

Optimal path for automated pedestrian detection: image deblurring algorithm based on generative adversarial network

Xiujian Dong¹, Jianping Lan²

School of Electrical and Information Engineering, Hubei University of Automotive Technology, Shiyan, Hubei, 442002, China

¹Corresponding author

E-mail: ¹xjus51@163.com, ²lanjianping2022@163.com

Received 2 November 2023; accepted 9 December 2023; published online 21 February 2024
DOI <https://doi.org/10.21595/jme.2023.23765>



Copyright © 2024 Xiujian Dong, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. The pedestrian detection technology of automated driving is also facing some challenges. Aiming at the problem of specific target deblurring in the image, this research built a pedestrian detection deblurring model in view of Generative adversarial network and multi-scale convolution. First, it designs an image deblurring algorithm in view of Generative adversarial network. Then, on the basis of image deblurring, a pedestrian deblurring algorithm in view of multi-scale convolution is designed to focus on deblurring the pedestrians in the image. The outcomes showcase that the peak signal to noise ratio and structural similarity index of the image deblurring algorithm in view of the Generative adversarial network are the highest, which are 29.7 dB and 0.943 dB respectively, and the operation time is the shortest, which is 0.50 s. The pedestrian deblurring algorithm in view of multi-scale convolution has the highest peak signal-to-noise ratio (PSNR) and structural similarity indicators in the HIDE test set and GoPro dataset, with 29.4 dB and 0.925 dB, 40.45 dB and 0.992 dB, respectively. The resulting restored image is the clearest and possesses the best visual effect. The enlarged part of the face can reveal more detailed information, and it is the closest to a real clear image. The deblurring effect is not limited to the size of the pedestrians in the image. In summary, the model constructed in this study has good application effects in image deblurring and pedestrian detection, and has a certain promoting effect on the development of autonomous driving technology.

Keywords: generative adversarial network, image deblurring, pedestrian detection, multi-scale convolution, feature pyramid.

1. Introduction

As the boost of technology, autonomous driving technology has gradually matured. However, the real traffic environment is too complex, so it is indispensable to improve the safety of automatic driving if you want to widely apply autonomous vehicle to actual scenes. Optimizing pedestrian detection (PD) is an important challenge [1]. Image deblurring (ID) involves utilizing algorithms or tools to enhance blurred images, ultimately leading to more precise detection of pedestrians. This, in turn, improves the performance and reliability of PD technology. It is of great significance for optimizing PD. Researchers such as Ngo TD have proposed a method for deconvoluting collected data to improve image quality. This method trains the Convolutional neural network to estimate the Point spread function parameters, and performs image deconvolution to obtain the improvement of image SNR and modulation transfer function [2]. Liu Y. Q. and other scholars proposed that using the generalized Gaussian function can more accurately describe the defocus blur, which is better than the existing algorithms deploying parametric or nonparametric cores, and is conducive to the ID process [3]. The Generative Adversarial Networks (GAN) contains two kinds of models. Its Generative model has a convincing ability to generate real examples from the existing sample distribution. It has expanded from the initial image generation to various fields of computer vision, promoting various

applications in many fields. Moreover, its Game theory optimization strategy has promoted privacy and security oriented research [4-5]. Multi-scale convolution (MC) means the addition of multiple convolution kernels in the convolution operation for extracting and fusing features at multiple scales, thereby improving the model [6]. GAN and MC have been widely applied in the field of ID. Tomosada H et al. proposed an algorithm based on convolutional neural network GAN for ID. In order to improve the quality of deblurred images and reduce computational time, a GAN ID method using discrete cosine transform was proposed [7]. Zhao Q. et al. proposed a new context-aware MC neural network for ID that uses multi-scale pyramid structures to eliminate blurring, increasing computational complexity [8]. However, it is difficult to restore blurred images to a level comparable to clear images, both in terms of traditional algorithms and deep learning algorithms that have emerged in recent years. There is still great room for progress in the field of ID. In addition, there is relatively little research on the application of GAN and MC in image specific target deblurring at home and abroad, and many problems are faced, especially the calculation time is long and cannot be clearly solved. Therefore, this study proposes ID algorithms based on GAN and pedestrian deblurring algorithms based on MC. The proposed algorithm is expected to be applied in PD in autonomous driving, as it enhances image clarity, minimizes noise, and enables the detection of pedestrians in complex environments with greater precision. Consequently, it is expected to improve traffic safety. There are two main innovations in the research. The first is to design an ID algorithm using the GAN. The second point is to use MC to blur specific targets of pedestrians in the image.

The study is divided into three. In the first, a PD model in view of GAN and MC is built. It includes the research of ID algorithm in view of GAN and the research of pedestrian deblurring algorithm in view of MC. The second analyzes the effect of the PD model in view of GAN and MC. The third summarizes the research results of this study and proposes shortcomings and prospects.

2. Construction of PD deblurring model in view of GAN and MC

2.1. Research on ID algorithm in view of GAN

ID is a classic issue in low-level computer vision. The purpose is to process fuzzy input images containing noise and restore them to clear images close to real scenes. The formation of a blurred image involves convolving a clear image with a fuzzy kernel before adding noise. Therefore, to deblur an image, it is necessary to first determine the size of the blur kernel, and then eliminate the noise as much as possible. This research uses the Conditional Generative adversarial network (cGAN) to achieve ID. The essence of GAN is in view of the model of confrontation form. Through the continuous game in the generator and the discriminator, the deblurring image is output. GAN typically includes alternating training phases for the generator and discriminator. During each training iteration, the generator and the discriminator compete with each other and gradually improve the generator's generation ability and discriminator's recognition ability through alternating training processes. This iterative competitive training helps to continuously improve the performance and generative ability of GAN. The cGAN can restore close to the real clear image, and is extensively utilized in image enhancement, inpainting, image transformation and other fields. The Feature Pyramid Network module has excellent feature fusion capabilities, generating multi-layer features that encode different semantics and contain higher quality information, striking a balance between efficiency and accuracy. It can be applied to deblurring networks to effectively improve the network's feature extraction ability [9]. The generator adopts an Feature Pyramid Network structure. The complete structure is shown in Fig. 1.

