LEARNING ILLUMINANT ESTIMATION FROM OBJECT RECOGNITION

Marco Buzzelli [a], Joost van de Weijer [b], Raimondo Schettini [a]*

[a] Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca
Viale Sarca 336, Milan 20126, Italy

[b] Computer Vision Center
Universitat Autònoma de Barcelona
Bellaterra, Spain

ABSTRACT

In this paper we present a deep learning method to estimate the illuminant of an image. Our model is not trained with illuminant annotations, but with the objective of improving performance on an auxiliary task such as object recognition. To the best of our knowledge, this is the first example of a deep learning architecture for illuminant estimation that is trained without ground truth illuminants. We evaluate our solution on standard datasets for color constancy, and compare it with state of the art methods. Our proposal is shown to outperform most deep learning methods in a cross-dataset evaluation setup, and to present competitive results in a comparison with parametric solutions.

Index Terms— Illuminant estimation, computational color constancy, semi-supervised learning, deep learning, convolutional neural networks

1. INTRODUCTION

Color constancy is the ability of human beings to recognize the colors of objects independently of the characteristics of the light source. Computational color constancy aims to first estimate the illuminant and subsequently use this information to correct the image, to display how it would appear under a canonical illuminant [1]. The class of "learned" methods is among the most successful illumination estimation methods to date [2, 3], and typically relies on a training set of images which are labeled with the respective scene illuminant. Although the human visual system is often compared to a machine learning algorithm, during evolution it was never presented with ground truth illuminants. Instead it is hypothesized that the ability of color constancy arose because it helped other crucial tasks, such as recognizing fruits, objects, and animals independently of the scene illuminant [4].

This observation motivates us to investigate to what extent we can learn illuminant estimation as a byproduct of an auxiliary task, for which we consider object recognition. Previous works highlighted the usefulness of object recognition in applications related to color theory, such as assessment of chromatic adaptation transforms through memory color matching
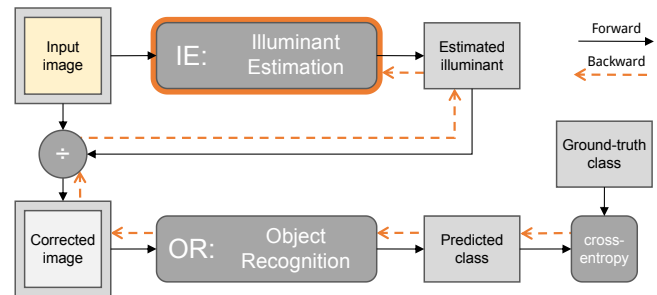
---

*Corresponding author: marco.buzzelli@disco.unimib.it

Fig. 1. Schematic representation of our proposal for computational color constancy. The Illuminant Estimation module is trained with the objective of optimizing the Object Recognition module. At inference time, only the Illuminant Estimation module is needed.

[5], performing color constancy with the assistance of faces [6] or via pre-training on generic object recognition before fine-tuning for illuminant estimation [7]. In our proposal we describe a training process which uses only the labels for the auxiliary task but no illuminant ground truth whatsoever. During training, the input image will first pass through an illuminant estimation network, which estimates the scene illuminant and corrects the image accordingly. The white-balanced image will then be processed by an object classification network, which produces an estimation of the classes that are present in the image. By training both networks in an end-to-end fashion, we can effectively train the illuminant estimation network without any illuminant ground truth. This illuminant estimation network can then be independently applied to other datasets, such as standard color constancy benchmarks, without using the object recognition network. To the best of our knowledge this is the first learned method for illuminant estimation which does not require illuminant annotations.

Approaches to color constancy can be divided into parametric and learned. Parametric solutions are typically based on handcrafted features which depend on few manually-tunable parameters, the most effective to date being Akbarinia et al. [3] and Cheng et al. [8]. For a complete review of parametric methods for illuminant estimation we refer to [2].

