

# Color Constancy Using CNNs

Simone Bianco<sup>1</sup> Claudio Cusano<sup>2</sup> Raimondo Schettini<sup>1</sup>

<sup>1</sup>University of Milan-Bicocca, Italy

<sup>2</sup>University of Pavia, Italy

{simone.bianco, raimondo.schettini}@disco.unimib.it, claudio.cusano@unipv.it

## Abstract

*In this work we describe a Convolutional Neural Network (CNN) to accurately predict the scene illumination. Taking image patches as input, the CNN works in the spatial domain without using hand-crafted features that are employed by most previous methods. The network consists of one convolutional layer with max pooling, one fully connected layer and three output nodes. Within the network structure, feature learning and regression are integrated into one optimization process, which leads to a more effective model for estimating scene illumination. This approach achieves state-of-the-art performance on a standard dataset of RAW images. Preliminary experiments on images with spatially varying illumination demonstrate the stability of the local illuminant estimation ability of our CNN.*

## 1. Introduction

Many computer vision problems in both still images and videos can make use of color constancy processing as a pre-processing step to make sure that the recorded color of the objects in the scene does not change under different illumination conditions. The observed color of the objects in the scene depends on the intrinsic color of the object (i.e. the surface spectral reflectance), on the illumination, and on their relative positions.

In general there are two methodologies to obtain reliable color description from image data: computational color constancy and color invariance [28]. Computational color constancy is a two-stage operation: the former is specialized on estimating the color of the scene illuminant from the image data, the latter corrects the image on the basis of this estimate to generate a new image of the scene as if it was taken under a reference light source. Color invariance methods instead represent images by features which remain unchanged with respect to specific imaging condition.

In this work we focus on computation color constancy, using a CNN to learn discriminant features for the illuminant estimation task. Recently, deep neural networks have

gained the attention of numerous researchers outperforming state-of-the-art approaches on various computer vision tasks [25, 26]. One of CNNs advantages is that it can take raw images as input and incorporate feature learning into the training process. With a deep structure, CNN can learn complicated mappings while requiring minimal domain knowledge.

To the best of our knowledge, this is the first work that investigates the use of CNNs for illuminant estimation. The main contribution of our paper is that we propose a novel method that allows learning and prediction of scene illuminant on local regions. Previous approaches typically accumulate features over the entire image to obtain statistics for estimating the overall illuminant, and only a few approaches have shown the ability to estimate spatially varying illuminant. By contrast, our method can estimate the illuminant on small patches (such as  $32 \times 32$ ). We show experimentally that the proposed method advances the state-of-the-art on a standard dataset of RAW images. In addition to the superior overall performance, we also show quantitative and qualitative results that demonstrate the quality of the local illuminant estimation of our method.

## 2. Problem formulation and related works

The image values for a Lambertian surface located at the pixel with coordinates  $(x, y)$  can be seen as a function  $\rho(x, y)$ , mainly dependent on three physical factors: the illuminant spectral power distribution  $I(x, y, \lambda)$ , the surface spectral reflectance  $S(x, y, \lambda)$  and the sensor spectral sensitivities  $\mathbf{C}(\lambda)$ . Using this notation  $\rho(x, y)$  can be expressed as

$$\rho(x, y) = \int I(x, y, \lambda)S(x, y, \lambda)\mathbf{C}(\lambda)d\lambda, \quad (1)$$

where  $\lambda$  is the wavelength,  $\rho$  and  $\mathbf{C}(\lambda)$  are three-component vectors and the integration is performed over the visible spectrum. The goal of color constancy is to estimate the color  $\mathbf{I}(x, y)$  of the scene illuminant, i.e. the projection of  $I(x, y, \lambda)$  on the sensor spectral sensitivities  $\mathbf{C}(\lambda)$ :

$$\mathbf{I}(x, y) = \int I(x, y, \lambda)\mathbf{C}(\lambda)d\lambda. \quad (2)$$

Usually the illuminant color is estimated up to a scale factor as it is more important to estimate the chromaticity of the scene illuminant than its overall intensity [21]. Since the only information available are the sensor responses  $\rho$  across the image, color constancy is an under-determined problem [14] and thus further assumptions and/or knowledge are needed to solve it.

Several computational color constancy algorithms have been proposed, each based on different assumptions. The most common assumption made a uniform light source color across the scene, i.e.  $\mathbf{I}(x, y) = \mathbf{I}$ .

State-of-the-art solutions can be divided into two main classes: statistic approaches, and learning-based approaches. Statistic approaches estimate the scene illumination only on the base of the content in a single image making assumptions about the nature of color images exploiting statistical or physical properties; learning-based approaches require training data in order to build a statistical image model, prior to estimation of illumination.

