DarkGAN: Night Image Enhancement using Generative Adversarial Networks

Prasen Alaspure[1][0000−0003−2199−2611], Praful Hambarde[1][0000−0002−7976−9095], Akshay Dudhane[1], and Subrahmanyam Murala[0000−0003−3384−4368] [1]
{2018eem1017, 2018eez0001, 2017eez0001, subbumurala}@iitrpr.ac.in

Computer Vision and Pattern Recognition Lab, IIT Ropar India

Abstract. Low light image enhancement is one of the challenging tasks in computer vision, and it becomes more difficult when images are very dark. Recently, most of low light image enhancement work is done either on synthetic data or on the images which are considerably visible. In this paper, we propose a method to enhance real-world night time images, which are dark and noisy. The proposed DarkGAN consists of two pairs of Generator - Discriminator. Moreover, the proposed network enhances dark shades and removes noise up to a much extent, with natural-looking colors in the output image. Experimental results evaluation of the proposed method on the "See In the Dark" dataset demonstrates the effectiveness of the proposed model compared with other state-of-the-art methods. The proposed method yields comparable better results on qualitative and quantitative assessments when compared with the existing methods.

Keywords: Low-light image enhancement · Generative Adversarial Networks

1 Introduction

The images captured in the day time look pretty, good and more colorful, but at night it isn't so. Images taken at night are generally very dark, highly suffered from noise and blurred. Due to a lack of visibility and low contrast, the images taken at night convey very little information. Nowadays, major work in the field of object detection [23], depth estimation [15], [13], [14] human action recognition [3], and image de-hazing [9], [8], [7] is done by considering day time vision only as working in night vision for given tasks is very challenging. These tasks become more difficult when images are incredibly dark.

Object detection, tracking, extraction and classification are the most important tasks to automate any robotic operation. To perform all these operations smoothly even in the dark, objects should be clearly visible without which the whole operation can be misguided. Hence first, we need to enhance dark shades so that objects can be seen even in a very dark environment. We also need to remove noise, blur effect and then proceed with further operations.

In computer vision, most of low light image enhancement work is done either on synthetic data or the images which considerably visible. In traditional methods of image enhancement, Histogram Equalization [28] is used to improve contrast by stretching pixel intensities but it often produces unrealistic effects in images, another method of Log transformation makes darker pixels brighter but this could lead to loss of information.

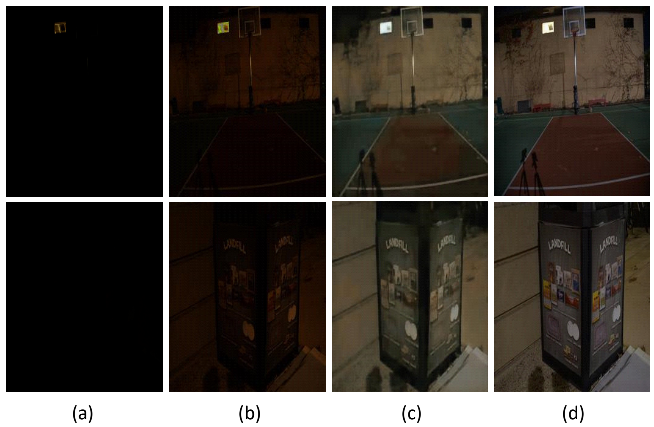

Fig. 1. Sony's short exposure image reconstructed as JPG using (a) Sony Imaging Edge application, (b) Rawpy library, (c) Proposed Method's output on (b), and (d) Ground truth image.

Dong *et al.* [6] take the inverse of dark channel prior [16] treat it as a hazy image then de-haze it and again inverted back which helps in removing dark shades. However, this method lacks the necessary physical model to support the imaging process of low-light images. The Retinex theory proposed by Land *et al.* [24] is based on the color image can be decomposed into two factors such as reflectance and illumination assumption. Single-scale Retinex (SSR) [22] and multiscale Retinex (MSR) [21] treats the reflectance as the final enhanced result which often looks unnatural and frequently appears to be over enhanced.

The recent development in the field of deep learning attracts many researchers to work on low light image enhancement. Multi-layer perceptrons [2], deep autoencoders [26], convolutional networks [19, 34], stacked sparse denoising autoencoders (SSDA) [1, 33] and trainable non-linear reaction diffusion (TNRD) [5] but unfortunately most of the methods are trained on synthetic data such as adding Gaussian or salt & pepper noise. In EnlightenGAN [20] researchers consider considerably visible images but the proposed method is focused on very dark images in which very little information is visible.

