PD-GAN: PERCEPTUAL-DETAILS GAN FOR EXTREMELY NOISY LOW LIGHT IMAGE ENHANCEMENT

Yijun Liu, Zhengning Wang, Yi Zeng, Hao Zeng, Deming Zhao*

University of Electronic Science and Technology of China

ABSTRACT

Extremely noisy low light enhancement suffers from high-level noise, loss of texture detail, and color degradation. When recovering color or illumination for images taken in a dark environment, the challenge for networks is how to balance the enhancement for noise and texture details for a good visual effect. A single network is not suitable for solving the ill-posed problem of mapping the input image's noise to the clear target in the ground truth. To solve the problems, we pro-pose perceptual-details GAN (PD-GAN) utilizing Zero-DCE to initially recover illumination and combine residual dense-block Encoder-Decoder structure to suppress noise while finely adjusting the illumination. Besides, fractional differential gradient masks are integrated into the discriminator to enhance details. Experiment results demonstrate that PD-GAN outperforms other methods on the extremely low-light image dataset.

Index Terms— image enhancement, GAN, fractional differential, noise, blur

1. INTRODUCTION

Low-light image enhancement research has gained growing popularity in recent years. For images taken in an extremely dark environment, they suffer from low SNR, loss of texture detail, and color degradation. However, most of the existing image enhancement methods rely on images with high SNR as input, which are not suitable for practical low-light images. Most methods can well enhance the details in the dark area but amplify the noise simultaneously, which does not match the human perception and can not be utilized to enhance the majority of the practical low-light images with noise. Existing enhancement methods may either enhance both the noise and scene details, which do not match human perception[1], or fail to recover low-light images' illuminance [2, 3].

Chen et al. [4] collected many raw pictures and provides an extremely low light dataset named (SID (See In the Dark) and utilized Unet to enhance the image and got an amazing performance. However, the model severely relies on the raw domain pictures, which can not be applied to widely used regular standard RGB (sRGB, 24bits/pixel) images[5].

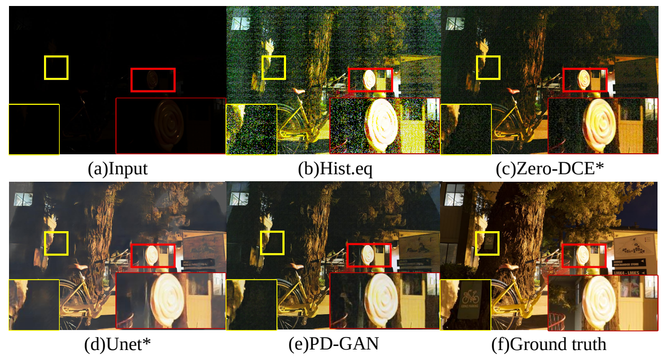

Fig. 1. Only using Unet* based on Unet can cause blur, as shown in the red box of (d) and (f). Because an image with extremely low illumination will completely lose part of the local details, and the corresponding position of the ground truth has a clear object, as shown in the yellow box.

The SID dataset has various images with different levels of real noise, loss of texture details, and color distortion. Nevertheless, they have the same ground truth. It is not easy for networks to complete such a complex mapping. For a dark image, there is no valid information in some parts of the input image, as shown in the yellow box in fig.1(b). Fig.1(b) is the result after histogram equalization, and fig.1(f) is the ground truth. It shows that there is only noise in the yellow box without any textures. However, in the ground truth image, the details are very clear in the corresponding position. If we use Unet-based network Unet* to enhance the image with much noise, the output image will have blurs as shown in fig.1 (d).

In this paper, we propose an extremely low image enhancement method named PD-GAN for general RGB images with different noise levels. To avoid blur, we propose a network combining residual denseblock and Encoder-Decoder to denoise and guide the enhancement of details by using a fractional-order differential mask[6]. What is more, we use a retrained Zero-DCE[2] net named Zero-DCE* to restore the colors of the image initially. Fractional order differential operation is much better for preserving the visual appearance

* Corresponding author.

E-mail address: zhengning.wang@uestc.edu.cn (Zhengning Wang).

This work was supported by grants from the National Foundation of China (No. 61872068, 61720106004) and Science & Technology Department of Sichuan Province of China (No. 21ZDYF3829, 2019YFG0426).

as compared to traditional integer-order methods. Generative adversarial networks famous for their ability to generator visual realistic images are adopted in PD-GAN.

