Low-light Image Enhancement Based on Conditional Generative Adversarial Network

Huaji Li$^1$, Jianghua Cheng$^{1,*}$, Tong Liu$^1$, Bang Cheng$^1$ and Zilong Liu$^2$

$^1$ College of Electronic Science, National University of Defense Technology, Changsha, Hunan, 410073, China

$^2$ 32065Troops, Shenyang, Liaoning, 110162, China

$^*$Corresponding author; *e-mail: jianghua_cheng@nudt.edu.cn

Abstract. We present an end-to-end low-light image enhancement learning method. This learning is based on the conditional generative adversarial networks (GAN) and realizes low-light images enhancement. Specifically, our method uses a convolutional neural network containing residual structures as a generator and WGAN-GP as a discriminator to generate an effective low-light enhancement model under the constraints of GAN loss, Perceptual loss and Structural similarity loss. The model can retain the detailed information of the original image, improve the brightness of the image without generating noise interference, while the generated images are more natural and have higher quality. Extensive experimental results show that our method has reached the state-of-art in multiple objective evaluation indicators of image quality, and the visual appearance is superior.

1. Introduction

In the past few decades, extensive algorithms were proposed to improve the quality of low light images objectively and subjectively. Compared to natural light images containing higher contrast and more detailed information, the low light image often leads to poor performance in high level visual tasks. There are two main methods: traditional methods and learning-based methods.

The traditional low light image enhancement methods are mainly divided into 2 categories. The first type of method based on the histogram equalization (HE) technology adds additional priority and constraints, and its purpose is to expand the range and improve image contrast. Arici et al. [1] analysed and punished image unnatural visual effects, which has certain effects on improving visual quality; Nakai et al. [2] introduced and used a differential grayscale histogram to expand the grayscale difference between adjacent pixels. HE and its variants can improve the contrast of the image to a certain extent, but the colour performance is poor. The second type of method is based on the retinal theory, the retinal theory is the brightness and colour sensing model of human visual, which assumes that the image can be divided into reflectivity and illumination. At early stages, Single-scale Retinex theory (SSR) and Multi-scale Retinex theory (MSR) output appear unnatural and excessive enhanced issues. Wang et al. [3] proposed a method called NPE, which enhances image contrast to maintain the naturalness of brightness to a certain extent, but the brightness is not improved obviously. Fu et al. [4] designed a weighted variation model SRIE with reflection and lighting at the same time to achieve a better reflection and illumination layer, and then generate the target image by adjusting the brightness. Typically, the traditional low light enhancement method is established above the specific statistical model and...
assumption, which is a certain effect in improving image illumination and eliminating noise, but there is a limitation in colour processing and image naturalness.

With the development of deep learning, the low-level visual task has achieved significant results from the depth model, for example noise reduction, super-resolution, image deraining and dehazing etc. A large number of methods for low-light image enhancement appeared at the same time. LLNET[5] used a depth automatic encoder to perform low-light image denoising, but its network structure is too simple, leading to a unsatisfactory results. Other CNN-based methods, such as LLCNN[6], cannot enhance image contrast and denoising at the same time. Shen et al. [7] believed that the multi-scale retinal model is equivalent to a feedforward convolutional neural network with different Gaussian convolution kernels, so they constructed a convolutional neural network (MSR-NET) to learn the end-to-end mapping between dark images and bright images. Wei et al. [8] designed a deep network called Retinex-Net based on the Retinex model. This network combines image decomposition and illumination mapping, and uses the noise removal tool (BM3D [9]) to process the reflection component, but the colour of results is severely distorted after enhancement. Wang et al. [10] proposes a depth network (GLADNET) for enhancing image global illumination, but these strategies do not consider the impact of noise on different brightness regions. In order to obtain more low-light image information, Chen et al. [11] proposed an end-to-end full convolutional network for processing low-light images based on raw image data, which handles noise and colour distortion. However, the illuminance factors need people to consider, and data specific to the RAW format, which limits its applicable scenes.