ID algorithms based on GAN usually consider how to design appropriate generator and discriminator network structures, as well as how to balance the training process between the generator and discriminator, to generate high-quality deblurring results. In the feature pyramid structure, the underlying features include less semantic information, but the target localization is

relatively accurate. The upper level features contain rich semantic information, but the target localization is relatively vague. Each layer of feature pyramid is equivalent to a Convolutional neural network. Different size of feature pyramid corresponds to different Receptive field, and different size of convolution kernel corresponds to different feature information in the extracted image [10-11]. The feature pyramid section has three plug and play frameworks to choose from: Inception ResNet v2, MobileNet v2, and MobileNet DSC. This study uses Inception ResNet v2 and MobileNet v2 frameworks. The former focuses on improving the overall restoration quality, while the latter focuses on model lightweight while also ensuring the restoration quality of the image.

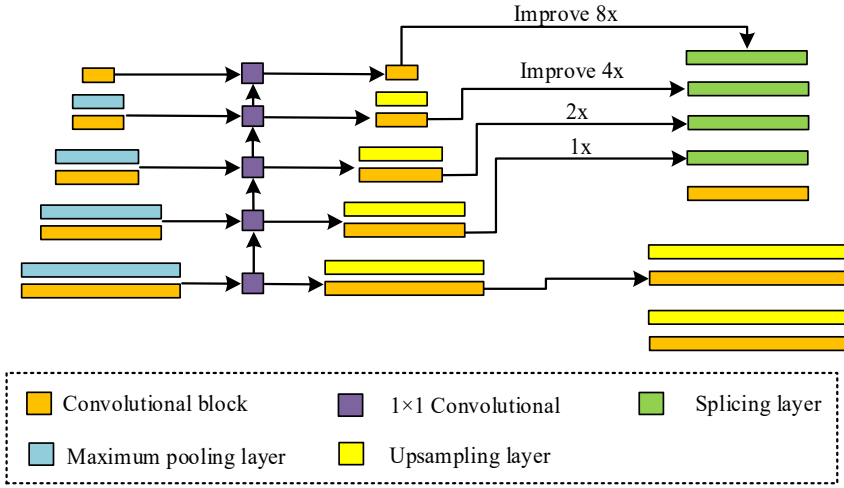


Fig. 1. Generator structure diagram

The discriminator in GAN is utilized for distinguishing the authenticity of generated image, mainly distinguishing the differences in the generated image and the real clear image, including the clarity of the image and whether there is distortion, etc. [12]. This study proposes a dual scale discriminator that targets both local and global aspects, enabling networks to obtain feature information from different dimensions of images, which is beneficial for processing larger and more complex fuzzy types. During the backpropagation process, it can promote the generator to generate more realistic and comprehensive restored images. A local discriminator is used to extract detailed information from images. For those containing complex target movements, a global discriminator helps integrate contextual information. In addition, compared to a single discriminator, the training speed of a dual scale discriminator is faster and smoother, while the deblurring effect has stronger perception and higher quality sharpness. Fig. 2 demonstrates the relevant details.

For image restoration, the commonly used loss structures include the loss in view of pixel space, and the setting of certain parameters in the training phase to compare the restored images and the original blurred image. This study fused three losses as the overall Loss function. It is showcased in Eq. (1):

$$LG = 0.5MSE + 0.006Lc + 0.01Ladv, \quad (1)$$

where, MSE represents the mean squared error (MSE). Lc represents perceptual loss, which can help correct color and texture deformations. $Ladv$ represents global and local losses. The loss function of traditional GAN is shown in Eq. (2):

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim P_Z(z)} [\log (1 - D(G(z)))], \quad (2)$$

where, G and D represent generators and discriminators, respectively. $E_{x \sim P_{data}(x)}$ represents the distribution of real samples. $E_{z \sim P_Z(z)}$ represents that z follows the distribution of a random noise. The loss function of the least squares countermeasure network is showcased in Eq. (3):

$$\begin{cases} \min_D V(D) = \frac{1}{2} E_{x \sim P_{data}(x)} [(D(x) - n)^2] + \frac{1}{2} E_{z \sim P_Z(z)} D(G(z) - m)^2, \\ \min_G V(G) = \frac{1}{2} E_{z \sim P_Z(z)} D[(G(z) - l)^2], \end{cases} \quad (3)$$

where, it is assumed that m and n are the generated restored images and the clear image in the actual scene, respectively. The purpose of the discriminator is to minimize the squared error between the two. It assumes that the image obtained by the generator is set to l , which means that the discriminator identifies it as a real image. In view of the relevant network, the study adapted relativistic, as shown in Eq. (4):

$$LD = E_{x \sim P_{data}(x)} [D(x) - E_{z \sim P_Z(z)} D(G(z) - 1)^2] + E_{z \sim P_Z(z)} [(D(G(z)) - E_{x \sim P_{data}(x)} D(x) + 1)^2]. \quad (4)$$

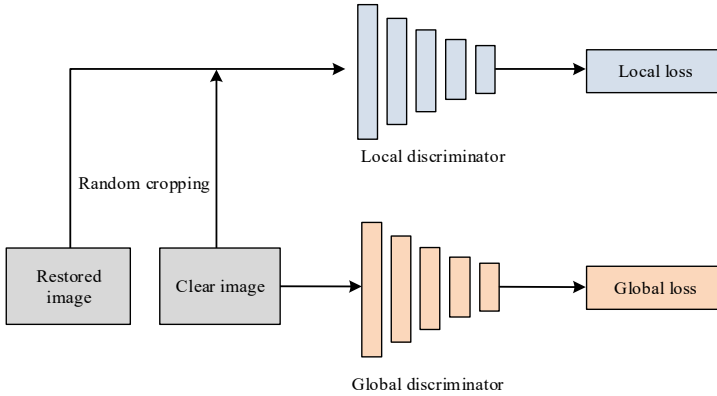


Fig. 2. Structure diagram of dual scale discriminator

In summary, the GAN based ID model constructed in this study mainly contains a generator and a dual scale discriminator. This model adds the results obtained by the generator to the original blurred image for obtaining a restored images. The program inputs the randomly cropped image blocks of the restored and clear image into a local discriminator, calculates the loss, and performs gradient updates based on local loss backpropagation, and continuously cycles until the generator produces a nearly real, clear image. It inputs the restored and clear image into the global discriminator, and the subsequent steps are the same as above. The ID model in view of GAN is shown in Fig. 3.

