Color-based Data Augmentation for Reflectance Estimation

Article *in* Color and Imaging Conference · November 2018

DOI: 10.2352/ISSN.2169-2629.2018.26.284

5 authors, including:

Hassan AHMED Sial
COMPUTER VISION AND ROBOTICS (VICOROB)
9 PUBLICATIONS **20** CITATIONS
SEE PROFILE

Ramon Baldrich
Autonomous University of Barcelona
49 PUBLICATIONS **673** CITATIONS
SEE PROFILE

Robert Benavente
Autonomous University of Barcelona
31 PUBLICATIONS **3,952** CITATIONS
SEE PROFILE

Maria Vanrell
Autonomous University of Barcelona
86 PUBLICATIONS **1,598** CITATIONS
SEE PROFILE

Color-based Data Augmentation for Reflectance Estimation

H. Sial, S. Sancho-Asensio, R. Baldrich, R. Benavente, M. Vanrell
Computer Vision Center/ Universitat Autnoma de Barcelona; Bellaterra (Barcelona), Spain

Abstract

Deep convolutional architectures have shown to be successful frameworks to solve generic computer vision problems. The estimation of intrinsic reflectance from single image is not a solved problem yet. Encoder-Decoder architectures are a perfect approach for pixel-wise reflectance estimation, although it usually suffers from the lack of large datasets. Lack of data can be partially solved with data augmentation, however usual techniques focus on geometric changes which does not help for reflectance estimation. In this paper we propose a color-based data augmentation technique that extends the training data by increasing the variability of chromaticity. Rotation on the red-green blue-yellow plane of an opponent space enable to increase the training set in a coherent and sound way that improves network generalization capability for reflectance estimation. We perform some experiments on the Sintel dataset showing that our color-based augmentation increase performance and overcomes one of the state-of-the-art methods.

Introduction

Intrinsic image decomposition is an inverse optics problem that seeks to separately estimate light and material properties such as shading, illumination, specularities or material reflectance from a single image. Barrow and Tenenbaum [1] were the first to introduce the decomposition of an image as the product of two intrinsic terms, such as

$$Im(x,y) = Re(x,y) \cdot Sh(x,y), \tag{1}$$

where image Im can be decomposed at each pixel (x,y) by the product of the intrinsic Reflectance, Re, and Shading, Sh. This is the basic model used in the majority of existing methods. A generalization of this model to multiple intrinsic terms can be found in [2]. Reflectance, or albedo, accounts for reflection of light from object surface[1] and it is material dependent only, while shading represents reflection variation due to surface geometry, light position and inter-reflections. Such a decomposition is an ill-posed problem as there are many possible solutions, that is why it is considered to be a challenging problem having applications in both vision and graphics, where examples varies from recognition, to segmentation, to shadow removal to re-rendering.

The classical approach to the problem was based on the assumption that smooth changes in the image values were due to shading variation whereas sharp changes were caused by differences on the reflectance of the surfaces in the scene. This assumption, known as the Retinex assumption in the intrinsic decomposition framework, was the basis of several early methods either on greyscale images [3, 4, 5] or color images [6, 7]. Later works added other constraints to the problem by assuming sparsity of reflectances in the scene [8, 9], by adding priors on shape and illumination [10], or by introducing color-based surface descriptors [11].

After re-birth of convolution neural networks (CNN) in the last years, this visual task is again getting researchers' interest and now the community is trying to solve it using the advantages of computation power and flexibility of CNNs. In this new framework, new problems arise, such as those related to the training of these deep architectures that require the availability of large datasets. Building datasets for this kind of application requires a big effort on controlling lights, positioning, and careful selection of objects and their materials, which is only possible in laboratory conditions like in the MIT dataset [12]. That becomes even harder when thousands of images are needed. To solve this handicap, alternative solutions are used, such as using synthetic datasets where reflectance and shading ground truth can be easily computed by the renderer [13], or using data augmentation techniques that provide dataset extensions that can add value to the training set, as in [14].

