channel-wise CNN, which solves the sub-equation in Eq. (4) disjointly. Moreover, we note that our formulation of the illuminant estimation task lifts the non-linearly of the problem. In fact, as it can be seen, in the sub-equations of Eq. (4), the connection between the inputs $\mathbf{I}_i$ and the unknown $\mathbf{e}_i$ is linear as opposite to the standard formation which contains a non-linear operator, i.e., element-wise multiplication.

The proposed model is presented in Figure 1. It is composed of two blocks, the disjoint block and the merging block. The disjoint block learns to solve each sub-equation separately. To this end, each color channel has a separate CNN sub-model. Moreover, we exploit the symmetry of the sub-problems by sharing the weights of 'filter blocks' of the three sub-models. In the merging block, we concatenate the outputs of each channel of the first block. Then, we use a model which acts on this mixed representation and aims to learn the optimal way to merge the feature maps of each channel and approximate the illuminant $\mathbf{e}$.

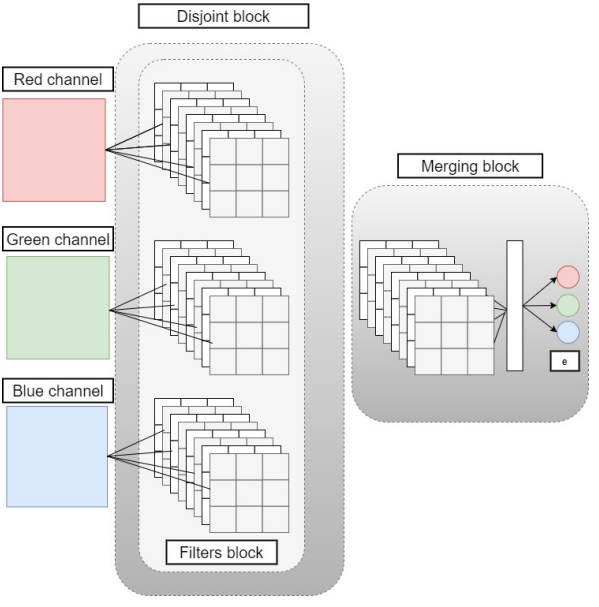

Figure 1: The full illustration of the proposed approach

Formally, given an image $\mathbf{I} = [\mathbf{I}_r, \mathbf{I}_g, \mathbf{I}_b]$, the outputs of the disjoint block are obtained as follows:

$$\mathbf{F}_r = f^1_\theta(\mathbf{I}_r), \quad \mathbf{F}_g = f^1_\theta(\mathbf{I}_g), \quad \mathbf{F}_b = f^1_\theta(\mathbf{I}_b), \tag{5}$$

where f^1_θ is a fully convolutional CNN-model with parameters θ. $\mathbf{F}_r$, $\mathbf{F}_g$, and $\mathbf{F}_b$ are the output feature maps for the red, green, blue channels, respectively. Using the same model $f^1_\theta(\cdot)$ over all the three channels reduces the number of trainable parameters of the total model. Intuitively, the model learns a feature extractor which is independent of the color channel. This constraint is inspired by channel-wise symmetry of the illuminant estimation

task. Next, the illuminant estimate is computed as follows:

$$\mathbf{e} = f^2_{\theta'}(GAP(\mathbf{F}_r \frown \mathbf{F}_g \frown \mathbf{F}_b)), \tag{6}$$

where $\frown$ denotes the concatenation of the feature maps of the different channels, *GAP* is the global average pooling operator compiling a global feature representation of the input image. Finally, the predicted illuminant is obtained through a second model, namely $f^2_{\theta'}$, with parameters θ'. This model takes a vector-representation of the scene and learns to approximate the illuminant and is formed of fully connected layers. The parameters θ and θ' of the inner models f^1 and f^2 can be jointly optimized in an end-to-end matter during the back-propagation.