2. RELATED WORK

2.1. Low-light image enhancement

A kind of low-light image enhancement methods [1, 7, 8, 9] is based on the Retinex theory. They decompose the input low-light image into illumination and reflectance and enhance the image's illumination and suppress noise. Another kind of methods adopts advanced learning-based techniques to enhance high-quality user retouched images or images taken using high-end cameras.

Like the curves adjustment used in photo editing software, the state-of-the-art work, Zero-DCE net [2] enhances both the image objectives and noise. To avoid obvious noise, Zero-DCE has the problem of insufficient illumination enhancement for images with extremely low illumination. EnlightenGAN[3] found that the VGG-based perceptual loss of low-light images and normal images is similar. Hence, the paper utilizes this principle to maintain the correlation between input and output, which results in EnlightenGAN not changing the image's texture and suppressing noise. Therefore, EnlightenGAN is not suitable for image enhancement that handles extremely low illumination.

Wu et al. [10] propose a histogram equalization-based method to improve global and local image contrast simultaneously. Yu et al.[11] propose joint enhancement and denoising of low light images via just noticeable difference transform. Method [10, 11] does not focus on extremely low light image enhancement. Zhang et al.[12] develop a principle-inspired multi-scale aggregation network to achieve the exposure enhancement and noises removal on extremely low-light conditions, but the effects are not very well.

2.2. Image denoising

According to the review[13], increasing the network's receptive field can capture more contextual information to improve denoising performance. Among them, increasing the network width and depth is the most common way to increase the receptive field. However, it demands high computing complexity and memory occupation. Dilate convolution can effectively solve this problem. Besides, the combination of CNN and dimensionality reduction method is commonly used for image denoising[14].

3. PROPOSED METHOD

3.1. Overview and Motivation

Since the images are obtained on extremely low light conditions and converted to sRGB format, a lot of details have been lost, and high-level noise appeared in the images. If we reconstruct images with a single-structure denoising network under normal lighting conditions, such as CBDNet[15] which is based on Unet, the results will be blurry. The reason for this phenomenon is that when an area of the input image is full of noise, i.e., the details are completely lost, the process of mapping the noise into a clear object corresponding to the ground truth is extremely ill-posed.

To solve the problems mentioned above, we propose a GAN-based framework PD-GAN. PD-GAN conducts adversarial training on the fractional differential gradient mask of the image to avoid blur appearance. Besides, since the dimensionality reduction network can extract high-level features, we introduce the Encoder-Decoder structure to recover the illuminance faster. Due to the data's particularity, a single network structure cannot achieve an ideal denoising result. So we introduce residual denseblock into PD-GAN for fine-tuning images. Additionally, as Zero-DCE runs fast and makes images brighter, we utilize retrained Zero-DCE to preprocess the image.

3.2. Architecture

Fractional Order Differential Gradient Mask. Inspired by various gradient-guided tasks [16, 17] for super-resolution or inpainting, we adopt fractional-order differential gradient mask in our work. In work[18], Pu et al. proposed six fractional order differential operators and discussed fractional-order differential masks' capability for texture enhancement.

The fractional-order differential mask of a picture can be obtained by convolution on the input image and the fractional order differential operator.

$$M_{frac} = I * F_{frac} \tag{1}$$

where M_{frac} is the fractional-order differential gradient mask, I is a 2-dimensional image of the input, F_{frac} is fractional order differential operator, $*$ represents convolution operation.

Generator. Figure.2shows the network architecture of the generator.Some low-light image local areas are completely noisy or dark but have a clear object in the ground truth image's corresponding positions. It is ill-posed for a network to complete such a map. A single structure such as Unet will be confused under the constraint of element-wise function. As a result, Unet-based methods can cause much blur even though they can recover illumination. Residual dense block performs relatively poorly in recovering illumination, but it can finely control noise.