## 2.1. Statistic-based algorithms

Van de Weijer et al. [33] have unified a variety of algorithms. These algorithms estimate the illuminant color  $\mathbf{I}$  by implementing instantiations of the following equation:

$$\mathbf{I}(n, p, \sigma) = \frac{1}{k} \left( \iint |\nabla^n \rho_\sigma(x, y)|^p dx dy \right)^{\frac{1}{p}}, \quad (3)$$

where  $n$  is the order of the derivative,  $p$  is the Minkowski norm,  $\rho_\sigma(x, y) = \rho(x, y) \otimes G_\sigma(x, y)$  is the convolution of the image with a Gaussian filter  $G_\sigma(x, y)$  with scale parameter  $\sigma$ , and  $k$  is a constant to be chosen such that the illuminant color  $\mathbf{I}$  has unit length (using the 2-norm). The integration is performed over all pixel coordinates. Different  $(n, p, \sigma)$  combinations correspond to different illuminant estimation algorithms, each based on a different assumption. For example, the Gray World algorithm [6] – generated setting  $(n, p, \sigma) = (0, 1, 0)$  – is based on the assumption that the average color in the image is gray and that the illuminant color can be estimated as the shift from gray of the averages in the image color channels; the White Point algorithm [7] – generated setting  $(n, p, \sigma) = (0, \infty, 0)$  – is based on the assumption that there is always a white patch in the scene and that the maximum values in each color channel are caused by the reflection of the illuminant on the white patch, and they can be thus used as the illuminant estimation; the Gray Edge algorithm [33] – generated setting for example  $(n, p, \sigma) = (1, 0, 0)$  – is based on the assumption that the average color of the edges is gray and that the illuminant color can be estimated as the shift from gray of the averages of the edges in the image color channels.

The Gamut Mapping assumes that for a given illuminant, one observes only a limited gamut of colors [13]. It has

a preliminary phase in which a canonical illuminant is chosen and the canonical gamut is computed observing as many surfaces under the canonical illuminant as possible. Given an input image with an unknown illuminant, its gamut is computed and the illuminant is estimated as the mapping that can be applied to the gamut of the input image, resulting in a gamut that lies completely within the canonical gamut and produces the most colorful scene. If the spectral sensitivity functions of the camera are known, the Color by Correlation approach could be also used [12].

## 2.2. Learning-based algorithms

The learning-based color constancy algorithms, that estimate the scene illuminant using a model that is learned on training data, can be subdivided into two main subcategories: probabilistic methods and fusion/selection based methods. Bayesian approaches [16] model the variability of reflectance and of illuminant as random variables, and then estimate illuminant from the posterior distribution conditioned on image intensity data.

Given a set computational color constancy algorithms, in [1] an image classifier is trained to classify the images as indoor and outdoor, and different experimental frameworks are proposed to exploit this information in order to select the best performing algorithm on each class. In [2] it has been shown how intrinsic, low level properties of the images can be used to drive the selection of the best algorithm (or the best combination of algorithms) for a given image. The algorithm selection and combination is made by a decision forest composed of several trees on the basis of the values of a set of heterogeneous features.

In [17] the Weibull parametrization has been used to train a maximum likelihood classifier based on mixture of Gaussians to select the best performing color constancy method for a certain image.

In [8] a statistical model for the spatial distribution of colors in white balanced images is developed, and then used to infer illumination parameters as those being most likely under their model. High level visual information has been used to select the best illuminant out of a set of possible illuminants [34]. This is achieved by restating the problem in terms of semantic interpretability of the image. Several color constancy methods are applied to generate a set of illuminant hypotheses. For each illuminant hypothesis, they correct the image, evaluate the likelihood of the semantic content of the corrected image, and select the most likely illuminant color. In [3, 4] the use of automatically detected objects having intrinsic color is investigated. In particular, it is investigated how illuminant estimation can be performed exploiting the color statistics extracted from the faces automatically detected in the image. When no faces are detected in the image, any other algorithm in the state-of-the-art can

be used. In [23, 24] the surfaces in the image are exploited and the color constancy problem is addressed by unsupervised learning of an appropriate model for each training surface in training images. The model for each surface is defined using both texture features and color features. In a test image the nearest neighbor model is found for each surface and its illumination is estimated by comparing the statistics of pixels belonging to nearest neighbor surfaces and the target surface. The final illumination estimation results from combining these estimated illuminants over surfaces to generate a unique estimate.

### 3. The proposed approach

The proposed framework of using CNN for illuminant estimation is as follows. Given color image, we sample non-overlapping patches from it and for each of them we perform a contrast normalization through histogram stretching. We use a CNN to estimate the illuminant for each patch and combine the patch scores to obtain an illuminant estimation for the image.

#### 3.1. Network architecture

The proposed network consists of five layers. Figure 1 shows the architecture of our network, which is a 32x32x3 - 32x32x240 - 4x4x240 - 40 - 3 structure. The input is contrast normalized 32x32 image patches. The first layer is a convolutional layer which filters the input with 240 kernels each of size 1x1x3 with a stride of 1 pixel. The convolutional layer produces 240 feature maps each of size 32x32, followed by a max-pooling operation with 8x8 kernels and stride of 8 pixels that reduces each feature map to a 4x4 feature map. These are reshaped into a 3840 (4x4x240) vector. One fully connected layer of 40 nodes come after the reshaping. The last layer is a simple linear regression with a three dimensional output that gives the illuminant estimate.