The quality evaluation of restored images in all image restoration methods follows some universal indicators, which are divided into subjective indicators and objective indicators. Subjective indicator means the utilization of human vision to determine the ID effect of algorithms, without specific standards, and therefore are not entirely reliable and are only used as auxiliary references. Objective indicators are usually divided into two categories: PSNR and structural similarity. The numerical value of PSNR is directly proportional to the quality of image restoration. The deblurring quality of the algorithm is evaluated in three respects: pixel, brightness, and structure, with a pixel range of 0 to 255 [13]. PSNR is shown in Eq. (5):

$$PSNR = 10 \times \lg \left(\frac{255^2}{MSE} \right). \quad (5)$$

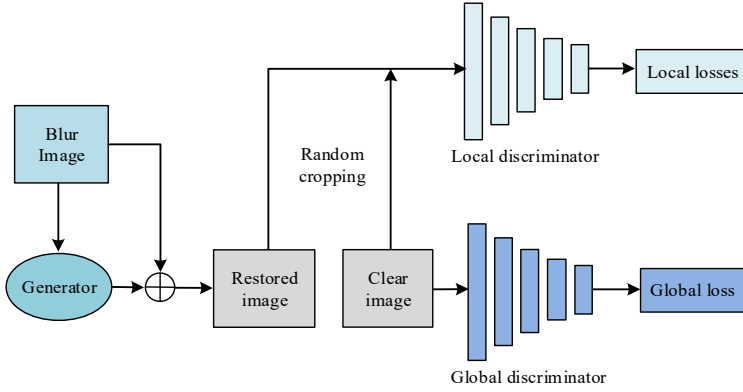


Fig. 3. ID model in view of GAN

The MSE represents the disparity in the restored images and the noisy image. In Eq. (5), it is known that the width of the deblurred image and noise image is n , the length is m , and the MSE is shown in Eq. (6):

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2, \quad (6)$$

where, I represents deblurring the image. K represents a noisy image. $I(i, j)$ and $K(i, j)$ serve as the pixel values of I and K in coordinates. Structural Similarity Index (SSIM) reflects the similarity between images in three respects: brightness, contrast, and structure. The value range is $[0, 1]$, and the larger the value, the greater the similarity in the restored images and the clear image, and the higher the restoration quality [14]. SSIM is shown in Eq. (7):

$$SSIM = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X + \sigma_Y + C_2)}, \quad (7)$$

where, μ_X and μ_Y serve as the mean values of image X and image Y . σ_X and σ_Y serve as variance. σ_{XY} represents covariance. C_1 and C_2 represent constants.

2.2. Research on pedestrian deblurring algorithm in view of MC

PD is an important technology in autonomous vehicle. This technology utilizes computer vision technology for determining the presence of pedestrians in an image or video sequence and provide precise positioning. As the characteristics of both rigid and flexible objects, as well as the susceptibility of pedestrians to factors such as wearing, occlusion, posture, and perspective, PD is a hot and challenging topic in the computer vision [15]. To optimize PD technology for autonomous driving, this study aims to focus on deblurring pedestrians in images in view of ID. YOLO-v3 incorporates current excellent detection framework ideas, such as residual networks and feature fusion, while maintaining speed, further improving detection accuracy, especially for small objects [16]. This study first uses the object detection network YOLO-v3 to detect pedestrians in the image, then saves their coordinates and inputs them into the deblurring network for deblurring the areas within the coordinates. Fig. 4 showcases the relevant details.

However, this pedestrian deblurring algorithm that utilizes the object detection network YOLO-v3 to preprocess pedestrians in images is not end-to-end, and the deblurring effect is not ideal. Therefore, to enable the network to focus on extracting pedestrian features from images, this study adopts MC for feature fusion. MC enables effective handling of local features with

greater precision. Through weighting and MC in backpropagation, the network can focus on extracting pedestrian features from images. The existing multi-scale feature fusion methods in view of DL cannot do without the most basic convolution, and the traditional convolution operation is shown in Eq. (8):

$$\begin{cases} H_{c,x,y} = \sum_{k=1}^{cin} \sum_{i=-\frac{K-1}{2}}^{\frac{K-1}{2}} \sum_{j=-\frac{K-1}{2}}^{\frac{K-1}{2}} g_o F_p, \\ o = c, k, i, j, \\ p = k, x + i, y + j, \end{cases} \quad (8)$$

where, Cin serves as the quantity of input channels in the network. H serves as the output feature map. g represents the convolutional kernel used by the network. F serves as the input feature map. Extended convolution expands the sampling interval in the spatial domain from covering larger objects, as shown in Eq. (9):

$$\begin{cases} H_{c,x,y} = \sum_{k=1}^{cin} \sum_{i=-\frac{K-1}{2}}^{\frac{K-1}{2}} \sum_{j=-\frac{K-1}{2}}^{\frac{K-1}{2}} g_o F_r, \\ r = k, x + id, y + jd, \end{cases} \quad (9)$$

where, d represents the expansion factor. MC disperses organized dilation rates across different kernels within a convolutional filter, which is beneficial for extracting global and local information. It can also integrate multi-scale features, give full play to the characteristics of cavity convolution to expand the local Receptive field, and does not add additional computing costs.

