Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Infrared visible image fusion algorithm based on double branching and resultant decomposition

Likun Hu[1,✉], Wei Zhou[2]

[1] *Electrical Engineering School of Guangxi University, Nanning Guangxi, 541000, China*

ABSTRACT

Currently, visible and infrared image fusion (VIF) technology has a wide range of applications in road safety monitoring, anti-surveillance, etc. However, the traditional image fusion algorithms in the feature fusion process will have limitations such as part of the information is lost, etc. For this reason, this paper proposes an infrared visible image fusion algorithm based on the double-branching and decomposition of the results. The algorithm firstly adopts the dense block method, extracts visible image features, and uses a feature pyramid network to extract infrared features. The algorithm firstly adopts the dense block method to extract the visible image features, and uses the feature pyramid network to extract the infrared features, then, based on the deep learning network structure to extract the image information of different modalities, and designs the fusion network constrained by the three loss functions of the gradient loss, intensity loss and decomposition loss, so as to obtain a good fusion effect of the image. The experimental results show that the proposed algorithm achieves the optimal value in five indexes, and reaches sub-optimal value in one index, indicating that the proposed algorithm fuses the images with the optimal value and sub-optimal value. At the same time, the proposed algorithm retains the main thermal radiation information of infrared images better than other algorithms such as DenseFuse and IFCNN, which is superior to some extent.

Keywords: infrared image fusion, result decomposition, feature pyramid, deep learning, loss function

1. Introduction

With the continuous development of artificial intelligence technology and image sensing technology, various types of image sensors are widely used in people's daily life. However, all kinds of image sensors acquire information both redundant information with the same target and complementary information, which are scattered in images of different modalities, resulting in the extraction of effective information becomes difficult [1-2]. Multimodal image fusion is to extract and integrate the information from different modal images of the same scene, aggregate the effective information,

✉ Corresponding author.
E-mail address: ziei4821@163.com (L. Hu).
Received 25 February 2024; Revised 05 May 2024; Accepted 10 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-430

remove the redundant information, and improve the information quality of the image and scene perception, which is conducive to a variety of subsequent advanced tasks [3-6]. Infrared and visible light image fusion is a typical example of multimodal image fusion. Visible light images have better resolution and better reflection effect for the texture edge information of the target image, but the imaging quality is greatly affected by external factors [7-10]. Infrared images can reflect the infrared band information beyond the human eye, converted to visible information mapped to the image, but the infrared image contrast and the ability to show details is insufficient [11-13]. Therefore, both infrared images and visible images have certain limitations, relying on a single type of image can not meet the actual needs of engineering, while the fusion of the two can make the target to be understood in the complex environment is still highlighted [14-15].

A variety of infrared and visible image fusion methods have been proposed by related researchers. They are broadly categorized into traditional image fusion methods and deep learning methods in terms of fusion framework. Traditional image fusion methods are usually highly interpretable. Li, Z. et al. proposed a visible and near-infrared image fusion algorithm that fuses the reflection weight model and transmission weight model for the problem of color distortion and halo artifacts caused by the loss of spectral features in the fused image, which effectively avoids color distortion while maintaining the naturalness of the fused image [16]. Zhao, L. et al. emphasized that multi-scale decomposition of infrared and visible images is the key to fusing images, and designed a new image fusion algorithm to perform multi-scale decomposition of images in the spatial frequency domain, and to obtain fused images with high clarity and rich texture by extracting and fusing the significant feature textures of the images [17]. Ma, J. and Zhou, Y. constructed a gradient wavelet filter based on a fuzzy gradient threshold function and global optimization, which was applied to the fusion of infrared and visible images, and was able to significantly improve the contrast of the fused images and play an important role in edge blurring and noise processing [18].