Data augmentation techniques focus on extending the training sets adding transformed versions of the own dataset images. Usually, image transformations are based on geometric properties such as scaling, translation, spatial rotation, flipping, adding noise, changing lighting or perspective conditions. In this work we propose a new data augmentation technique based on color to extend the dataset from a photometric point of view instead of the classical geometric changes, which usually do not have any effect for reflectance estimation. We propose a color transformation based on chromaticity rotation that fits with the constrains of the reflectance estimation problem since it can be applied to the original image and to reflectance without affecting shading. In what follows we review some CNN architectures that have been used to deal with reflectance estimation. Afterwards we explain our color-based augmentation method and the CNN architecture we have used in our experiments. Such architecture has been trained on the Sintel dataset [13] that is the largest available. Finally, we present the results showing that our color-based augmentation improves the performance on the tested architecture and overcomes the state of the art network on the dataset used.

Related works

Here we review deep learning architectures which have been proposed for intrinsic image estimation using a single image as input. Narihira *et al.* [15] were the first to introduce end-to-end multi-scale deep regression networks to predict reflectance and shading using a single image. Their two-stage architecture, inspired by the work of [16], allows the network to learn features at both local and global stages. First, the network learns features at coarse

[1] Reflectance estimation in computer vision seeks to classify image pixels in different colored materials from a single uncalibrated image, without physical considerations of the true acquired material properties.

level which on a later stage are merged with a second finer-scale network. The last layers of the second network are divided in two branches to learn reflectance and shading independently without assuring the fulfillment of the basic intrinsic model definition of Eq.1. All the layers of the network are fully convolutional making it scale-invariant.

Shelhamer *et al.* [17] decompose a single image into depth, shading and reflectance images. They first infer depth from the input image using a fully convolutional network (FCN) [18] and then decompose the input image and the estimated depth into reflectance and shading by joint optimization, similarly to what Barron and Malik [10] and Chen and Koltun [19] had previously done using depth from Kinect. It is not an end-to-end architecture to predict intrinsic images but it was the first approach to decompose the input image into shape, shading and illumination.

Zhou *et al.* [20] used a deep network to predict relative reflectance between patches based on human annotated data and then enforced energy minimization with a conditional random field (CRF) to disintegrate the image in reflectance and shading.

Kim *et al.* [21] presented a joint convolution neural field (JCNF) to jointly predict depth and intrinsic images from a single image. JCNF design shares layers between depth, shading and albedo pipelines and merge a gradient scale network for each task.

More recently, Shi and Dong introduced in [22] a large dataset based on ShapeNet [23] models and trained a mirror-link encoder-decoder CNN to predict reflectance, shading and specular components of images. It is the first deep-learning model not based on the assumption of diffuse reflectance or Lambertian surfaces by considering $Im(x,y) = Re(x,y) \cdot Sh(x,y) + Sp(x,y)$. The proposed architecture has an encoder to progressively encode and down-sample the features from the input image, and then three separate decoders provide outputs for reflectance, shading and specularities.

Lettry *et al.* [24] presented a deep residual network based on the powerful and simple-to-train generative adversarial networks (GAN) [25]. The network first predicts shading. Reflectance is defined to be the element-wise division between the original image and the shading. The interesting point of this approach is the definition of the loss which considers a data term (difference between estimations and ground truth), an edge energy term (difference between the gradients of estimations and ground truth), and a perceptually-motivated adversarial term.

Most recently, Fan *et al.* [26] presented a modified CNN to force reflectances surfaces to be uniform. It is based on the direct intrinsic [15] work, adding a new pathway that extract the sensitive areas where the output should be smooth. At the end they apply a learn difussion operation to overcome the non uniformity of the direct intrinsic output.

Finally, Baslamisli *et al.* [27] introduced a new synthetic dataset based on ShapeNet [28] by assigning random color to each homogeneous part of objects. They also introduced an approach based on two deep learning architectures with a new image formation loss. The first network, called IntrinsicNet, is a traditional end-to-end encoder/decoder network. In the second one they exploit the concept of classical Retinex theory. It is a network trained in two stages. In the first stage they learn reflectance and shading gradients. In the second stage, the network is used to obtain the intrinsic decomposition by combining both the gradient outputs of the first stage and the input image.

Color-based data augmentation

Deep convolutional architectures have become a flexible tool to solve problems that has provoked methodological changes in the design of computer vision solutions. One of these changes is that important research interests have shifted from the design of the solutions towards the setting of adequate experimental setups to find the best hyper-parameters to reach best performances. As we mentioned before, data augmentation in the training stage is one of these methodological aspects that attracts some attention.

