Deep Multi-Illumination Fusion for Low-Light Image Enhancement

Wei Zhong[1](✉), Jie Lin[2], Long Ma[2], Risheng Liu[1], and Xin Fan[1]

[1] DUT-RU International School of Information Science and Engineering, Dalian University of Technology, Dalian 116024, China
zhongwei@dlut.edu.cn

[2] School of Software, Dalian University of Technology, Dalian 116024, China

Abstract. In recent years, improving the visual quality of low-light images has attracted tremendous attention. Most of the existing deep learning approaches estimate the single illumination and then obtain the enhanced result according to the Retinex theory. However, only estimating the single illumination limits the solution space of the enhanced result, causing the unideal performance, e.g., color distortion, details loss, etc. To overcome the issues, we design a new Deep Multi-Illumination Fusion (denoted as DMIF) network to effectively handle low-light image enhancement. Specifically, we first construct an illumination estimation module to generate multiple illuminations to enlarge the solution space. We fuse these illuminations and aggregate their advantages by an illumination fusion algorithm to produce a final illumination. Finally, the enhanced result is obtained according to the Retinex theory. Plenty of experiments are conducted to fully indicate our effectiveness and superiority against other state-of-the-art methods.

Keywords: Low-light image enhancement · Image fusion · Deep network

1 Introduction

As for many computer vision and multimedia applications, high visibility images with clear targets are urgent. Limited to the adverse imaging conditions, the low-quality images with low illumination are frequent and inevitable. In recent years, there emerge many algorithms to enhance low-light images.

A common method is histogram equalization, which enlarges the dynamic range and increases the image contrast, but its limitations are obvious and the results tend to be over enhancement. Based on the Retinex theory [13], there exists an assumption, i.e., the low-light image can be decomposed into two parts, illumination and reflectance, where illumination represents the intensity of light

This work is partially supported by National Natural Science Foundation of China (NSFC) under Grant 61906029, the Fundamental Research Funds for the Central Universities.

H. Ma et al. (Eds.): PRCV 2021, LNCS 13021, pp. 140–150, 2021.
https://doi.org/10.1007/978-3-030-88010-1_12

Fig. 1. Visual comparison of low-light image enhancement.

exposure, reflectance denotes the physical properties of the object itself. This model can be formulated as: $\mathbf{S} = \mathbf{I} \odot \mathbf{R}$, where $\mathbf{S}$ denotes the low-light input and "$\odot$" is the pixel-wise multiplication. $\mathbf{I}, \mathbf{R}$ are illumination and reflectance, respectively. Numerous approaches are currently aimed at removing or adjusting the illumination map to reduce the impact of the illumination. Early attempts include Single-scale Retinex [11] and Multi-scale Retinex [10]. Their results tend to look unnatural, and over-exposure in some cases. In RRM [15], the procedure of noises suppression was also considered in the designed model derived from the Retinex theory. In LIME [7], the illumination was estimated by the structure-aware prior and the reflectance was further obtained by utilizing the Retinex theory. SRIE [5] and JIEP [2] built their model by defining the physical priors of different components, to simultaneously estimate the reflectance and illumination. In [20], an enhancement algorithm for inhomogeneous illumination images was proposed, to balance the details and naturalness. Although these traditional methods get better results in some cases, they are limited since the regularization capacity. The reason is that the exact distribution of these potential components (especially the illumination) is hard to certain by a simple regularization constraint.

Recently, many approaches based on CNN have been proposed to solve low-light image enhancement task. LightenNet [14] is an early method utilized CNN to estimate the illumination from low-light image and remove it to obtain the