With the development of deep learning techniques, various deep learning methods are applied to image fusion tasks. Li, H. and Wu, X. J. established a deep learning architecture for the infrared and visible image fusion problem, which is capable of extracting the relevant features in the source image and fusing them, and subsequently reconstructing the fused image through a decoder to achieve better infrared and visible image fusion [19]. Junwu, L. et al. introduced the LSWT-NSST fusion algorithm, which utilizes the lifting smooth wavelet transform (LSWT) to decompose the infrared and visible images and the non-subsampled shear wavelet transform (NSST) to extract the target and detail features of the images, which shows good visual perception in the fusion of infrared and visible images [20]. Tang, L. et al. simultaneously considered the visual quality of infrared and visible image fusion and the high-level visual task requirements, and proposed a semantic-aware real-time image fusion network algorithm (SeAFusion), which utilizes semantic loss to guide the high-level semantic information back to the image fusion module, complementing the gap between image fusion and high-level visual tasks [21]. Zhang, L. et al. proposed a ResNet-152 based fusion algorithm for infrared and visible images, which utilizes the ResNet-152 network to process the multilayer features of the high-frequency portion on the basis of the original image decomposition, and the reconstructed fused image effectively reduces the artifacts and noises and presents more details of the fused image [22].

In addition, the image fusion algorithm based on double branching and result decomposition consists of intra-channel branching and cross-channel branching, and then the results of the two branches are superimposed, and the fused image can retain more image information, which guarantees the diversity of the fused image features and realizes better fusion effects. However, there are few studies on the fusion of infrared and visible images using image fusion algorithms based on double branching and result decomposition, and further research is urgently needed.

Although methods based on deep learning have made great progress in theory and application in recent years, there are still some shortcomings. Firstly, using fixed network structure to extract image features of different modes will lead to partial information loss. Many algorithms based on deep

learning use the same network structure to extract information in different modal images, which cannot obtain the most effective features and may lead to information loss. Secondly, previous algorithms have never considered using the images obtained after fusion to guide the fusion process. As the output result, the quality of the fused image is closely related to the source image and the fusion process. The previous algorithm did not try to use the fusion result to add to the training of image fusion.

2. The Algorithm Model

2.1. Network Model

The proposed algorithm network is mainly divided into four parts: encoder, fusion layer, decoder and decomposition network. Firstly, infrared and visible images are input to the encoder. After feature extraction, the features fused by the fusion layer are input to the decoder to obtain the fusion image. Finally, the fused image is input into the decomposition network to obtain the decomposed infrared and visible image, and the decomposed image will be input into the loss function to guide the training of the network. The overall framework structure of the network is shown in Figure 1.