The generator takes a low light image as input and reconstructs a brightened image. The input and output image sizes are expected to be 512×512 to extract enough features. First, utilize Zero-DCE* to preprocess the input image to perform preliminary illumination enhancement. Then, input the initially enhanced image into two branch networks. The first

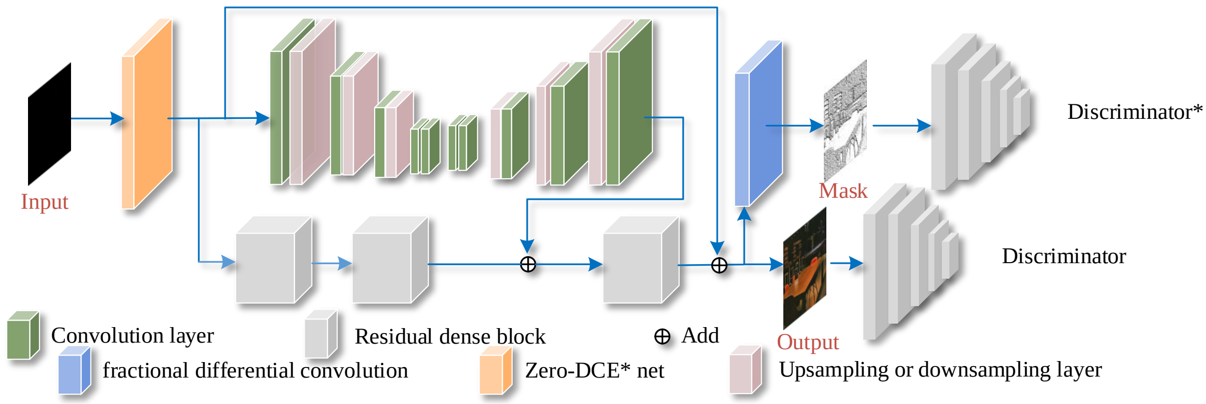

Fig. 2. Overview of the PD-GAN. Zero-DCE* is the retrained Zero-DCE net. Two discriminators are responsible for discriminating the image and the fractional order differential gradient mask.

branch is the Encoder-Decoder structure formed by Unet removing the skip connection. Remove the jump connection so that the network can extract high-level features to enhance the illumination without causing blur. Besides, we added the dilated convolution and gated convolution to expand receptive fields to enhance the denoising ability. The second branch is two serially connected residual dense blocks. Residual dense block can complete the image's refinement and denoise because of its dense connection. Then add the feature maps obtained by the two branch networks pixel by pixel, and fine-tune the image through the third residual dense block.

Discriminator. As illustrated in fig.2, two discriminators D and D^*, which use the same structure as PathchGAN[19] are applied in the proposed network. D is a global discriminator that determines whether the generated image's global information is similar to the ground truth. D^* is a novel fractional order discriminator, which is responsible for judging whether the fractional order differential gradient masks are true or not. Fractional differential gradient masks are famous for their excellent capabilities for texture detail detection.

3.3. Loss Function

Loss of discriminator. The discriminator and generator are alternatively trained with the discriminator loss function L_D and the generator loss function L_G. L_D uses the loss function of WGAN-GP[20].

Loss of generator. The generator loss function includes 4 parts, as shown in Eq. 2.

$$L_G = L_p + \lambda_1 L_{gan} + \lambda_2 L_1 + \lambda_3 L_{frac} \tag{2}$$

where L_p stands for the Perceptual Loss, WGAN Loss L_{gan}, element-wise loss function L_1 and fractional loss function L_{frac}. Specifically, referring to the perceptual loss in DeblurGAN[21], we utilize the third convolutional layer in VGG-19 as a feature map.

Inspired by WGAN-GP, PD-GAN has two discriminators D and D^*, the content loss L_{gan} is calculated by Eq.3.

$$L_{gan} = -\mathbb{E}_{\tilde{\mathbf{x}} \in \mathbb{P}_g}[D(\tilde{\mathbf{x}})] - \mathbb{E}_{\tilde{\mathbf{x}} \in \mathbb{P}_g}[D^*(\tilde{\mathbf{x}})] \tag{3}$$

Considering that L_p uses a high-level feature map, we also adopt L_1 to measure the element-wise difference between the generated image and the target image. L_{frac} is the loss function component from the fractional order differential, which complements L_p and L_1 to get better performance. The fractional order differential loss function is shown in Eq. 4.

$$L_{frac} = \|\tilde{\mathbf{x}}_{frac} - \mathbf{x}_{frac}\|_1 \tag{4}$$

$\tilde{\mathbf{x}}_{frac}$ and $\mathbf{x}_{frac}$ are the fractional order differential gradient maps of the restored image and ground truth image.