Here we propose a data augmentation technique that seeks to extend the color diversity of datasets. It is based on the idea that for any given image we can apply a chromaticity rotation without affecting intensity. This is a transformation that in the intrinsic model affects reflectance but does not affect shading. This chromatic change should be equally applied to the ground truth reflectance and to the original image to hold the intrinsic decomposition property. In this way we achieve an extension of the image dataset from a photometric point of view instead of the most common geometric extensions. This augmentation increases the dataset color variability and we prove it increases the generalization capabilities of the network architecture.

To this end we use an opponent-like (CieLab-like) transformation where intensity and chromaticity are separated. Afterwards, we apply a rotation on the plane formed by the red-green and blue-yellow axes. The transform to this opponent space is a linear transform on the RGB color space given by the following equations:

$$\begin{array}{rcl} O_1 & = & (R+G+B-1.5)/1.5, \\ O_2 & = & (R-G), \\ O_3 & = & (R+G-2B)/2. \end{array} \quad (2)$$

It is based on the one proposed by Platanoits *et al.* [29] but normalizing and shifting the three axes within the range $\left[-1,1\right]$. This space was conceived to achieve certain physiological inspiration on uncalibrated RGB, and has provided interesting results in computer vision.

Augmentation can be reached by random rotations which are computed on the O_2-O_3 plane, followed by a stretching transform, denoted with the function S, that keeps the original contrast of the image which could be reduced by the new range after the rotation

$$\begin{array}{rcl} O_2^\theta & = & S(O_2\cos(\theta)+O_3\sin(\theta)), \\ O_3^\theta & = & S(O_3\cos(\theta)-O_2\sin(\theta)), \\ S(O_i) & = & (O_i-\min(O_i)\frac{\max(O_i^\theta)-\min(O_i^\theta)}{\max O_i-\min O_i}+\min(O_i^\theta). \end{array} \quad (3)$$

We can see an example of the effects of this chromaticity rotation in figure 1, where we also can see how the transformation holds the hypothesis of preserving the shading while changing reflectance.

Experiments

In order to evaluate the effect of the proposed data augmentation on the reflectance estimation problem we have selected a UNet-like architecture proposed in [30] which has been trained on the Sintel dataset [13]. In next sections we explain all the experimental details.

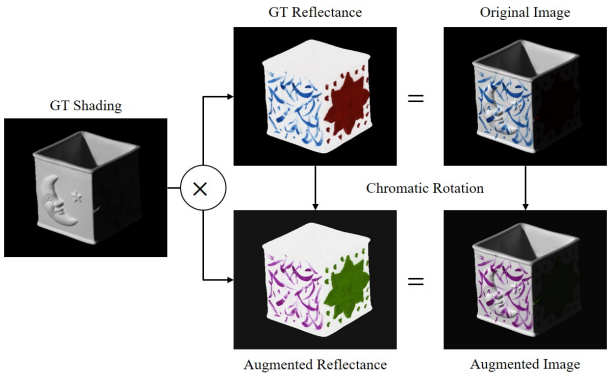

Figure 1. *Example of chromaticity rotation for color-based augmentation.*

CNN architecture

Reflectance estimation is a regression problem, which establishes the sort of network architecture needed to tackle the problem. Consequently, it needs the output to be of the same dimensions as the input. One of the first proposals that successfully dealt with this kind of pixel-wise architectures was Segnet [31], which was applied to a segmentation task. It is based on an information contracting stream followed by an expanding one with a final softmax layer to produce a label per pixel. This contracting and expanding scheme has been called encoder/decoder. The architecture in the contracting stream is a VGG16 network where the fully connected layers have been removed, while the expanding stream upsamples the input taking into account the indexed results in the corresponding pooling operation of its counterpart in the contracting stream.

When the problem is segmentation, the lack of resolution due to the upsampling is not as evident as in regression problems where the output space presents a higher range of values. In our case we have continuous values in a 3D color space. To overcome this problem [32] proposed the U-Net architecture. They concatenated the upsampled version of a given layer with the before-downsampling version of that layer in the encoding phase. Another mechanism to reduce the coarseness of the output is to introduce hypercolumn information in the pipeline [33]. Given a pixel, its hypercolumn is the concatenation of all the corresponding activations at each convolutional layer. The mechanism up samples the activation of its layer outputs, merging all of them without any convolutional operation. Once the information is gathered, the system applies a convolutional scheme to figure out how to achieve the proper regression output. At the end, the system will have two outputs that should recover the same reflectance from different infromation sources. The final loss will weight the differences between the groundtruth reflectance and the reflectances on both outputs, decoder stream and the hyper-column process.