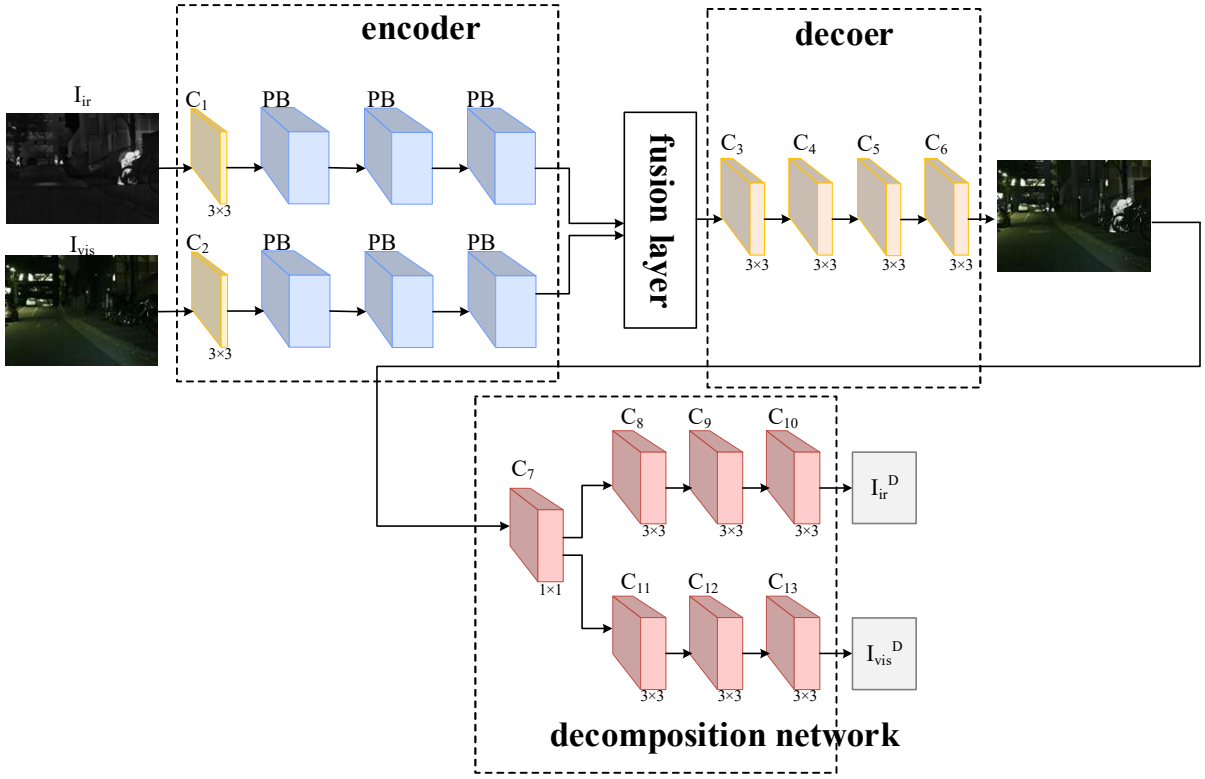


Fig. 1. Overall Network Structure

In the network, the infrared image and visible image are input to the encoder, processed by the fusion layer, and then transmitted to the decoder. The fusion image is input to the decomposition network, and the decomposed infrared visible image is obtained through the decomposition network. In the two feature extraction branches of the encoder, two convolution layers with the size of 3×3 and activation function of ReLU are first designed to extract shallow features. Then, the infrared feature extraction branch and visible light feature extraction branch are designed. The two branches include three pyramid blocks (PB) and three dense blocks (DB) respectively, which are used to extract features from images of different modes. These two modules are described in sections 2.2 and 2.3.

Then the extracted features are sent to the fusion layer, and the two features obtained are fused by means of splicing in the fusion stage. Then a decoder is used to process the fusion features. In the decoder, the features are reconstructed using four convolution layers with the size of 3×3 and the activation function is ReLU to obtain the fusion image.

After that, the purpose of using the result decomposition network is to decompose the fused image to obtain the sum of the external and visible images. First, a common 1×1 convolution layer is used to extract features from the fused images, and then decomposition is achieved using two branches, each of which uses three convolution layers with 3×3 convolution kernels. In this part, except for the last convolutional layer using tanh as the activation function, other convolutional layers use ReLU as the activation function. The difference between the decomposed infrared and visible images and the source images is smaller and smaller by comparing the two resulting images with the source images and inputting the comparison results into the loss function for training. It allows the decomposition results to better align with the source image. This consistency of decomposition can make the fusion result contain more details and reduce the information loss in the fusion process.

To prevent information loss, the step size of all convolutional layers in the network is set to 1. Except for the first and last layers, the fill of all convolutional layers is set to 1, and the fill of the first and last layers is set to 0.

2.2. Dense Block Structure

Because the visible image and infrared image contain different mode information, different structures are used in the two feature extraction branches according to their mode characteristics. The use of dense structures in image fusion was first proposed in DenseFuse, where each layer is connected to all previous layers. This connection enables information to be transmitted quickly and allows the network to fuse the features of different layers at an earlier stage, which is conducive to reducing the loss of details when extracting the features of visible light images. In this paper, three dense blocks are designed to extract visible light features. The structure of dense blocks is shown in Figure 2.