Learning-based solutions recently proved to be particularly effective when trained in a cross-validation setup: Bianco et al. [9] developed a patch-wise illuminant estimation neural network and generate a global estimation through consensus. Lou et al. [7] propose the fine-tuning of a network that was pre-trained for both object recognition, and illuminant regression based on the predictions of other color constancy methods. Oh et al. [10] reformulate the problem of illuminant estimation as a classification task by clustering the target illuminants. All these solutions require an illuminant ground truth, and are designed to be trained on the same type of images that were seen during the test phase.

2. PROPOSED METHOD FOR COLOR CONSTANCY

In this section, we present our learning approach to estimate the scene illuminant in the absence of any illumination ground truth data, but with label information for an auxiliary task. Here we consider object recognition as the auxiliary task, but other objectives such as object detection or semantic segmentation could be used as well. We propose an Illuminant Estimation / Object Recognition network (IEOR), which is composed of two parts: an Illuminant Estimation module (IE) and an Object Recognition module (OR), as shown in Fig. 1. Following this structure, IE is learning to predict, for any given image, a color correction that would improve OR.

2.1. Object Recognition network

The Object Recognition network performs the auxiliary supervised task described in our approach, which is used to indirectly train the Illuminant Estimation network. OR takes an input RGB image, and produces a prediction of class presence. For the purpose of this paper, we will focus on vegetable recognition [11], a task for which color is important as highlighted in Fig. 2. Other problems for which a correct white balancing is expected to improve results might be recognition of painting styles [12], or assessment of image aesthetics [13].

In order to obtain the desired separation of the network's internal logic into IE and OR modules, with the intermediate representation being the estimated illuminant, it is necessary to make OR overly-sensitive to color variations. This would then effectively drive IE to work as a useful preprocessing step, i.e. to learn to perform the necessary color correction. To fulfill this requirement, the following strategy is adopted:

1. We pre-train OR alone on the chosen auxiliary task. No color jittering is applied in this phase.

2. We connect IE and train the whole system end-to-end with color jittering as described in Sec. 2.2.
   During this second phase the gradients flow through OR, but we update only the weights in IE.

If we do not constrain the training process as explained, it might lead to unusable solutions, such as always producing a neutral illuminant, or spreading the processing without a clear distinction of the intermediate "estimated illuminant". For the current implementation of our method, we adopt an architecture based on AlexNet [14]. The network weights are initialized on the task of object classification from ImageNet [14], with the final fully-connected layer re-instantiated in order to match the different cardinality and task. Before feeding the image to the network, we subtract 0.5 as an approximation of the mean-image subtraction technique that allows to speed up training and fine-tuning of the neural model.

2.2. Illuminant Estimation network

The Illuminant Estimation network takes an input RGB image and predicts the scene illuminant. For the purpose of this paper, we choose to limit the illuminant model to a diagonal matrix, i.e. $\rho^e = (\rho_R^e, \rho_G^e, \rho_B^e)^\mathrm{T}$, although alternative combinations are possible [15]:

$$\begin{bmatrix} R_\mathrm{out} \\ G_\mathrm{out} \\ B_\mathrm{out} \end{bmatrix} = \begin{bmatrix} 1/\rho_R^e & 0 & 0 \\ 0 & 1/\rho_G^e & 0 \\ 0 & 0 & 1/\rho_B^e \end{bmatrix} \cdot \begin{bmatrix} R_\mathrm{in} \\ G_\mathrm{in} \\ B_\mathrm{in} \end{bmatrix} \quad (1)$$

This representation also matches the annotation associated to most datasets for color constancy, which provide ground truth illuminants in terms of triplets [16, 17, 8].

During the end-to-end training of the whole IEOR network, we artificially augment the input data by applying a random illuminant extracted from a Gaussian distribution with mean 1 and standard deviation 0.3. Once training is completed as described in Sec. 2.1, IE can be used as a standalone network for color correction on any image, which is not necessarily depicting the classes seen during the training phase.

The IE module is also based on AlexNet, which was shown in the past to perform well on color constancy [7], and images are preprocessed by subtracting 0.5 as with the OR module. The color-corrected image obtained by applying the estimated illuminant to the input image is clipped between 0 and 1. This operation is necessary to avoid feeding to the pre-trained OR values in a range that was never encountered during training, and is also recreating what would happen by displaying the color-corrected image.