2. Methods
In this section, we describe the specific details of methods. Firstly, the overall structure of the proposed residual conditional generative adversarial network is discussed. Then, the loss function is designed in detail to resolve the limitations of simple constraints problem during the training.

2.1. Network architecture
The entire network frame includes two parts of the generator and discriminator. The generator network structure is similar to the network structure proposed by Kupyn et al. [12] for image deblurring, including two strided convolution blocks with a step length of 2, nine residual blocks and two transposed convolution blocks, as shown in Figure 1. Each residual block consists of convolution layer, an instance normalized layer and RELU activation layer. Dropout regularization with a probability of 0.5 is added after the first roll layer in each residual block. In addition, the skipped connection is introduced to retain more original image details globally. CNN learns the residual correction($I_r$) of the dark light image($I_l$) in the training, and finally outputs the image of normal light($I_s=I_l+I_r$). We find that such formulation makes training faster and the resulting model has better generalization.

In the training phase, we define a discrimination network D of Wasserstein GAN with gradient punishment, referred to as WGAN-GP [13]. Its discrimination network architecture is the same as PATCHGAN [14]. In addition to the last convolution layer, all the convolution layers are followed by InstanceNorm layer and LeakyReLU with $\alpha=0.2$, as shown in Figure 2. Different from the original GAN discriminator, the output of this discriminator is an N×N matrix, which is not a simple two-classifier of true and false. Consequently, the discriminator can differentiate images at the feature patch level, which reduces the amount of parameters and calculation and increases convergence speed.

![Figure 1. Generator architecture.](image-url)
2.2. Loss function

The loss function is composed of GAN loss $L_{\text{gan}}$, perceptual loss $L_{\text{per}}$ and structural losses $L_{\text{ssim}}$. In the experiment, $\lambda_1$ is equal to 50, $\lambda_2$ is equal to 50.

$$L_{\text{total}} = L_{\text{gan}} + \lambda_1 \cdot L_{\text{per}} + \lambda_2 \cdot L_{\text{ssim}}$$ (1)

**GAN loss.** A large number of methods involving conditions GAN generally use the original GAN target as a loss function. We use WGAN-GP as the critic function, which is shown to be robust to the choice of generator architecture in different architectural experiments, so that a lighter network architecture can be used. The loss is calculated as follows:

$$L_{\text{gan}} = \frac{1}{N} \sum_{i=1}^{N} -D(G(I_i))$$ (2)

Where $I_i$ represents dark images, $G(.)$ represents the generator, $D(.)$ represents the discriminator.

**Perceptual loss.** Two classic selections of content loss function are MAE loss and MSE loss on raw pixels. However, using these functions as a unique optimization goal can result in the blurry artifacts on generated images due to the pixel wise average of possible solutions in the pixel space. Therefore, we have adopted recently proposed Perceptual loss [15]. Perceptual loss is a simple MSE loss based on the difference of the enhanced image and ground truth CNN feature maps, which is defined as follows:

$$L_{\text{per}} = \frac{1}{W_i \cdot H_i \cdot C_i} \sum_{y=1}^{W_i} \sum_{z=1}^{H_i} \sum_{c=1}^{C_i} || \phi_{i,j}(I_h)_{x,y,z} - \phi_{i,j}(G(I_i))_{x,y,z} ||^2$$ (3)

Where $G(I_l)$ and $I_h$ are the enhanced image and ground truth, $W_i$, $H_i$ and $C_i$ represent the dimensions of the respective feature maps within the VGG network, and $\phi_{i,j}$ indicates the feature map obtained by $j$-th convolution layer in $i$-th block in the VGG-19 Network, which is pre-trained in the ImageNet. In this work, $\phi_{i,j}$ is used as the extraction layer. Deeper extraction layer represents the image feature of the higher level. Perceived loss focuses on restoring general content while GAN loss is concentrated on restoring texture details. If there is no perceptual loss during training, or only MSE loss is used, the training cannot converge to the best result.