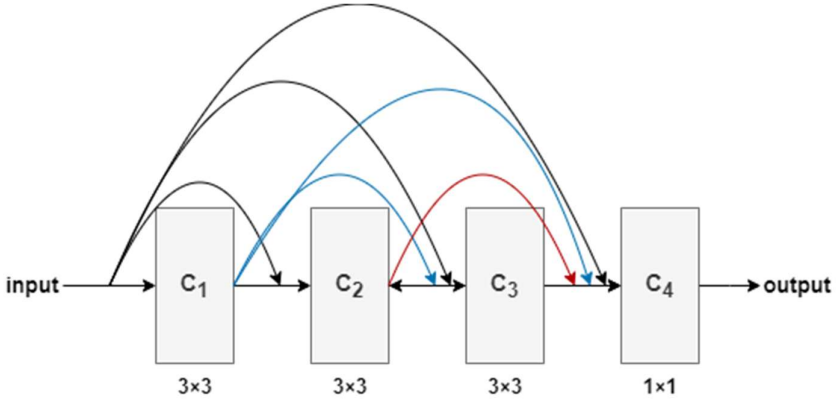


Fig. 2. Dense Block Structure

The dense block contains four convolution layers, where the first three convolution layers, and the size are 3×3, and the last convolution layer is 1×1 for dimensionality reduction. In these four convolutional layers, the input of each layer is a cascade of the output of all the previous layers. This structure can realize inter-layer interconnection, reduce information loss, and is suitable for extracting detailed features of visible light images.

2.3. Pyramid Block Structure

For infrared images, more attention should be paid to the significance information, so it is necessary to use a structure that is more conducive to extracting edge features to extract features. In the infrared

feature extraction branch, this paper adopts the pyramid block structure for feature extraction, which was first proposed and is used in semantic segmentation tasks.

The pyramid block structure is shown in Figure 3.

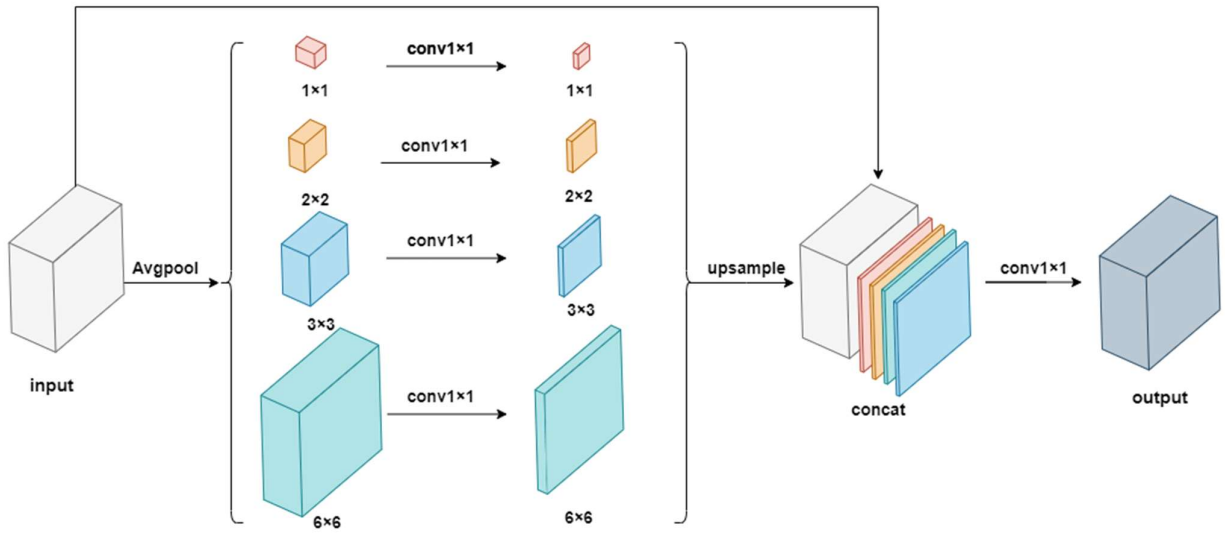


Fig. 3. Pyramid Structure

The pyramid module can divide the input feature map to obtain grids of different sizes, and carry out average pooling in each grid respectively. Such a structure can obtain global information by aggregating information from different regions. In the pyramid module, the input features are first divided into 1×1, 2×2, 3×3 and 6×6 grid areas, where the 1×1 division represents the average pooling of the entire input features. Similarly, the other three divisions represent dividing the input features into 2×2, 3×3, and 6×6 subregions, and then average pooling of each subregion. After this, each pooled output is convolved 1×1, reducing the number of channels to the original 1/N, where N is the level of the original pyramid, 1, 2, 3, and 6 respectively. Then bilinear interpolation is used to upsample the convolution low-dimensional feature graph, and the feature graph with the same size as the original feature is obtained. Finally, the features of different levels are spliced to get the final features.