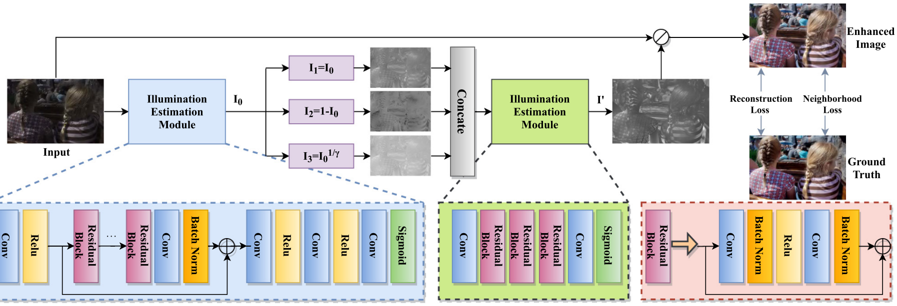


Fig. 2. The overall flowchart of DMIF. The basic illumination $\mathbf{I}_0$ is predicted from the low-light image by the illumination estimation module, $\mathbf{I}_1$, $\mathbf{I}_2$ and $\mathbf{I}_3$ are obtained by different formulas, and their concatenation is used as the input of the illumination fusion network to predict the final illumination. Finally, the final result is obtained by dividing the input by the $\mathbf{I}'$. Below the flowchart are the illumination estimation networks, illumination fusion networks and residual blocks.

enhanced result. RetinexNet [21] estimated both illumination and reflectance in a two-stage way that consists of decomposition and enhancement. However, this work did not take into account the different effects of noise on different areas of light. DeepUPE [19] increased network complexity, taking samples on a bilateral grid and presenting multiple loss functions. However, it also estimated single illumination which is lacked complementary information. Chen *et al.,* solved the extremely low light imaging problem by using the new data set to directly manipulate the original sensor data [3]. Jiang *et al.,* proposed Enlightening GAN [9], which can be trained without low/normal light image pairs. In SSIENet [23] the result of histogram equalization is used as reference, and the reflection and illumination images are decomposed simultaneously. In [22], a deep recursive band network (DRBN) is proposed to recover a linear band representation of an enhanced normal-light image with paired low normal-light images.

In summary, most of existing methods only estimate the single illumination, which limits its solution space to influence the enhanced performance. To settle this issue, we in this paper design an end-to-end Deep Multi-Illumination Fusion (DMIF) network. We provide a group of visual comparison in Fig. 1. Obviously, The result of SSIENet shows excessive brightness while DeepUPE produces color deviation. In contrast, our method addresses the color distortion and non-uniform brightness, providing a more comfortable visual expression. In brief, our contributions can be described as three-folds:

– We devote ourselves to estimate multiple illuminations to provide a larger range of solution space for more effectively handling low-light image enhancement.
– We design a deep multi-illumination fusion network to cater to our demands of estimating multiple illuminations and fuse these outputs by a simple fusion module.

– Comprehensive and elaborate experiments are conducted to illustrate our effectiveness and superiority against existing state-of-the-art approaches.

2 Deep Multi-Illumination Fusion

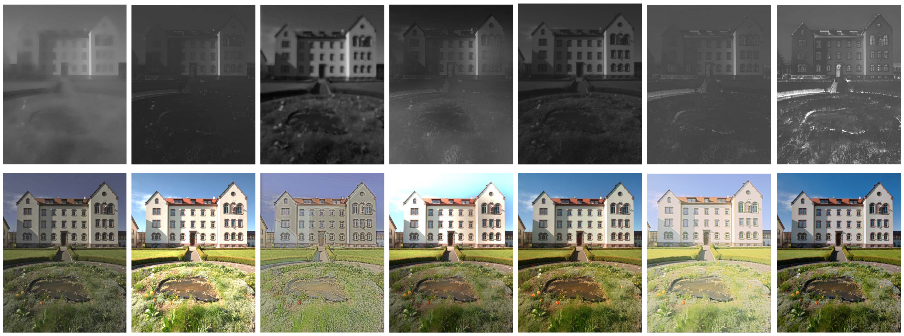
In this section, we clearly introduce our proposed algorithm (DMIF), including network architecture and training loss functions. The flow chart is shown in Fig. 2.

2.1 Network Structure

In fact, the issue of existing Retinex-based methods (maybe iterative algorithms, maybe deep networks) lies in the inaccurate illumination estimation. This is to say, we need to make sure that the illumination estimation module can generate an effective enough illumination map. Otherwise, the model inference based on illumination brings about the deviates. Hence, we first carefully design the illumination estimation module. In order to preserve more details in the illumination, we keep the resolution of the illumination map without down sampling. Specifically, 16 residual blocks are applied which contains Convolutional layer (Conv), Batch Normalization (BN), Rectified Linear Units (ReLU) and a skip connection, and the kernel size of the Conv is 3×3, 64 channels. A sigmoid is added at the end of the module to normalize the value. This module outputs a basic illumination $\mathbf{I}_0$.