4. EXPERIMENT

4.1. Dataset and Training Details

We prepare our training data based on the SID[4] dataset, which consists of raw data and grand truth image pairs. The low light images are taken with short exposure time and the ground truth image data are collected with long exposure time. We convert the raw data to sRGB data as the dataset of our paper. We use 1/4 images as the testing dataset.

We implemented the proposed model in the Pytorch framework with a GTX 1080Ti GPU. We randomly crop patches of resolutions 1024x1024 and then resize them to 512x512 as input and adopt the ADAM optimizer with an initial learning rate of 0.0001, and the learning rate decay to zero in 100 epochs. During the test, we resize the resolution to 1056x720.

4.2. Comparing to Existing Methods

Visual comparisons. We compare the visual quality of PD-GAN with several classical or recent competing methods.

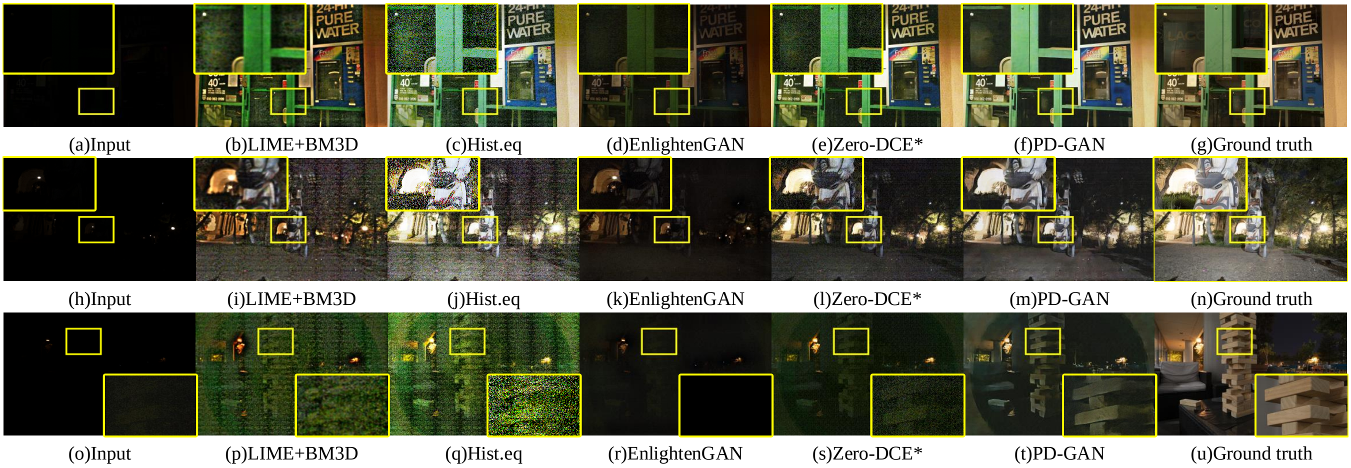

Fig. 3. Visual results of state-of-the-art methods and PD-GAN on input low-light images. Images in the first row have a low level of noise, images in the second row have a medium level of noise, and images in the third row has a high level of noise.

Table 1. Objective evaluation of different methods. SID* is retrained SID net. Zero-DCE* is retrained Zero-DCE net. Our PD-GAN get the best result.

Methods	NIQE	PSNR	SSIM
Hist.eq[22]	13.954	13.99	0.160
LIME+BM3D[1]	20.651	14.36	0.394
SID*[4]	18.731	14.57	0.397
EnlightenGAN[3]	11.233	13.74	0.183
Zero-DCE*[2]	12.767	13.99	0.226
PD-GAN	**10.569**	**17.20**	**0.492**
Ground truth	8.255		

The results are demonstrated in fig.3, where the first column shows the original low-light images, and the second to fifth columns are the images enhanced by LIME+BM3D, Histogram equalization, EnlightenGAN, Zero-DCE* net, and PD-GAN. Besides, the last column shows the ground truth images. We next zoom in on some details in the bounding boxes. The histogram equalization method can obtain relatively good illumination enhancement, but there is much noise. EnlightenGAN can obtain high-quality images for enhancing images with sufficient details. However, they have the problem of insufficient illumination enhancement when dealing with extremely low illumination images. Zero-DCE* net is retrained Zero-DCE net with original data and SID data. It can enhance illumination well, but there is conspicuous noise.