Some state-of-the-art methods Dong *et al.* [6], SRIE [10], NPEA [29], LIME [12], RRM [25] and ALSM [30] we have compared in this paper. It is observed that these methods work well when low light images have very low or negligible

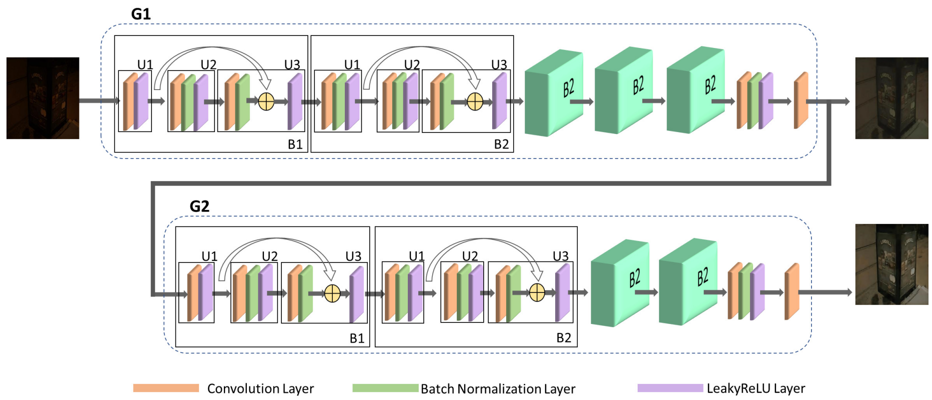

Fig. 2. The proposed DarkGAN (Generator 1 and Generator 2) architecture.

noise, which is not a real-world scenario. On the other hand, when images are very dark and noisy, these methods failed to produce real colors in the output images. Fig.1 demonstrates the overall view of this research paper.

2 Proposed method

In Generative adversarial networks (GANs) [11], Generator G tries to learn to map from the random noise vector z to output image y. Generator G always produces 'fake' images and Discriminator D tries to discriminate between 'real' and 'fake' images. This is unsupervised learning in which both Generator G and Discriminator D are trained together in an adversarial manner until Generator G fools the Discriminator D. This concept further extended from unsupervised to supervised by conditional GANs [27] in which Generator G tries to learn to map from observed image x and random noise vector z to y. After the success of Image-to-Image Translation with conditional adversarial networks [18], we are further extending this concept. We have proposed cGAN based Dark Night Image Enhancement using Generative Adversarial Networks (DarkGAN) for night time image enhancement.

2.1 The Network Architecture

The proposed DarkGAN architecture consists of two pairs of Generator *i.e* (G1 and G2) and Discriminator *i.e* (D1 and D2) models. In which the first pair is for coarse-level enhancement and the second pair for fine-level enhancement. The architecture of G1 and G2 are consists of residual blocks [17] as shown in Fig. 2. Each residual block [17] consists of 3 units as shown in Fig. 2 and each unit has convolutional, batch normalization and the LeakyReLU layer. The features of the first unit are added to third unit pixel-wise except for the last block in each Generator.

The architecture of D1 and D2 are the same and consists of 4 blocks where each block has convolutional, batch normalization and LeakyReLU layer. We used 'PatchGAN' based Discriminator [18] in which we found that patch size of 64×64 to be very effective compared to patch sizes of 1×1, 32×32, 128×128 and 256×256.

2.2 The Loss Function

The main objective of cGAN [11] is given by

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \tag{1}$$

where G tries to minimize this objective against an adversarial D that tries to maximize it, i.e.

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) \tag{2}$$

The L1 loss function combined with GAN loss defined as.

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{3}$$

where L1 loss is

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[||y - G(x, z)||_1] \tag{4}$$

We have tried replacing L1 with L2 but it did not work well, output images were showing unwanted sharpening which was making images look unrealistic. We also found that the combined loss of cGAN and L1 was creating much blurry output images. To overcome this problem, we have decided to go with a combination of cGAN, L1 and SSIM loss for G1 and combination of cGAN, L1 and multiscale SSIM loss for G2. The main objective for the proposed DarkGAN is defined as given below:

$$G_1^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) + SSIM \tag{5}$$

$$G_2^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) + MS\_SSIM \tag{6}$$

where, SSIM *et al.* [31] is structural similarity index and MS SSIM *et al.* [32] is multi scale structural similarity index.

2.3 Training

We train the networks from scratch using the combined loss of cGAN + λ*(L1) + SSIM as G1 loss function and the combined loss of cGAN + λ*L1 + MS SSIM as G2 loss function. While training, input to the network is a short exposure image from "See In the Dark (SID) Dataset" [4].