Subsequently, we generate multiple illumination with different meanings to learn complementary information of the basic illumination. We empirically explore three explicit forms. The following three models are calculated based on $\mathbf{I}_0$: $\mathbf{I}_1 = \mathbf{I}_0$, $\mathbf{I}_2 = 1 - \mathbf{I}_0$, $\mathbf{I}_3 = \mathbf{I}_0^{1/\gamma}$, where $\mathbf{I}_1$, $\mathbf{I}_2$ and $\mathbf{I}_3$ represent different illumination maps. It is worth noting that these three formulas have three characteristics: 1) they can normalize the illumination value to 0–1; 2) the curve these formulas should be monotonous to preserve the differences local region; 3) they should be differentiable in the process of gradient backpropagation. To be specific, we first preserve the basic illumination map. Moreover, we consider the second formula. As we all know, low-light images have extremely low pixel values. From this formula, we obtain the inverse illumination map, which looks like the image with fog. What's more, with this illumination, we can address the problem of overexposure to a certain extent. The third formula is inspired by the gamma correction, we utilize the third formula, and we set γ as 2.2. Gamma correction is used for smooth extended dark details. To increase the nonlinearity and contrast of illumination, we perform gamma correction on the illumination. We find that the quality of the final illumination map could be improved by fusing the gamma corrected illumination map.

In order to aggregate the above three illumination advantages, we integrate and optimize illumination maps to generate the better illumination. The concatenation operation is utilized for integration. Then, we develop a network to optimize the fusion result and further aggregate their advantages. As showed in



16.58/0.66 JIEP | 16.03/0.67 LIME | 9.25/0.69 RetinexNet | 12.66/0.67 LightenNet | 21.95/0.65 DeepUPE | 10.66/0.73 SSIENet | **22.85/0.91 Ours**

Fig. 3. Comparison of illumination components and the enhanced results. The top row is the estimated illumination, and the bottom row is the enhanced result. PSNR/SSIM scores are reported below each image.

Fig. 2, we only use three residual blocks but still be able to predict the desired results. the kernel size of the Conv is also 3×3, 64 channels. This module outputs a final illumination $\mathbf{I}'$. It's worth noting that, the pixel values of $\mathbf{I}'$ are in definitely ranges, $\{\mathbf{I}'|\mathbf{S}_i \leq \mathbf{I}'_i \leq 1, i = 1, ..., N\}$, thus preventing colors from going beyond the range indicated.

2.2 Loss Function

Our loss function $\mathcal{L}$ contains two parts, i.e., the reconstruction loss $\mathcal{L}_{recon}$ and neighborhood loss $\mathcal{L}_{nb}$ formulated as:

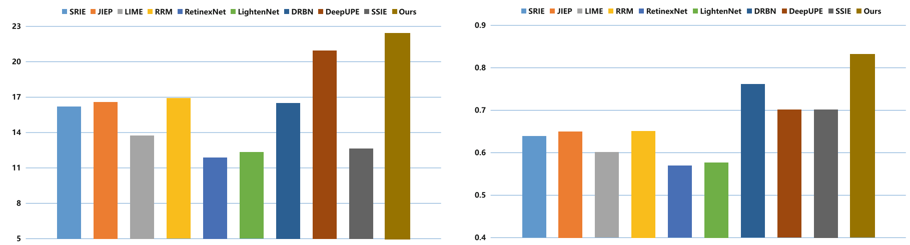
$$\mathcal{L} = \mathcal{L}_{recon} + \lambda \mathcal{L}_{nb}. \tag{1}$$

