Double-channel cascade-based generative adversarial network for power equipment infrared and visible image fusion

Jihong Wang[1] and Haiyan Yu[1,*]

[1]School of Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou 450000 China

Abstract

At present, visible light imaging sensor and infrared imaging sensor are two commonly used sensors, which are widely used in aviation, navigation and other military fields of detection, monitoring and tracking. Due to their different working principles, their performance is different. The infrared imaging sensor records the infrared radiation information of the target itself by acquiring the infrared radiation of the ground target. It identifies the target by detecting the thermal radiation difference between the target and the background, so it has special recognition and camouflage ability, such as finding people, vehicles and artillery hidden in the woods and grass. Although the infrared imaging sensor has a good detection performance for thermal targets, it is insensitive to the brightness changes of the scene and has low imaging resolution, which is not conducive to human eyes interpretation. Visible light imaging sensor is sensitive to the reflection of the target scene and has nothing to do with the thermal contrast of the target scene. The obtained image has high clarity and can provide the details of the target scene. Therefore, the fusion of infrared and visible images will be beneficial to the combination of infrared image's better target indication characteristics and visible image's scene clearing information. In this paper, we propose a double-channel cascade-based generative adversarial network for power equipment infrared and visible image fusion. The experimental results show that the fusion image not only retains the target information of the infrared image, but also retains more details of the visible image, and achieves better performance in both subjective and objective evaluation.

*Corresponding author. Email: zkdywjh@126.com; zxcvfdsa5024@foxmail.com

1. Introduction

Image fusion is a process in which two or more sensors obtain the image or image sequence information about a specific scene at the same time or at different times to synthesize it to generate new information about the scene interpretation [1-3]. In here, visible and infrared image fusion is one of the research fields of multi-source sensor information fusion [4], and its application involves target tracking [5], target detection [6], medical imaging [7], etc. Visible image can provide the color and texture details of the scene, but it is susceptible to the change of ambient light; infrared image reflects the thermal radiation information, although not susceptible to the change of light, but can not provide details. By using their

complementary information, infrared and visible images can be fused to obtain more comprehensive and accurate information of the target and scene, and it has better visual effects.

At present, the fusion methods of infrared image and visible image mainly include traditional method and image fusion method based on deep learning. Traditional image fusion methods can be divided into two categories: one is image fusion method based on Transform domain, such as multi-scale transform (MST) [8,9] and sparse representation (SR) [10,11]. The other is spatial domain based image Fusion, such as guided filtering-based fusion (GFF) [12,13], etc. However, most fusion methods require manual design of complex fusion rules, and the calculation is more and more complicated. In recent years, methods based on deep learning have been widely used in the field of image fusion. According to the Network structure, the image fusion methods based on deep learning include Convolutional Neural Network (CNN) [14-17] and Generative Adversarial Network (GAN) [18]. Reference [19] proposed a method based on convolutional neural network, which could effectively extract source image features, but this method ignored the middle layer, resulting in information loss and still required manual design of activity level measurement and fusion rules.

Reference [20] proposed a generative adversarial network for infrared and visible image fusion method. The final fusion image is obtained through the adversarial training between generator network and discriminator network, but only relying on adversarial training to increase the details leads to insufficient local information of the fused image, and the target edge information in the fusion result is often fuzzy.

To solve the above problems, this paper proposes a double-channel cascade-based generative adversarial network (DCGAN) for power equipment infrared and visible image fusion. DCGAN is an end-to-end model

without the need to manually design complex fusion rules. The balance is achieved through the game confrontation between generator and discriminator to fuse infrared and visible images. At the same time, the evaluation index is introduced into the loss function part to guide the model training and get higher quality fusion image.

2. Proposed DCGAN

2.1. DCGAN framework

The framework of infrared and visible image fusion method based on DCGAN is shown in figure 1. The fusion image of infrared and visible images with rich information is generated by the generator model. In the generator model, feature extraction and transmission from source images and feature information sharing are carried out by cascade method. Discriminator network plays an antagonistic role with generator network in the whole framework, distinguishing whether the target image is a real image or one generated by generator through training, and obtaining the final fusion image through game antagonistic balance with generator network. In identifying two discriminant model design, namely discriminant infrared image and visible light image discriminant, in the process of generator networks and discrimination against, fused images through infrared image discriminant for more target information, and through the visible light image discriminant for more detail information, until the discriminant cannot distinguish between generate images and real images.
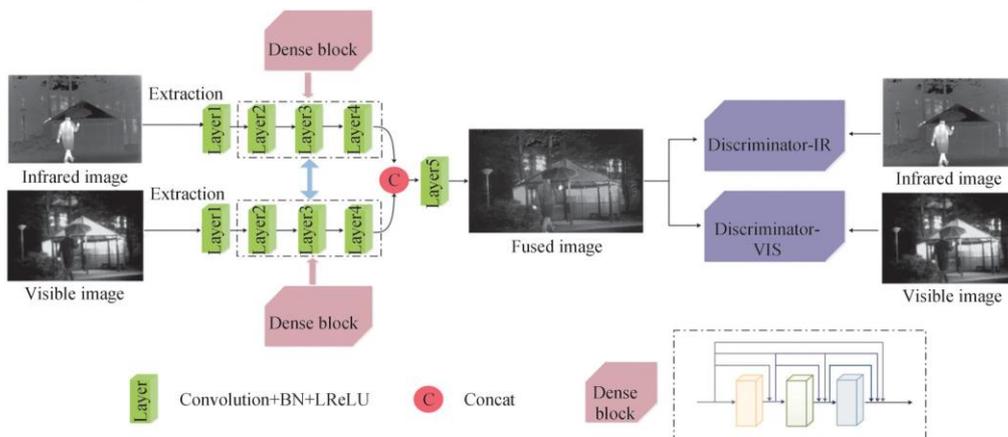
Figure 1. DCGAN framework