#### 3.2. Contrast normalization

In order to be robust across lighting condition and since in color constancy the illuminant has to be estimated up to a scale factor, all the extracted patches are contrast normalized. Among the different contrast enhancement techniques we have chosen the global histogram stretching as does not change the relative contributions of the three color channels.

#### 3.3. Pooling

In the convolution layer, the contrast normalized image patches are convolved with 240 filters and each filter generates a feature map. We then apply max-pooling on each feature map to reduce the filter responses to a lower dimension. In contrast with object recognition scenario, where pooling tends to be performed on small neighborhoods, we observe that even in case of spatially varying illuminants, these are

locally homogeneous, i.e. the same illuminant tends to be present on all the locations of a 32x32 patch. This permits the use of larger pooling kernels.

#### 3.4. ReLU nonlinearity

Instead of traditional sigmoid or tanh neurons, we use Rectified Linear Units (ReLUs) [29] in the fully connected layer. Krizhevsky et al. demonstrated that that ReLUs enable the network to train several times faster compared to using tanh units while achieving almost identical performance [26].

#### 3.5. CNN features

Together with learning an ad-hoc CNN for the color constancy problem, we also investigate how a pre-trained one works on this problem. To this end, we extract a 4096-dimensional feature vector from each image using the Caffe [22] implementation of the deep CNN described by Krizhevsky et al. [26]. Features are computed by forward propagation of a mean-subtracted  $227 \times 227$  RGB RAW image through five convolutional layers and two fully connected layers. More details about the network architecture can be found in [26, 22]. The CNN was discriminatively trained on a large dataset (ILSVRC 2012) with image-level annotations to classify images into 1000 different classes. Features are obtained by extracting activation values of the last hidden layer. The extracted features are then used as input to a linear Support Vector Regressor (SVR) [10] to estimate the illuminant color for each image. We refer to this method as AlexNet+SVR in the experimental results.

### 4. Experimental Setup

The aim of this section is to investigate if the proposed algorithm can outperform state-of-the-art algorithms in the illuminant estimation on a standard dataset of RAW images.

#### 4.1. Image Datasets and Evaluation Procedure

To test the performance of the proposed algorithm, a standard dataset of RAW camera images having a known color target are used. It is captured using high-quality digital SLR cameras in RAW format, and is therefore free of any color correction. The dataset [16] was originally available in sRGB-format, but Shi and Funt [31] reprocessed the raw data to obtain linear images with a higher dynamic range (14 bits as opposed to standard 8 bits). The dataset has been acquired using a Canon 5D and a Canon 1D DSLR cameras and consists of a total of 568 images. The Macbeth ColorChecker (MCC) chart is included in every scene acquired, and this allows to accurately estimate the actual illuminant of each acquired image. Examples of images within the RAW dataset are reported in Figure 2.

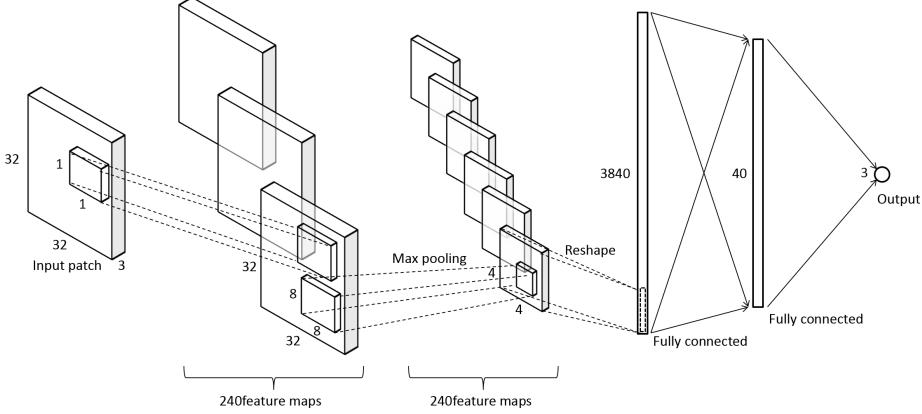


Figure 1. The architecture of our CNN.

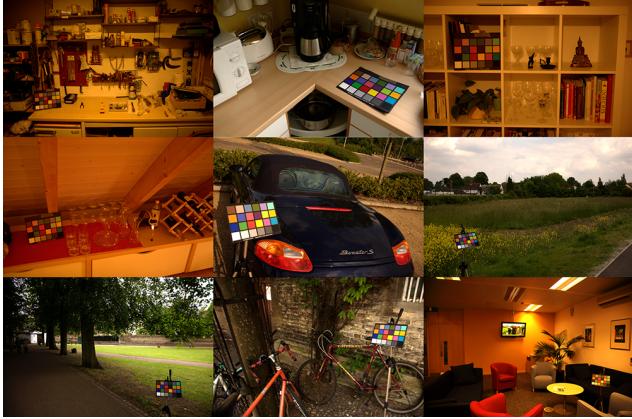


Figure 2. Example of images within the RAW dataset.