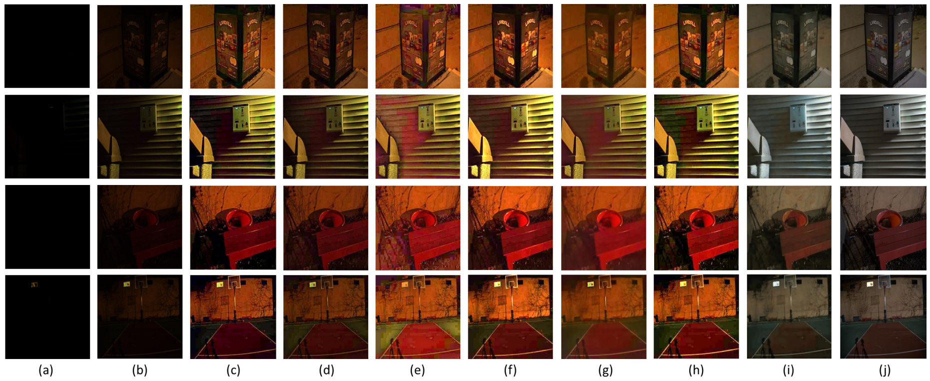

Fig. 3. Comparison of results over different state-of-the-art methods (a) raw image constructed as JPG, (b) input image, (c) Dong [6], (d) SRIE [10], (e) NPEA [29], (f) LIME [12], (g) RRM [25], (h) ALSM [30], (i) Proposed Method and (j) Ground truth

We train the G1-D1 pair for 40 epochs with a learning rate of 0.0002 and while G1-D1 training, the G2-D2 pair is kept in an ideal state. After 40 epochs, training of G1-D1 pair stopped and G2-D2 pair training starts with using G1 in a testing mode such that, the first input image is given to G1 and then applied to G2. Here we assume that G1 is trained at its best and further enhancement is provided by G2. G2-D2 pair is trained for 60 epochs with learning rate of 0.0001. We used Adam optimizer and λ is fixed at 100 in both the pairs during training.

3 Experiments and Results

3.1 The Dataset

In this paper, we use "See In the Dark" (SID) dataset [4] for both training and testing purposes. The dataset contains short exposure and long exposure raw images from Sony and Fuji cameras where all images in the dataset are real-world night time images. Short exposure images are input images and long exposure images are their corresponding ground truth images. Short exposure images (input images) are so dark such that when they are constructed as JPG, we saw almost nothing. So, to use them for training and testing purposes, we first reconstructed them using the Rawpy library. We used images from both Sony and Fuji cameras for training and testing but we keep training and testing images separately. The target of the proposed method is night time dark image enhancement and we did not consider any day time images here.

3.2 Experimental Set-up

Since the prime focus of the proposed method is to enhance real-world dark images of the night time, we prefer to use the SID dataset [4] as it contains real-world very dark images of the night time. The dataset has very fewer images, we choose 200 images for training and 20 images for testing from both sony and fuji camera. We avoid common images from both of these while training. At a later stage, we found that training images are not sufficient and hence we go for data augmentation. We increase the dataset by rotating, flipping and transposing images which finally turn out to be a set of 1400 images.

3.3 Qualitative Analysis

We have compared the results of the proposed method with several state-of-the-art methods as shown in Fig. 3. As we can see in Fig. 3 (a) images are so dark such that we saw absolutely nothing when raw images constructed as JPG images. We reconstructed raw images as JPG using RAWPY library, even after this reconstruction very little information is visible as shown in Fig. 3 (b) and these images are used as the input images in this proposed method for training and testing.

When tested on these input images as shown in Fig.3. (b), all methods are enhancing dark shades but they failed to produce natural colors as input images are very noisy. NPEA [29] introduces many artifacts in the output images, outputs of SRIE [10] are bit dark and slightly blurred, RRM [25] is producing more blurred output images, when viewed in zoom mode artifacts can be easily observed in Dong [6], LIME [12] and ALSM [30] .

The proposed DarkGAN method when applied to input image it is enhancing dark shades with preserving the natural color, removing most of the noise, reducing blurred effect and removing most of the artefacts.

We have used official codes released by the respective authors to implement existing methods.

