RDGAN: RETINEX DECOMPOSITION BASED ADVERSARIAL LEARNING FOR LOW-LIGHT ENHANCEMENT

Junyi Wang, Weimin Tan, Xuejing Niu and Bo Yan

School of Computer Science
Shanghai Key Laboratory of Intelligent Information Processing,
Fudan University, Shanghai, China
Email: byan@fudan.edu.cn

ABSTRACT

Pictures taken under the low-light condition often suffer from low contrast and loss of image details, thus an approach that can effectively improve low-light images is demanded. Traditional Retinex-based methods assume that the reflectance components of low-light images keep unchanged, which neglect the color distortion and lost details. In this paper, we propose an end-to-end learning-based framework that first decomposes the low-light image and then learns to fuse the decomposed results to obtain the high quality enhanced result. Our framework can be divided into a RDNet (Retinex Decomposition Network) for decomposition and a FENet (Fusion Enhancement Network) for fusion. Specific multi-term losses are respectively designed for the two networks. We also present a new RDGAN (Retinex Decomposition based Generative Adversarial Network) loss, which is computed on the decomposed reflectance components of the enhanced and the reference images. Experiments demonstrate that our approach is good at color and detail restoration, which outperforms other state-of-the-art methods.

Index Terms— Low-light Enhancement, Adversarial Learning, Retinex Decomposition, Fusion Enhancement, Deep Learning

1. INTRODUCTION

Pictures taken under the low-light condition not only bring unpleasant user experience, but degrade the performance on other computer vision tasks (e.g. object detection and person re-identification) as well, since most of the solutions to these tasks are designed for well-exposed images. Thus, an approach that can effectively improve the quality of low-light images is demanded.

Traditional single low-light image enhancement methods include histogram-based methods, dehaze-based methods and Retinex-based methods. Histogram-based methods [1] redistribute the histogram to the uniform distribution and adjust the Gamma curve index. Dehaze-based methods [2] utilize the similarity between low-light enhancement and dehazing on the inverse images. Retinex-based methods [3, 4, 5, 6, 7, 8, 9] generally decompose the low-light images into illumination and reflectance components, which can recompose better enhanced results. However, most of the Retinex-based approaches assume that the reflectance components keep unchanged during the enhancement procedure regardless of the color distortion and lost details.

Simple learning-based low-light enhancement methods w/o decomposition [10, 11] have limited enhanced results. SID [12] only improves RAW images and cannot be a post-processing to normal sRGB images. SICE [13] decomposes low-light imaged into smooth and texture components instead regardless of the noises. RetinexNet [14] also learns Retinex decomposition, but it mainly deals with the decomposed illumination and only uses BM3D [15] to denoise the decomposed reflectance, thus the enhanced results are not satisfactory.

In this paper, we propose an end-to-end framework RDGAN (Retinex Decomposition based Generative Adversarial Network), which first decomposes the low-light images and then learns to fuse the decomposed results. In summary, the major contributions of this paper are as follow:

- The RDNet (Retinex Decomposition Network) is proposed for decomposing the input low-light image into illumination and reflectance components. Our RDNet is trained with a multi-term decomposition loss consisting of four losses: L_{ini} loss, L_{wtv} loss, L_{com} loss and L_{err} loss. Unlike in the traditional methods, the RDNet can directly predict the decomposed results w/o iterations.

- The FENet (Fusion Enhancement Network) is proposed for combining the input low-light image and its decomposed results to generate the final enhanced result, which outperforms other methods that only deal with

This work is supported in part by the National Key Research and Development Plan (Grant No.: 2016YFC0801005), and NSFC (Grant No.: 61772137).