The generator network consists of four parts: input source image, feature extraction module, feature connection module and fusion output module. The network structure is shown in figure 2. The feature extraction module contains four layers of convolutional neural networks. Each convolutional layer uses a 3×3 convolutional kernel with step size set to 1. A 1×1 convolutional kernel is used in the fusion output module. Compared with infrared image, visible image contains more details and infrared image contains more target information. In order to retain more effective information of source image, the idea of dense network is introduced and information exchange learning is designed between the two channels to build a generator model. The exchange information is generated through the method of connection and convolution between the two paths, and the exchange information is connected with the output of the previous layer as the input of the next convolution layer. In the process of designing the network structure, in order to avoid problems such as gradient dispersion, Batch Norm was used to normalize the data after the convolution kernel at the first four layers according to reference [21], and then an LReLu activation function was used to improve the network effect, and Tanh activation function was used at the last layer.

Figure 2. Generator structure

The main function of discriminator network is to judge whether the images generated by generator obey the real sample distribution. In order to improve the quality of fusion image, the generated fusion image can retain more texture details through game confrontation training between generator and discriminator. In this paper, Discriminator-IR (D-IR) and Discriminator-VIS (D-VIS) have the same network structure design, as shown in figure 3. In here, 3×3 convolution kernel is used in the convolution layer, and the number of filters in the convolution layer is set to 32, 64, 128 and 256. In order to avoid introducing noise, the convolution layer with step size of 2 is used to replace the pooling layer, which makes the classification effect of discriminator better. In order to prevent the image distribution from being destroyed leading to the training instability, according to the reference [22], the discriminator input layer does not use data normalization. All layers of the discriminator use activation functions to improve the effect of image generation, and the last layer is linear layer classification.
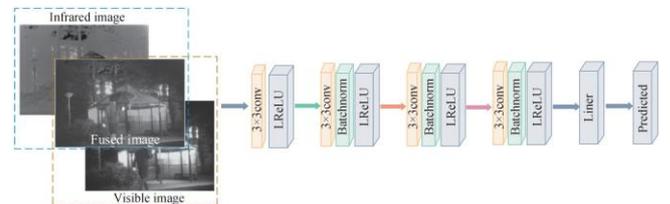
Figure 3. Network structure of the discriminator

2.2. Loss function

The loss function L_G of the generator includes multi-antagonistic loss L_{adv} and content loss L_{con} between the generator and the discriminator, which can be expressed as:

$$L_G = L_{adv} + \alpha L_{con} \qquad (1)$$

In the formula, α aims to balance L_{adv} and L_{con}. The purpose of L_{adv} is to obtain more detail information of visible image and target information of infrared image through multiple antagonism between generator and two discriminators.

The adversarial loss function is designed in the form of mean square error loss to calculate the error between the reconstructed image and the source image. The smaller L_{adv} denotes the better generated image fusion by the generator, i.e.,

$$L_{adv} = \frac{1}{N}\sum\sum[D_i(I_{fused}^n) - a]^2 \qquad (2)$$

Where N represents the number of fused images, I_{fused}^n represents the fused images. $D_i(\cdot)$ represents the discriminator classification result. a represents the false data value that the generator hopes the discriminator to believe. The designed content loss function in the generator network structure is expressed as:

$$L_{con} = \frac{1}{HW}(\parallel I_f - I_r \parallel_F^2 + \xi \parallel \nabla I_f - \nabla I_v \parallel_F^2) \qquad (3)$$

In the formula, H represents the height of the input source image. W represents the width of the input source image. $\parallel \cdot \parallel_F$ is the norm. I_r represents the infrared image of the input. I_f represents the fused image. ξ aims to balance the positive parameters between the two terms, and ∇ represents the gradient calculation.

According to the subjective evaluation, it can be seen that the fusion results only pay attention to the visible light information and lose the thermal target information of the infrared image. Therefore, in order to retain more effective information of the source image and the correlation between the source image and the fusion image, the content loss function designed in this paper includes gradient loss and similarity loss, namely,

$$L_{con} = \beta_1 L_{grad} + \beta_2 L_{SSIM} \qquad (4)$$

Where β_1 and β_2 are the parameters controlling the balance between the two items. L_{grad} aims to retain thermal target information of infrared image and detail information of visible image. L_{SSIM} aims to retain similarity information between fusion result and source image, and comprehensively measure consistency between brightness, contrast and structure.