One last procedure to speed-up the operations and provide more flexibility to the model is the substitution of the current convolutional layers by inception modules where the $n \times n$ filters are substituted by $n \times 1$ and $1 \times n$ filters [34]

The three above-mentioned ideas have been recently combined in the EURNet architecture proposed in [30] for reflectance estimation. Figure 2 shows an schematic overview of the architecture used in our experiments.

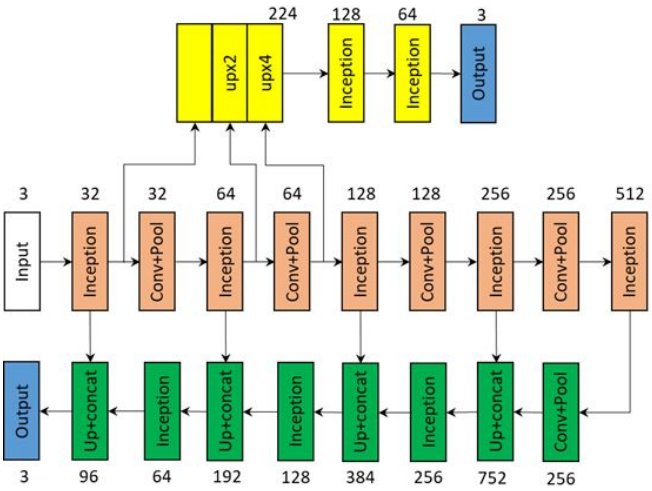

Figure 2. *EURNet architecture scheme. The number of channels is depicted on the top of each box. Reddish boxes are the encoding stream. Green boxes are the decoding stream. Yellow boxes represent hyper-column layers. Blue boxes represent the outputs, reflectance estimations that are going to be linked by the loss.*

Datasets

Our Network is trained on MPI Sintel dataset [13] and we followed the same methodology as in [15, 19] where they only used the *clean-pass* images as *final-pass* images have some graphic effects and do not fulfill the property of Eq.1 which is an essential condition. We used a total of 890 images, from 18 scenes having 50 frames each (one of the scenes has only 40 frames), and as previous works we used a two-fold cross validation, that means, our network is trained on 50% of the images and tested on the rest, this is called Image Split set. We also used the scene split introduced in [15] where half of the scene are used for training and the rest of scenes for testing.

Evaluation metrics

There are three usual metrics for performance evaluation in intrinsic image decomposition. Two of them are data-related metrics, namely, the mean-squared error and the local mean-squared error, and the third is a perceptually-motivated metric, the structural dissimilarity index.

Mean-squared error (MSE) measures the average of the squares of the pixel-wise errors between the estimated reflectance and the ground truth. As in previous works[12, 19, 15], we remove intensity effects and compute MSE as:

$$MSE(x,\hat{x}) = \sum_{i=1}^{N} \frac{||x_i - \hat{\alpha}\hat{x}_i||^2}{N}, \tag{4}$$

where N is the number of pixels in x, and $\hat{\alpha} = argmin_\alpha ||x_i - \alpha\hat{x}_i||^2$ is a parameter which adjusts the absolute brightness of the estimation to minimize the error because the ground truth is only defined up to a scale factor, that is, $\hat{\alpha}$ adds scale invariance to the measure.

Local mean-squared error (LMSE) measures the average of the

scale-invariant MSE computed on overlapping square windows of size 10% of the image size along its larger dimension. The overlap between neighboring windows is 50%.

Structural dissimilarity index (DSSIM) is a distance metric derived from structural similarity index (SSIM). SSIM characterizes image similarity as perceived by human observers and accounts for multiple independent structural and luminance differences. It is defined by

$$SSIM(x,\hat{x}) = \frac{(2\mu\hat{\mu}+c_1)(2\sigma_{x\hat{x}}+c_2)}{(\mu^2+\hat{\mu}^2+c_1)(\sigma^2+\hat{\sigma}^2+c_2)}, \quad (5)$$

where μ is the mean of x, $\hat{\mu}$ is the mean of $\hat{x}$, σ^2 is the variance of x, $\hat{\sigma}^2$ is the variance of $\hat{x}$, $\sigma_{x\hat{x}}$ is the covariance between x and $\hat{x}$. c_1 and c_2 stabilize the division when denominator approaches zero. They depend on the dynamic range of the pixel-values.Based on this definition of SSIM, DSSIM is defined as