For f^1_θ, we use a SqueezeNet-like [23] fully convolutional architecture composed as follows: First, we have a convolutional layer with 64 3×3 kernels, then a 3×3 maxpooling layer. Next, we use two fire modules [23] with a size of 64 followed by a 3×3 maxpooling layer. At the end, we have two fire modules with a size of 128 followed by a 3×3 maxpooling layer. For the second inner-model of the merging block $f^2_{\theta'}$ we use a fully connected model containing a fully connected layer with 40 units and ReLu activation and a dropout regularizer with rate of 10%. The output of this layer is connected to the output layer composed of 3 units.

3 Uncertainty estimation block

For the practical use of illuminant estimation techniques, it is important to be able to identify when the model will fail and when its prediction for a given scene is not reliable. This can be seen as an uncertainty estimation problem [16, 29]. We propose to augment our trained illuminant estimation model to predict the model uncertainty. We add an additional branch linked to the last intermediate layer which aims to learn to predict the error based on the feature representation. Intuitively, the model learns which feature combinations are robust and are thus likely to yield low errors and which combinations result in erroneous estimates. The predicted error can be seen as an uncertainty estimate as it directly quantifies to expected loss. Similar to an uncertainty measure, it is expected to have high values in the case of high errors and lower values in the case of low errors. Compared to the existing uncertainty estimation approaches in color constancy [27], we note that our approach requires only a single forward pass of the same image to produce an uncertainty estimate, which enables us to save time and energy. The proposed approach can be incorporated also inside any other illuminant estimation method to measure uncertainty.

The full illustration of the proposed scheme is presented in Figure 2. Given an input image, we generate two outputs: the main illuminant prediction and the predicted error using an auxiliary branch. As we have access to the ground-truth illuminations of our training samples, we can construct a training set for the additional branch by computing the true errors obtained by the trained illuminant estimation model. While training the uncertainty estimation block, we freeze the prediction part of the network to ensure a 'fixed' representation of every input sample and fine-tune only the additional branch of the network. As the topology of this additional model, we use two fully connected layers with Relu activation of sizes 40 and 15, respectively, and one-dimensional fully connected output layer. This additional model is trained using the mean square error to approximate the error, namely the Recovery error $error_{recovery}$.

4 Experimental evaluation

4.1 Experiments on channel-wise color constancy

In this section, we validate the performance of our proposed approach empirically. To this end, we use INTEL-TAU dataset [28], which is the largest publicly available dataset for computational color constancy with 7022 total images split to 10 folds. As in prior works with this dataset [1, 26, 27], the models are evaluated using 10-fold cross validation and the average performance is reported.

As error metric, we use the Recovery angular error $error_{recovery}$ [21] which measures the cosine similarity between the ground truth and prediction:

$$error_{recovery}(\mathbf{e}^{gt}, \mathbf{e}^{est}) = \cos^{-1}\left(\frac{\mathbf{e}^{gt}\mathbf{e}^{est}}{\|\mathbf{e}^{gt}\|\|\mathbf{e}^{est}\|}\right) \tag{7}$$

For more quantitative insights, we also use the Reproduction angular error $e_{reproduction}$ [15] defined as follows:

$$error_{reproduction}(\mathbf{e}^{gt}, \mathbf{e}^{est}) = \cos^{-1}\left(\frac{r(\mathbf{e}^{gt}, \mathbf{e}^{est})\ \mathbf{o}}{\|r(\mathbf{e}^{gt}, \mathbf{e}^{est})\|}\right), \tag{8}$$

where $\mathbf{e}^{gt}$ is the ground truth illumination, $\mathbf{e}^{est}$ is the estimated illumination, $r(\mathbf{e}^{gt}, \mathbf{e}^{est}) = \mathbf{e}^{gt}/\mathbf{e}^{est}$ is the element-wise division of $\mathbf{e}^{gt}$ by $\mathbf{e}^{est}$, and $\mathbf{o}$ is the normalized unit vector, i.e., $\mathbf{n} = [1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}]^T$. For both error metrics, we report the average of the best 25%, the average, the median, the trimean, and the average of the worst 25% of the test errors.