The edge parts of the features passing through the pyramid module can be extracted effectively, which is conducive to the feature extraction of infrared images, so it is used in the infrared branch of the feature extraction network.

2.4. Result Decomposition Network

Since the information in the fusion image obtained by infrared visible image fusion is completely from the source image, the gap between the two images obtained by decomposition of the fusion image and the source image to some extent represents the information loss in the fusion process. Therefore, a result decomposition network is designed to decompose the fusion results to obtain infrared and visible images. The gap between the source image and the decomposed image is then input into the loss function. The loss function can reduce the gap through continuous training, and the information loss in the fusion process can be continuously reduced through the requirement of consistency between the decomposed image and the source image. The result is an image rich in information. The result decomposed network structure is shown in Figure 4.

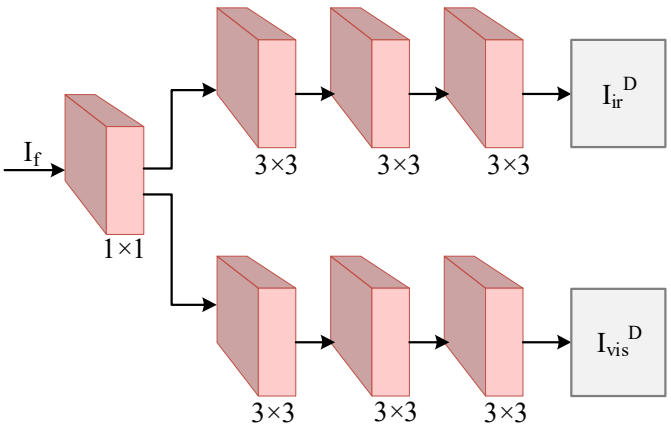


Fig. 4. Result Decomposition Network

In the result decomposition network, the input fusion image first passes through a 1×1 convolution layer, which is used to extract features from the fusion image. Then the extracted features are input into two branches consisting of three 3×3 convolution blocks, and the two branches will decompose the features to generate the decomposed infrared image and visible image. With the exception of the last convolution layer which uses the tanh activation function, all the other convolution layers use LeakyReLU as the loss function. In the initial stages of training, the output of the two branches may be exactly the same, but after that the results are compared with the source image and the comparison results are entered into the loss function. With the progress of training, the loss function becomes smaller and smaller, and the decomposition result will be closer and closer to the source image, so the information loss of the source image will be smaller and smaller in the process of image fusion.

2.5. Loss Function Design

The loss function of the fusion network in this paper uses gradient loss (L_{grad}), intensity loss (L_{int}) and decomposition loss (l_{dc}) to constrain the training process. The formula of the loss function is defined as follows:

$$L = \lambda_1 L_{grad} + \lambda_2 L_{\text{int}} + L_{dc} \tag{1}$$

where, λ_1 and λ_2 are the weights of gradient loss and strength loss.

Gradient loss can enrich the texture information contained in the fused image, and the calculation formula of gradient loss L_{grad} is

$$L_{grad} = \frac{1}{2} L_G\left(I_f, I_1\right) + \frac{1}{2} L_G\left(I_f, I_2\right) \tag{2}$$

$$L_G\left(x, y\right) = \frac{1}{HW} \left\| \nabla x - \nabla y \right\|_2^2 \tag{3}$$

where, I_f is the fusion image, I_{ir} is the infrared image, I_{vi} is the visible image, H and W represent the height and width of the image, ∇ is the Sobel gradient operator, and $\left\| \cdot \right\|_2^2$ is the L2 norm.