$$DSSIM(x,\hat{x}) = \frac{1-SSIM(x,\hat{x})}{2}. \quad (6)$$

Experimental setup

Our implementation was based on Keras [35] with Theano backend [36]. We used Adam to optimize the network with an initial learning rate of 0.0002, which is updated with a factor of 0.1 when reaching a plateau. Our images were not cropped but reduced to a resolution of 192×448 and with a batch size of 8. We used 5 dropout layers with 50% dropout rate at regular intervals (more details are given in [30]). As the loss has to measure the difference between the output and the real image, the back-propagation will focus on isolated large errors if the MSE loss is applied. Applying Huber loss alleviates this problem [37]. The weights of the decoder and hyper-column losses are 0.8 and 0.2 respectively.

In order to evaluate the effects on the final performance of our proposed color-based augmentation we performed a two-steps experiment, first we trained EURNet from scratch and then it was fine tunned adding color-based augmentation, we refer to this trained version as EURNet+CA. On the test stage we only took the output from the decoder stream. We used the same methodology for both Image Split and Scene Split.

Results and Discussion

The results of our four experiments are summarized in tables 1 and 2. In the first one we show the performance of several previous methods as reported by Narihira *et al.* [15] and the performance of the two networks we trained, EURNet and EURNet+CA. This first table corresponds to the Image Split experiment, providing the three usual metrics: MSE, LMSE and DSSIM explained in a previous section. We can see that our data augmentation clearly improves the performance with respect to EURNet and overcomes Direct intrinsics [15] that, to the best of our knowledge, is the state of the art on the Sintel dataset. This increase happens for the three evaluated metrics.

In the second table (table 2) we show the results for the Scene Split experiment. Again, we can see that color augmentation improves EURNet on all metrics, although in this case we do not achieve better results than Direct Intrinsics.

In figure 3 we show several examples obtained of the estimated reflectances by the two trained networks EURNet and EURNet+CA (figure 3.(c) and (d), respectively). Although both estimated reflectances still present some degree of blurring in the edges we can see that edges and global color is better in (d) (see the right side of girl hair in first row). Another improvement can be seen in segments corresponding to the same reflectance. In the bottom row we zoom some windows from previous images in columns (b), (c) and (d), where we can better observe these improvements. In figure 3.(e) the face skin is more homogeneous for the EURNEt+CA estimation and in (f) color augmentation has removed all the shading effects that appear in the EURNet estimation. Finally, in (g) we can see that the reflectance of the dark green foliage is also enhanced by the color augmentation.

Table 1. Quantitative results in the Image Split experiment ($\times 100$), lower score is better.

Method	MSE	LMSE	DSSIM
Baseline: reflectance constant	3.69	2.40	22.80
Retinex [12]	6.06	3.66	22.70
Lee et al. [38]	4.63	2.24	19.90
Barron et al. [10]	4.20	2.98	21.00
Chen and Koltun [19]	3.07	1.85	19.60
Direct intrinsics [15]	1.00	0.83	20.14
EURNet	1.09	0.836	21.85
EURNet+CA	**0.77**	**0.65**	**19.05**

Table 2. Quantitative results in the Scene Split experiment ($\times 100$), lower score is better.

Method	MSE	LMSE	DSSIM
Direct intrinsics [15]	**2.01**	**1.31**	**20.73**
EURNet	2.56	1.66	26.31
EURNet+CA	2.31	1.33	24.06

Conclusions

In this paper we present a new data augmentation technique based on color changes instead of on the usual geometric transformations. We seek to overcome the problem of lack of data for training deep neuronal architectures for reflectance estimation. Our proposal is based on a color transformation that changes image chromaticity while preserving intensity and global contrast, these properties perfectly fit for estimating intrinsic reflectance from a single image. Augmentation can be randomly applied on the image dataset, both on the image and the ground truth reflectance.

To evaluate the performance of the proposed approach we performed several experiments on the Sintel dataset. We show that our proposed augmentation noticeable increase the performance overcoming the state of the art on this dataset.

Acknowledgments

This work has been supported by MILETRANS project (TIN2014-61068-R) and FPI Predoctoral Grant (BES-2015-073722) of Spanish Ministry of Economy and Competitiveness.

