2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE) | 979-8-3503-1344-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICISCAE59047.2023.10392948

One improve CycleGAN network for image enhancement under low illumination environment

Long Xiao
College of Sciences
Shanghai Institute of Technology
Shanghai, China
216182115@mail.sit.edu.cn

Fengchao Wang
College of Sciences
Shanghai Institute of Technology
Shanghai, China
fcwang@sit.edu.cn

Jin Chen
College of Sciences
Shanghai Institute of Technology
Shanghai, China
jinchenxl@sit.edu.cn

Yu Sun
College of Sciences
Shanghai Institute of Technology
Shanghai, China
yusunw@sit.edu.cn

Jing Yang
College of Sciences
Shanghai Institute of Technology
Shanghai, China
Yangjingxqq@sit.edu.cn

***Abstract*—The low-light conditions were poor, resulting in low-quality photos and image recognition errors. To address this issue, a image enhancement method was proposed based on an improved amplitude Cycle-Consistent Generative Adversarial Networks (CycleGAN) network in low light. A transformer focusing mechanism is added to the emitter CycleGAN generator network to enhance the information of images in low-light and improve image exposure rates, while identity loss is used to guarantee the enhanced quality and resolution of fuzzy images. The improved model was conducted a comparative analysis of CycleGAN, MBLLEN, Kind and DeepUPE, using the license plate dataset and found that it is more stable than the existing model in both SSIM and PSNR.**

Keywords—CycleGAN network, transformer, low illumination, image recognition, image Enhance

I. Introduction

This paper is about the license plate enhancement technology under low illumination. In the past few years, computer vision in low illumination image intensifier has achieved great development, from the Histogram Equalization [1] and Retinex [2] to RetinexNet [3], EnlightenGAN[4], etc. Under the influence of CycleGAN [5] network in image-to- image translation, we consider image enhancement in low illumination as a special form of image translation, animage enhancement method in low-light based on CycleGAN is proposed. It's different from my previous job, for license plate data set Residual network [6] and a Convolutional Block Attention Module (CBAM) [7] have been added to the generator toavoid color distortion during image transmission. In addition, color information is output via a decoder to preserve the integrity of color information. Transformer modules [8] is also added to the Unet network to address the problem of edge detail loss in image sampling. The improved model effectively enhances image brightness while preserving imagecolor and detail information.

This article has three contributions. First, we added CBAM to the input side of CycleGAN, and residual module to the generator, with better color and brightness compared to some of the current models. Afterwards, we design the transformer module to replace the jump connection in Unet and narrow the feature difference between codec and codec, reducing detail loss during image sampling and allowing images to retain better detail and resolution. Finally, we adopt a new combination of loss functions and make a night license plate dataset, in which the model is validated.

II. Related work

We design the structure of generator and discriminator, and select a suitable combination of loss functions. Finally, we make the data set of license plate under low illumination, text the model and get a good result.

A. Generator

The Fig.1 illustrates the process of enhancing an image using the proposed method. Firstly, the image is transformed into grayscale using the weighted average method. Next, the CBAM attention mechanism is employed to extract the grayscale's detail features, comprising both channel and spatial attention. This approach effectively reduces information loss during image enhancement. The enhanced image is processed using the Unet feature extraction network, which convolves, pools and convolves the image to generate a feature image. The encoding and decoding operations of the transformer module are then applied to the feature image before the encoding and decoding features are combined. The merged information is subsequently convolved, upsampled and upsampled again to recover the dimension and feature information. To enhance accuracy during the convolution bottleneck layer sampling, a residual module is utilized. This method facilitates precise segmentation of light and dark areas, resulting in superior feature extraction. By effectively integrating with the transformer, the network can produce more detailed image output. Unlike recursive loops, transformer uses only the attention mechanism to establish global dependencies between input and output. By replacing Unet's hop connection with the transformer codec, the proposed method minimizes information loss during image sampling.