Since the thermal radiation information of the infrared image is characterized by its pixel intensity and the texture detail information of the visible image is characterized by its gradient [23], the image intensity and gradient are calculated and can be expressed as:

$$L_{grad} = \frac{1}{HW}[\lambda_1(\parallel I_{fused} - I_{ir} \parallel_F^2 + \parallel I_{fused} - I_{vis} \parallel_F^2) + \lambda_2(\parallel I_{fused} - I_{ir} \parallel_F^2 + \parallel I_{fused} - I_{vis} \parallel_F^2] \qquad (5)$$

Where λ_1 and λ_2 are the positive parameters controlling the balance between the two terms. I_{ir} and I_{vis} represent the infrared image and visible image respectively. ∇I_{ir} and ∇I_{vis} represent the values of the two source images calculated by the gradient operator

In the design of content loss function, the idea of reference [24] is introduced. Image structural similarity is an image quality measure that calculates the differences between images. In order to retain valid information in the source image, the similarity weight between the fusion image and the source image is defined, which can be expressed as:

$$L_{SSIM} = 1 - SSIM \qquad (6)$$

Where SSIM represents the structural similarity between the output fusion image and the input source image.

In the image fusion based on generative adjudgment network proposed, only a single visible discriminator is designed to fight against the generator, forcing the fusion image to keep more details of the visible image. It can be seen that the fusion results only focus on visible light information and lose the target information of the infrared image. In this paper, two discriminators are designed to make the fusion image not only retain more detailed information of visible image, but also contain thermal target information of infrared image. The loss function of discriminator D_IR and discriminator D_VIS can be expressed as:

$$L_{D\_IR} = \frac{1}{N}\sum_{n=1}^{N}[D_{IR}(I_{ir}) - b]^2 + \frac{1}{N}\sum_{n=1}^{N}[D_{IR}(I_{fused}) - d]^2 \qquad (7)$$

$$L_{D\_VIS} = \frac{1}{N}\sum_{n=1}^{N}[D_{VIS}(I_{vis}) - c]^2 + \frac{1}{N}\sum_{n=1}^{N}[D_{VIS}(I_{fused}) - d]^2 \qquad (8)$$

Where b, c, d represent the truth data of infrared image, visible image and fusion image respectively.

3. Experimental results and analysis

3.1. Experimental platform and parameter setting

From the TNO Image Fusion Dataset [25] and VIFB public Dataset [26], 41 and 21 groups of registered visible and infrared images in different scenes are selected as experimental training data, and then 22 and 21 groups are selected as testing data. In order to expand the training data set, the sliding window method is adopted, the clipping step is set to 10, the size of the clipping image block is 120×120, and the training parameters are set as

follows: the batch size is 32, the number of training iterations is 20, and the initial learning rate of gradient descent is 10. In the generator, α is set as 20, $\beta_1 = 1$, $\beta_2 = 5$, $\lambda_1 = 15$ and $\lambda_2 = 500$. The parameter settings are close to 1. The truth value of infrared image b and visible image c are set as random numbers between 0.7 and 1.2. The truth value of fusion image d is set as between 0 and 0.3. The training step of discriminator is set to 2, and Adam optimizer is used to optimize network model parameters [27-30].

DCGAN network is trained based on TNO and VIFB open data sets. Firstly, a single infrared and visible image is taken as the input image of the network respectively, and the fusion image is generated through the generator network with double cascaded channels. Under the constraint of generator loss function, the information of the two source images is preserved as much as possible. Then, the discriminator network discriminates and classiates the image and the visible and infrared images, and repeats in the number of iterations, so as to realize the antagonistic training of generator network and discriminator network and get the ideal model parameters. Parameter optimization in DCGAN is updated by Adam optimizer.

Experimental hardware platform: the operating system is Win10, CPU: Intel Core I5, 3.2GHz, Memory 24 GB, Software platform MATLAB2019a, VS Code, training environment: Tensorflow 1.14, Python version 3.5.2.

3.2. Subjective evaluation

In order to prove the effectiveness of the double discriminator, four groups of images ("Men", "Helicopter", "Steamboat" and "Solider") are set up for detailed experimental comparison and analysis, and the results are shown in figure 4. Figure 4(a) (b) are four groups of registered infrared and visible images respectively. Figure 4(c) (d) are the results of infrared image discriminator and visible image discriminator respectively. Taking the "Helicopter" group as an example, it can be seen that a single infrared discriminator training model can retain the pixel intensity information of infrared images, but lacks the details of visible images. Although the single visible discriminator training model can retain the pixel intensity information of helicopter, the spatial contrast of the whole image is low. Meanwhile, some visible light detail is missing from the bottom of the helicopter. Figure 4(e) is the result of the double discriminator. It is not difficult to see that the fused image has higher visual information fidelity and looks more clear and natural.

