ISSN 2616-5775 Vol. 8, Issue 2: 20-25, DOI: 10.25236/AJCIS.2025.080203

Research on portability and effectiveness based on MIMO-Unet

Jingyao Zhang*

School of Optical Engineering, China Academy of Space Technology, Beijing, 100094, China
**Corresponding author: 673675215@qq.com*

Abstract: *In the field of computer vision, many excellent image restoration algorithms have emerged, and MIMO-Unet model has achieved a good effect in image deblurring. In this paper, MIMO-Unet model is taken as the research object to introduce the internal principle of MIMO-Unet, and the data set adopted is different from the open GOPRO data set collection means, and the self-made binocular camera is used to construct the data set. Among them, the blurry images and sharp images in the data set are captured by binocular camera. On this basis, the MIMO-Unet model is applied to deblur the images, so as to verify the effectiveness and portability of the MIMO-Unet model. In addition, compared with the public GOPRO data set, the image feature degradation of the data set made by binocular cameras is more serious, and the training effect of MIMO-Unet model on this data set is significantly reduced. In this paper, PSNR and SSIM are selected as two indicators of image defuzzing effect. The training results of the MIMO-Unet model on the binocular camera data set are 0.8028 (SSIM) and 28.87dB (PSNR), which are 0.15 and 6.93dB less than the training results on the GOPRO data set, respectively.*

Keywords: *MIMO-Unet, image Restoration, binocular camera, GOPRO*

1. Introduction

In the field of computer vision, image deblurring has always been a difficult problem, so many scholars have devoted themselves to studying various algorithms to try to solve this kind of problem better. The difficulty of this kind of problem is that it needs to find the mapping relationship between degraded images and sharp images in some way, but this mapping is often very complex, and it is difficult to solve this problem through traditional calculation methods, so the focus of solving this problem is shifted to the field of deep learning. [6] The mapping relationship is perceived by the neural network [3] and the mapping weight set of a specific data set is constructed.

The image is a high-dimensional matrix in the computer, and the extraction of image features is very important. By perceiving the image features of the data set, the neural network can generate a judgment similar to that of human beings, which is conducive to image deblurring. However, it is incomplete to extract image features only through the network of a single stage and a single scale, which needs to be input into the model through multi-scale images. Coarse-to-fine strategy is used to sense image features at different levels to achieve image deblurring effect. The famous model is MIMO-Unet. [1]

In short, the MIMO-Unet model uses a single encoder-decoder [5][7] as the main skeleton, three different scale images as inputs, and utilizes compression and downsampling characteristics for Feature extraction, while utilizing Asymmetric Feature Fusion (AFF) [4] to effectively combine multi-scale features.

In this paper, a binocular camera is built by ourselves, and a dataset of a certain scale is constructed with this device to test the restoration effect of the MIMO-Unet model and prove its portability and usability. In addition, the results are compared with the restoration effect of the MIMO-Unet model on the GOPRO dataset, so as to comprehensively evaluate the performance of the MIMO-Unet model. Finally, the MIMO-Unet model is summarized, hoping to improve the model and promote its development.

ISSN 2616-5775 Vol. 8, Issue 2: 20-25, DOI: 10.25236/AJCIS.2025.080203

2. Related work

2.1 Binocular camera Introduction

The binocular camera consists of two commercial detectors, A and B, The physical object of the binocular camera is as shown in the Figure 1, and the performance parameters of the detector are as shown in the Table 1.