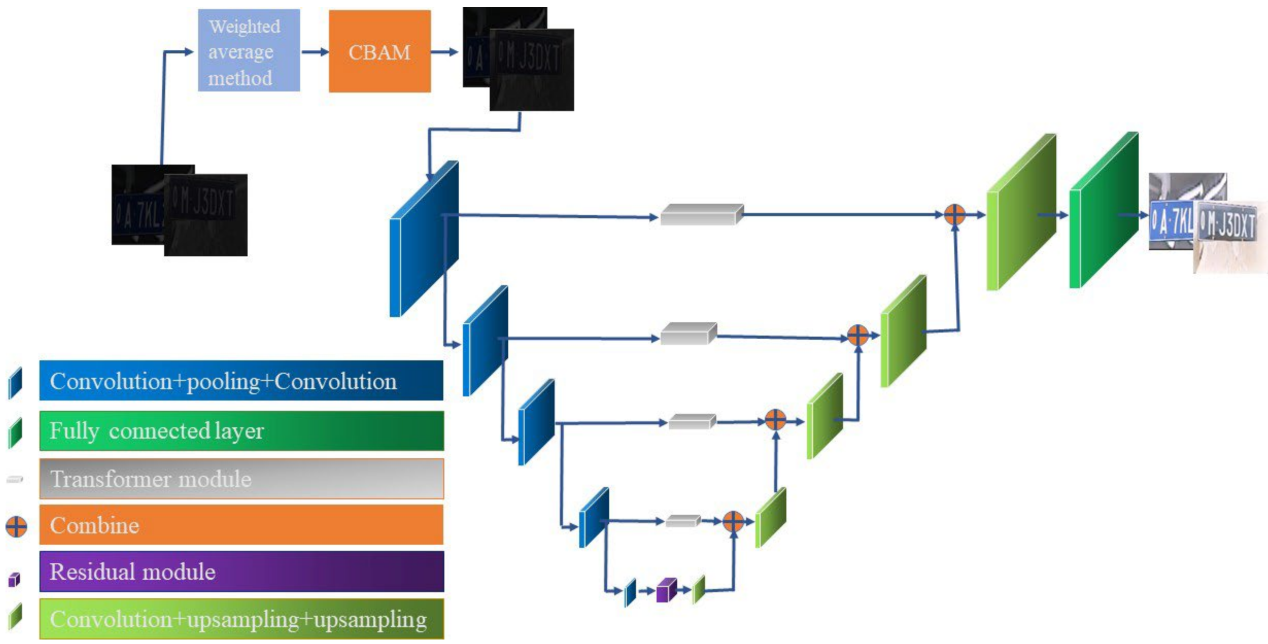

Fig.1.Generator

In Fig.2, the transformer module consists of subblock stacking, subblock fusion and two swin transformers [9]. This module replaces the jump connection in the original Unet network and helps narrow the semantic gap between encoding and decoding features, which reduces the problem of image information loss and retains the configuration details of this photo. Transformer module first encodes its input image into sub-blocks for stacking operation. The processed sub-blocks are then fed into two consecutive swin transformer modules for attention mechanism calculation. The first is normal window attention, and the second is moving window attention. The moving window attention mechanism can interact with other windows so that global information characteristics can be recorded. Finally, the sub-block is restored to the same semantics as the input feature coding through sub-block fusion and deconvolution operation.

The swin transformer comprises several layers, including the layer standardization layer, multiple attention mechanism layers, residual connection layer, and multi-layer sensing unit. At the coding end, the feature subblocks undergo nested operations, where they are classified and vector information about their location is embedded. Subsequently, the subblocks are fed into two swin transformers that operate the attention mechanism. The first swin transformer calculates the attention of common windows, while the second one calculates the attention of the displacement window. Finally, subblock merging is performed to upsample and restore the resolution to match that of the input. The two Swin transformers employ different attention modules, with the first one using the conventional window multiple self-attention module (WSA) to divide the input into non-overlapping windows. After this series of steps, the details of feature extraction by the encoder are retained, and the weight of the network is automatically adjusted by the attention mechanism during training. In this way, the feature fusion is effective and the semantic difference between the two ends of the codec is avoided. Enhanced images are achieved while preserving more detail. The second one uses the shift-window multiple self-attention (SW-MSA), which calculates the attention score using the moving window approach.