978–1–5386–1737–3/18/$31.00 ©2019 IEEE

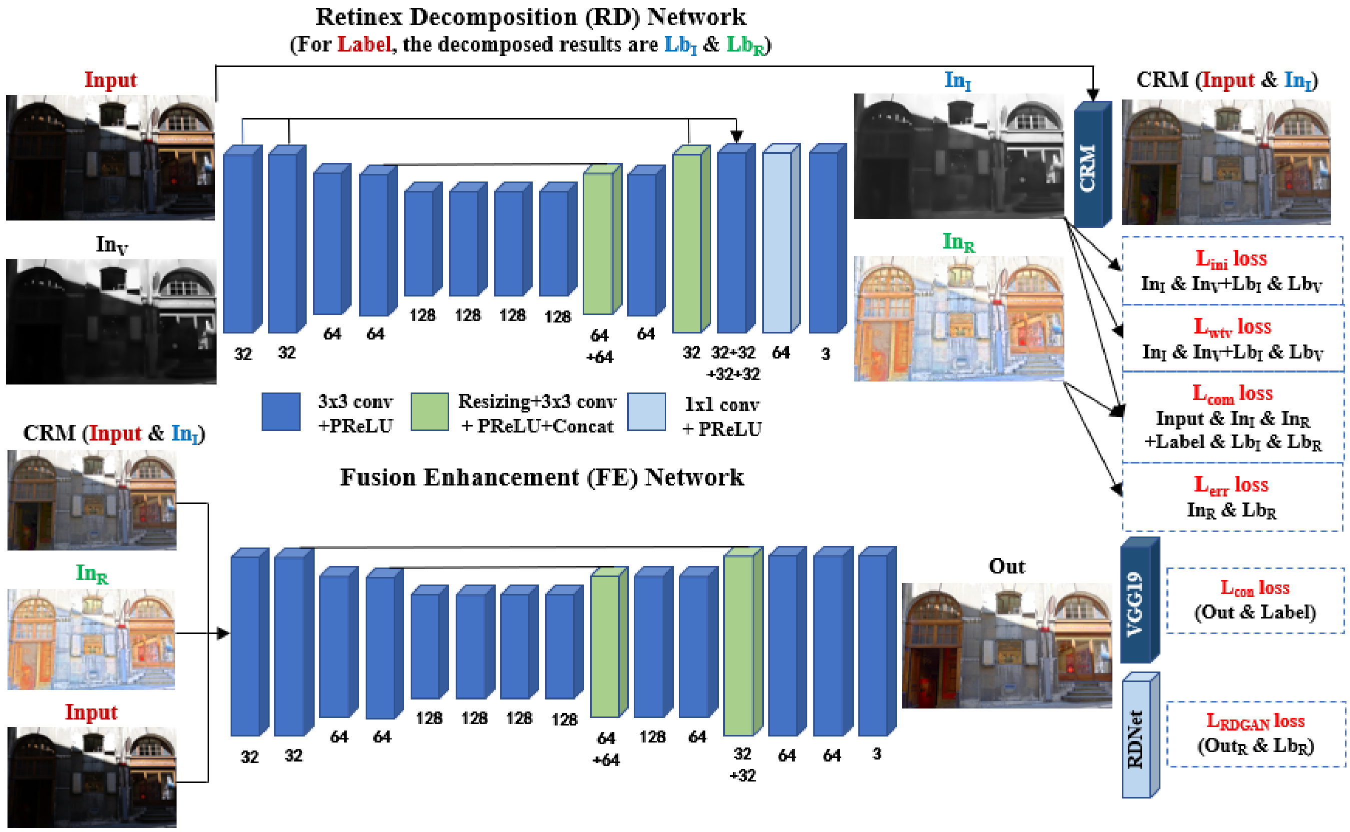


Fig. 1. The architecture of our proposed framework RDGAN, consisting of the RDNet for Retinex decomposition (the upper part), the FENet for fusion enhancement (the lower part) and a discriminator network which is not shown here. The RDNet inputs a low-light image as well as its estimated illumination and outputs the decomposed illumination and reflectance components. FENet takes the input low-light image, the decomposed illumination based CRM [8] result and the decomposed reflectance as inputs to generate the final enhanced result.

the illumination component. We adopt the CRM (Camera Response Model) function for mapping the decomposed illumination to a rough well-exposed result.