In Table 1, we provide the results for the following unsupervised approaches: White-Patch [34], Grey-World [9], Color-PCA [10], Shades-of-Grey [14], Weighted Grey-Edge [17], Greyness Index 2019 [32], Color Tiger [2], PCC_Q2 [26], and the method reported in [37]. For the supervised approaches, we report the results of Fast Fourier Color Constancy (FFCC) [4] and the following five CNN-based approaches: Fully Convolutional Color Constancy (FC^4) [22], Bianco [5], C3AE [24], and BoCF [25] along with our approach CWCC.

As can be seen, the supervised approaches consistently outperform the unsupervised methods across all metrics. This is due to the fact that unsupervised approaches typically rely on strong assumptions regarding the content of the scene. Thus, when these assumptions are violated, the methods fail. On the contrary, to learning-based approaches where the illuminant estimation is learned end-to-end without any prior assumptions lead to better performance.

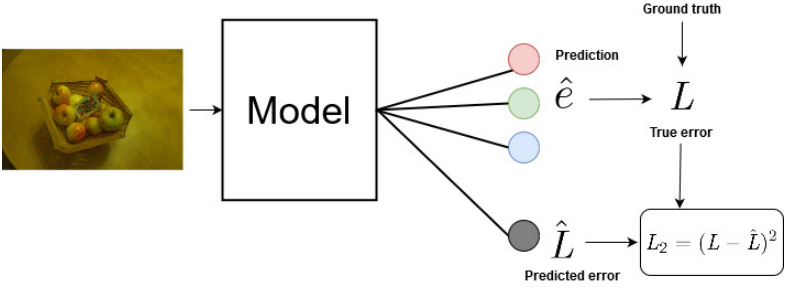

Figure 2: The full illustration of the proposed uncertainty estimation scheme

Table 1: Results of benchmark methods on INTEL-TAU dataset using cross-validation protocol.

	$error_{recovery}$					$error_{reproduction}$				
Method	Best25%	Mean	Med.	Tri.	W.25%	Best25%	Mean	Med.	Tri.	W.25%
Grey-World	1.0	4.9	3.9	4.1	10.5	1.2	6.1	4.9	5.2	13.0
White-Patch	1.4	9.4	9.1	9.2	17.6	1.8	10.0	9.5	9.8	19.2
Grey Edge	1.0	5.9	4.0	4.6	13.8	1.2	6.8	4.9	5.5	13.5
Grey Edge 2	1.0	6.0	3.9	4.8	14.0	1.2	6.9	4.9	5.6	15.7
Shades-of-Grey	0.9	5.2	3.8	4.3	11.9	1.1	6.3	4.7	5.1	13.9
Cheng et al. 2014	0.7	4.5	3.2	3.5	10.6	0.9	5.5	4.0	4.4	12.7
Weighted GE	0.8	6.1	3.7	4.6	15.1	1.1	6.9	4.5	5.4	16.5
Yang et al. 2015	0.6	3.2	2.2	2.4	7.6	0.7	4.1	2.7	3.1	9.6
Color Tiger	1.0	4.2	2.6	3.2	9.9	1.1	5.3	3.3	4.1	12.7
Greyness Index	0.5	3.9	2.3	2.7	9.8	0.6	4.9	3.0	3.5	12.1
PCC_Q2	0.6	3.9	2.4	2.8	9.6	0.7	5.1	3.5	4.0	11.9
FFCC	0.4	2.4	1.6	1.8	5.6	0.5	3.0	2.1	2.3	7.1
Bianco	0.9	3.5	2.6	2.8	7.4	1.1	4.4	3.4	3.6	9.4
C3AE	0.9	3.4	2.7	2.8	7.0	1.1	3.9	3.3	3.5	8.8
BoCF	0.7	2.4	1.9	2.0	5.1	0.8	3.0	2.3	2.5	6.5
FC4 (VGG16)	0.6	2.2	1.7	1.8	4.7	0.7	2.9	2.2	2.3	6.1
CWCC	0.7	2.4	1.9	2.0	4.9	0.8	3.0	2.3	2.7	6.3