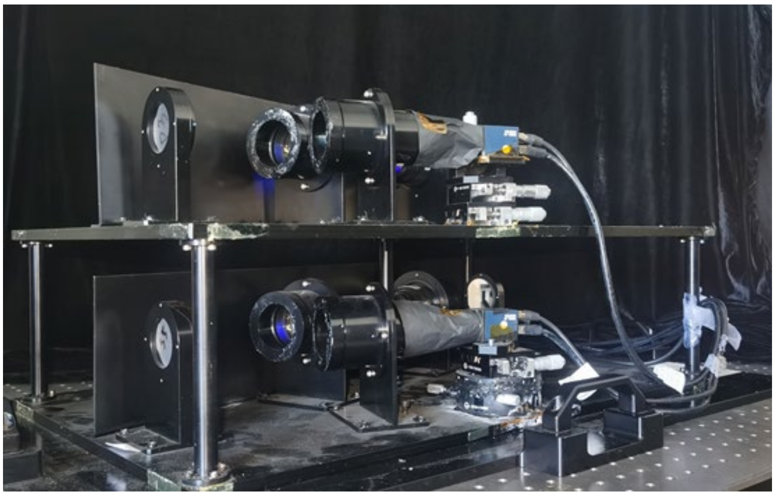


Figure 1: Binocular camera

Table 1: Performance specification of detector

Performance specification	Parameter
Model	MER-530-20GM-P NIR
Interface	GigE
Resolution	2592(H)×2048(V)
Frame frequency	20fps
Size of pixel	4.8μm×4.8μm
Depth of pixel	8bit,10bit
Time of exposure	5μs~1s

In the B camera, a custom anamorphic mirror is added to the first optical element to simulate the error of a low-precision mirror, resulting in a degraded image. Parameters of the deformable mirror are shown in the Table 2.

Table 2: Performance specification of anamorphic mirror

Performance specification	Parameter
Model	50
Range of voltage/V	-250~+300
Reflectance	99.9%
Diameter of effective area/mm	51
Thickness of mirror/mm	1.6
Weight/kg	0.85

2.2 Data set creation

Camera A and camera B took pictures under the premise that illumination, overlooking Angle, overhead Angle and atmospheric path are consistent. [8] The sharp images and blurry images are captured by camera A and camera B respectively, and the ORB algorithm [2] is used to complete the calibration of image pairs, forming a data set of 1200 image pairs. A screenshot of a partial homemade dataset is as shown in the Figure 2.

ISSN 2616-5775 Vol. 8, Issue 2: 20-25, DOI: 10.25236/AJCIS.2025.080203

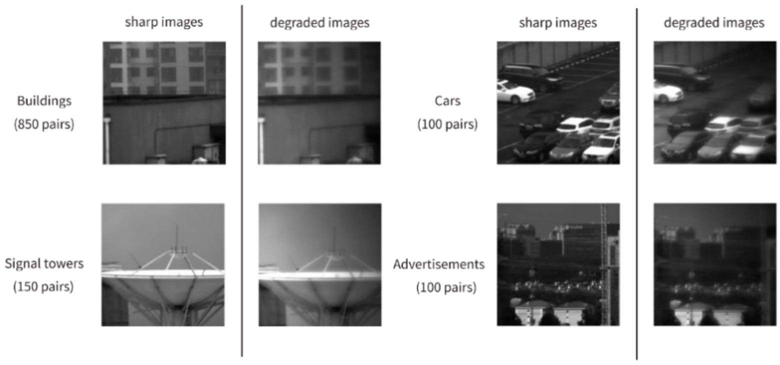


Figure 2: Self-made data set

The data set is divided into train set, test set and validation set, and the specific number of partitions are shown in the Table 3:

Table 3: Segmentation of data set

Data set	Amount
The number of train set	960
The number of validation set	120
The number of test set	120

2.3 MIMO-Unet

The structure of the MIMO-Unet model is as shown in the Figure 3.