- We also present a multi-term adversarial loss for our FENet. The VGG-based content loss guarantees the perceptual quality, and the new RDGAN loss is computed on the decomposed reflectance components of the enhanced and the reference images. Our RDGAN loss helps the adversarial learning with color and detail restoration.

2. OUR APPROACH

In this section, our proposed framework RDGAN can be divided into two new networks: the RDNet for decomposing the input low-light image into illumination and reflectance components, and the FENet for combining the input low-light image and its decomposed results. We first introduce our RDNet as well as its loss functions in detail, and then our FENet.

2.1. Retinex Decomposition Network

In the Retinex theory, the illumination and reflectance components respectively represent the illumination dynamic range and the intrinsic property of objects. The Retinex theory can be formulated as:

$$P = I \circ R \tag{1}$$

where P denotes the input image, I and R respectively denote the illumination and reflectance components and $\circ$ represents the element-wise multiplication.

As shown in Fig.1, our RDNet shares the same model weights for the Retinex decomposition of the low-light and the reference images, and the corresponding decomposed results are In_I & In_R and Lb_I & Lb_R. Inspired by [6], we initialize the estimated illumination P_V with the V channel of P in HSV domain, which equals to the maximum value of RGB channels at each pixel x:

$$P_V(x) = \max P^{c \in \{R,G,B\}}(x) \tag{2}$$

To fuse narrow breaks and fill gaps on the contours, we further apply a morphologically closing operation to the initial P_V. In detail, we first make a 3x3 max-pooling operation, and then another min-pooling (max-pooling on the inverse image). With the final P_V, we can formulate our RDNet as:

$$P_I, P_R = RDNet(P_V, P) \tag{3}$$

Deconvolutional layers in the RDNet are replaced by nearest-neighbor upsampling and 3x3 convolutional layers to relieve the checkerboard artifacts [16]. PReLU [17] is adopted as activation to take the negative coefficients into consideration.

2.2. Decomposition Loss

The multi-term decomposition loss consists of L_{ini} loss, L_{wtv} loss, L_{com} loss and L_{err} loss. The L_{ini} loss is proposed to initialize the decomposed illumination P_I, which calculates MSE (mean square error) between P_I and the estimated illumination P_V:

$$\mathcal{L}_{ini} = \frac{1}{N}\sum_{j=1}^{N} ||In_I^j - In_V^j||^2 + ||Lb_I^j - Lb_V^j||^2 \tag{4}$$

The L_{wtv} loss utilizes TV (total variation) to smoothen the decomposed illumination. Inspired by [7], we use a weighted TV to avoid over smoothing and halo artifacts. In detail, we first define a weight matrix P_W, in which $\nabla_{h,v}$ represents TV on horizontal and vertical directions and $\omega(x)$ is a 3x3 window centered at pixel x:

$$P_W(x) = |\sum_y \nabla_{h,v} P_I(y \in \omega(x))| \tag{5}$$

Then the L_{wtv} loss can be formulated as:

$$\mathcal{L}_{wtv} = \frac{1}{N}\sum_{j=1}^{N} \frac{||\nabla In_I^j||^2}{In_W^j |\nabla In_V^j| + \epsilon} + \frac{||\nabla Lb_I^j||^2}{Lb_W^j |\nabla Lb_V^j| + \epsilon} \tag{6}$$

where ϵ denotes a small constant to avoid the denominator close to zero.

The L_{com} loss is proposed to guarantee the decomposition accuracy, which calculates MSE between $P_I \circ P_R$ and the input image P:

$$\mathcal{L}_{com} = \frac{1}{N}\sum_{j=1}^{N} ||In_I^j \circ In_R^j - Input^j||^2 + ||Lb_I^j \circ Lb_R^j - Label^j||^2 \tag{7}$$

Since the Retinex theory assumes that the reflectance component is relatively consistent, it is an intuitive idea to calculate MSE between In_R and Lb_R. The L_{err} loss is similar to [14] and can be formulated as:

$$\mathcal{L}_{err} = \frac{1}{N}\sum_{j=1}^{N} ||In_R^j - Lb_R^j||^2 \tag{8}$$

The final multi-term decomposition loss is a weighted sum of the above four losses:

$$\mathcal{L}_{RD} = \mathcal{L}_{ini} + \alpha * \mathcal{L}_{wtv} + \beta * \mathcal{L}_{com} + \gamma * \mathcal{L}_{err} \tag{9}$$

In our work, we set α to 0.5, β to 10 and γ to 0.5. The parameter selection keeps a good balance between the estimation of illumination and reflectance components.