3. EXPERIMENTS

In these experiments we set out to evaluate how our illuminant estimation network, which does not require any illuminant ground truth, compares to other illuminant estimation methods on standard datasets. We consider two possible normalization techniques (global and channel-wise), and compare the results with state of the art methods.

3.1. Experimental setup

The training phase requires a classification task where color is a discriminating feature, in order to properly drive the learn-

Fig. 2. Samples from VegFru dataset, showing the importance of color in vegetable class discrimination

ing process. To this extent, the vegetables subset from the **VegFru** dataset [11] fits the desired criteria, as shown in Fig. 2. It contains more than 90000 images belonging to 200 vegetable classes. At test time we only apply the IE module (the branch related to object recognition is not used) and we evaluate our solution on two widely-adopted benchmarks for color constancy. **Shi-Gehler** [16] is a linear reprocessing of the original RAW images from the Gehler dataset [18]. It contains 568 images of different scenarios with a ColorChecker in each picture as a support for the ground truth extraction. At test time, the area corresponding to the ColorChecker is edited out to prevent the illuminant estimation algorithm from directly using it. Results on the Shi-Gehler datasets are computed with the ground truth from [9]. **NUS** from National University of Singapore [8] contains a total of 1853 images coming from 9 digital cameras of various brands. The type of content is similar to Shi-Gehler, with both indoor and outdoor scenes, and ground truth extracted from a ColorChecker. A unique characteristic of this dataset is the presence of the same scenes taken from different cameras. Results on NUS are reported as the average performance on each camera.

For evaluation we use the well-established angular recovery error [2, 8], which compares the estimated (ρ^e) and reference (ρ^r) illuminant triplets regardless of their magnitude:

$$err_{ang} = \arccos\left(\frac{\rho^{e\mathrm{T}}\,\rho^r}{||\rho^e||\,||\rho^r||}\right) \qquad (2)$$

3.2. Input normalization

In order to account for the potential discrepancies between training and test images, we adopt several techniques for input normalization. We apply gamma correction whenever a linear dataset is used at test time, since our training classification datasets are already processed for gamma. The estimated illuminant is then re-corrected before the final evaluation with the provided ground truth. We then devise two alternative ways to ensure the range of image values is stable, based on

the generation of a support diagonal illuminant:

1. Global normalization, which brings the overall average of the input image to a fixed value without changing the relationship between single channels.

2. Channel normalization, that modifies each channel independently, thus affecting the original illuminant.

3.3. Results

Table 1 reports the results of the two variants of our method on Shi-Gehler and NUS datasets, compared to different baselines and different algorithms from the state of the art.

Global normalization brings to very similar results on the two datasets, providing a robust performance assessment of our solution. Channel normalization leads instead to contrasting conclusions on each dataset, suggesting a more challenging nature of NUS over Shi-Gehler. As channel normalization destroys the relationship between channels, in fact, it essentially discards any potentially misleading information about the original illuminant. In the case of NUS, our model benefits from this preprocessing, as it is forced to estimate the proper illuminant directly from the image content.

In order to assess the effective advantage that comes from using a classification-based loss, we train the Illuminant Estimation module for direct regression (third row) over the color-augmented examples of VegFru, i.e. we train without OR. Performance on both sets shows that our classification-based strategy can lead up to a 19% improvement in angular error, thus highlighting the relevance of the proposed approach.

The "Learned (cross-dataset)" block includes data-driven models that are trained on one dataset and tested on a different one, instead of adopting a more traditional training-test split of the same dataset. Our solution was trained on the VegFru dataset, and is therefore best compared with methods from this category, which are either trained on the GrayBall dataset [20, 7] or on SFU-Lab [21]. Our IEOR network outperforms the solutions by Joze et al. [20] and Gao et al. [21], while producing comparable results to Lou et al. [7]. For completeness, we also include other state of the art methods that are trained on the training-set portion of the same dataset used for evaluation ("Learned (in-dataset)"), although direct comparison with these solutions is not applicable. Parametric methods [8] and [19] present different results with varying parameters and configurations. For such solutions we only report the best configuration, which was directly selected from the test set performance.