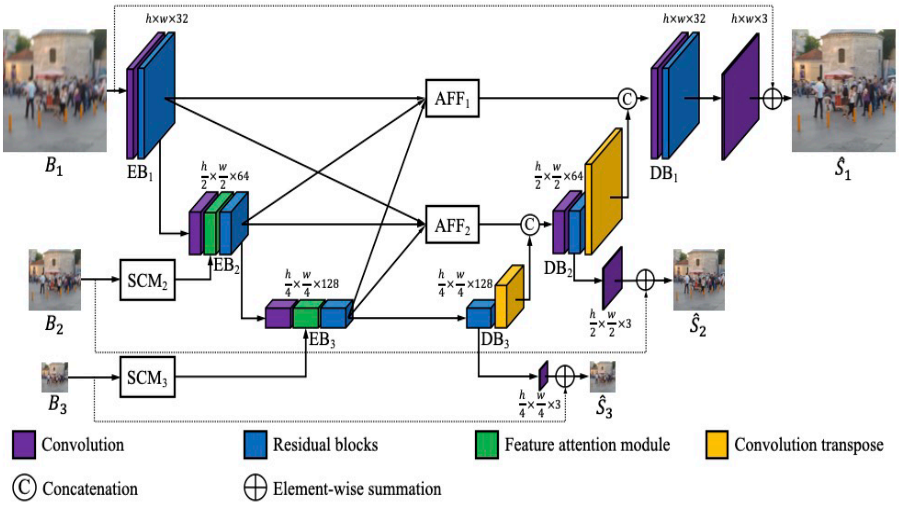


Figure 3: The structure of MIMO-Unet

Where, the SCM module completes the extraction of image features in the process of Down-Sample. It consists of two sets of stacked 3*3 and 1*1 convolution layers, and then concats the output of the previous step, through the 1*1 convolution layer to complete the output, resulting in $\mathrm{SCM}_{\mathrm{k}}^{\mathrm{out}}$. The SCM module structure is as shown in the Figure 4.

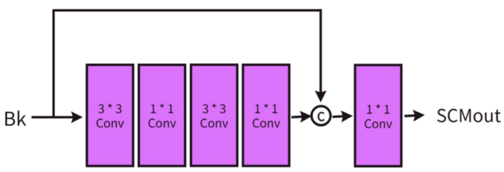


Figure 4: SCM

The FAM module operates on the previously extracted image features, including two ways of

ISSN 2616-5775 Vol. 8, Issue 2: 20-25, DOI: 10.25236/AJCIS.2025.080203

emphasizing features and suppressing features. The specific operation process is as follows:

Step 1: $\mathrm{SCM}_{\mathrm{k}}^{\mathrm{out}}$ and $(\mathrm{EB}_{\mathrm{k-1}}^{\mathrm{out}})^{\downarrow}$ are multiplied by pixels.

Step 2: The output from step 1 is processed through a 3 × 3 convolution layer.

Step 3: Combine the output from Step 2 with $(\mathrm{EB}_{\mathrm{k-1}}^{\mathrm{out}})^{\downarrow}$ to make Residual-Connection.

The FAM module structure is as shown in the Figure 5.

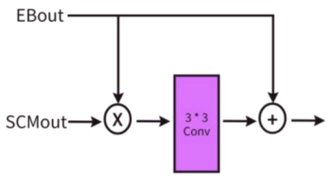


Figure 5: FAM

AFF module integrates Multi-Scale features effectively. Multi-Scale information flows can be obtained by embedding multiple AFF modules, linking the previously separate Enconder process with the Decoder process to improve the Deblurring effect. The specific operation process is as follows:

Step 1: Resize the $\mathrm{EB}_{\mathrm{i}}^{\mathrm{out}}$ generated by the Encoder procedure.

Step 2: Concat operation is performed on $(\mathrm{EB}_{\mathrm{i}}^{\mathrm{out}})_{\mathrm{resized}}$.

Step 3: Pass the output from step 2 through a set of 1*1 and 3*3 convolution layers.

The AFF module structure is as shown in the Figure 6.

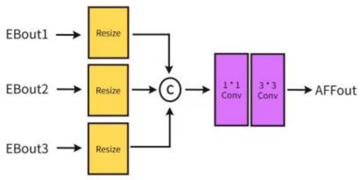


Figure 6: AFF

The loss function adopted by the MIMO-Unet model is as follows:

$$L = \sum_{k=1}^{K} \frac{1}{t_k} \parallel \hat{S_k} - S_k \parallel_1 + \lambda \sum_{k=1}^{K} \frac{1}{t_k} \parallel \mathcal{F}(\hat{S_k}) - \mathcal{F}(S_k) \parallel_1 \tag{1}$$