The DCGAN method in this paper is compared with eight classical image fusion methods, including Cross Bilateral Filter (CBF) [31], Convolutional Sparse Representation (CSR) [32], Fourth order Partial Differential Equation (FPDE) [33], A Fusion Approach to Infrared and Visible Images (DenseFuse) [34], generative adversarial network (FusionGAN), a Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion (DDCGAN) [35], A Novel Decomposition Method for Infrared and Visible Image Fusion (MDLatLRR) [36]. The "Men" group is selected for experimental analysis, and the experimental results are shown in figure 5. It can be seen that the experimental results of CBF algorithm and ConvSR algorithm show that the tree background is over-blurred and contains noise artifacts. The experimental results of FPDE algorithm, CNN algorithm and DenseFuse algorithm improve the above problems, and there is no artifact phenomenon in the part with branches, but the results lack the texture details of branches.
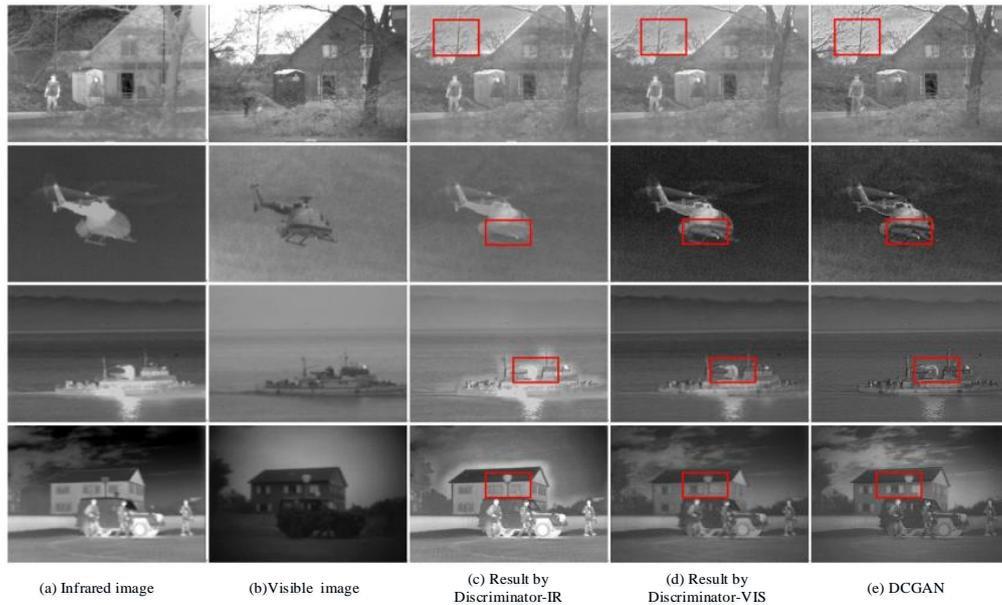
Figure 4. Subjective fusion results of infrared and visible images of four different scenes
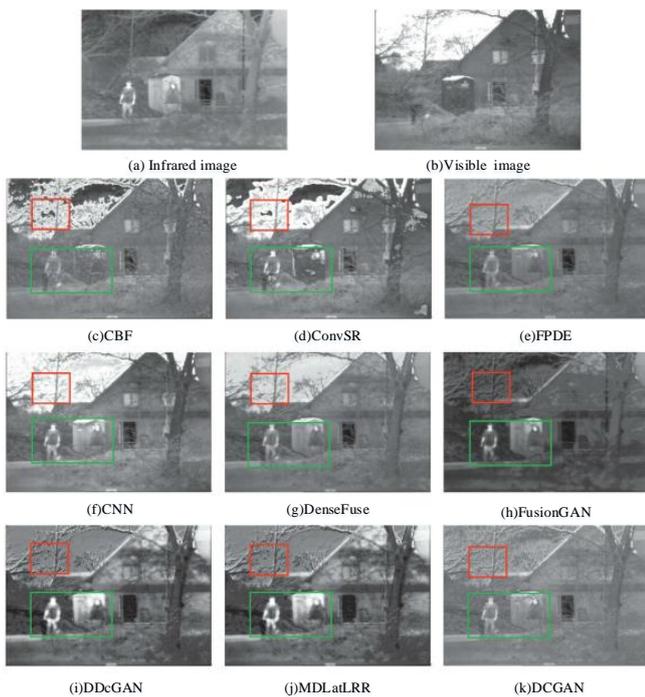
Figure 5. Comparison of fusion results of 'Two men in front of house'

The experimental results of FusionGAN algorithm, DDGAN algorithm and MDLatLRR algorithm show that the hierarchical contrast between the target and the whole scene is weak and more edge and detail information of infrared and visible images is missing. In the experimental results of DPCAM algorithm proposed in this paper, the texture structure features of the upper left branches are rich and obvious, and the fusion image retains the edge information of the infrared image well and contains more detailed information of the visible image.

The fusion results of some scenes may be accidental. Another six groups of images of different scenes are selected from the TNO Image Fusion Dataset and VIFB for an expanded comparative experiment, and the experimental results are shown in figure 6. It is not difficult to see that the fusion results of CBF and ConvSR algorithms are not clear and have very serious noise phenomenon. Compared with CBF and ConvSR algorithms, the fusion results of FPDE, CNN, DenseFuse and MDLatLRR algorithms contain clearer contour information of source images, but some information is not obvious. The fusion results of FusionGAN and DCGAN algorithms are obtained through the antagonistic game between generator network and discriminator network. Most of the results only retain the useful information of visible images, but lack the target characteristic information of infrared images. However, the proposed method can obtain high quality fusion image with more detailed information and clear target.

Figure 6. Results of infrared and visible image fusion based on TNO and VIFB dataset with different algorithms

The subjective evaluation of fusion images is based on the subjective judgment of human eyes, which is one-sided and random to some extent. Therefore, the fusion image quality evaluation needs to be combined with objective evaluation comprehensive analysis and comparison.

3.3. Objective evaluation of fused images

Information entropy (EN), edge information retention ($Q^{AB/F}$), difference correlation coefficient sum (SCD), structural similarity (SSIM), visual information fidelity (VIFF) and spatial frequency (SF) are selected for comparative experimental analysis of fusion image quality [37].

1) Information entropy is an evaluation index based on information theory, which measures the average amount of information contained in the fused image and is defined as:

$$EN = -\sum_{l=0}^{L-1} p_l \log_2 p_l \qquad (9)$$

Where L is the gray level of the image. P is the normalized histogram of the corresponding gray level in the fusion image. The larger entropy value denotes the more information contained in the fused image.