We empirically set $\lambda = 0.5$. Since we have low light/normal paired images, a following reconstruction loss is defined:

$$\mathcal{L}_{recon} = \sum_{i,j} \|\mathbf{R}(i,j) - \widetilde{\mathbf{R}}(i,j)\|^2, \tag{2}$$

where $\mathbf{R}$ represents the final enhanced result, and $\widetilde{\mathbf{R}}$ is the ground truth, i and j represent pixel position in x and y direction, respectively. Note that restoring an illumination map from a single low-light image is a highly ill-posed problem, and hence the final enhanced result, that is, the reflection map is unideal. Without the guidance of the illumination ground truth, it is necessary to add additional loss constraints on reflection map. Hence, inspired by EPS [4], we also put forward a following neighborhood loss to explicitly encourage the network to learn the local information:

$$\mathcal{L}_{nb} = \sum_{i,j} \sum_{(p,q) \in \mathcal{N}(i,j)} \|(f(i,j) - f(p,q)\|_1, \tag{3}$$



(a) PSNR (b) SSIM

Fig. 4. Quantitative results of low-light image enhancement.

Input LightenNet DeepUPE DRBN SSIENet **Ours**

Fig. 5. More visual results. The first three lines are the results of MIT-Adobe FiveK dataset, the last three lines are the results of LIME, MEF and NPE datasets.

where $f(x, y) = \mathbf{R}(x, y) - \widetilde{\mathbf{R}}(x, y)$ and $\mathcal{N}_{i,j}$ denotes the 5×5 neighborhood centered at pixel (i, j). This term explicitly penalizes deviations in the gradient domain and enhances the image contrast. We add this neighborhood term to the weighted ℓ_1 loss since smoothing involves evident gradient changes. It also can prevent color deviation and improve the generalization ability of the model.

3 Experimental Results

3.1 Implementation Details

We utilize the MIT-Adobe FiveK dataset [1] as the training dataset with 5000 low/normal image pairs, each with five retouched images produced by different experts (A/B/C/D/E). We follow previous methods [6,8,18] to use only the output by Expert C, randomly select 500 images for testing, and train on the remaining 4500 images. We randomly cropped the image to a small size of 48×48 and used the Adam optimizer to iterate at a rate of 500. The learning rate was 0.001 for the first 50 epoched and 0.0001 for the next. Our code is implemented on a NVIDIA Titan X GPU based on tensorlfow.

3.2 Performance Evaluation

We compared our algorithm with advanced low-light enhancement methods including LIME [7], JIEP [2], SRIE RRM [15], LightenNet [14], RetinexNet [21], DeepUPE [19], SSIENet [23] and DRBN [22]. We evaluated them in five widely-used datasets, including MIT-Adobe FiveK [1], LIME [7], MEF [16], NPE [20], and VV[1] datasets.

Results on MIT-Adobe FiveK Dataset. We first evaluate our approach in this dataset including quantitative and qualitative results. We calculate the PSNR and SSIM in 500 testing. As is shown in Fig. 4, obviously our approach has the best PSNR/SSIM. To evaluate the performance of the illumination map, we choose six representative methods for comparison. As is shown in Fig. 3, LIME and SSIENet have the sensible issues, causing some details cannot be recovered. In addition, Fig. 5 provides more visual comparisons with other better performing methods, it is not difficult to find that these methods are deficient in brightness and color restoration, but our method can address these problems well. Our results have more comfortable brightness and saturation.

Results on Other Datasets. We directly test the trained model in other datasets to verify the generalization ability of our method. Through quantitative and qualitative comparison, it is found that our method also performs well in other datasets. Then we show more visual comparisons in LIME, MEF and NPE datasets, and the result is shown in Fig. 5, In these data sets, some methods have similar problems. For example, SSIENet still has uneven brightness, and DeepUPE's results are significantly less bright in these datasets (Table 1).

[1] https://sites.google.com/site/vonikakis/datasets.

Table 1. Quantitative comparison in terms of NIQE.