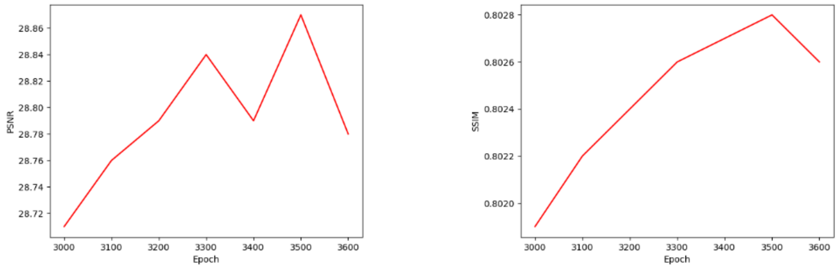
Where, K is the number of layers in the network, $\mathrm{t_k}$ is the number of elements to achieve normalization, $\mathcal{F}$ is FFT, to achieve the transformation of the image from the time domain to the frequency domain, λ is a hyperparameter with a value of 0.1.

The research shows that only considering the difference between the time domain of the image pair is not comprehensive, but also needs to take the difference in the frequency domain into consideration. The theoretical basis is that the conversion process from blurry images to sharp images is to restore the high-frequency signal part and minimize the difference in the frequency domain . The emergence of Multi-scale Frequency reconstruction (MSFR) loss function can be used to quantify the frequency domain difference between image pairs, so as to reduce the L1 distance between sharp and blurry images.

3. Results

The PSNR and SSIM indicators of the MIMO-Unet model after 3000 rounds are as shown in the Figure 7.

ISSN 2616-5775 Vol. 8, Issue 2: 20-25, DOI: 10.25236/AJCIS.2025.080203



(Left side is test set PSNR, right side is test set SSIM)

Figure 7: Test set PSNR and SSIM on self-made data set

After 3500 rounds, both PSNR and SSIM decreased, which are attributed to overfitting. Therefore, The evaluation index values corresponding to 3500 rounds are used: PSNR is 28.87dB, SSIM is 0.8028. The effect diagram of image restoration is as shown in the Figure 8(three pairs are randomly selected).