## 4.2. Error metric

The error metric considered, as suggested by Hordley and Finlayson [21], is the angle between the RGB triplet of estimated illuminant ( $\rho_w$ ) and the RGB triplet of the measured ground truth illuminant ( $\hat{\rho}_w$ ):

$$e_{\text{ANG}} = \arccos \left( \frac{\rho_w^T \hat{\rho}_w}{\|\rho_w\| \|\hat{\rho}_w\|} \right). \quad (4)$$

## 4.3. Benchmark algorithms

Different benchmarking algorithms for color constancy are considered. Since each image of the dataset contains only one MCC, only global color constancy algorithms based on the assumption of uniform illumination can be compared. Six of them are generated varying the three variables ( $n, p, \sigma$ ) in Equation 3, and correspond to well known and widely used color constancy algorithms. The values chosen for  $(n, p, \sigma)$  are reported in Table 1 and set as in [19]. The algorithms are used in the original authors' implementation which is freely available online (<http://lear.inrialpes.fr/people/vandeweijer/code/ColorConstancy.zip>).

The seventh algorithm is the pixel-based Gamut Mapping [18]. The value chosen for  $\sigma$  is also reported in Table 1. The other algorithms considered are illumination chromaticity estimation via Support Vector Regression (SVR [15]); the Bayesian (BAY [16]); the Natural Image Statistics (NIS [17]); the High Level Visual Information [34]: bottom-up (HLVI BU), top-down (HLVI TD), and their combination (HLVI BU&TD); the Spatio-Spectral statistics [8]: with Maximum Likelihood estimation (SS ML), and with General Priors (SS GP); the Automatic color constancy Algorithm Selection (AAS) [2] and the Automatic Algorithm Combination (AAC) [2]; the Exemplar-Based color constancy (EB) [23]; the Face-Based (FB) color constancy algorithm [3] using GM or SS ML when no faces are detected.

Table 1. Values chosen for  $(n, p, \sigma)$  for the state-of-the-art algorithms which are instantiations of Eq.3.

Algorithm	$n$	$p$	$\sigma$
Gray World (GW)	0	1	0
White Point (WP)	0	$\infty$	0
Shades of Gray (SoG)	0	4	0
general Gray World (gGW)	0	9	9
1st-order Gray Edge (GE1)	1	1	6
2nd-order Gray Edge (GE2)	2	1	1
Gamut Mapping (GM)	0	0	4

The last algorithm considered is the Do Nothing (DN) algorithm which gives the same estimation for the color of the illuminant ( $\mathbf{I} = [1 1 1]$ ) for every image, i.e. it assumes that the image is already correctly balanced.

## 4.4. CNN learning

We train our CNN on 32x32 random patches taken from images in RAW format. Images have been resized to  $\max(w, h) = 1200$ . The net is learned using a three-fold

cross validation on the folds provided with the dataset: for each run one is used for training, one for validation and the remaining one for test. For training, we assign each patch with the illuminant groundtruth associated to the image to which it belongs. At testing time, we generate a single illuminant estimation per image by pooling the predicted patch illuminants. By taking image patches as input, we have a much larger number of training samples compared to using the whole image on a given dataset, which particularly meets the needs of CNNs. Net parameters have been learned using Caffe [22] with euclidean loss. The learned net is then fine-tuned by using as loss the angular error and adding knowledge about the way local estimates are pooled to generate a single global estimate for each image.

## 5. Results and Discussion

In Table 2 the minimum, the 10<sup>th</sup>-percentile, the median, the average, the 90<sup>th</sup>-percentile, and the maximum of the angular errors obtained by the considered state-of-the-art algorithms and the proposed approach on the RAW dataset are reported. The table is divided into three blocks and for each of them the best result for each statistic is reported in bold. The first block includes statistic-based algorithms, the second one learning-based algorithms, and the third one the different variants of the proposed approach.

From the results it is possible to see that the deep CNN pre-trained on ILSVRC 2012 [26] coupled with SVR (i.e. AlexNet+SVR) is already able to outperform most statistic-based algorithms and some learning-based ones. The next entry is the angular error made by our CNN on the patches. It is possible to see that, with respect to the median error, the proposed approach is able to outperform half of the learning-based algorithms considered. The next two entries are the results obtained by our approach by pooling patch-based illuminant estimations over the whole image. Two very simple pooling strategies have been considered, i.e. average and median pooling. It is possible to see that both outperform most of the state-of-the-art algorithms in terms of median error, with the average-pooling having a maximum error just 0.3% worse than the best algorithm in the state-of-the-art. The last entry reported is the fine-tuned CNN with median-pooling and angular error loss. We can notice that it reaches both a median, an average and a maximum angular error better than all the state-of-the-art algorithms considered. The improvement is of 1.5%, 5.1% and 0.2% respectively but it is remarkable that they have been achieved by the same algorithm, while the best values in the state-of-the-art for the same statistics were obtained by three different learning-based algorithms. i.e. FB+GM, EB and SS GP respectively.

Figure 3 reports some examples of images on which the fine-tuned CNN makes the largest estimation errors. From left to right we report the original RAW image, the image

corrected with the groundtruth illuminant, the image corrected with the CNN estimate, and the image corrected with the algorithm in the state-of-the-art making the best estimate on that image. Once we have an estimate of the global illuminant color  $\mathbf{I}$ , each pixel in the image is color corrected using the von Kries model [35], i.e.:  $\rho_{out}(x, y) = \text{diag}(\mathbf{I}^{-1})\rho_{in}(x, y)$ .