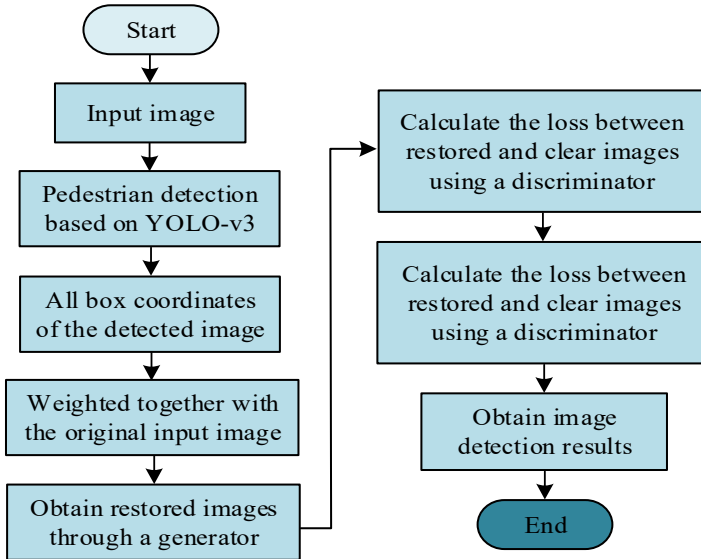


Fig. 4. Structure diagram of a specific target deblurring network in view of PD

Therefore, it can replace the traditional convolution in many commonly used Convolutional neural network backbones, without introducing other parameters or requiring complex computation, while achieving superior Feature learning [17]. In addition, MC collects multi-scale information from various input channels for linear summation operations, and the expansion rate

varies at different kernels of a convolutional filter. As the axis of the input and output channels of the filter periodically changes, the expansion rate also changes, and features are concentrated on a larger scale in a neat style. To process input channels with different Receptive field, the rate of various filters for a specific expansion channel is also different, as shown in Eq. (10):

$$Hc, x, y = \sum_{k=1}^{cin} \sum_{i=-\frac{K-1}{2}}^{\frac{K-1}{2}} \sum_{j=-\frac{K-1}{2}}^{\frac{K-1}{2}} g_{c,k,i,j} F_{k,x+iD(c,k),y+jD(c,k)}, \quad (10)$$

where, the difference between MC and traditional kernel extended convolution is that its expansion rate periodically changes with the axis of the input channels and output channels of the filter. It blends a dilation rate spectrum with a unified kernel size, economically compressing multi-scale features in one convolution operation. MC is periodically dispersed across input channels and output channels, allowing a single channel to gather diverse feature resolutions from the perspectives of both input and output channels [18]. The pedestrian deblurring algorithm based on MC needs to consider how to design MC structures to capture image information at different scales, and how to effectively integrate these multi-scale information to improve the deblurring effect. The MC adopted by this research institute has stronger feature extraction ability compared to other networks, while having similar complexity. Because it applies convolutions of different scales to the same calculation, convolution operation of different scales are alternately performed in the channel, exploring finer granularity. However, this finer grained convolution operation will increase the computing complexity of the network. Therefore, this study improves the training efficiency of the network through group convolution. The principle of group convolution is shown in Fig. 5 [19].

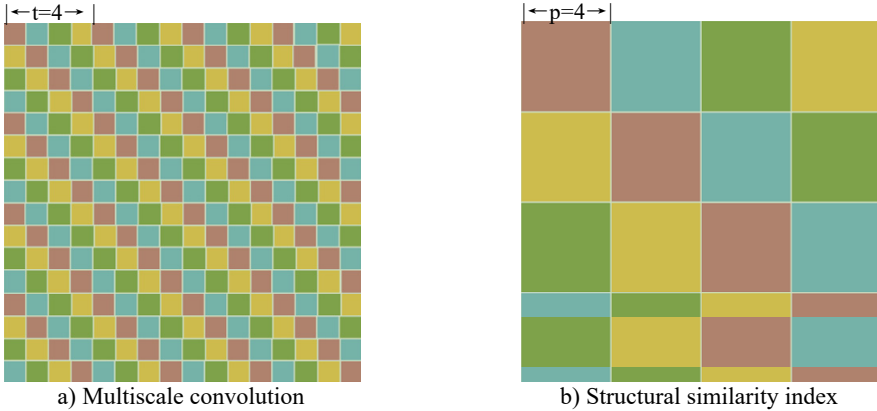


Fig. 5. The principle of group convolution

Multiscale convolution and group convolution are Equivalence relation. To facilitate network calculation, group convolution will replace the MC form in Fig. 5(a). MC is a generalized expansion, where the interval between self cycles determines how many types of expansion rates are included in a partition, and all kernels may share the same expansion rate. This study adopts this MC method to obtain contour information in images. Furthermore, the dataset contains a substantial number of pedestrians, resulting in more acquired pedestrian contour data. This leads to enhanced focus towards the pedestrian part in the subsequent deblurring networks. To achieve better feature fusion results, this study adds MC to the lower layer of the network. In summary, the main workflow of this research method is shown in Fig. 6.

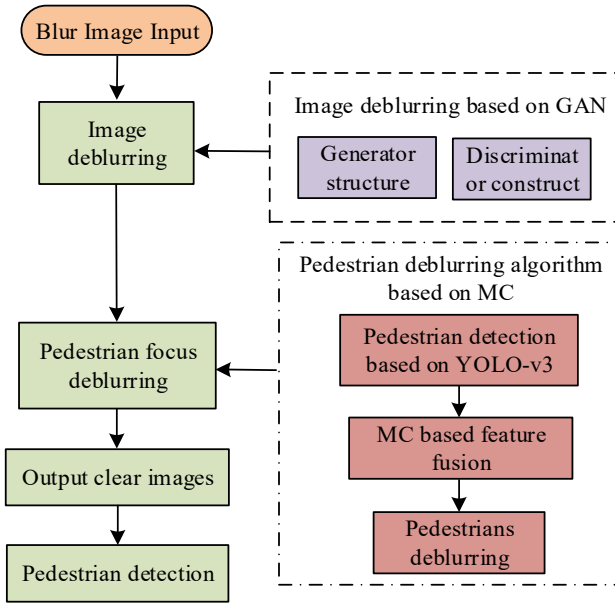


Fig. 6. Main workflow diagram of pedestrian deblurring based on GAN and MC

3. Effect analysis of PD deblurring algorithm in view of GAN and MC

3.1. ID algorithm in view of GAN

The current fuzzy synthesis methods for most fuzzy datasets are relatively single, and there are significant differences between the synthesized fuzzy images and the real fuzzy images. Due to the lack of a large quantity of paired fuzzy and clear image datasets, this study used the GoPro dataset for experimentation, which is also convenient for comparison with other algorithms. The Github address for the GoPro dataset is: <https://seungjunnah.github.io/Datasets/gopro.html>. The study used a 16 GB computer for the experiment, and the graphics card used was GTX1660Ti. This study compared the Inception ResNet v2, MobileNet v2, and MobileNet DSC frameworks of the feature pyramid [20], and the outcomes are demonstrated in Table 1. Table 1 shows that the PSNR and SSIM indicators of Inception ResNet v2 are both the highest, at 29.67 dB and 0.945 dB, respectively. MobileNet-DSC has the shortest computation time, which is 0.07 seconds. But the PSNR index is the lowest at 27.98 dB, and the SSIM index of MobileNet v2 is the lowest at 0.911. Therefore, considering the indicators of the experiment outcomes, this study adopts the Inception ResNet v2 framework, which is suitable for portable terminals such as mobile phones.