Methods	LIME	MEF	VV	NPE	AVG
SRIE	4.0502	3.4513	3.0164	3.1845	3.4251
JIEP	**3.7192**	3.4351	2.9704	3.2415	**3.3406**
LIME	4.1291	3.7663	3.2012	3.6384	3.6844
RRM	4.8661	5.0626	3.8002	3.2473	4.2437
RetinexNet	4.5923	4.4104	4.0776	4.2126	4.3231
LightenNet	3.7312	**3.3823**	*2.9524*	3.3884	3.3631
DRBN	3.9645	4.0579	3.2036	**2.8772**	3.4839
SSIENet	4.8260	4.3510	4.2368	2.9609	4.0936
DeepUPE	3.9282	3.5342	3.0082	3.7977	3.5668
Ours	*3.6124*	*3.3731*	**2.9530**	*2.6322*	*3.1430*

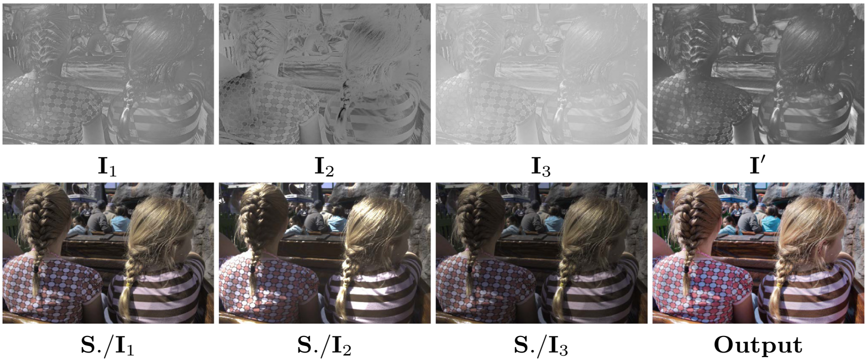


Fig. 6. Visual comparison of different illumination maps and the corresponding enhanced results.

3.3 Ablation Analysis

To analyze the effect of each illumination, we show the visual comparison of different illumination in our designed network. In the Fig. 6, the first row shows three illuminations and the fused illumination, and the second row shows the corresponding enhancement results. This experiment reflects that the necessity and effectiveness of our designed multiple illumination estimation module.

We also conduct the ablation study to verify our effectiveness about the main components. According to the different combinations of the three formulas, we consider four networks and evaluated their quantitative results in the MIT-Adobe FiveK dataset. All the cases include "$I_1 + I_2 + I_3$", "$I_1 + I_2$", "$I_1 + I_3$", and the original single illumination "I_1". For "I_1", since they only have single illumination, we removed the illumination fusion algorithm. In the experiment, we find that the result of "I_1" is slightly brighter than that of "$I_1 + I_2 + I_3$",

Table 2. Ablation study of different cases of our DMIF.

Methods	$\mathbf{I}_1$	$\mathbf{I}_1 + \mathbf{I}_2$	$\mathbf{I}_1 + \mathbf{I}_3$	$\mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3$
PSNR	20.3170	21.0910	20.5995	**21.9697**
SSIM	0.8017	0.8210	0.8116	**0.8315**

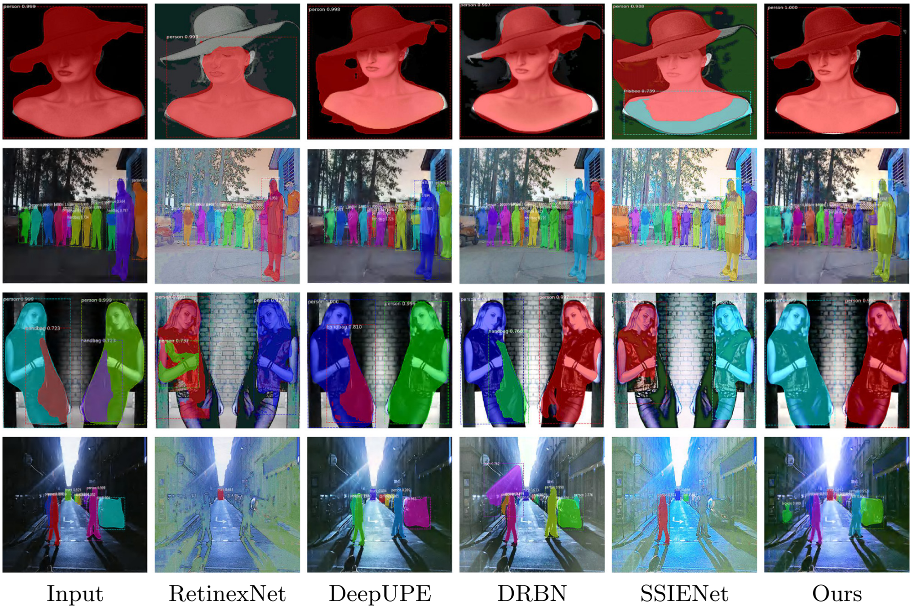


Fig. 7. Visual comparison of instance segmentation **(Best viewed with zoom)**.