Table 2: Number of parameters of different CNN-based approaches

Method	# parameters
Bianco	154k
Fc4(SqueezeNet)	1.9M
FC4 (AlexNet)	3.8M
DS-Net	17.3M
CWCC	155k

The proposed approach CWCC outperforms the state-of-the-art unsupervised approaches. For example, in terms of the worst 25%, CWCC yields better results compared to the method in Yang et al. 2015 by 2.7°. Compared to learning-based approaches, we note that CWCC outperforms Bianco and C3AE across all metrics and achieves comparable results compared to FFCC and FC4. In fact compared to FFCC, which is not a CNN-based approach our methods performs better in the extreme cases. This is clear in terms of the worst 25% metric where our approach yields 0.7° and 0.8° improvement in the Recovery and the Reproduction errors, respectively. Compared to the CNN-based approach FC4, we note that our achieves competitive results across all the metrics while using less than 10% of the number of parameters, as illustrated in Table 2. This supports our assumptions that channel-wise decomposition is reasonable in the illuminant estimation context. Figure 3 illustrates some visual result of our approach on four different test samples from INTEL-TAU. As it can be seen, our approach generalizes well for different environments.

To further illustrate the usefulness of exploiting the symmetry between the sub-equations of our problem formulation in Eq. (4) via weight sharing, we perform an ablation study by comparing the performance of our method CWCC to a variant of our model, called CWCC*, that uses a different feature extractor for each channel. Formally, the shared model f^1_θ in Eq. (5) is replaced as follows:

$$\mathbf{F}_r = f^1_{\theta_r}(\mathbf{I}_r), \quad \mathbf{F}_g = f^1_{\theta_g}(\mathbf{I}_g), \quad \mathbf{F}_b = f^1_{\theta_b}(\mathbf{I}_b), \tag{9}$$

The empirical result of this model on INTEL-TAU are presented in Table 3. As can be seen, even though removing the weight sharing constraint gives the model more flexibility, the performance of the model declines in all metrics. This is clear in terms of the worst

Figure 3: Visual results on four samples of INTEL-TAU. From left to right: Input image, CWCC output, and ground truth image. The corresponding $error_{recovery}$ errors from top to down are: 2.95, 3.53, 1.41, and 5.92.

25%, where CWCC* yields a higher error by 0.5° compared to the standard CWCC. This can be explained by the data scarcity. In fact, removing the weight sharing constraint almost triples the number of trainable parameters to be optimized. As the training data is limited, this leads CWCC* to overfit to the training data and to fail to generalize well for the unseen test samples.

Table 3: Ablation study results of CWCC on INTEL-TAU dataset using cross-validation protocol.

	$error_{recovery}$					$error_{reproduction}$				
Method	Best25%	Mean	Med.	Tri.	W.25%	Best25%	Mean	Med.	Tri.	W.25%
CWCC	0.7	2.4	1.9	2.0	4.9	0.8	3.0	2.3	2.7	6.3
CWCC*	0.8	2.7	2.1	2.2	5.4	0.9	3.3	2.6	2.8	6.9

4.2 Experiments on uncertainty estimation

The results for uncertainty estimation on the test samples of the different INTEL-TAU folds are presented in Figure 4. For most of the samples, the predicted error correlates well with the true error and the model is able to tell how confident it is about its illuminant prediction. However, it is worth noting that for some extreme examples, the model is over-confident,

i.e., predicting low error even though the true errors are high ($> 6^\circ$). Eliminating these cases can be achieved by setting a lower threshold on the predicted loss. For example, we could set our threshold on when to rely on the model at 2.5° predicted error. Then, we can guarantee with a high probability that the true errors will not exceed 5°. However, using this strategy can lead to many false negatives, i.e., deciding not to rely on the model because the predicted uncertainty is higher than the threshold while the model prediction is actually good. So there is a trade-off to be made depending on the application.