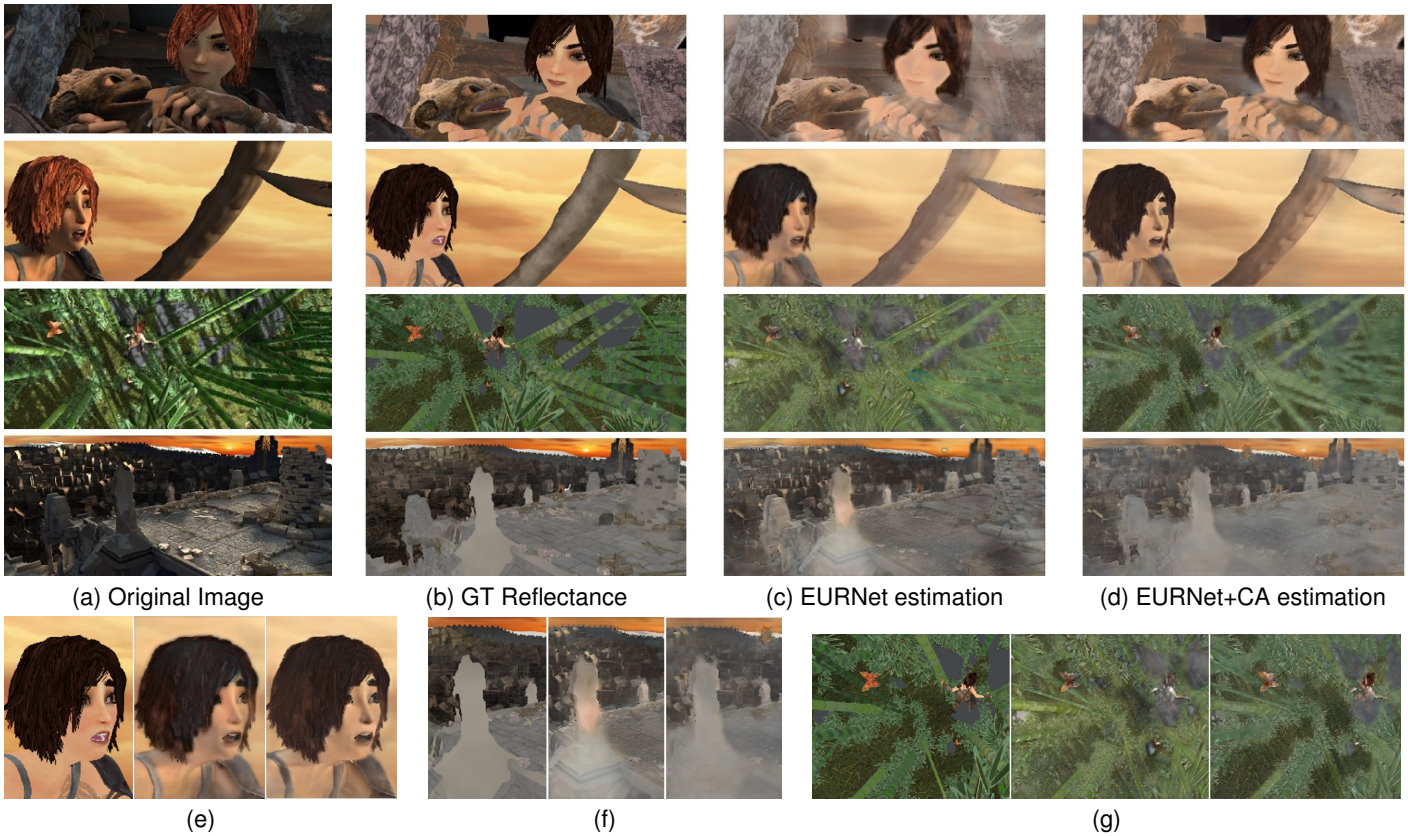

Figure 3. *Reflectance estimation results on Sintel dataset. (a) and (b) are dataset original image and ground-truth reflectance respectively. (c) Estimated reflectance with EURNet network. (d) Estimated reflectance with EURNet trained using our color-based augmentation. In bottom row some detail windows of the previous images.*

References

[1] H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics," *Comput. Vis. Syst*, vol. 2, 1978.

[2] M. Serra, O. Penacchio, R. Benavente, M. Vanrell, and D. Samaras, "The photometry of intrinsic images," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1494–1501, 2014.