2.3. Fusion Enhancement Network

To simplify the fusion procedure, we apply the CRM (Camera Response Model) function in [8] for mapping the decomposed illumination of the low-light image to a rough well-exposed result. The CRM function can be formulated as:

$$k = (\frac{1}{In_I + \epsilon})^a \tag{10}$$

$$CRM(In_I, Input) = e^{b(1-k)} Input^k \tag{11}$$

where ϵ denotes a small constant and the fixed camera parameters a and b are respectively set to -0.3293 and 1.1258 for all exposure degrees.

Since the CRM function neglects the reflectance component, there is room for further enhancement. As shown in Fig.1, our FENet learns to combine the input low-light image, the decomposed illumination based CRM [8] result and the decomposed reflectance to generate the final enhanced result, which can be formulated as:

$$Out = FENet(Input, CRM(In_I, Input), In_R) \tag{12}$$

where Out denotes the enhanced result. The structure of our FENet is similar to the RDNet except for the number of convolutional layers.

2.4. Adversarial Enhancement Loss

During the pre-training process of the FENet, only the VGG-based [18] content loss is used:

$$\mathcal{L}_{con} = \frac{1}{N}\sum_{j=1}^{N} ||V_{5\_4}(Out^j) - V_{5\_4}(Label^j)||^2 \tag{13}$$

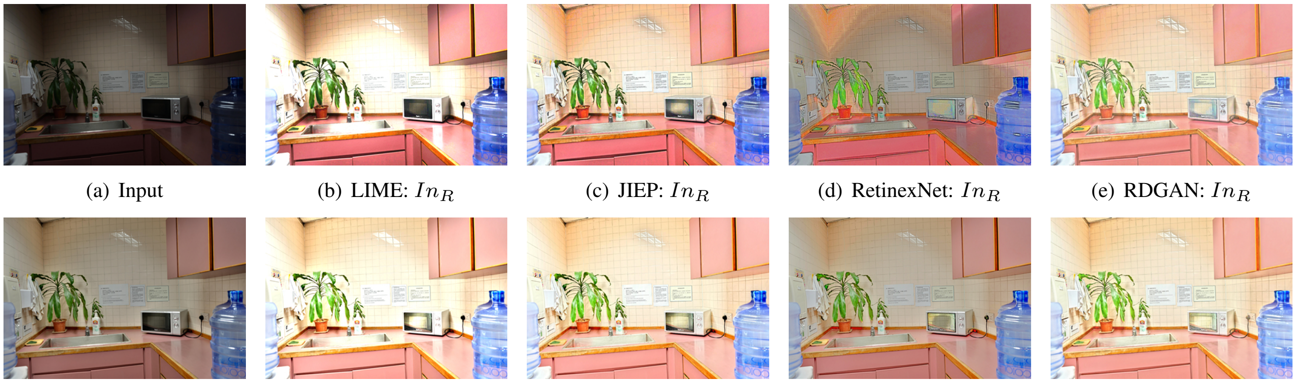
where V_{5_4} represents the feature maps from the $relu5\_4$ layer in a pre-trained VGG19 model.