### 5.1. Effects of parameters

Several parameters are involved in the CNN design. In this section, we examine how these parameters affect the network performance. The network architecture has been chosen by starting from a 7-layer deep CNN similar to [26] and removing layers until no further improvement in performance was possible. In Figure 4 we report how the different parameters affect the illuminant estimation performance. Each point on the graphs represents the best result that can be obtained by fixing a parameter at the value indicated and searching over all the possible combinations of the other ones.

**Kernel size** Figure 4.a shows how the performance varies with the width of the convolution kernels. We can see from Figure 4.a that the estimation error decreases by decreasing the kernel size. At first this could be surprising, since in different domains larger kernels are preferred. However, it is not the first time that such small kernels are used, see [32]. From the color constancy point of view, this choice of kernel size confirms the finding of Cheng et al. [9], where they showed that spatial information does not provide any additional information that cannot be obtained directly from the color distributions.

**Number of kernels** Figure 4.b shows how the performance varies with respect to the number of convolution kernels. It is possible to see that the CNN tends to prefer an intermediate number of kernels.

**Pooling size** Figure 4.c shows how the performance varies with respect to the number of convolution kernels. It is possible to see that the CNN tends to prefer an intermediate pooling size.

**Number of fully connected units** Figure 4.d shows how the performance varies with respect to the number of fully connected units. The plot shows that better performance can be reached with a number of fully connected units around 40.

**Patch size** Since in our experiment the illuminant is estimated as the median illuminant of all patches sampled, we examine how the patch size affects performance. For every patch size, the same number of patches is randomly extracted from each image being sure that none of them contained the reference color chart. Figure 4.e shows the change of performance with respect to patch size. From the plot we see that a larger patch size results in better performance.

Table 2. Angular error statistics obtained by the state-of-the-art algorithms considered on the RAW dataset.

Algorithm	Min	10 <sup>th</sup> prc	Med	Avg	90 <sup>th</sup> prc	Max
DN	3.72	10.38	13.55	13.62	16.45	27.37
GW	0.18	1.88	6.30	6.27	10.12	24.84
WP	0.08	1.38	5.61	7.46	15.68	40.59
SoG	0.18	1.04	4.04	4.85	9.71	19.93
gGW	<b>0.03</b>	0.82	3.45	4.60	9.68	22.21
GE1	0.16	1.82	4.55	5.21	9.78	19.69
GE2	0.26	2.06	4.43	5.01	<b>8.93</b>	<b>16.87</b>
GM	0.05	<b>0.40</b>	<b>2.28</b>	<b>4.10</b>	11.08	23.18
SVR	0.66	3.36	6.67	7.99	14.61	26.08
BAY	0.10	1.17	3.44	4.70	10.21	24.47
NIS	0.08	0.93	3.13	4.09	8.57	26.20
HLVI BU	0.06	0.75	2.54	3.30	6.59	17.51
HLVI TD	0.11	0.85	2.63	3.65	7.53	25.24
HLVI BU&TD	0.13	0.77	2.47	3.38	6.97	25.24
SS ML	0.06	0.85	2.93	3.55	7.23	15.25
SS GP	0.07	0.82	2.90	3.47	7.00	<b>14.80</b>
AAS	<b>0.03</b>	0.77	3.16	4.18	9.15	22.21
AAC	0.05	0.90	2.90	3.74	7.93	14.98
EB	0.14	0.73	2.24	<b>2.77</b>	<b>5.52</b>	19.44
FB+GM	0.05	<b>0.40</b>	<b>2.01</b>	3.67	9.50	23.18
FB+SS GP	0.08	0.75	2.57	3.18	6.67	<b>14.80</b>
AlexNet+SVR	0.12	0.98	3.09	4.74	11.18	29.15
CNN per patch	<b>0.00</b>	0.99	2.69	3.67	7.79	30.93
CNN average-pooling	0.04	0.99	2.44	3.18	6.37	14.84
CNN median-pooling	0.06	0.97	2.32	3.07	6.15	19.04
CNN fine-tuned	0.06	<b>0.69</b>	<b>1.98</b>	<b>2.63</b>	<b>5.54</b>	<b>14.77</b>

## 5.2. Local illuminant estimation

Our CNN predicts the illumination on small image patches, so it can be easily used to predict local illuminants as well as giving a global illuminant estimate for the entire image. Given the performance of the per patch error in Table 2 we expect our CNN to perform well even on local estimation. We perform here a preliminary test by using our learned CNN as-is on a dataset of synthetic images: the images are taken from the previous RAW dataset [31] and on half of each image we manually changed the illuminant, resulting in two illuminants for each image.

Among the algorithms in the state-of-the-art able to deal with non-uniform illumination, e.g. [27, 30, 11, 5, 24, 4] we report as comparison the results of the Multiple Light Sources (MLS) [20] using White Point (WP) and Gray World (GW) algorithms, grid based sampling, in the clustering version setting the number of clusters equal to the number of lights in the scene, i.e. two.