2) Edge information retention is an evaluation index based on source image and fusion image, which evaluates the fusion quality of significant information expression in the fusion image, and is defined as:

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[Q^{AF}(i,j)\omega^A(i,j)+Q^{BF}(i,j)\omega^B(i,j)]}{\sum_{i=1}^{M}\sum_{j=1}^{N}[\omega^A(i,j)+\omega^B(i,j)]}$$

(10)

$$Q^{AF}(i,j) = Q_a^{AF}(i,j)Q_g^{AF}(i,j) \qquad (11)$$

$$Q^{BF}(i,j) = Q_a^{BF}(i,j)Q_g^{BF}(i,j) \qquad (12)$$

Where $\omega^A(i,j)$ and $\omega^B(i,j)$ represent the edge information intensity of source image A and B at position (i,j). ω^X represents the weight of importance of each source image to the fused image. $Q_a^{BF}(i,j)$ and $Q_g^{BF}(i,j)$ represent the fused image respectively. The larger the edge information retention value is, the more edge information is contained in the fused image.

3) The information correlation coefficient is an evaluation index based on the source image and the fusion image. According to the correlation between image differences and the difference correlation sum of image quality measures, the false information contained in the fusion image can be evaluated, which is defined as:

$$SCD = r(D_1, I_1) + r(D_2, I_2) \qquad (13)$$

$$r(D_X, I_X) = \frac{\sum_i \sum_j [D_X(i,j)-\overline{D}_X][I_X(i,j)-\overline{I}_X]}{\sqrt{\sum_i \sum_j [D_X(i,j)-\overline{D}_X]^2[I_X(i,j)-\overline{I}_X]^2}}$$

(14)

Where X=1,2. $D_1(D_1 = I_f - I_r)$ and $D_2(D_2 = I_f - I_{vi})$ represent the information of infrared and visible images respectively. $\overline{D}_X$ and $\overline{I}_X$ represent the average pixel values of D_X and I_X respectively. The larger the information correlation coefficient denotes

the better the fusion performance and the less false information.

4) Visual information fidelity is an evaluation index based on human visual perception, which reflects image distortion characteristics by fitting real human eyes. Multi-scale method is used to divide source image and fusion image into different blocks, and then evaluate the visual information of each block. Finally, the mutual information between images is calculated to measure the quality of the image [38-41].

5) Structural similarity is an evaluation index based on structural similarity, which measures the degree of similarity with source image brightness, contrast and structure by calculating the amount of information change of image structure. A new unreferenced image fusion performance measure (MS-SSIM) is defined as:

$$SSIM_a(I_F) = 0.5[SSIM(I_F, I_1) + SSIM(I_F, I_2)] \quad (15)$$

Where I_1 and I_2 represent the infrared and visible images respectively. I_F represents the fusion image.

6) Spatial frequency is an evaluation index based on image features, which is used to measure the structural texture of the fusion result. The larger the value of spatial frequency is, the richer the edge and texture information of the fused image is, which is defined as:

$$SF = \sqrt{RF^2 + CF^2} \quad (16)$$

$$RF = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}[F(i,j) - F(i,j-1)]^2} \quad (17)$$

$$CF = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}[F(i,j) - F(i-1,j)]^2} \quad (18)$$

Where M and N represent the height and width of the fused image respectively. RF and CF represent the frequency of the row and the frequency of the column respectively. $F(i,j)$ represents the value of the pixel point of the i-th row and j-th column in the fused image. The larger the SF value denotes the clearer image.

Firstly, a quantitative analysis is conducted on the experiment group of "Two men in front of house", and the results are shown in Table 1. As can be seen from the objective evaluation index data in Table 1, the VIFF index of DPCAM algorithm is 40.31% higher than that of FusionGAN algorithm and has good visual information fidelity. At the same time, $Q^{AB/F}$, SCD, MS-SSIM and SF are superior to FusionGAN, increasing by 0.72%, 69.58%, 40.26%, 18.07% and 71.14% respectively. Although DCGAN algorithm is lower than contrast algorithm in some objective evaluation indexes, most indexes are better than contrast algorithm, which is consistent with subjective evaluation.

Table 1. 'Two men in front of house' objective evaluation of comparative experiments

Index	Method								
	CBF	ConvSR	FPDE	CNN	DenseFuse	FusionGAN	DDCGAN	MDLatLRR	Proposed
EN	6.487	6.461	5.630	6.428	5.916	6.261	**7.353**	4.939	6.938
$Q^{AB/F}$	0.395	0.431	0.331	0.338	0.339	0.134	0.237	**0.451**	0.439
SCD	0.655	0.320	1.343	0.563	1.024	1.006	1.577	1.499	**1.819**
VIFF	0.431	0.896	0.565	0.723	0.719	0.449	**0.923**	0.469	0.753
MS-SSIM	0.556	0.827	0.929	0.910	0.895	0.784	0.619	0.924	**0.956**
SF	0.042	0.046	0.047	0.044	0.020	0.001	0.033	0.018	**0.053**

Ten groups of infrared and visible images of different scenes are selected from the TNO dataset and fused images are obtained by different algorithms. The fusion images are quantitatively analyzed and compared, and the obtained index scores are visualized in the form of broken line graph. Figure 7 is the objective evaluation index broken line graph of nine different algorithms [42-46]. It can be seen that, as a hot research method in recent years, deep learning method achieves higher index scores compared with the indexes obtained by traditional

methods. In addition, the comparison between various indicators shows that DCGAN method has good visual information fidelity while improving fusion image quality, which is consistent with subjective evaluation.