On the other hand, we present a new RDGAN loss for the alternate training process, which is an adversarial loss computed on the decomposed reflectance components of the enhanced and the reference images, which can be formulated as:

$$\mathcal{L}_{RDGAN\_d} = -\log \mathcal{D}_{real}(Lb_R^j) - \log \mathcal{D}_{fake}(Out_R^j) \tag{14}$$

$$\mathcal{L}_{RDGAN\_g} = -\log \mathcal{D}_{real}(Out_R^j) \tag{15}$$

where $\mathcal{D}_{fake}$ and $\mathcal{D}_{real}$ represent the probability that the input image belongs to enhanced images or reference images



(a) Input (b) LIME: In_R (c) JIEP: In_R (d) RetinexNet: In_R (e) RDGAN: In_R

(f) Label (g) LIME: Lb_R (h) JIEP: Lb_R (i) RetinexNet: Lb_R (j) RDGAN: Lb_R

Fig. 2. Examples of the decomposed reflectance components of the low-light and the reference images by different approaches, from left to right: Retinex-based approach LIME [6] and JIEP[7], learning-based approach RetinexNet [14], our RDGAN.

Table 1. Ablation study on the two proposed networks and the RDGAN loss. Average PSNR and FSIMc are computed on the enhanced results of 53 testing images.

RDNet	FENet	L_{GAN}	L_{RDGAN}	PSNR	FSIMc
✓	×	×	×	22.05	0.9386
×	✓	×	×	19.88	0.9496
✓	✓	×	×	21.91	0.9547
✓	✓	✓	×	21.76	0.9570
✓	✓	✓	✓	**22.34**	**0.9583**

judged by the discriminator network, and Out_R denotes the decomposed reflectance of Out through the RDNet.

The final adversarial enhancement loss is the sum of the above two losses:

$$\mathcal{L}_{FE} = \mathcal{L}_{con} + \mathcal{L}_{RDGAN\_g} \tag{16}$$

3. EXPERIMENT

Our framework is implemented with TensorFlow and TensorLayer packages. We first train the RDNet with the initial learning rate 1e-4 and then pre-train the FENet w/o adversarial learning. After that, we adversarially train the FENet as well as the discriminator network and the learning rate is set to 4e-4. The mini-batch size is set to 16. MSRA initialization [17] is adopted to initialize the model weights, which are optimized by Adam [19] . All experiments are conducted on a desktop with Intel Xeon E5-2630 CPU and NVIDIA GTX 1080 Ti GPU.

3.1. Low-light Dataset

The SICE dataset [13] is selected as the training dataset, which includes 589 sequences of high-resolution pictures taken of the same scene in different exposure degrees. Each image sequence is paired with a well-exposed reference image. However, we found that the low-light and the reference images in the SICE dataset are not pixel-wise aligned due to dynamic objects and camera shakes, so we abandon 92 image sequences and remain a training set with 444 image sequences as well as a testing set with 53 image sequences.

We select no more than three images from each image sequence and crop patches in size 128x128 from the downscaled images. Our training set is further augmented by randomly flipping horizontally and vertically as well as rotating by 90 degrees.

3.2. Ablation Study

We investigate the performance of enhanced results w or w/o the two proposed networks and the RDGAN loss. The first baseline model only includes the RDNet and outputs the decomposed illumination based CRM result. The second one only includes the FENet w/o the decomposed results. The third one includes both networks but trained w/o adversarial learning, while the fourth one trained with the standard adversarial loss. PSNR and FSIMc [20] are adopted as the objective assessment measures.

As shown in Table1, the experimental results support our contributions. Furthermore, as shown in Fig.2, low-light enhancement methods LIME [6], JIEP [7] and RetinexNet [14] have some distortion in the decomposed reflectance, while the decomposed reflectance of our RDGAN is relatively consistent.

3.3. Enhancement Comparison

We evaluate several state-of-the-art low-light enhancement methods, including Retinex-based approaches NPEA [3], SRIE [4], MF [5], LIME [6], JIEP [7], CRM [8], RobustRetinex [9], and learning-based approaches GLADNet [11],

(e) Input (f) SRIE [4] (g) MF [5] (h) LIME [6]

(m) JIEP [7] (n) CRM [8] (o) RobustRetinex [9] (p) RDGAN

Fig. 3. Examples of the enhanced results of 53 testing images from the SICE [13] dataset by different approaches.