Images in the first row have a low level of noise, images in the second row have a medium level of noise, and images in the third row have a high noise level. As shown in fig.3, when the illumination decreases from the first row to the third row, the noise becomes more conspicuous. Compared with other methods, PD-GAN can better handle images with different illumination intensities and different noise levels.

Table 2. Ablation Study. RDB and FDGM are the abbreviations of residual denseblock and Fractional differential gradient mask respectively.

Methods	NIQE	PSNR	SSIM
RDB+Encoder-Decoder+FDGM	10.569	17.20	0.492
RDB+FDGM	10.742	16.01	0.419
RDB+Encoder-Decoder	10.684	16.65	0.470

Quantitative comparisons. We adopt Natural Image Quality Evaluator (NIQE), PSNR, and SSIM to provide quantitative comparisons. As shown in Table 1, compared to existing methods, PD-GAN obtains a lower NIQE, higher PSNR, and higher SSIM, indicating better performance.

4.3. Ablation Study

We implement an ablation study to demonstrate the importance of fractional differential gradient mask (FDGM), encoder-decoder, and residual dense block (RDB) in PD-GAN. By comparing the first row with the second row and the third row in Table.2, respectively, it shows that both FDGM and encoder-decoder have an excellent effect.

5. CONCLUSION

In this paper, we propose a perceptual-details generative adversarial network named PD-GAN for extremely low light image enhancement. The proposed method has a good denoising effect while preserving the texture details from blur. It achieves better performance compared to the existing methods. In the future work, we will focus on inpainting low light images with more degradation.

6. REFERENCES

[1] 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, 2016.

[2] Chunle Guo, Chongyi Li, Jichang Guo, et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[3] Yifan Jiang, Xinyu Gong, Ding Liu, et al., "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[4] Chen Chen, Qifeng Chen, Jia Xu, et al., "Learning to see in the dark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.

[5] Ke Xu, Xin Yang, Baocai Yin, et al., "Learning to restore low-light images via decomposition-and-enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2281–2290.

[6] Yi-Fei Pu, Ni Zhang, Zheng-Ning Wang, et al., "Fractional-order retinex for adaptive contrast enhancement of under-exposed traffic images," *IEEE Intelligent Transportation Systems Magazine*, 2019.

[7] Bolun Cai, Xianming Xu, Kailing Guo, et al., "A joint intrinsic-extrinsic prior model for retinex," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4000–4009.

[8] Ruixing Wang, Qing Zhang, Chi-Wing Fu, et al., "Underexposed photo enhancement using deep illumination estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6849–6857.

[9] Chen Wei, Wenjing Wang, Wenhan Yang, et al., "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[10] X. Wu, T. Kawanishi, and K. Kashino, "Reflectance-guided, contrast-accumulated histogram equalization," in *ICASSP 2020*, 2020, pp. 2498–2502.

[11] L. Yu, H. Su, and C. Jung, "Joint enhancement and denoising of low light images via jnd transform," in *ICASSP 2020*, 2020, pp. 2658–2662.

[12] J. Zhang, R. Liu, L. Ma, et al., "Principle-inspired multi-scale aggregation network for extremely low-light image enhancement," in *ICASSP 2020*, 2020, pp. 2638–2642.

[13] Chunwei Tian, Lunke Fei, Wenxian Zheng, et al., "Deep learning on image denoising: An overview," *Neural Networks*, 2020.

[14] Runsheng Yu, Wenyu Liu, Yasen Zhang, et al., "Deep-exposure: Learning to expose photos with asynchronously reinforced adversarial learning," in *Advances in Neural Information Processing Systems*, 2018, pp. 2149–2159.

[15] Shi Guo, Zifei Yan, Kai Zhang, et al., "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1712–1722.

[16] Cheng Ma, Yongming Rao, Yean Cheng, et al., "Structure-preserving super resolution with gradient guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7769–7778.

[17] Jianfu Zhang, Li Niu, Dexin Yang, et al., "Gain: Gradient augmented inpainting network for irregular holes," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1870–1878.

[18] Yi-Fei Pu, Ji-Liu Zhou, and Xiao Yuan, "Fractional Differential Mask: A Fractional Differential-Based Approach for Multiscale Texture Enhancement," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 491–511, Feb. 2010.

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, et al., "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[20] Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," in *NIPS*, 2017.

[21] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, et al., "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8183–8192.

[22] Stephen M Pizer, E Philip Amburn, John D Austin, et al., "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.