The objective evaluation experiment was expanded and 21 groups of fusion results were selected from TNO and VIFB data sets for objective comparison experiment. The results are shown in Table 2.

It can be seen that the proposed method has obtained the average quantitative values of 6 evaluation indexes in 21

groups of different scenarios, and the results of 3 evaluation indexes are optimal. Although the method is not optimal in the results of the other 3 evaluation indexes, it is only second to the individual algorithm, which is consistent with the subjective evaluation results on the whole. In other words, the fusion result of the proposed algorithm has less false information, and the fusion result is naturally clear and has a good sense of hierarchy.

Table 2. The average quantitative value of the fusion results of 21 groups

Index	Method								
	CBF	ConvSR	FPDE	CNN	DenseFuse	FusionGAN	DDCGAN	MDLatLRR	Proposed
EN	6.711	6.907	6.798	7.128	6.919	6.373	**7.489**	6.210	6.918
$Q^{AB/F}$	0.333	**0.404**	0.265	0.307	0.278	0.130	0.348	0.372	0.363
SCD	1.408	1.164	1.564	1.443	1.665	1.325	1.626	1.659	**1.756**
VIFF	0.531	0.708	0.705	0.782	0.741	0.365	0.721	0.653	**0.760**
MS-SSIM	0.699	0.784	0.876	0.891	0.829	0.588	0.767	0.897	**0.899**
SF	0.522	**0.063**	0.059	0.062	0.027	0.019	0.049	0.028	0.053
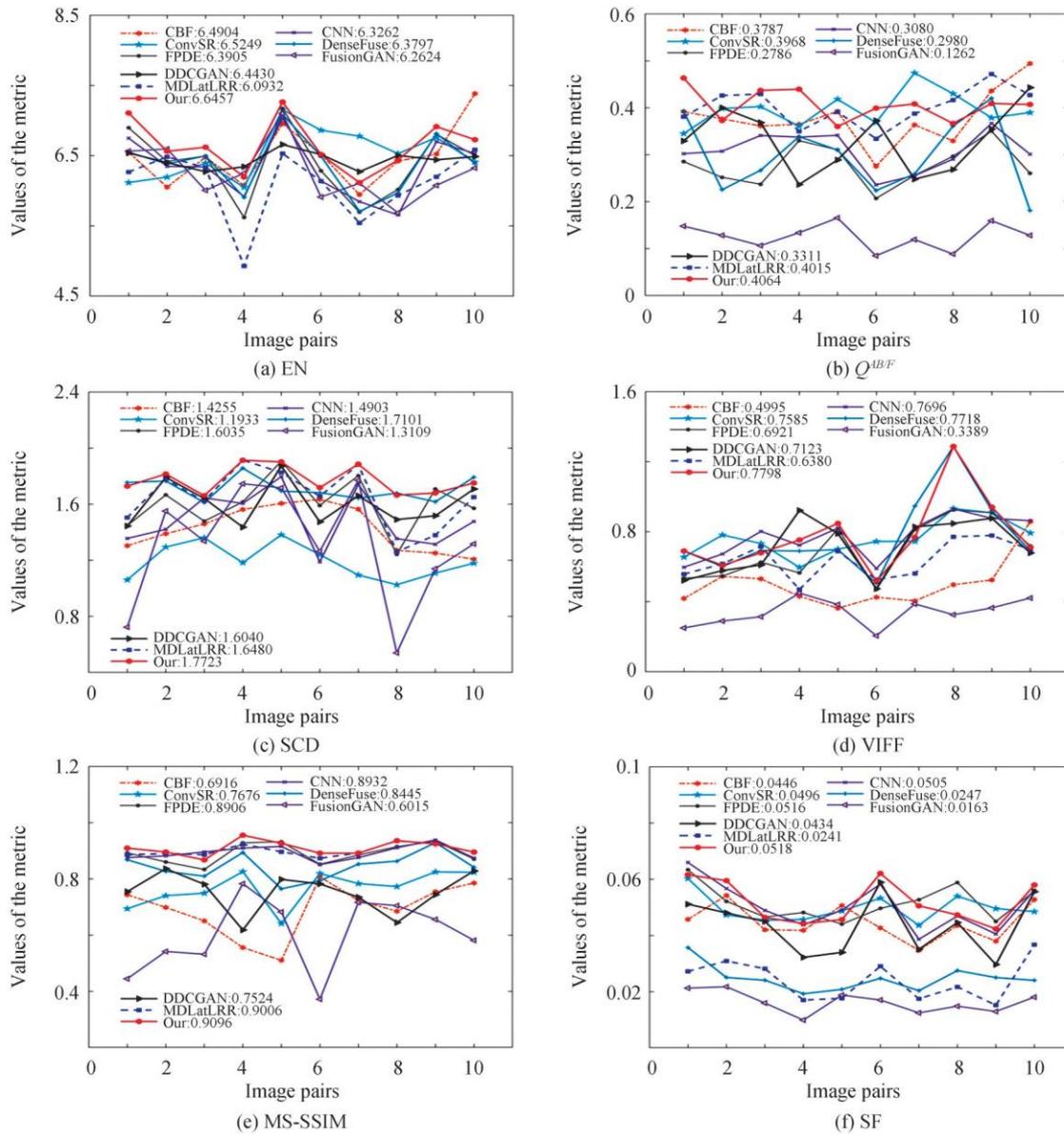
Figure 7. Comparison of six objective metrics for different algorithms