Figure 3 shows some examples of our predictions with the corresponding angular error, in order to provide a visual guide for the interpretation of the reported performance values.

3.4. Analysis on the training-set color bias

The illuminant distribution present in the classification training dataset plays an important role for our proposed method.

Table 1. Performance of different algorithms for illuminant estimation. Our solution can be directly compared with cross-dataset learned methods. Values reported for parametric solutions [8] and [19] are selected with the best configuration of test parameters. Learned methods with in-dataset training are also reported, although a direct comparison cannot be performed.

	Method	Angular error (Shi-Gehler [16])				Angular error (NUS [8])			
		Mean	Median	Std	Max	Mean	Median	Std	Max
Baselines	Unchanged	13.62°	13.55°	2.85°	27.37°	19.50°	18.82°	1.90°	25.83°
	Greyworld	7.35°	6.70°	3.78°	25.84°	4.59°	3.64°	3.57°	22.61°
	Regression	5.96°	5.31°	3.47°	19.88°	5.19°	3.90°	4.16°	22.07°
Parametric	Akbarinia et al. [3]	3.8 °	2.4 °	-	-	-	-	-	-
	Cheng et al. [8]	3.52°	2.14°	-	28.35°	3.02°	2.12°	-	17.24°
	Funt et al. [19]	3.2 °	2.3 °	-	21.7 °	-	-	-	-
Learned (cross-dataset)	**Ours (global norm.)**	4.84°	4.12°	3.22°	20.80°	4.88°	4.17°	3.11°	18.70°
	Ours (channel norm.)	5.48°	4.81°	3.21°	19.88°	4.32°	3.37°	3.56°	22.36°
	Joze et al. [20]	6.5 °	5.1 °	-	-	-	-	-	-
	Gao et al. [21]	5.03°	3.39°	-	-	-	-	-	-
	Lou et al. [7]	4.7 °	3.3 °	5.3 °	-	-	-	-	-
Learned (in-dataset)	Chakrabarti et al. [22]	3.59°	2.96°	-	21.58°	3.04°	2.40°	-	15.38°
	Bianco et al. [9]	2.63°	1.98°	-	14.77°	-	-	-	-
	Oh et al. [10]	2.16°	1.47°	-	-	2.41°	2.15°	-	-

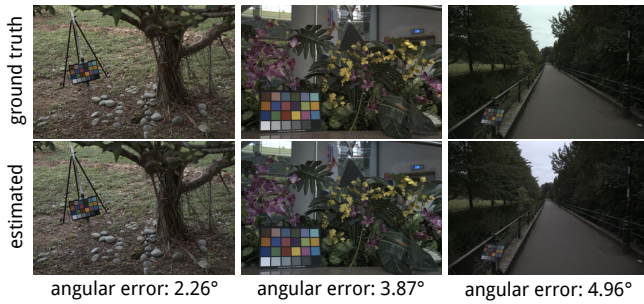angular error: 2.26° angular error: 3.87° angular error: 4.96°

Fig. 3. Example color corrections of our method at different levels of angular error

IE implicitly learns to replicate the average illuminant of the training set. If this average is not neutral, the predicted illuminants on the test set will not, in general, match the provided ground truth. If we use an oracle to shift our predictions to the average ground truth illuminant, we can expect to further reduce the angular error on Shi-Gehler from $4.84°$ to $4.60°$ (5%), and from $4.32°$ to $4.23°$ on NUS (2%).

Secondly, a high variability in the illuminant during the initial pre-training of OR may make it excessively robust to color variations, thus compromising the consequent training of IE. This condition is particularly hard to avoid: since dataset heterogeneity typically results in better classification performance [23], dataset curators tend to implicitly or explicitly encourage a certain degree of illuminant variability. To this extent, in the future we might consider preprocessing the classification-oriented training set itself for unsupervised white balancing, reducing the intra-class variability with the help of methods for saliency estimation [24].