but the overall image quality is reduced. The results in Table 2 show that the complete three models improve the enhancement.

3.4 Object Instance Segmentation

In order to further demonstrate our superiority, we directly apply our trained models to enhance low-light images of some real-world scenes. We enhance the low-light image in UFDD dataset [17] by utilizing our method without any fine-tuning and then execute Mask-RCNN [12] to achieve the object instance segmentation. We consider the segmentation results of the original low-light image RetinexNet, DeepUPE, DRBN and SSIENet for comparison. The visual comparison is shown in Fig. 7. Obviously, the results of other methods have the problems of unsatisfactory segmentation and inaccurate object detection, but our proposed algorithm realizes a more superior performance with high accuracy, since it recovers more semantic information in the image.

4 Conclusion

In this paper, we proposed a deep multi-illumination fusion network for low-light enhancement. We applied different formulas and well-designed network to predict multiple illumination with different meanings. We developed a fusion module to fuse them, to overcome details loss and color deviation caused by using single illumination. The low/normal image pairs and the proposed loss function are used to train the whole network. Experimental results showed that our method was superior to advanced low-light enhancement methods. In the future, we will incorporate more illumination maps to further improve performance.

References

1. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2011)
2. Cai, B., Xu, X., Guo, K., Jia, K., Hu, B., Tao, D.: A joint intrinsic-extrinsic prior model for retinex. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
3. Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
4. Zhu, F., Liang, Z., Jia, X., Zhang, L., Yu, Y.: A benchmark for edge-preserving image smoothing. IEEE Trans. Image Process. **28**, 3556–3570 (2019)
5. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
6. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM Trans. Graph. **36**(4), 118 (2017)
7. Guo, X., Li, Y., Ling, H.: LIME: low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. **26**(2), 982–993 (2017)
8. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: a white-box photo post-processing framework. ACM Trans. Graph. (TOG) **37**(2), 1–17 (2018)
9. Jiang, Y., et al.: EnlightenGAN: deep light enhancement without paired supervision. arXiv preprint arXiv:1906.06972 (2019)
10. Jobson, D.J., Rahman, Z., Woodell, G.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans. Image Process. **6**(7), 965–976 (1997)
11. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE Trans. Image Process. **6**(3), 451–462 (1997)
12. Kaiming He, Georgia Gkioxari, P.D., Girshick, R.: Mask R-CNN. In: IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
13. Land, E., McCann, J.: Lightness and retinex theory. J. Opt. Soc. Am. **61**, 1–11 (1971). https://doi.org/10.1364/JOSA.61.000001
14. Li, C., Guo, J., Porikli, F., Pang, Y.: LightenNet: a convolutional neural network for weakly illuminated image enhancement. Pattern Recogn. Lett. **104**, 15–22 (2018)
15. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process. **27**(6), 2828–2841 (2018)

16. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE Trans. Image Process. **24**(11), 3345–3356 (2015)
17. Nada, H., Sindagi, V.A., Zhang, H., Patel, V.M.: Pushing the limits of unconstrained face detection: a challenge dataset and baseline results. In: IEEE International Conference on Biometrics Theory, Applications and Systems, pp. 1–10. IEEE (2018)
18. Park, J., Lee, J.Y., Yoo, D., So Kweon, I.: Distort-and-recover: color enhancement using deep reinforcement learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2018
19. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2019
20. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. **22**(9), 3538–3548 (2013)
21. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference (2018)
22. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: a semi-supervised approach for low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2020
23. Zhang, Y., Di, X., Zhang, B., Wang, C.: Self-supervised image enhancement network: training with low light images only (2020)