3.4 Quantitative Analysis

We have evaluated the results of the proposed and existing methods using PSNR, MSE and multi-scale SSIM evaluation. Here, Fig 4 shows a comparative analysis of the different state-of-the-art methods including the proposed method over standard evaluation parameters. The proposed method outperforms than others on each parameter. From the qualitative and quantitative analysis, it is clear that the proposed DarkGAN gives outstanding results over state-of-the-art methods. This evaluation is done on 20 testing images, we show only a few images here and others are enclosed in the supplementary material.

3.5 Ablation Study

Analysis using different loss functions We have tested different loss functions to achieve better results. We begin with cGAN + L1 loss in G1 and cGAN + L2 loss in G2 but it is observed that L2 loss is producing unwanted sharpening of images that do not look realistic. The combination of cGAN + L1 + SSIM loss always performs better with lambda multiplies with L1 loss when used as loss

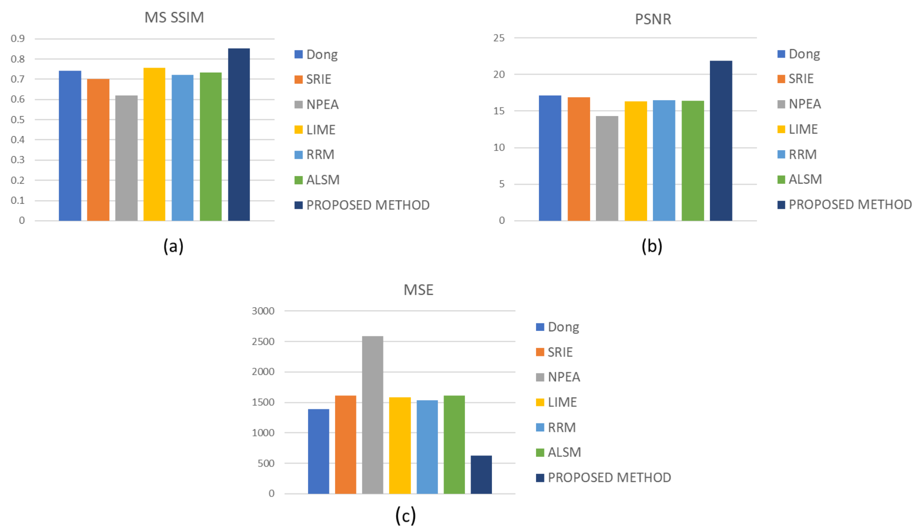

Fig. 4. Comparison of different state-of-the-art methods with proposed method on quantitative parameters (a) Multi Scale SSIM, (b) PSNR and (c) MSE tested on 20 images.

functions for both G1 and G2. But, it is observed that we got a slightly better improvement when SSIM is replaced by MS SSIM as loss for G2. The comparison using several quantitative parameters such as multi-scale SSIM, PSNR and MSE of various loss functions used for G1 and G2 is shown in Table 1. A total of 20 testing images are used for these evaluations.

Table 1. Quantitative analysis of different loss functions for G1 and G2

Loss fo G1	Loss for G2	MS SSIM	PSNR	MSE
L1	**L2**	0.8185	21.58	694.6
L2	**L1**	0.8437	20.64	719.79
L1 + SSIM	**L2 + SSIM**	0.8466	21.42	659.17
L1 + SSIM	**L1 + SSIM**	0.8495	21.77	**620.58**
L1 + SSIM	**L1 + MS SSIM**	**0.8537**	**21.87**	630.68
L1 + SSIM	**L2 + MS SSIM**	0.8422	21.79	647.38
L1 + MS SSIM	**L1 + MS SSIM**	0.8258	21.75	724.94

Analysis using different Generator architectures We perform an ablation analysis using different Generator architectures to evaluate the performance of the second pair (G2-D2) in the proposed network. We consider two networks, the first one with a single pair (G1-D1) and another network with combined of both pairs (G1-D1, G2-D2). We perform qualitative and quantitative analysis using 20 test images for both network output results. The comparative qualitative analysis is shown in Fig. 5. The quantitative analysis over standard evaluation parameters like PSNR, MSE and multi-scale SSIM is shown in Table 2. In Fig. 5 we can clearly see that output of G1 is a bit darker, noisy and less information is visible compared with the combined output of G1-G2.