The numerical results are reported in Table 3, while a couple of examples are given in Figure 5. For the proposed approach we report three different entries: the first one is the error on each patch; the second and third ones are the patch-by-patch errors by taking into account the spatial arrangement of the patches to perform a spatial filtering of the

local estimates: the former employs a 3x3 Gaussian filter, the latter a 3x3 median filter.

## 6. Conclusions

In this work we have developed a CNN for color constancy. Our algorithm combines feature learning and regression as a complete optimization process, which enables us to employ modern training techniques to boost performance. The experimental results showed that our algorithm achieves state of the art performance on a standard dataset of RAW images outperforming 21 algorithms in the state-of-the-art belonging to both statistic-based and learning-based classes. Furthermore, a preliminary test, shows that our algorithm can be adapted to estimate local illuminants.

As future work we plan to investigate other pooling strategies to combine patch-based illuminant estimations into a global one. We plan also to investigate if additional information can be fed to the CNN to further improve the performance. We will also conduct a more thorough study about the extension of the proposed approach to local illuminant estimation, conducting the experiments on larger datasets and comparing with more algorithms in the state-of-the-art.

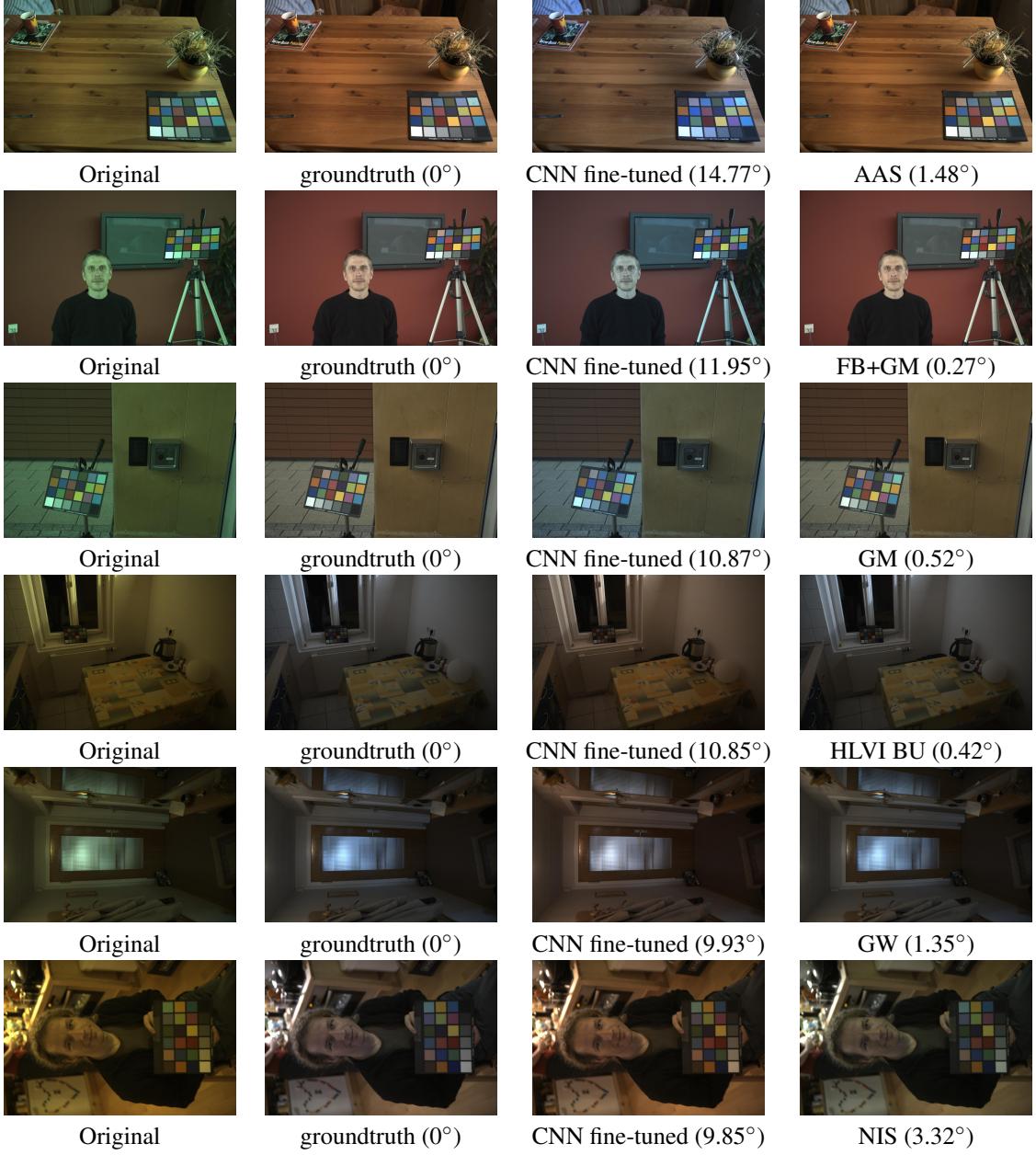


Figure 3. Examples of images on which the fine-tuned CNN makes the largest estimation errors. Left to right: original RAW image, correction with the groundtruth illuminant, correction with the CNN estimate, and correction with the algorithm in the state-of-the-art making the best estimate on the given image.