4. Conclusion

In this paper, combining the generative adversarial network and dense network, an infrared and visible image fusion method based on double channel cascade adversarial mechanism is proposed. In the generator model, feature extraction is carried out by double-path cascade to retain more information of the source image, and the real fusion image is obtained by game confrontation between the generator model and the discriminator model. The proposed method is compared with eight classical image fusion methods by qualitative and quantitative analysis. The results show that the proposed method can not only reduce the noise, but also keep the thermal target information of infrared image and more texture details of visible image, and the fusion results are clear and natural. The subsequent work will continue to improve the network structure, seek the optimal parameter value, make each evaluation index achieve the ideal effect, and apply it in different fields.

References

[1] Zhang L, Zeng G, Wei J. Adaptive Region-Segmentation Multi-Focus Image Fusion Based on Differential Evolution[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(32).

[2] Shoulin Yin, Ye Zhang. Singular value decomposition-based anisotropic diffusion for fusion of infrared and visible images[J]. International Journal of Image and Data Fusion, 10(2), pp: 146-163, 2019.

[3] Shoulin Yin, Hang Li, Lin Teng, Man Jiang & Shahid Karim. An optimised multi-scale fusion method for airport detection in large-scale optical remote sensing images [J]. International Journal of Image and Data Fusion, vol. 11, no. 2, pp. 201-214, 2020. DOI: 10.1080/19479832.2020.1727573

[4] Zhou Z, Bo W, Sun L, et al. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters[J]. Information Fusion, 2016, 30, pp.15-26.

[5] X. Zhang, P. Ye, S. Peng, J. Liu, K. Gong and G. Xiao, "SiamFT: An RGB-Infrared Fusion Tracking Method via Fully Convolutional Siamese Networks," in IEEE Access, vol. 7, pp. 122122-122133, 2019, doi: 10.1109/ACCESS.2019.2936914.

[6] Shoulin Yin, Hang Li, Asif Ali Laghari, et al. A Bagging Strategy-Based Kernel Extreme Learning Machine for Complex Network Intrusion Detection[J]. EAI Endorsed Transactions on Scalable Information Systems. 21(33), e8, 2021. http://dx.doi.org/10.4108/eai.6-10-2021.171247

[7] Li Yan, Xiaowei Wang, and Shoulin Yin. Campus Garbage Image Classification Algorithm Based on New Attention Mechanism [J]. International Journal of Electronics and Information Engineering. Vol. 13, No. 4, pp. 131-141, 2021.

[8] Li S, Kang X, Fang L, et al. Pixel-level image fusion: A survey of the state of the art[J]. Information Fusion, 2017, 33, pp.100-112.

[9] Qingwu Shi, Shoulin Yin, Kun Wang, Lin Teng and Hang Li. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems (2021). https://doi.org/10.1007/s12530-021-09392-3

[10] Yang Y, Song S, Liu D, et al. Robust Background Feature Extraction Through Homogeneous Region-Based Joint Sparse Representation for Hyperspectral Anomaly Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, PP(99):1-15.

[11] Jisi A and Shoulin Yin. A New Feature Fusion Network for Student Behavior Recognition in Education [J]. Journal of Applied Science and Engineering. vol. 24, no. 2, pp.133-140, 2021.

[12] S. Li, X. Kang and J. Hu, "Image Fusion With Guided Filtering," in IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2864-2875, July 2013, doi: 10.1109/TIP.2013.2244222.

[13] Rajalingam B, Al-Turjman F, Santhoshkumar R, et al. Intelligent multimodal medical image fusion with deep guided filtering[J]. Multimedia Systems, 2020(1):1-15.

[14] Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation [J]. Multimedia Tools and Applications. Vol. 79, pp. 31049-31068, 2020.

[15] S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.

[16] Yin, S., Li, H. & Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images [J]. Sensing and Imaging, vol. 21, 2020. https://doi.org/10.1007/s11220-020-00314-2

[17] Shoulin Yin, Jie Liu, and Lin Teng. A Sequential Cipher Algorithm Based on Feedback Discrete Hopfield Neural Network and Logistic Chaotic Sequence [J]. International Journal of Network Security. Vol. 22, No. 5, pp. 869-873, 2020.

[18] Bharti V, Biswas B, Shukla K K. EMOCGAN: a novel evolutionary multiobjective cyclic generative adversarial network and its application to unpaired image translation[J]. Neural Computing and Applications, 2021:1-15.

[19] R. Lai, Y. Li, J. Guan and A. Xiong, "Multi-Scale Visual Attention Deep Convolutional Neural Network for Multi-Focus Image Fusion," in IEEE Access, vol. 7, pp. 114385-114399, 2019, doi: 10.1109/ACCESS.2019.2935006.

[20] Ma J, Wei Y, Liang P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48:11-26.

[21] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. 2015. arXiv:1511.06434

[22] Yu Y., Gong Z., Zhong P., Shan J. Unsupervised Representation Learning with Deep Convolutional Neural Network for Remote Sensing Images. International Conference on Image and Graphics. ICIG 2017. Lecture Notes in Computer Science, vol 10667, pp. 97-108. Springer, Cham. https://doi.org/10.1007/978-3-319-71589-6_9

[23] A. Nawaz et al., "Fusion of color and infrared images using gradient transfer and total variation minimization," 2017 2nd International Conference on Computer and

Communication Systems (ICCCS), 2017, pp. 82-85, doi: 10.1109/CCOMS.2017.8075272.