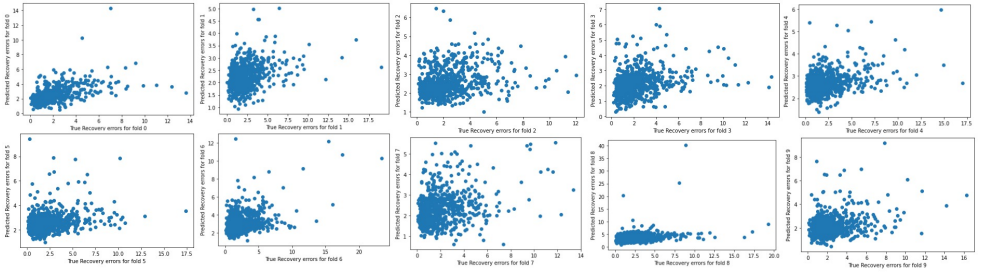

Figure 4: Predicted loss vs true loss using the proposed approach on the different folds of INTEL-TAU. The correlation coefficients from fold 0 to 10 are: 0.47, 0.34, 0.24, 0.25, 0.34, 0.30, 0.45, 0.28, 0.33, and 0.31.

5 Conclusion

In this paper, we proposed a channel-wise decomposition of the color constancy problem into three independent sub-problems. Based on this decomposition, we proposed a novel CNN-based illuminant estimation approach, called Channel-Wise Color Constancy (CWCC), which leverages this formulation. The proposed method substantially reduces the number of parameters needed to solve the task by up to 90% while achieving competitive experimental results compared to state-of-the-art methods. Moreover, we proposed a novel efficient color constancy uncertainty estimation approach that augments the trained model with an auxiliary branch that learns to predict the error based on the feature representation. We showed empirically that the proposed technique can indeed be used to avoid several extreme error cases and, thus, improves the practicality of the proposed technique. Our future research directions include improving the uncertainty estimation approach to generalize better on the extreme cases.

Acknowledgment

This work has been supported by the NSF-Business Finland Center for Visual and Decision Informatics (CVDI) project AMALIA. The work of Jenni Raitoharju was funded by the Academy of Finland (project 324475).

References

[1] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun-Ta Tsai, and Francois Bleibel. Cross-camera convolutional color constancy. *arXiv preprint arXiv:2011.11890*, 2020.

[2] Nikola Banić, Karlo Koščević, and Sven Lončarić. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*, 2017.

[3] Jonathan T. Barron. Convolutional color constancy. *IEEE International Conference on Computer Vision*, pages 379–387, 2015.

[4] Jonathan T. Barron and Yun-Ta Tsai. Fast fourier color constancy. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[5] S. Bianco, C. Cusano, and R. Schettini. Color constancy using CNNs. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 81–89, 2015.

[6] Simone Bianco and Raimondo Schettini. Color constancy using faces. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 65–72. IEEE, 2012.

[7] Simone Bianco and Raimondo Schettini. Adaptive color constancy using faces. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1505–1518, 2014.

[8] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017.

[9] J. Cepeda-Negrete and R.E. Sanchez-Yanez. Gray-world assumption on perceptual color spaces. In *Image and Video Technology*, pages 493–504, 2014.

[10] D. Cheng, D. Prasad, and M. Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 2014.

[11] Partha Das, Anil S Baslamisli, Yang Liu, Sezer Karaoglu, and Theo Gevers. Color constancy by gans: an experimental survey. *arXiv preprint arXiv:1812.03085*, 2018.

[12] Marc Ebner. *Color constancy*, volume 7. John Wiley & Sons, 2007.