The ID algorithm in view of the GAN is compared with the CNN deblurring algorithm in view of the real dynamic scene, the cGAN based deblurring algorithm and the generalized mathematical L0 sparse expression based deblurring algorithm [21-23]. The index outcomes are demonstrated in Fig. 7. The figure shows that compared to the other three algorithms, the algorithm designed in this study has the highest PSNR and SSIM indicators, with 29.7 dB and 0.943 dB, respectively. The PSNR index of the L0 algorithm is the lowest, at 24.6 dB. The SSIM index of the CNN algorithm is the lowest, at 0.807 dB. Furthermore, the CNN algorithm takes 4.36 s to operate, whereas the cGAN algorithm takes 0.87 s. The L0 algorithm, on the other hand, has the longest operation time at 20 minutes, and the ID algorithm records the shortest operation time with respect to the GAN, at just 0.50 s. Therefore, the ID algorithm in view of the GAN has a good deblurring effect, and the obtained image has a good effect in terms of objective indicators, with high efficiency. It has certain effectiveness and superiority.

Table 1. Comparison of indicators for three frameworks

Model	PSNR / dB	SSIM / dB	Time / s
Inception-ResNet-v2	29.67	0.945	0.49
MobileNet-v2	28.20	0.911	0.13
MobileNet-DSC	27.98	0.919	0.07

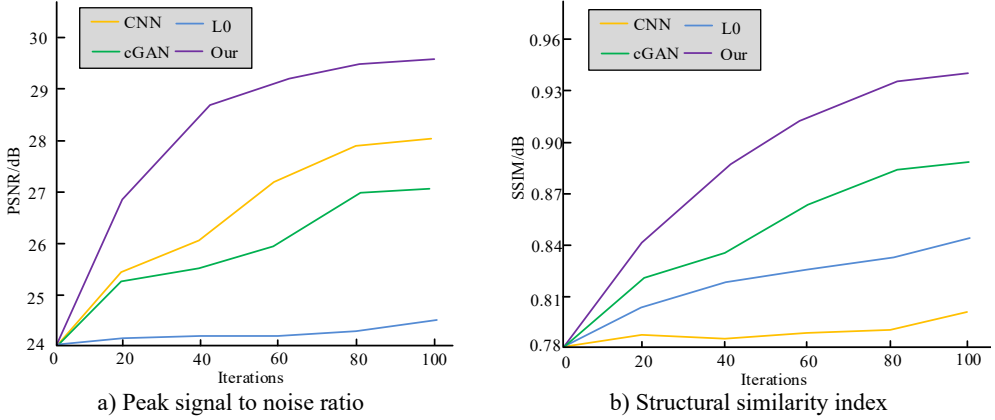


Fig. 7. PSNR and SSIM results of four algorithms

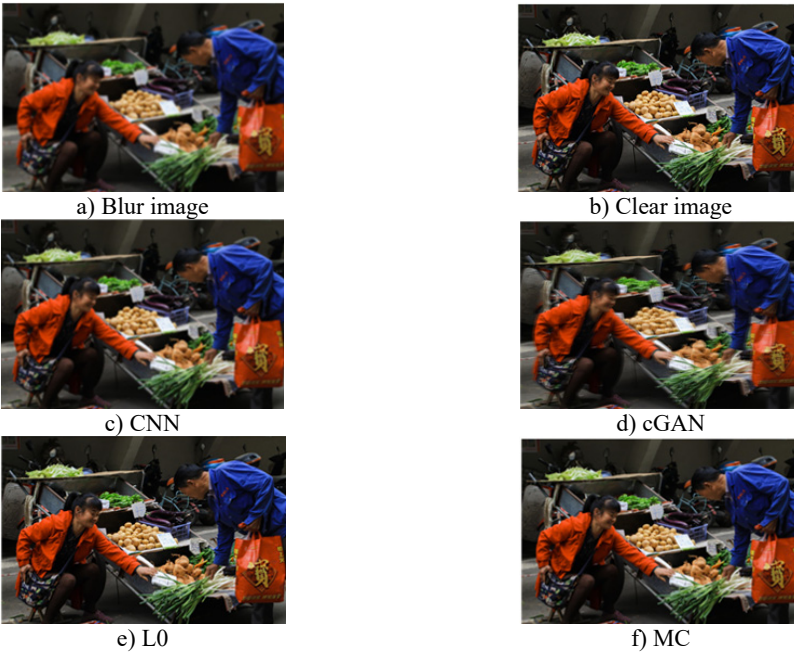


Fig. 8. Comparison of deblurring visual effects of four algorithms

The comparing of the deblurring visual effects of the four algorithms is shown in Fig. 8. Fig. 8 shows that, compared with the other three algorithms, the pedestrian image obtained by the ID algorithm in view of the GAN designed by the research is the clearest in visual effect and the closest to the real clear image. The deblurring effect of the CNN algorithm is poor, and the details of pedestrians are not clear. The image obtained by the cGAN algorithm has artifacts. The L0 algorithm has the worst visual effect, with pedestrians experiencing distortion. Therefore, the deblurring image obtained by the ID algorithm in view of the GAN has a better visual effect, which is closer to the true and clear image, and has achieved good results.