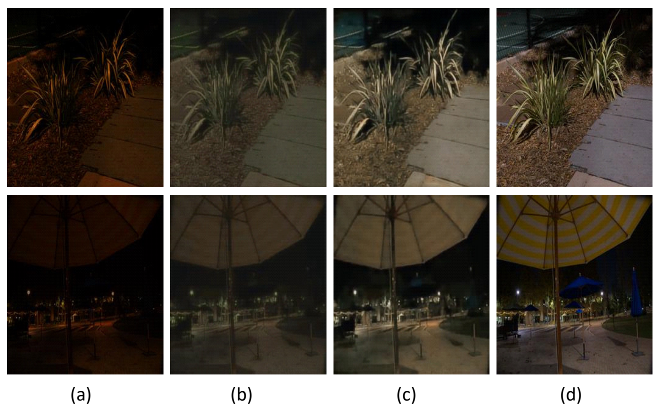

Fig. 5. (a) Input image, (b) Output of G1, (c) Output of combined G1-G2 and (d) Ground truth

Table 2. Ablation Analysis

	MS-SSIM	PSNR	MSE
G1	0.6911	18.05	1385.81
G1-G2	**0.8537**	**21.87**	**630.68**

4 Conclusion

In this paper, we have proposed a novel deep network DarkGAN for very dark and noisy image enhancement. The proposed DarkGAN consists of the duel Generator-Discriminator networks. Among which the first Generator-Discriminator pair designed for the major visibility improvement while the second pair aimed at precise fine enhancement. The performance of the proposed DarkGAN has been compared both qualitatively and quantitatively with the existing state-of-the-art methods for low-light image enhancement. We are also looking forward to working on high-resolution images as the proposed network does not yields much better results on high-resolution images.

References

1. Agostinelli, F., Anderson, M.R., Lee, H.: Adaptive multi-column deep neural networks with application to robust image denoising. In: Advances in Neural Information Processing Systems. pp. 1493–1501 (2013)
2. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2392–2399. IEEE (2012)
3. Chaudhary, S., Murala, S.: Deep network for human action recognition using weber motion. Neurocomputing **367**, 207–216 (2019)
4. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3291–3300 (2018)
5. Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(6), 1256–1272 (2016)
6. Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., Lu, Y.: Fast efficient algorithm for enhancement of low lighting video. In: 2011 IEEE International Conference on Multimedia and Expo. pp. 1–6. IEEE (2011)
7. Dudhane, A., Biradar, K.M., Patil, P.W., Hambarde, P., Murala, S.: Varicolored image de-hazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4564–4573 (2020)
8. Dudhane, A., Hambarde, P., Patil, P., Murala, S.: Deep underwater image restoration and beyond. IEEE Signal Processing Letters **27**, 675–679 (2020)
9. Dudhane, A., Murala, S.: Ryf-net: Deep fusion network for single image haze removal. IEEE Transactions on Image Processing **29**, 628–640 (2019)
10. 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. pp. 2782–2790 (2016)
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
12. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing **26**(2), 982–993 (2016)
13. Hambarde, P., Dudhane, A., Murala, S.: Single image depth estimation using deep adversarial training. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 989–993. IEEE (2019)
14. Hambarde, P., Dudhane, A., Patil, P.W., Murala, S., Dhall, A.: Depth estimation from single image and semantic prior. In: 2020 IEEE International Conference on Image Processing (ICIP). pp. 1441–1445. IEEE (2020)
15. Hambarde, P., Murala, S.: S2dnet: Depth estimation from single image and sparse samples. IEEE Transactions on Computational Imaging (2020)
16. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(12), 2341–2353 (2010)

17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
18. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134 (2017)
19. Jain, V., Seung, S.: Natural image denoising with convolutional networks. In: Advances in neural information processing systems. pp. 769–776 (2009)
20. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. arXiv preprint arXiv:1906.06972 (2019)
21. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image processing **6**(7), 965–976 (1997)
22. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing **6**(3), 451–462 (1997)
23. Kuang, H., Chen, L., Chan, L.L.H., Cheung, R.C., Yan, H.: Feature selection based on tensor decomposition and object proposal for night-time multiclass vehicle detection. IEEE Transactions on Systems, Man, and Cybernetics: Systems **49**(1), 71–80 (2018)
24. Land, E.H.: The retinex theory of color vision. Scientific american **237**(6), 108–129 (1977)
25. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Transactions on Image Processing **27**(6), 2828–2841 (2018)
26. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61**, 650–662 (2017)
27. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
28. Pizer, S.M.: Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In: Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia, May 22-25, 1990. p. 337. IEEE Computer Society Press (1990)
29. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing **22**(9), 3538–3548 (2013)
30. Wang, Y.F., Liu, H.M., Fu, Z.W.: Low-light image enhancement via the absorption light scattering model. IEEE Transactions on Image Processing **28**(11), 5679–5690 (2019)
31. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
32. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2, pp. 1398–1402. Ieee (2003)
33. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in neural information processing systems. pp. 341–349 (2012)
34. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017)