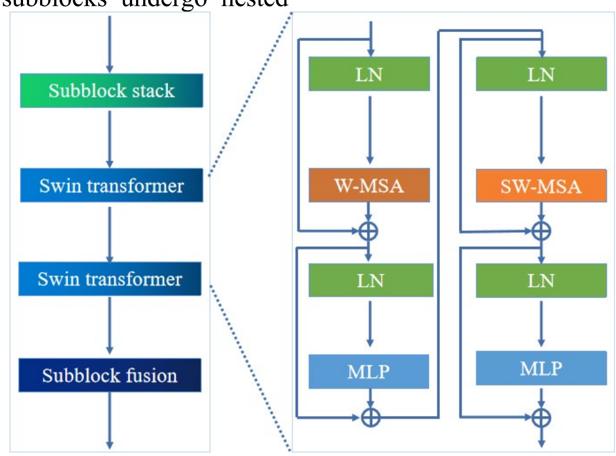

Fig.2 Swin Transformer

The After this series of steps, the details of feature extraction by the encoder are retained, and the weight of the network is automatically adjusted by the attention mechanism during training. In this way, the feature fusion is effective and the semantic difference between the two ends of the codec is avoided. Enhanced images are achieved while preserving more detail.attention mechanism. The first swin transformer calculates the attention of common windows, while the second one calculates the attention of the displacement window.

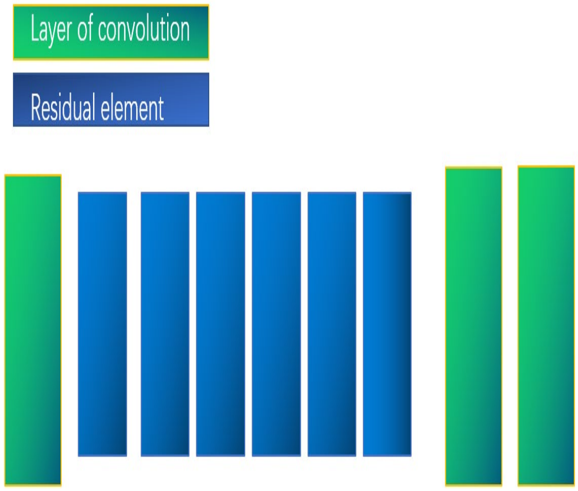

Fig. 3. Residual module

The residual network is based on the Plain VGG [10] network, so that can expend network layers and enhance model accuracy. In figure 3, the residual network is make up three modules. First module is convolution layer, which combines images and performs convolutional operations. The convolution has a step size of 2, resulting in an image that is half the size of the original. The second module is the residual element module, which enhances the brightness of the image through multiple residual elements and restores the image's dimensions to the same as the input while enhancing the image.

B. Discriminator

Under low illumination, the license plate images collected by the discriminator can be affected by environmental lighting, resulting in significant differences in local illumination. Therefore, local and global illumination are both important while enhancing the image's brightness. We used a global discriminator to ensure consistent overall brightness and a local discriminator to ensure consistent local brightness. This approach helps to avoid both overexposure and low illumination in specific regions of the image. Employed PatchGAN for both the global and local discriminators [11].

C. Loss Function

In order to generate image details closer to the input image, the combination form of relative mean GAN and least square GAN is used to resist loss [12]. The global loss function is:

$$L_G^{Global} = E_{x_r} \sim p_{real}\left[\left(D_{R_a}\left(x_r, x_f\right) - 1\right)^2\right] + E_{x_f \sim} P_{fake}[D_{R_a}(x_f, x_r)^2 \quad (1)$$

Where x_r and x_f represents sampling from the distribution of real and false images.

Local discriminator, on the other hand, randomly cuts 5 small pieces from the input and output images each time, and LSGAN is its anti-loss [13], which is defined as follows:

$$L_G^{lobal} = E_{x_f \sim} P_{fake \sim patchs}[D(x_f - 1)^2] \quad (2)$$

The mapping invariance loss function can reduce the difference of low frequency between generated photo and true. The formula is shown in 6.

$$L_I = E_{x \sim P_{data}(x)}[||G(x) - x||_1] = E_{y \sim P_{data}(y)}[||H(y) - y||_1] \quad (3)$$

For the local discriminator, the input and output images are randomly clipped into small blocks, and then the small blocks are regularized by self-preserving feature loss. In addition, after the feature mapping of VGG19, an instance will be added to the end of the layer, and then it will be added to and. Finally, the total loss function can be written as follows [11,12,13]:

$$loss{=}\lambda_1 L_I + \lambda_2 L_G^{Global} + \lambda_3 L_G^{local} + L_{CycleGAN} \quad (4)$$

Where λ_1 ,λ_2 and λ_3 are loss adjustment coefficients of corresponding items.