In summary, the three frameworks of the feature pyramid focus on different effects, among which Inception ResNet has the best performance and can markedly enhance the operational efficiency of the network. The MobileNet framework has a shorter running time and is suitable for portable terminals such as mobile phones. The deblurring effect of the ID algorithm in view of the GAN designed in this research has reached excellent outcomes in both, and has higher efficiency. The pedestrian image obtained is the clearest in the visual effect, and also the closest to the real clear image.

3.2. Pedestrian deblurring algorithm in view of MC

This study used a fuzzy dataset HIDE that only includes pedestrians for experimentation (the Github address for the HIDE dataset is: <https://github.com/pp00704831/BANet-TIP-2022>). This dataset is designed for multiple fuzzy attention problems, including 8422 images of different pedestrians in various real scenes, covering a wide variety of fuzzy image types. The training set consists of 6397 images, including two types of high and low blur rates [24]. It divides the 2025 test set images in HIDE into two types: close range pedestrians and distant pedestrians. Compare the pedestrian deblurring algorithm in view of MC with three classic deblurring algorithms: CNN deblurring algorithm in view of real dynamic scenes, cGAN deblurring algorithm in view of cGAN, and generalized mathematical L0 sparse representation deblurring algorithm in view of generalized mathematical L0 sparse representation. The indicator results are shown in Table 2. Table 2 shows that the PSNR indicators of the four algorithms when processing close range pedestrian images are all lower than those when processing distant pedestrian images. The research designed algorithms have the highest PSNR and SSIM indicators, with 29.4 dB and 0.925 dB respectively. Therefore, the pedestrian deblurring algorithm in view of MC designed in this study has reached excellent outcomes in objective indicators, and the deblurring effect is better.

Table 2. Comparison of indicators for four algorithms in the HIDE dataset

HIDE dataset classification	HIDE 1 (Close shot)		HIDE 2 (Vision)	
Algorithm	PSNR / dB	SSIM / dB	PSNR / dB	SSIM / dB
CNN	25.9	0.850	26.9	0.889
cGAN	24.8	0.837	25.2	0.776
L0	21.9	0.759	23.3	0.738
MC	29.3	0.911	29.4	0.925

This study used the most widely used dataset GoPro, which is very close to blurred images in real life, for experiments [25]. The indicator results of the four deblurring algorithms are shown in Fig. 9. Fig. 9 shows that compared to the other three algorithms studied and designed, the PSNR and SSIM indicators are the highest, at 40.45 dB and 0.992 dB, respectively. The PSNR index of the deblurring algorithm in view of the generalized mathematical L0 sparse expression is the lowest at 24.6 dB, while the SSIM index of the CNN deblurring algorithm in view of real dynamic scenes is the lowest at 0.807 dB. Therefore, the pedestrian deblurring algorithm in view of MC has achieved good results in objective indicators, and has certain feasibility and superiority.

Due to the more complex real traffic conditions, the proportion of pedestrians in the obtained detection images often varies in size. Therefore, verifying the algorithm's deblurring effect on pedestrians of different sizes in images and more comprehensively reflecting the advantages and disadvantages of the MC based pedestrian deblurring algorithm is to better optimize the detection of autonomous pedestrians and improve the safety of autonomous driving. Therefore, this study added a deblurring comparative analysis for pedestrian targets of different sizes in the image, as showcased in Fig. 10. From Fig. 10, the algorithm designed in this study achieved good deblurring effects for pedestrians of three distinct sizes. For larger pedestrian targets, facial details were restored more clearly and were close to real clear image. For smaller pedestrian targets, the algorithm was also able to recover their limb movements well. Therefore, the deblurring effect of

the pedestrian deblurring algorithm in view of MC is not limited to the size of pedestrians in the image.