Table 3. Angular error statistics obtained on the synthetic RAW dataset of images with spatially varying illumination.

Algorithm	Min	10 <sup>th</sup> prc	Med	Avg	90 <sup>th</sup> prc	Max
DN	5.90	9.99	13.38	13.62	17.09	27.71
MLS+GW	0.12	3.35	8.03	8.72	14.66	32.98
MLS+WP	0.23	2.59	6.09	7.03	13.15	33.58
CNN per patch	0.00	1.01	2.83	3.72	7.78	31.78
CNN gaussian filtering	0.01	0.99	2.71	3.50	7.16	27.03
CNN median filtering	0.01	0.96	2.66	3.39	6.88	23.13

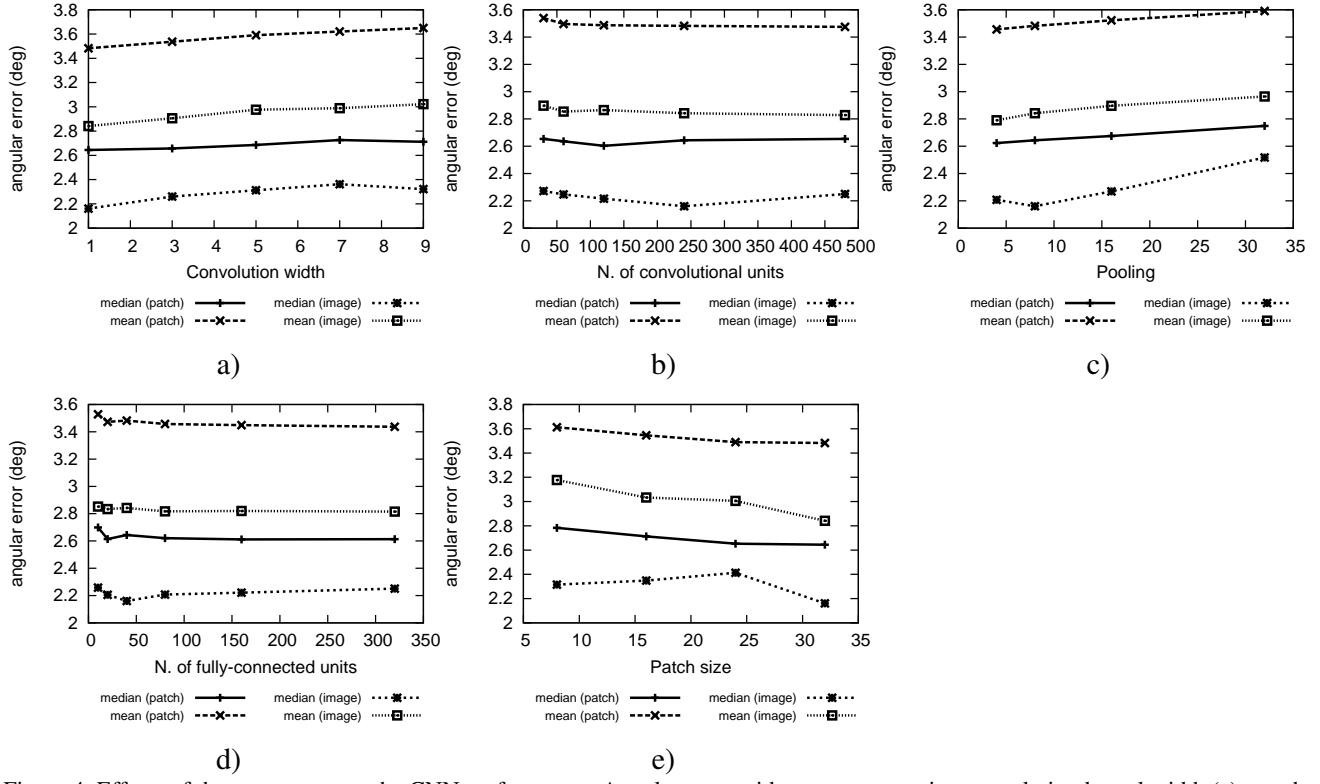


Figure 4. Effects of the parameters on the CNN performance. Angular error with respect to varying convolution kernel width (a), number of convolutional kernels (b), pooling size (c), number of fully connected units (d) and input patch size (e). Each point corresponds to the best performance that can be obtained by fixing a single parameter at the value indicated by trying all the combinations for the other ones.