**Structural loss.** The SSIM metric focuses on the low-level information of image to measure the difference between the enhanced image and ground truth. Different from the perceptual loss, this loss function is aimed at the brightness, contrast and structure of the image. Therefore, we use the image quality assessment algorithm SSIM to establish structural losses, which is used to guide the generation and learning process of the model. This loss is intended to improve the visual quality of the output image, which is defined as follows:

$$L_{\text{ssim}} = 1 - \frac{1}{N} \sum_{y=1}^{W_i} \sum_{z=1}^{H_i} \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$ (4)

Where $\mu_\times$ and $\mu_\gamma$ represent the average of the image pixels, the $\sigma_\times^2$ and $\sigma_\gamma^2$ represent the variance, represents the covariance, $C_1$ and $C_2$ are constants to prevent denominator from zero.
2.3. Implementation details
We implemented all of our models using Pytorch depth learning framework. Training is conducted on a single GTX Tesla100 GPU, which is based on the public dark light dataset (LOL). Since the model is completely convolved and is trained on the image block, they can be applied to an image of any size. For the optimization process, the 5 gradient drop steps are executed on the discriminator, and then 1 step is performed on the generator. Adam is used as an optimizer. The learning rate of the generator and the discriminator is initially set to 10^-4. After 150 epochs of training, we decay rate to zero in the next 150 epochs training. All of our models have been trained with batch size=16, which showed empirically better results on validation. The training phase has been trained for 1 day.

3. Experimental evaluation

3.1. Indicator evaluation
We used a plurality of objective evaluation indicators peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), Natural Image Quality Evaluator (NIQE), and Lightness Order Error (LOE) to evaluate the enhancement effect of our methods. Table 1 reports the evaluation results of different methods on the LOL dataset. In order to facilitate readers to compare and read, we measured the corresponding PSNR and SSIM values of 15 images in the evaluation set. The line graphs are shown in Figures 3 and 4.

Table 1. Quantitative performance comparison of our method with those state-of-the-arts on LOL dataset

| Metrics       | PSNR↗ | SSIM↗ | LOE↘ | NIQE↘ |
|---------------|-------|-------|------|-------|
| BIMEF[15]     | 13.88 | 0.59  | 739.6| 7.51  |
| LIME[16]      | 16.76 | 0.44  | 1084.1| 8.37 |
| RetinexNet[8] | 16.77 | 0.42  | 1705.9| 8.87 |
| EnlightGan[17]| 17.48 | 0.65  | 1115.1| 4.68 |
| KinD[18]      | 17.65 | 0.77  | 1048.8| 4.71 |
| Zero-DCE[19]  | 14.80 | 0.56  | 873.7 | 7.79 |
| Ours          | 18.67 | 0.79  | 670.0| 4.35  |

After comparison of objective evaluation indicators, the PSNR and SSIM indicators of our methods are 5.77% and 2.59% higher than several state-of-the-art methods respectively, which indicates that our model has a better ability to retain image details. Moreover, the LOE and NIQE indicators values are the best among similar methods, indicating that the images enhanced by our model have more natural appearance and better quality.

3.2. Ablation experiment
In order to demonstrate the effect of perceptual loss and structure loss on model training, we keep the same epochs of training and other optimizer parameters unchanged, and performs experiments without...
perceptual loss and structure loss in turn. The visual comparison of the different results is shown in Figure 5. We can clearly observe that the results without Lssim have a relatively low contrast, and model lacking Lper fails to recover the colour variation and the contextual details. Therefore, we could conclude that both loss components have played an important role in the proposed model.

Figure 5. Visual comparison from the loss ablation study. (b)–(d) demonstrates the effectiveness of each component in the whole loss function; (e) and (a) represent the original input and ground truth.

3.3. Visual evaluation
Figure 6 shows the visual performance of different methods on LOL dataset. We can observe that some of the existing methods have problems such as chromatic aberration, large noise interference and low brightness. The result of our methods is closer to the ground truth, of which the visual performance is more natural and the outline structure is clearer.