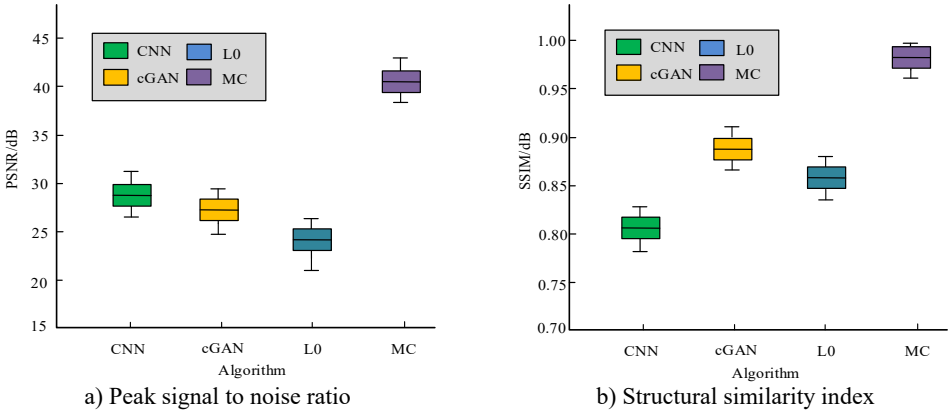


Fig. 9. Indicator results of four deblurring algorithms on the GoPro dataset

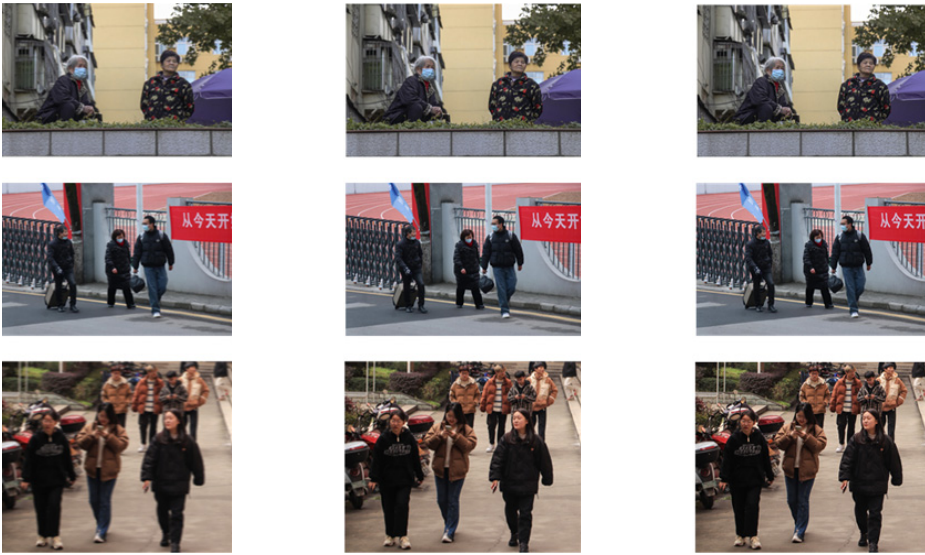
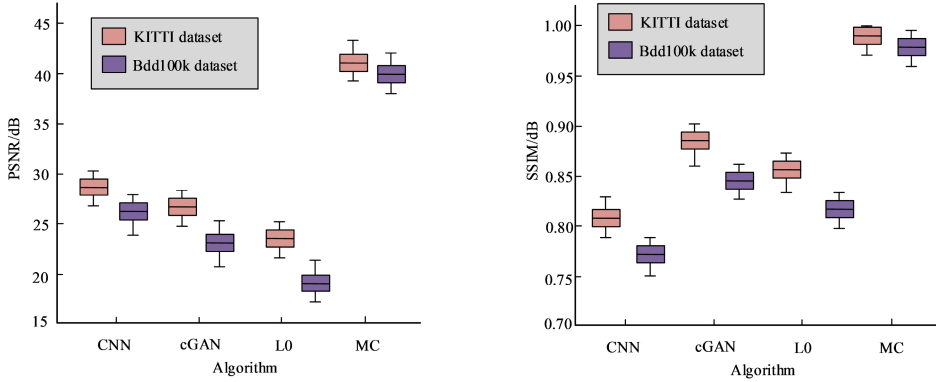


Fig. 10. Comparison of deblurring of pedestrian targets of different sizes in the GoPro dataset

To further verify the PD performance of the MC based pedestrian deblurring algorithm in real driving scenarios, experiments were conducted using the KITTI dataset. This dataset emphasizes the algorithm's robustness and accuracy in actual road scenes, as well as the PD dataset Bdd100k which covers various extreme weather conditions. The comparison results of the indicators of the four algorithms are shown in Fig. 11. From Fig. 11, it can be seen that compared to the other three algorithms, the PSNR and SSIM metrics of the algorithm designed in the study perform better on both datasets. The results show that the pedestrian deblurring algorithm based on MC presents a notable level of accuracy and robustness in real driving conditions.

In summary, the pedestrian deblurring algorithm in view of MC designed in this study has reached excellent outcomes in both in different datasets. It has also reached excellent outcomes in restoring pedestrian details, surpassing several traditional deblurring algorithms and producing images closer to reality. In addition, for pedestrians of different sizes, the deblurring effect of the algorithm is maintained at a high level, and the deblurring effect is not limited to the size of

pedestrians in the image.



a) Peak signal to noise ratio
b) Structural similarity index
Fig. 11. Comparison results of indicators on the KITTI dataset and the Bdd100k dataset

4. Conclusions

As the boost of the society, autonomous vehicle attracts the public’s concern. Security is one of the most important issues of concern. Strengthening pedestrian target detection is beneficial for improving the safety of autonomous driving. To deblur specific targets of pedestrians in images, this study established a PD deblurring model in view of GAN and MC, and designed ID algorithms in view of GAN and pedestrian deblurring algorithms in view of MC. The results showed that the PSNR and SSIM indicators of Inception ResNet v2 were both the highest, at 29.67 dB and 0.945 dB, respectively. The PSNR and SSIM metrics of the GAN based ID algorithm are both the highest, at 29.7 dB and 0.943 dB, respectively, with the shortest operation time of 0.50 s. The pedestrian deblurring algorithm in view of MC has the highest PSNR and SSIM indicators in the HIDE test set, reaching 29.4 dB and 0.925 dB respectively. The PSNR and SSIM indicators in the GoPro dataset are both the highest, at 40.45 dB and 0.992 dB, respectively. The restored image obtained is the clearest and visually effective, with more detailed information visible in the enlarged part of the face, which is very close to the real clear image. The restoration effect of the enlarged part of the pedestrian is also good, and the deblurring effect is not limited to the size of the pedestrian in the image. In summary, the model constructed by the research institute has certain feasibility and effectiveness, and has good results in ID and pedestrian detail restoration. However, the fuzzy dataset containing pedestrians collected in this study lacks sufficient richness, which may affect the practical application effect of the model. Therefore, it is necessary to collect more pedestrian fuzzy data for experiments to demonstrate the deblurring effect of the model in different scenarios. This would enable the model to be suitably applied to enhance the safety of autonomous driving.

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Xiujuan Dong contributed the central idea, analysed most of the data, and wrote the initial