[3] Y. Weiss, "Deriving intrinsic images from image sequences," in *International Conference on Computer Vision*, pp. 68–75, 2001.

[4] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1459–1472, 2005.

[5] M. F. Tappen, E. H. Adelson, and W. T. Freeman, "Estimating intrinsic component images using non-linear regression," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1992–1999, 2006.

[6] B. Funt, M. Drew, and M. Brockington, "Recovering shading from color images," in *European Conference on Computer Vision*, pp. 124–132, 1992.

[7] G. Finlayson, S. Hordley, C. Lu, and M. Drew, "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59–68, 2006.

[8] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf, "Recovering intrinsic images with a global sparsity prior on reflectance," in *Neural Information Processing Systems*, pp. 765–773, 2011.

[9] L. Shen, C. Yeo, and B.-S. Hua, "Intrinsic image decomposition using a sparse representation of reflectance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2904–2915, 2013.

[10] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1670–1687, 2015.

[11] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell, "Names and shades of color for intrinsic image estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 278–285, 2012.

[12] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground-truth dataset and baseline evaluations for intrinsic image algorithms," in *International Conference on Computer Vision*, pp. 2335–2342, 2009.

[13] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *European Conf. on Computer Vision (ECCV)* (A. Fitzgibbon et al. (Eds.), ed.), Part IV, LNCS 7577, pp. 611–625, Springer-Verlag, Oct. 2012.

[14] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2018.

[15] T. Narihira, M. Maire, and S. X. Yu, "Direct intrinsics: Learning albedo-shading decomposition by convolutional regression," in *International Conference on Computer Vision (ICCV)*, 2015.

[16] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," *CoRR*,

vol. abs/1406.2283, 2014.

[17] E. Shelhamer, J. T. Barron, and T. Darrell, "Scene intrinsics and depth from a single image," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[18] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," *CoRR*, vol. abs/1411.6387, 2014.

[19] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, pp. 241–248, IEEE, 2013.

[20] T. Zhou, P. Krähenbühl, and A. A. Efros, "Learning data-driven reflectance priors for intrinsic image decomposition," *CoRR*, vol. abs/1510.02413, 2015.

[21] S. Kim, K. Park, K. Sohn, and S. Lin, "Unified depth prediction and intrinsic image decomposition from a single image via joint convolutional neural fields," *CoRR*, vol. abs/1603.06359, 2016.

[22] J. Shi, Y. Dong, H. Su, and S. X. Yu, "Learning non-lambertian object intrinsics across shapenet categories," *CoRR*, vol. abs/1612.08510, 2016.

[23] A. X. Chang, T. A. Funkhouser, L. J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "Shapenet: An information-rich 3d model repository," *CoRR*, vol. abs/1512.03012, 2015.

[24] L. Lettry, K. Vanhoey, and L. V. Gool, "DARN: a deep adversial residual network for intrinsic image decomposition," *CoRR*, vol. abs/1612.07899, 2016.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 2672–2680, Curran Associates, Inc., 2014.

[26] Q. Fan, D. P. Wipf, G. Hua, and B. Chen, "Revisiting deep image smoothing and intrinsic image decomposition," *CoRR*, vol. abs/1701.02965, 2017.

[27] A. S. Baslamisli, H. Le, and T. Gevers, "CNN based learning using reflection and retinex models for intrinsic image decomposition," in *Computer Vision and Pattern Recognition*, 2018.

[28] A. X. Chang, T. A. Funkhouser, L. J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "Shapenet: An information-rich 3d model repository," *CoRR*, vol. abs/1512.03012, 2015.

[29] K. Plataniotis and A. Venetsanopoulos, *Color Image Processing and Applications*. Springer, 2000.

[30] "Reflectance estimation using an eurnet deep network architecture," in *(Under review)*, 2018.

[31] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[33] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 447–456, 2015.

[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.

[35] F. Chollet *et al.*, "Keras." `https://github.com/fchollet/keras`, 2015.

[36] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints*, vol. abs/1605.02688, May 2016.

[37] P. J. Huber, "Robust estimation of a location parameter," *The annals of mathematical statistics*, pp. 73–101, 1964.

[38] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin, "Estimation of intrinsic image sequences from image depth video," in *European Conference on Computer Vision*, pp. 327–340, 2012.