Figure 6. Visual comparison with state-of-the-art methods on the LOL dataset.
4. Conclusions
We present a conditional generative adversarial networks with WGAN-GP, in which a series residual block structures are used to increase illuminance of the image while maintaining its quality. Moreover, we have made multiple adjustments in the loss function design and basic architecture to establish a more complete learning framework and better convergence. The experimental results show the superiority of our method, which exhibit better performance than other methods of low-light image enhancements in qualitative and quantitative. In future work, we plan to develop more efficient low-light enhancement frameworks through non-supervised learning.

Acknowledgments
This research was supported by Natural Science Foundation of Hunan Province (2020JJ4670).

References
[1] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. (2009) A histogram modification framework and its application for image contrast enhancement. IEEE Transactions on image processing, 18(9):1921–1935.
[2] Keita Nakai, Yoshihiko Hoshi, and Akira Taguchi. (2013) Color image contrast enhancement method based on differential intensity/saturation gray-levels histograms. Intelligent Signal Processing and Communications Systems (ISPACS), IEEE International Symposium, pp.445–449.
[3] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. (2013) Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 22(9):3538–3548.
[4] Xueyang Fu, Delu Zeng, Yu Huang, Xiao-Ping Zhang, and Xinghao Ding. (2016) A weighted variational model for simultaneous reflectance and illumination estimation. IEEE Conference on Computer Vision and Pattern Recognition, pp.2782–2790.
[5] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. (2017) Llnc: A deep auto encoder approach to natural low-light image enhancement. Pattern Recognition, 61:650–662.
[6] Li Tao, Chuang Zhu, Guoqing Xiang, Yuan Li, Huizhu Jia, and Xiaodong Xie. (2017) Llenn: A convolutional neural network for low-light image enhancement. IEEE Visual Communications and Image Processing (VCIP), pp.1–4.
[7] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma. (2017) MSR-net: Low-light Image Enhancement Using Deep Convolutional Network. https://arxiv.org/abs/1711.02488.
[8] C. Wei, W. Wang, W. Yang, and J. Liu. (2018) Deep Retinex Decomposition for Low-Light Enhancement. British Machine Vision Conference.
[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. (2007) Image Denoising by Sparse 3D Transform-Domain Collaborative Filtering. IEEE Transactions on Image Processing, 16(8):2080–2095.
[10] W. Wang, W. Chen, W. Yang, and J. Liu. (2018) GLADNet: Low-Light Enhancement Network with Global Awareness. IEEE International Conference on Automatic Face & Gesture Recognition, pp. 751-755.
[11] C. Chen, Q. Chen, J. Xu, and V. Koltun. (2018) Learning to See in the Dark. IEEE Conference on Computer Vision and Pattern Recognition. pp. 3291–3300.
[12] Kupyn O., Budzan V., Mykhailych M., et al. (2018) DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE pp. 8183-8192.
[13] Gulrajani I., Ahmed F., Arjovsky M., et al. (2017) Improved Training of Wasserstein GANs. ArXiv e-prints, Mar.p. pp: 1,3,4,5.
[14] Isola P., Zhu J Y., Zhou T., et al. (2016) Image-to-Image Translation with Conditional Adversarial Networks. IEEE Conference on Computer Vision & Pattern Recognition. IEEE.
[15] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. (2017) A new image contrast enhancement algorithm using exposure fusion framework. International Conference on Computer Analysis of Images and Patterns, Springer, pp.36–46.
[16] Xiaojie Guo, Yu Li, and Haibin Ling. (2017) Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, pp.26(2):982–993.
[17] Jiang Y., Gong X., Liu D., et al. (2021) EnlightenGAN: Deep Light Enhancement Without Paired Supervision. IEEE Transactions on Image Processing, pp. (99):1-1.
[18] Zhang, Yonghua; Zhang, Jiawan; Guo, Xiaojie. (2019) Kindling the Darkness: A Practical Low-light Image Enhancer. 27th ACM International Conference on Multimedia, pp.1632-1640.
[19] Guo C., Li C., Guo J., et al. (2020) Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp.1777-1786.