Table 2. Comparison on the enhanced results of 53 testing images by different approaches. **Boldface** and $\underline{\text{Underline}}$ respectively represent the highest and the second highest score.

Method	PSNR	FSIMc	Time(s)
NPEA [3]	19.81	0.9160	18.61
SRIE [4]	19.33	0.9221	18.35
MF [5]	21.17	0.9371	0.58
LIME [6]	19.93	0.8979	0.33
JIEP [7]	20.70	0.9375	9.48
CRM [8]	21.33	$\underline{0.9450}$	0.51
RobustR [9]	**23.45**	0.9296	41.54
GLADNet [11]	21.65	0.9357	0.31(GPU)
RetinexNet [14]	18.67	0.8535	0.27(GPU)
RDGAN	$\underline{22.34}$	**0.9583**	0.58(GPU)

RetinexNet [14]. Though we use the same training set with SICE [13] , neither its trained model nor its executable codes is released, thus it is not available for a fair comparison.

As shown in Fig.3 and Table2, our approach has the second highest score on PSNR and the highest score on FSIMc. FSIMc measures the structural similarity on color channels, which is more convincing than PSNR. To further demonstrate the advantages of our approach, we release our codes at `https://github.com/WangJY06/RDGAN/`.

However, our proposed framework may amplify noises and JPEG artifacts that do not look obvious in the original low-light images, which to some extent even degrade the image quality. One possible reason is that our approach, as well as most of the existing low-light enhancement methods, cannot perfectly deal with the noises and JPEG artifacts that are common in low-light images. The combination work of low-light enhancement and image denoising is one of our future work.

4. CONCLUSION

In this paper, we propose an end-to-end learning based framework RDGAN for single low-light image enhancement, which includes two new networks: the RDNet for decomposition and the FENet for fusion. The RDNet decomposes a single low-light image into illumination and reflectance components. The multi-term decomposition loss for our RDnet consists of four different losses, namely L_{ini} loss, L_{wtv} loss, L_{com} loss and L_{err} loss. The FENet combines the input low-light image and its decomposed components to obtain the final enhanced result. We also present a new RDGAN loss as a part of the adversarial enhancement loss for our FENet. Our RDGAN loss is computed on the decomposed reflectance components of the enhanced and the reference images, which helps the adversarial learning with color and detail restoration. Qualitative and quantitative experimental results as well as the ablation study demonstrate the advantages of our approach over other state-of-the-art methods.

5. REFERENCES

[1] S. C. Huang, F. C. Cheng, and Y. S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, 2013.

[2] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiangtao Wen, Wei Meng, and Yao Lu, "Fast efficient algorithm for enhancement of low lighting video," *IEEE Transactions on Image Processing*, 2011.

[3] Shuhang Wang, Jin Zheng, Hai Miao Hu, and Bo Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[4] Xueyang Fu, Delu Zeng, Yue Huang, Xiao Ping Zhang, and Xinghao Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Computer Vision and Pattern Recognition*, 2016, pp. 2782–2790.

[5] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, X-inghao Ding, and John Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, no. C, pp. 82–96, 2016.

[6] Xiaojie Guo, Yu Li, and Haibin Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.

[7] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao, "A joint intrinsic-extrinsic prior model for retinex," in *IEEE International Conference on Computer Vision*, 2017, pp. 4020–4029.

[8] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang, "A new low-light image enhancement algorithm using camera response model," in *IEEE International Conference on Computer Vision Workshop*, 2017, pp. 3015–3022.

[9] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[10] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[11] Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu, "Gladnet: Low-light enhancement network with global awareness," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2018, pp. 751–755.

[12] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun, "Learning to see in the dark," *in Computer Vision and Pattern Recognition*, 2018.

[13] Jianrui Cai, Shuhang Gu, Lei Zhang, and Lei Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[14] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference*, 2018.

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[16] Augustus Odena, Vincent Dumoulin, and Chris Olah, "Deconvolution and checkerboard artifacts," 2016.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *in IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.

[18] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[19] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," 2015.

[20] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378, 2011.