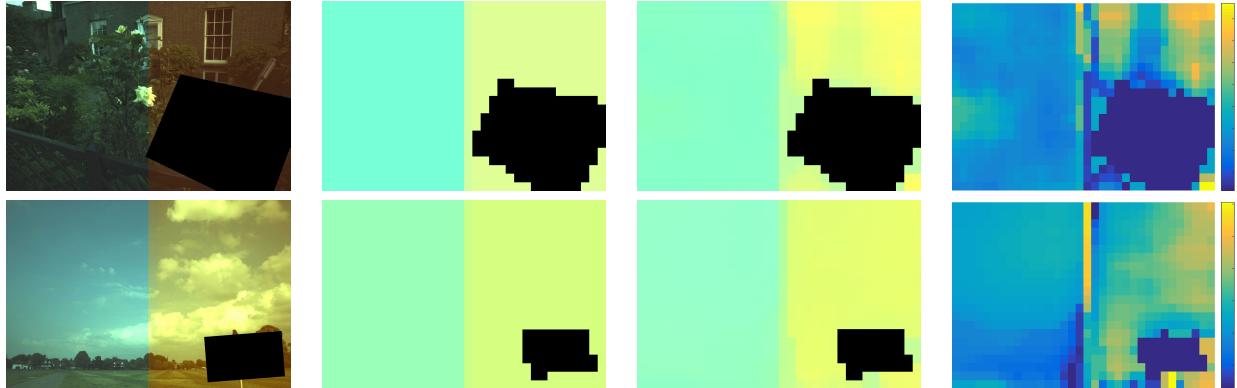


Figure 5. Examples of local illumination estimation. Left to right: original image, groundtruth illumination, local estimation, local angular error map.

## References

- [1] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Improving color constancy using indoor-outdoor image classification. *IEEE Transactions on Image Processing*, 17(12):2381–2392, 2008. 2
- [2] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Automatic color constancy algorithm selection and combination. *Pattern recognition*, 43(3):695–705, 2010. 2, 4
- [3] S. Bianco and R. Schettini. Color constancy using faces. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 65–72. IEEE, 2012. 2, 4
- [4] S. Bianco and R. Schettini. Adaptive color constancy using faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1505–1518, 2014. 2, 6
- [5] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troger, and A. Kaup. Color constancy and non-uniform illumination: Can existing algorithms work? In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 774–781. IEEE, 2011. 6

- [6] G. Buchsbaum. A spacial processor model for object color perception. *J. Franklin Inst.*, 310:1–26, 1980. 2
- [7] V. Cardei, B. Funt, , and K. Barnard. White point estimation for uncalibrated images. *Proceedings of the IS&T/SID 7th Color Imaging Conference*, pages 97–100, 1999. 2
- [8] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012. (DOI:10.1109/TPAMI.2011.252). 2, 4
- [9] D. Cheng, D. K. Prasad, and M. S. Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 5
- [10] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, V. Vapnik, et al. Support vector regression machines. *Advances in neural information processing systems*, 9:155–161, 1997. 3
- [11] M. Ebner. Color constancy based on local space average color. *Machine Vision and Applications*, 20(5):283–301, 2009. 6
- [12] G. Finlayson, S. Hordley, and P. Hubel. Color by correlation: a simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001. 2
- [13] D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990. 2
- [14] B. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough? *Proceedings of the 5th European Conference on Computer Vision*, pages 445–459, 1998. 2
- [15] B. Funt and W. Xiong. Estimating illumination chromaticity via support vector regression. In *Color and Imaging Conference*, volume 2004, pages 47–52. Society for Imaging Science and Technology, 2004. 4
- [16] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pages 1–8, 2008. 2, 3, 4
- [17] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(33):687–698, 2011. 2, 4
- [18] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2–3):127–139, 2010. 4
- [19] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: survey and experiments. *IEEE Transactions on Image Processing*, 20(9):2475–2489, 2011. 4
- [20] A. Gijsenij, R. Lu, and T. Gevers. Color constancy for multiple light sources. *Image Processing, IEEE Transactions on*, 21(2):697–707, 2012. 6
- [21] S. Hordley and G. Finlayson. Re-evaluating color constancy algorithms. *Proc. of the 17th International Conference on Pattern Recognition*, pages 76–79, 2004. 2, 4
- [22] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014. 3, 5
- [23] H. R. V. Joze and M. S. Drew. Exemplar-based colour constancy. In *BMVC*, pages 1–12, 2012. 3, 4
- [24] H. R. V. Joze and M. S. Drew. Exemplar-based color constancy and multiple illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(5):860–873, 2014. 3, 6
- [25] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. L. Cun. Learning convolutional feature hierarchies for visual recognition. In *Advances in neural information processing systems*, pages 1090–1098, 2010. 1
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1, 3, 5
- [27] E. H. Land et al. *The retinex theory of color vision*. Scientific America., 1977. 6
- [28] D. Lee and K. N. Plataniotis. A taxonomy of color constancy and invariance algorithm. In *Advances in Low-Level Color Image Processing*, pages 55–94. Springer, 2014. 1
- [29] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010. 3
- [30] E. Provenzi, C. Gatta, M. Fierro, and A. Rizzi. A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1757–1770, 2008. 6
- [31] L. Shi and B. V. Funt. Re-processed version of the gehler color constancy database of 568 images. [online]. available: <http://www.cs.sfu.ca/colour/data/>. last access: Nov. 1, 2011. 3, 6
- [32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014. 5
- [33] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007. 2
- [34] J. van de Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. *IEEE International Conference on Computer Vision (ICCV 2007)*, pages 1 –8, 2007. 2, 4
- [35] J. von Kries. Chromatic adaptation. *Festschrift der Albrecht-Ludwigs-Universität*, pages 145–158, 1902. 5