draft of the paper. Jianping Lan contributed to refining the ideas, carrying out additional analyses, and finalizing this paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] K. Zhang et al., “Deep image deblurring: A survey,” *International Journal of Computer Vision*, Vol. 130, No. 9, pp. 2103–2130, Jun. 2022, <https://doi.org/10.1007/s11263-022-01633-5>
- [2] T. D. Ngo, T. T. Bui, T. M. Pham, H. T. B. Thai, G. L. Nguyen, and T. N. Nguyen, “Image deconvolution for optical small satellite with deep learning and real-time GPU acceleration,” *Journal of Real-Time Image Processing*, Vol. 18, No. 5, pp. 1697–1710, May 2021, <https://doi.org/10.1007/s11554-021-01113-y>
- [3] Y.-Q. Liu, X. Du, H.-L. Shen, and S.-J. Chen, “Estimating generalized gaussian blur kernels for out-of-focus image deblurring,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, No. 3, pp. 829–843, Mar. 2021, <https://doi.org/10.1109/tcsvt.2020.2990623>
- [4] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, “Generative adversarial networks,” *ACM Computing Surveys*, Vol. 54, No. 6, pp. 1–38, Jul. 2022, <https://doi.org/10.1145/3459992>
- [5] H. Maeda, T. Kashiya, Y. Sekimoto, T. Seto, and H. Omata, “Generative adversarial network for road damage detection,” *Computer-Aided Civil and Infrastructure Engineering*, Vol. 36, No. 1, pp. 47–60, Jun. 2020, <https://doi.org/10.1111/mice.12561>
- [6] B. Wang, Y. Lei, N. Li, and W. Wang, “Multiscale convolutional attention network for predicting remaining useful life of machinery,” *IEEE Transactions on Industrial Electronics*, Vol. 68, No. 8, pp. 7496–7504, Aug. 2021, <https://doi.org/10.1109/tie.2020.3003649>
- [7] H. Tomosada, T. Kudo, T. Fujisawa, and M. Ikehara, “GAN-based image deblurring using DCT loss with customized datasets,” *IEEE Access*, Vol. 9, pp. 135224–135233, Jan. 2021, <https://doi.org/10.1109/access.2021.3116194>
- [8] Q. Zhao, D. Zhou, and H. Yang, “Cdm-net: context-aware image deblurring using a multi-scale cascaded network,” *Neural Processing Letters*, Vol. 55, No. 4, pp. 3985–4006, Jul. 2022, <https://doi.org/10.1007/s11063-022-10976-6>
- [9] F. Wen, R. Ying, Y. Liu, P. Liu, and T.-K. Truong, “A simple local minimal intensity prior and an improved algorithm for blind image deblurring,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, No. 8, pp. 2923–2937, Aug. 2021, <https://doi.org/10.1109/tcsvt.2020.3034137>
- [10] Y. Yu et al., “Capsule feature pyramid network for building footprint extraction from high-resolution aerial imagery,” *IEEE Geoscience and Remote Sensing Letters*, Vol. 18, No. 5, pp. 895–899, May 2021, <https://doi.org/10.1109/lgrs.2020.2986380>
- [11] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, and J. Yang, “Road segmentation for remote sensing images using adversarial spatial pyramid networks,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 6, pp. 4673–4688, Jun. 2021, <https://doi.org/10.1109/tgrs.2020.3016086>
- [12] J. Wang, Y. Chen, Z. Dong, and M. Gao, “Improved YOLOv5 network for real-time multi-scale traffic sign detection,” *Neural Computing and Applications*, Vol. 35, No. 10, pp. 7853–7865, Dec. 2022, <https://doi.org/10.1007/s00521-022-08077-5>
- [13] B. Li, X. Qi, P. H. S. Torr, and T. Lukasiewicz, “Lightweight generative adversarial networks for text-guided image manipulation,” *Advances in Neural Information Processing Systems*, Vol. 33, pp. 22020–22031, Jan. 2020, <https://doi.org/10.48550/arxiv.2010.12136>
- [14] K. Wu et al., “Improvement in signal-to-noise ratio of liquid-state NMR spectroscopy via a deep neural network DN-Unet,” *Analytical Chemistry*, Vol. 93, No. 3, pp. 1377–1382, Jan. 2021, <https://doi.org/10.1021/acs.analchem.0c03087>
- [15] M. Dehshiri, S. Ghavami Sabouri, and A. Khorsandi, “Structural similarity assessment of an optical coherence tomographic image enhanced using the wavelet transform technique,” *Journal of the Optical Society of America A*, Vol. 38, No. 1, pp. 1–9, Jan. 2021, <https://doi.org/10.1364/josaa.401280>
- [16] Y. Xiao et al., “Deep learning for occluded and multi-scale pedestrian detection: A review,” *IET Image Processing*, Vol. 15, No. 2, pp. 286–301, Dec. 2020, <https://doi.org/10.1049/ipr2.12042>

- [17] H. R. Alsanad, A. Z. Sadik, O. N. Ucan, M. Ilyas, and O. Bayat, "YOLO-V3 based real-time drone detection algorithm," *Multimedia Tools and Applications*, Vol. 81, No. 18, pp. 26185–26198, Mar. 2022, <https://doi.org/10.1007/s11042-022-12939-4>
- [18] Y. Qin, X. Wang, Q. Qian, H. Pu, and J. Luo, "Multiscale transfer voting mechanism: A new strategy for domain adaption," *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 10, pp. 7103–7113, 2020.
- [19] M.-A. Li, J.-F. Han, and J.-F. Yang, "Automatic feature extraction and fusion recognition of motor imagery EEG using multilevel multiscale CNN," *Medical and Biological Engineering and Computing*, Vol. 59, No. 10, pp. 2037–2050, Aug. 2021, <https://doi.org/10.1007/s11517-021-02396-w>
- [20] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8878–8887, Aug. 2019.
- [21] Y. Wu, P. Qian, and X. Zhang, "Two-level wavelet-based convolutional neural network for image deblurring," *IEEE Access*, Vol. 9, pp. 45853–45863, Jan. 2021, <https://doi.org/10.1109/access.2021.3067055>
- [22] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8183–8192, Apr. 2018.
- [23] Y. Zhang, H. Pu, and J. Lian, "Quick response barcode deblurring via L_0 regularisation based sparse optimisation," *IET Image Processing*, Vol. 13, No. 8, pp. 1254–1258, May 2019, <https://doi.org/10.1049/iet-ipr.2018.5738>
- [24] B. M. N. Smets, J. Portegies, E. J. Bekkers, and R. Duits, "PDE-based group equivariant convolutional neural networks," *Journal of Mathematical Imaging and Vision*, Vol. 65, No. 1, pp. 209–239, Jul. 2022, <https://doi.org/10.1007/s10851-022-01114-x>
- [25] Z. Zhong, Y. Gao, Y. Zheng, B. Zheng, and I. Sato, "Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network," *International Journal of Computer Vision*, Vol. 131, No. 1, pp. 284–301, Oct. 2022, <https://doi.org/10.1007/s11263-022-01705-6>



Xiujuan Dong received her master's degree from Northwest Normal University in 2012. She is currently a Lecturer at the School of Hubei Institute of Automobile Industry. Her research interests include Intelligent traffic digital image processing, signal and information processing.



Jianping Lan received his master's degree from Northwest Normal University in 2012. He is currently a Lecturer at the School of Hubei Institute of Automobile Industry. His research interests include automotive electronic and electrical architecture, active safety and intelligent driving.