4. CONCLUSIONS

We have presented a deep learning method for illuminant estimation that does not require any ground truth illuminant for training. In order to fulfill this objective, we indirectly train our neural model with the task of improving the performance of an auxiliary object recognition network. Our strategy allows the use of powerful deep learning models but without the need of collecting illuminant information, and relying instead on widely available annotations from other tasks. In experiments our method was, as expected, outperformed by learned methods which use ground truth data. However, surprisingly, in cross-dataset experiments our method obtains similar results as other learned based methods. As future research we plan to train Illuminant Estimation / Object Recognition networks for more complex illumination scenarios, e.g. with multiple illuminants. These scenarios are currently hardly investigated with learned based methods because of the lack of large datasets with ground truth information. However, in the proposed end-to-end learning of illuminant estimation the presence of ground truth data would not be required.

5. REFERENCES

[1] David A Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–35, 1990.

[2] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer, "Computational color constancy: Survey and experiments," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2475–2489, 2011.

[3] Arash Akbarinia and C Alejandro Parraga, "Colour constancy beyond the classical receptive field," *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[4] Inês Bramão, Luís Faísca, Karl Magnus Petersson, and Alexandra Reis, *The contribution of color to object recognition*, InTech, 2012.

[5] Kevin AG Smet, Qiyan Zhai, Ming R Luo, and Peter Hanselaer, "Study of chromatic adaptation using memory color matches, part i: neutral illuminants," *Optics express*, vol. 25, no. 7, pp. 7732–7748, 2017.

[6] Simone Bianco and Raimondo Schettini, "Color constancy using faces," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 65–72.

[7] Zhongyu Lou, Theo Gevers, Ninghang Hu, Marcel P Lucassen, et al., "Color constancy by deep learning.," in *BMVC*, 2015, pp. 76–1.

[8] Dongliang Cheng, Dilip K Prasad, and Michael S Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," *JOSA A*, vol. 31, no. 5, pp. 1049–1058, 2014.

[9] Simone Bianco, Claudio Cusano, and Raimondo Schettini, "Color constancy using cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 81–89.

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

[11] Saihui Hou, Yushan Feng, and Zilei Wang, "Vegfru: A domain-specific dataset for fine-grained visual categorization," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 541–549.

[12] Simone Bianco, Davide Mazzini, and Raimondo Schettini, "Deep multibranch neural network for painting categorization," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 414–423.

[13] Simone Bianco, Luigi Celona, Paolo Napoletano, and Raimondo Schettini, "Predicting image aesthetics with deep learning," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2016, pp. 117–125.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] Simone Bianco, Claudio Cusano, Flavio Piccoli, and Raimondo Schettini, "Artistic photo filter removal using convolutional neural networks," *Journal of Electronic Imaging*, vol. 27, no. 1, pp. 011004, 2017.

[16] Lilong Shi, "Re-processed version of the gehler color constancy dataset of 568 images," *http://www. cs. sfu. ca/~ color/data/*, 2000.

[17] Florian Ciurea and Brian Funt, "A large image database for color constancy research," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2003, vol. 2003, pp. 160–164.

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

[19] Brian Funt and Milan Mosny, "Removing outliers in illumination estimation," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2012, vol. 2012, pp. 105–110.

[20] Hamid Reza Vaezi Joze and Mark S Drew, "Exemplar-based color constancy and multiple illumination," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 5, pp. 860–873, 2014.

[21] Shao-Bing Gao, Kai-Fu Yang, Chao-Yi Li, and Yong-Jie Li, "Color constancy using double-opponency," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 10, pp. 1973–1985, 2015.

[22] Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler, "Color constancy with spatio-spectral statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 8, pp. 1509–1519, 2012.

[23] Simone Bianco, Marco Buzzelli, Davide Mazzini, and Raimondo Schettini, "Deep learning for logo recognition," *Neurocomputing*, vol. 245, pp. 23–30, 2017.

[24] Simone Bianco, Marco Buzzelli, and Raimondo Schettini, "Multiscale fully convolutional network for image saliency," *Journal of Electronic Imaging*, vol. 27, pp. 27 – 27 – 10, 2018.