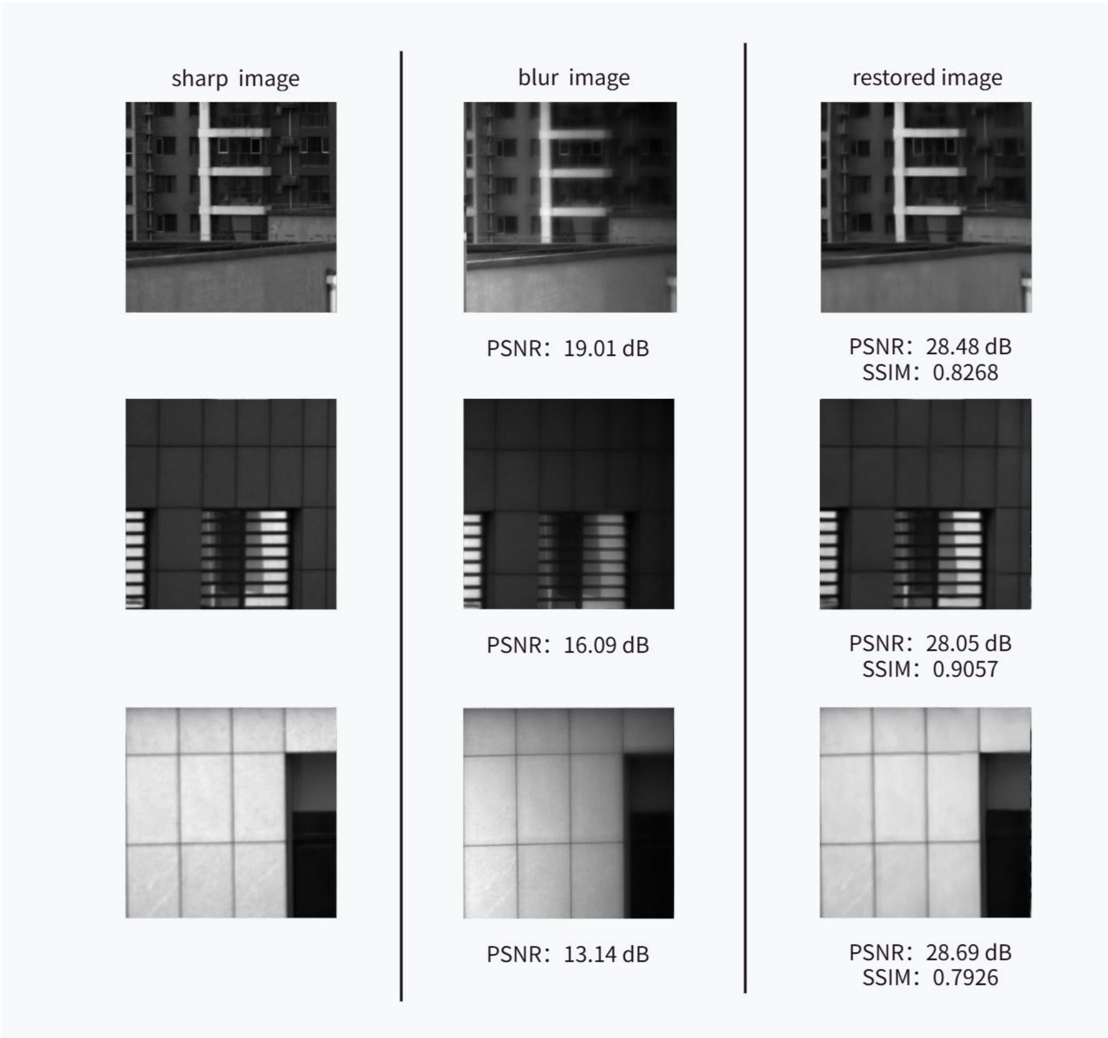


Figure 8: Restored Images

The MIMO-Unet model can basically achieve the image restoration effect in the self-made data set, which initially verifies the portability and effectiveness of the MIMO-Unet model. At the same time, it also verifies that the self-made data set is effective and can be used as a data source for subsequent research.

However, the restoration effect of the MIMO-Unet model on the public GOPRO dataset is: PSNR is 35.8dB, SSIM is 0.9528, 5.96dB and 0.2 more than the self-made dataset respectively, which proves that the restoration effect of the MIMO-Unet model is correlated with the image quality of the dataset.

ISSN 2616-5775 Vol. 8, Issue 2: 20-25, DOI: 10.25236/AJCIS.2025.080203

4. Conclusions

The MIMO-Unet model can not only achieve the deblurring effect on the public GOPRO data set, but also achieve the similar effect on the self-made data set with binocular camera equipment in this paper, which verifies the portability and effectiveness of the MIMO-Unet model. However, in the face of severely degraded images, the image deblurring effect of the MIMO-Unet model will be reduced, so it is necessary to further improve the model and add other modules to carry out further tests, so as to improve its restoration effect..

References

[1] Anton Baumann, Thomas RoBberg, Michael Schmitt. Probabilistic MIMO U-Net: Efficient and Accurate Uncertainty Estimation for Pixel-wise Regression[C]. arXiv:2308.07477, 2023.

[2] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, ORB: an efficient alternative to SIFT or SURF [C]. In ICCV 2011.

[3] Chen Chao, Qi Feng. Review on the development of Convolutional neural networks and their applications in computer vision [J]. Computer science, 2019, 46 (3): 11. DOI: 10.11896 / j.i SSN. 1002-137 - x. 2019.03.008.

[4] Saeed Anwar and Nick Barnes. Real image denoising with feature attention[C]. ICCV, 2019.

[5] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: convolutional networks for biomedical image segmentation[C]. In MICCAI, 2015.

[6] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration[C]. In ICLR, 2019.

[7] Kong J Y, Zhang H S, Improved U-Net network and its application of road extraction in remote sensing image [J]. Chinese Science and Technology, 2022, 42(3): 105-113. (in Chinese)

[8] Chen Ruilin, Zhang Bo, Duan Xikai, et al. Remote Sensing Image Resolution Enhancement Technology Based on Single-Pixel Imaging [J]. Spacecraft Recovery & Remote Sensing, 2023, 44(6): 130-139. (in Chinese)