[13] Andreas Ess, Tobias Mueller, Helmut Grabner, and Luc Van Gool. Segmentation-based urban traffic scene understanding. In *BMVC*, volume 1, page 2. Citeseer, 2009.

[14] G. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color Imaging Conference*, pages 37–41, 2004.

[15] G. Finlayson and R. Zakizadeh. Reproduction angular error: An improved performance metric for illuminant estimation. *perception*, 2014.

[16] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[17] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Physics-based edge evaluation for improved color constancy. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 581–588. IEEE, 2009.

[18] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[19] Daniel Hernandez-Juarez, Sarah Parisot, Benjamin Busam, Ales Leonardis, Gregory Slabaugh, and Steven McDonagh. A multi-hypothesis approach to color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[20] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. Deep outdoor illumination estimation. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2373–2382, 2017.

[21] S. Hordley and G. Finlayson. Re-evaluating colour constancy algorithms. In *International Conference on Pattern Recognition*, 2004.

[22] Y. Hu, B. Wang, and S. Lin. FC4: Fully convolutional color constancy with confidence-weighted pooling. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085 – 4094, 2017.

[23] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

[24] Firas Laakom, Jenni Raitoharju, Alexandros Iosifidis, Jarno Nikkanen, and Moncef Gabbouj. Color constancy convolutional autoencoder. In *Symposium Series on Computational Intelligence*, 2019.

[25] Firas Laakom, Nikolaos Passalis, Jenni Raitoharju, Jarno Nikkanen, Anastasios Tefas, Alexandros Iosifidis, and Moncef Gabbouj. Bag of color features for color constancy. *IEEE Transactions on Image Processing*, 29:7722–7734, 2020.

[26] Firas Laakom, Jenni Raitoharju, Alexandros Iosifidis, Uygar Tuna, Jarno Nikkanen, and Moncef Gabbouj. Probabilistic color constancy. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 978–982. IEEE, 2020.

[27] Firas Laakom, Jenni Raitoharju, Alexandros Iosifidis, Jarno Nikkanen, and Moncef Gabbouj. Monte carlo dropout ensembles for robust illumination estimation. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2021.

[28] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Intel-tau: A color constancy dataset. *IEEE Access*, 9:39560–39567, 2021.

[29] Antonio Loquercio, Mattia Segu, and Davide Scaramuzza. A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, 5(2): 3153–3160, 2020.

[30] Zhongyu Lou, Theo Gevers, Ninghang Hu, Marcel P Lucassen, et al. Color constancy by deep learning. In *BMVC*, pages 76–1, 2015.

[31] Uzair Nadeem, Syed Afaq Ali Shah, Ferdous Sohel, Roberto Togneri, and Mohammed Bennamoun. Deep learning for scene understanding. In *Handbook of Deep Learning Applications*, pages 21–51. Springer, 2019.

[32] Yanlin Qian, Joni-Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8062–8070, 2019.

[33] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449, 2017.

[34] Al. Rizzi, C. Gatta, and D. Marini. Color correction between gray world and white patch. In *The International Society for Optical Engineering*, 2002.

[35] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, pages 2207–2214, 2007.

[36] Feng Xie, Linmi Tao, Guangyou Xu, and Huijun Di. Estimating illumination parameters in real space with application to image relighting. In *Asian Conference on Computer Vision*, pages 490–499. Springer, 2006.

[37] Kai-Fu Yang, Shao-Bing Gao, and Yong-Jie Li. Efficient illuminant estimation for color constancy using grey pixels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2254–2263, 2015.

[38] Shun Yang, Wenshuo Wang, Chang Liu, and Weiwen Deng. Scene understanding in deep learning-based end-to-end controllers for autonomous vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1):53–63, 2018.

[39] Xiaotong Yuan, Stan Z Li, and Ran He. Color constancy via convex kernel optimization. In *Asian Conference on Computer Vision*, pages 728–737. Springer, 2007.