[24] Dongdong Xu, Yongcheng Wang, Shuyan Zhu, et al. Infrared and Visible Image Fusion with a Generative Adversarial Network and a Residual Network[J]. Applied Sciences, 2020, 10(2):554.

[25] TNO image fusion dataset[EB/OL]. https://flgshare. com/articles/TNO Image Fusion Dataset/1008029, September 15, 2018.

[26] X. Zhang, P. Ye and G. Xiao, "VIFB: A Visible and Infrared Image Fusion Benchmark," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 468-478, doi: 10.1109/CVPRW50498.2020.00060.

[27] Xiaowei Wang, Shoulin Yin, Hang Li. A Network Intrusion Detection Method Based on Deep Multi-scale Convolutional Neural Network[J]. International Journal of Wireless Information Networks. 27(4), 503-517, 2020.

[28] Xiaowei Wang, Shoulin Yin, Ke Sun, Hang Li, Jie Liu and Shahid Karim. GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition [J]. Journal of Applied Science and Engineering, vol. 23, no. 3, pp. 555-561, 2020.

[29] Yang Sun, Shoulin Yin, Hang Li, Lin Teng, Shahid Karim. GPOGC: Gaussian Pigeon-Oriented Graph Clustering Algorithm for Social Networks Cluster [J]. IEEE Access. Volume: 7, Page(s): 99254 - 99262, 03 July 2019.

[30] Shoulin Yin, Ye Zhang and Shahid Karim. Region search based on hybrid convolutional neural network in optical remote sensing images[J]. International Journal of Distributed Sensor Networks, Vol. 15, No. 5, 2019.

[31] Kumar S, B. K. Image fusion based on pixel significance using cross bilateral filter[J]. Signal Image & Video Processing, 2015, 9(5):1193-1204.

[32] Y. Liu, X. Chen, R. K. Ward and Z. Jane Wang, "Image Fusion With Convolutional Sparse Representation," in IEEE Signal Processing Letters, vol. 23, no. 12, pp. 1882-1886, Dec. 2016, doi: 10.1109/LSP.2016.2618776.

[33] GAO Xueqin, LIU Gang, XIAO Gang, et al. Fusion algorithm of infrared and visible images based on FPDE [J]. Journal of Automation, 2020, 46(4): 796-804. (In Chinese)

[34] H. Li and X. Wu, "DenseFuse: A Fusion Approach to Infrared and Visible Images," in IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2614-2623, May 2019, doi: 10.1109/TIP.2018.2887342.

[35] J. Ma, H. Xu, J. Jiang, X. Mei and X. -P. Zhang, "DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion," in IEEE Transactions on Image Processing, vol. 29, pp. 4980-4995, 2020, doi: 10.1109/TIP.2020.2977573.

[36] H. Li, X. -J. Wu and J. Kittler, "MDLatLRR: A Novel Decomposition Method for Infrared and Visible Image Fusion," in IEEE Transactions on Image Processing, vol. 29, pp. 4733-4746, 2020, doi: 10.1109/TIP.2020.2975984.

[37] TANG Lili, LIU Gang, XIAO Gang. Infrared and Visible Image Fusion Method Based on Dual -path Cascade Adversarial Mechanism[J]. Acta Photonica Sinica, 2021, 50(9):0910004

[38] Yin Shoulin, Liu Jie, Li Hang. A Self-Supervised Learning Method for Shadow Detection in Remote Sensing Imagery[J]. 3D Research, vol. 9, no. 4, December 1, 2018.

[39] Yin Shoulin, Liu Jie, Teng Lin. A new krill herd algorithm based on SVM method for road feature extraction[J]. Journal of Information Hiding and Multimedia Signal Processing, v 9, n 4, p 997-1005, July 2018.

[40] Shoulin Yin, Ye Zhang, Shahid Karim. Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model[J]. IEEE Access. volume 6, pp: 26069 - 26080, 2018.

[41] Shoulin Yin, Jie Liu, Ye Zhang, Lin Teng. Cuckoo search algorithm based on mobile cloud model[J]. International Journal of Innovative Computing, Information and Control. Volume 12, Number 6. pp.1809-1819. 2016.

[42] Laghari, A.A., Wu, K., Laghari, R.A. et al. A Review and State of Art of Internet of Things (IoT). Arch Computat Methods Eng (2021). https://doi.org/10.1007/s11831-021-09622-6

[43] Laghari A A, Laghari M A. Quality of experience assessment of calling services in social network[J]. ICT Express, 2021(2).

[44] Laghari A A, Laghari K, Memon K A, et al. Quality of Experience (QoE) Assessment of Games on workstations and Mobile[J]. Entertainment Computing, 2020, 34:100362.

[45] A. A. Laghari, H. He, A. Khan, N. Kumar and R. Kharel, "Quality of Experience Framework for Cloud Computing (QoC)," in IEEE Access, vol. 6, pp. 64876-64890, 2018, doi: 10.1109/ACCESS.2018.2865967.

[46] Laghari, A.A., Jumani, A.K. & Laghari, R.A. Review and State of Art of Fog Computing. Arch Computat Methods Eng 28, 3631–3643 (2021). https://doi.org/10.1007/s11831-020-09517-y