III. EXPERIMENT AND RESULTS

The hardware environment of our experiment is AMD Processor 3.59GHZ, 16 GB running memory, NVIDIA GEFORCE RTX 2070super graphics, and win 11 operating system. The experimental environment was python 3.85, and the GPU accelerated software was CUDA 11.6 and CUDNN8.60. In the training stage, Adam optimizer was used, the learning rate is set to 1e-4, and a total of 200 rounds of training were conducted. The training set consisted of 752 license plate images in low illumination and 1000 license plate images in normal illumination.

Finally, the improved module was compared the proposed model with the classical low illumination enhancement models CycleGAN, Kind, DeepUPE, MBLLEN, on six groups of images The results are as shown in the following figure. Obviously, our model is stronger than other models in license plate color restoration and brightness enhancement. Also compared several models from the two digital indexes of PSNR and SSIM.

Fig.4.Comparison of model effect

Obviously, according to figure 4, the improved model gets a better result than others. As shown in Table I, our model is more stable than the other model in PSNR and SSIM.

TABLE I. PSNR AND SSIM

Methods	**evaluation metrics**	
	PSNR	***SSIM***
CycleGAN	28.35	0.489
DeepUPE	30.16	0.669
Kind	29.86	0.675
MBLLEN	29.35	0.509
Our	32.51	0.755

IV. CONCLUSION

To solve the difficult of low-light recognizing images , an image enhancement network in low light is proposed based on the CycleGAN model. The approach incorporates a CBAM attention mechanism into the generator to reduce transmission loss by calculating spatial and co-channel attention. Additionally, a transformer attention mechanism is employed to replace skip connections in the feature extraction network, thereby enhancing image details. Finally, we add residual module to the network to increase its depth, resulting in more precise segmentation of light and dark areas in the image. In experiments, the improved model is compared to other models, results demonstrate superior result in detail, color restoration, and brightness enhancement. Nevertheless, for low-light image recognition, improving brightness is only the beginning, and future research should explore other areas such as artifact removal, image correction, and fuzzy image recognition.

ACKNOWLEDGMENT

The project was sponsored by Shanghai sailing pro- gram, China (No. 20YF1447600 and No. 18YF1422500), Research start-up project of Shanghai Institute of Technology (No. YJ2021-60), Collaborative innovation project of Shanghai Institute of Technology (No. XTCX2020-12), and Science and technology talent development fund for young and middle-aged teachers of Shanghai Institute of Technology (No. ZQ2022-6).

REFERENCES

[1] Y. Xu, J. Wen, L. Fei and Z. Zhang, “Review of Video and Image Defogging Algorithms and Related Studies on Image Restoration and Enhancement,” in IEEE Access, vol. 4, pp. 165-188, 2016.

[2] Land, E H, and J J McCann. “Lightness and retinex theory,” Journal of the Optical Society of America, vol. 61, pp. 1-11, 1971.

[3] Wei, Chen, et al. “Deep retinex decomposition for low-light enhancement,” arXiv preprint arXiv:1808.04560, 2018.

[4] Y. F. Jiang, et al. “Enlightengan: Deep light enhancement without paired supervision,” IEEE transactions on image processing, vol. 30, pp. 2340-2349, 2021.

[5] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 2242-2251, 2017.

[6] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, 2016

[7] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon. “Cbam: Convolutional block attention module,” Proceedings of the European conference on computer vision (ECCV), pp. 3-19, 2018.

[8] V. Ashish, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, et al. “Attention is all you need,” Advances in neural information processing systems 30, pp. 6000-6010, 2017.

[9] Z. Liu, Y. T. Lin, et al. “Swin transformer: Hierarchical vision transformer using shifted windows,” Proceedings of the IEEE/CVF international conference on computer vision, pp. 9992-10002, 2021.

[10] Simonyan, Karen, and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[11] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. “Yolov4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.1093, 2020.

[12] A. Aggarwal, M. Mittal, G. Battineni. “Generative adversarial networks: An overview,” IEEE signal processing magazine, vol. 35, no.1, pp. 53-65, 2018

[13] Qi, Guo-Jun. “Loss-sensitive generative adversarial networks on lipschitz densities,” International Journal of Computer Vision vol, 128, no. 5, pp. 1118-1140, 2020.