Color Constancy by Deep Learning

Zhongyu Lou
z.lou@uva.nl
Theo Gevers
Th.Gevers@uva.nl
Ninghang Hu
n.hu@uva.nl
Marcel P Lucassen
marcel@lucr.nl

Intelligent System Laboratory Amsterdam, University of Amsterdam,
Science Park 904, 1098 XH Amsterdam, The Netherlands

We propose a new framework using Deep Neural Networks (DNNs) to obtain accurate light source estimation. Color constancy is formulated as a DNN-based regression approach to estimate the color of the light source.

Traditional color constancy algorithms are constrained to a number of imaging assumptions or shallow learning models based on hand-crafted features. In this paper, instead of using traditional feature representations, we exploit deep learning architectures by means of Convolutional Neural Networks (CNN). Different from existing methods which rely on predefined low-level features, we propose to use CNNs to learn feature hierarchies to achieve robust color constancy. The deep CNN model used in this work consists of eight (hidden) layers. Such a deep model will yield multi-scale image features composed of pixels, edges, object parts and object models. The proposed deep learning approach needs large amounts of data with ground-truth for training. Unfortunately, there are no such datasets available for color constancy. Therefore, we propose a different training approach. Firstly, we propose a new training strategy which contains three steps to learn hierarchical features for color constancy. Secondly, we propose a method to generate more training images with ground-truth labels.

The model used in this work consists of eight layers as defined in [3]. The first five layers are convolutional layers. The last three layers are fully collected layers. Combining all the layers, the total number of parameters in this model is very large (around 60M). Therefore, large scale datasets with ground-truth light source labels are required to directly apply this model to the color constancy problem. However, such large scale dataset are not available. To this end, we propose an alternative training procedure consisting of different training steps in the following section. Further, we proposed a data augmentation method to generate more training images with ground truth labels.

The training contains three steps. In the first step, we train the model on ImageNet to derive features for object description. In this way, a rich and generic feature hierarchy is learnt to capture the complex visual patterns in generic, real-world images. The last layer of the model is replaced by a 1000 dimensional vector. The soft-max loss function is used for training. The aim of the first training step is to obtain a pre-trained feature model representing general images. Since the ImageNet dataset contains 1000 object categories, it provides abundant back-propagation information for training. We denote the parameters obtained by training on ImageNet as the $Net1$ network.

In the second step of training, the aim is to retrain and adjust the parameters of $Net1$ for the purpose of color constancy. We perform retraining of the (initial) parameters of $Net1$ based on light source estimation obtained by existing color constancy algorithms as labels. Although any other or combination of color constancy algorithms can be used to generate the labels for the ImageNet dataset, the gray-shades algorithm is used due to its efficiency and good performance. Specifically, we use $Net1$ as the initial weights and retrain the network on ImageNet dataset with light source labels generated by gray-shades. The resulting sets of parameters are denoted by $Net2$. The obtained feature representations $Net2$ are merely adopted (color constancy) versions of $Net1$. It is hypothesized that the obtained models $Net2$ replicate the performance of the color constancy methods used to provide the labels i.e. the gray-shades algorithm.

In the third step of training, the parameters of the model obtained in the previous step are retrained using existing (publicly available) datasets with (real) ground-truth label sets (e.g. Grayball [1] and ColorChecker [2]). In these datasets, the ground-truth color of the light source is given for each image under which it has been recorded. We use $Net2$ as initial parameters to retrain the model with a Euclidian loss. The parameters of this network obtained after this step is the final model, denoted by $Net3$.

Deep learning approaches greatly benefit from large training datasets. Since ground-truth labels are expensive to get (especially for the color constancy problem), data augmentation is widely exploited by different deep learning approaches. For example, in image denoising, extra training data is generated by applying simulated noise. In deep learning-based image classification, Krizhevsky et.al [3] use image translations and horizontal reflections to generate more training images.

Inspired by these data augmentation methods, in this paper, we use data augmentation to obtain more training data for color constancy. Specifically, for each training image, we correct for the color of the light source using the diagonal model of the ground-truth. Using Eq. 1 ,

$$\begin{pmatrix} R_u \\ G_u \\ B_u \end{pmatrix} = \begin{pmatrix} e_R & 0 & 0 \\ 0 & e_G & 0 \\ 0 & 0 & e_B \end{pmatrix} \begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix}, \tag{1}$$

the canonical image is obtained. Then, simulated light sources can be applied to the corrected image using the diagonal model in Eq. 1. Any simulated light source color can be used i.e. the Spectral Power Distribution (SPD) of different light sources such as tungsten halogen, fluorescent lamp, high pressure sodium, or daylight. However, many of them are less frequently present than others depending on the scenes from which the images are recorded. Therefore, in this paper, simulated light sources are derived from the training dataset. By clustering the ground-truth light color of the training set into k clusters, we obtain k simulated light sources by collecting the means of each cluster. In this way, per image, k additional training images are obtained with different ground-truth illuminant ($k = 10$ in our experiments). Note that data augmentation is only performed on the training images. We test the algorithm on the original testing images. We denote the proposed model trained on the extended Grayball dataset as $NetFinal$. Fig. 1 shows a number of images obtained by the proposed data augmentation method. Experiments show that the proposed algorithm ($NetFinal$) outperforms the state-of-the-art by 9%. Especially in cross dataset validation, our approach reduces the median angular error by 35%.

(a) GT(0°) (b) 4.14° (c) 6.80° (d) 11.22° (e) 22.78°

Figure 1: Data augmentation. (a) canonical image obtained by removing the effect of the light source color using Eq. 1. (b-e) images generated by applying simulated light sources using Eq. 1. The numbers are the angles between the simulated light sources and the canonical (white) light source.

[1] F. Ciurea and B. Funt. A large image database for color constancy research. *Proceedings of the Eleventh Color Imaging Conference*, 2003.

[2] P.V. Gehler, C. Rother, A. Blake, T. Sharp, and T. Minka. Bayesian color constancy revisited. *CVPR*, 2008.

[3] A. Krizhevsky, I. Krizhevsky, and G. Hinton. Imagenet classification with deep convolutional neural networks. *NIPS*, 2012.