The intensity loss can make the fused image have a similar intensity distribution to the source image. The calculation formula of strength loss L_{int} is

$$L_{\text{int}} = \frac{1}{2} L_I\left(I_f, I_1\right) + \frac{1}{2} L_I\left(I_f, I_2\right) \tag{4}$$

$$L_I(x,y)=\frac{1}{HW}\|x-y\|_2^2 \tag{5}$$

where, H and W are the height and width of the image.

Decomposition loss can compare the decomposed image with the source image, and calculate the gap between the two. The calculation formula of decomposition loss l_{dc} is

$$l_{dc}=\frac{1}{HW}\sum_i\sum_j\left(I_{1_{de_{i,j}}}-I_{1_{i,j}}\right)^2+\left(I_{2_{de_{i,j}}}-I_{2_{i,j}}\right)^2 \tag{6}$$

where, $I_{1_{de}}$ and $I_{2_{de}}$ are the decomposition results of the fused images, and I_1 and I_2 are the source images.

Gradient loss can preserve rich texture details in the fused image. Intensity loss preserves the target distribution in the source image. Decomposition loss can minimize the information loss of source image. Good fusion results can be obtained by using these three loss functions to constrain fusion networks.

3. Experimental Results and Analysis

3.1. Experimental Parameters and Environment Configuration

1) Data set: The training data set in this chapter also uses the SSIM data set mentioned in the previous chapter, and 20 pairs of images selected from MSRS are used as the test data set.

2) Training details: The training process in this chapter is run in the Pytorch framework in Windows on the NVIDIA GeForce GTX3070 GPU. The programming language is Python3.7, the batch size is 8, the learning rate is set to 0.0002, and then it decays exponentially. A total of 100 cycles were trained, and the model parameters were updated using the Adam optimizer. The hyperparameters λ_1 and λ_2 in the loss function are set to 1 and 10, respectively.

3.2. Ablation Experiment

In order to verify the rationality of the algorithm in this chapter, three experiments are set up to compare with the algorithm in this chapter. In the feature extraction stage, Experiment I does not use pyramid blocks and only uses dense blocks. In Experiment II, dense blocks are not used in the feature extraction stage, only pyramid blocks are used. The third experiment removes the decomposition loss in the loss function and removes the influence of the decomposition network during training. The specific experimental settings are as follows as Table 1:

Table 1. Experimental Setting Table

	Dense Block	Pyramid Block	Decomposition Network
Experiment 1	-	✓	✓
Experiment 2	✓	-	✓
Experiment 3	✓	✓	-
Algorithm in This Chapter	✓	✓	✓

(a) Infrared Imagery (b) Visible Image (c) Experiment 1

(d) Experiment 2 (e) Experiment 3 (f) Algorithms in This Chapter

Fig. 5. Fusion Results of Ablation Experiment

Figure 5 shows the fusion results of Experiments I, II and III with the algorithm in this chapter on the same image. Among them, Figure 5 (a) and Figure 5 (b) are infrared and visible images. Figure 5 (c) is the result of Experiment I, which only uses dense blocks in the feature extraction stage, and dense blocks have better retention of visible light information but weak ability to extract infrared features, so the infrared features in the fused image are not obvious enough. Figure 5 (d) shows the results of Experiment II, where only the pyramid module was used to extract features. Because the pyramid block has better extraction ability for infrared features, the infrared features are more obvious in the fusion result. In experiment 3, dense blocks are used to extract visible image features and pyramid blocks are used to extract infrared image features, but no decomposition network is used, which has better retention of infrared features and visible light details. However, the results of the complete algorithm are significantly better than the experimental results.

Next, the three experiments will be compared from the objective evaluation indicators, and the comparative analysis results of the objective evaluation indicators are shown in Table 2.

Table 2. Quantitative Analysis of Ablation Experiments

	EN	*SD*	*SF*	*VIF*	*MI*	*SCD*	Q_{abf}
Experiment 1	6.535	*45.457*	*9.911*	0.757	3.898	*1.556*	0.46
Experiment 2	6.576	42.714	8.395	0.678	3.478	1.509	0.426
Experiment 3	*6.582*	42.623	9.664	*0.767*	*3.935*	1.447	*0.532*
Algorithms in This Chapter	**6.854**	**50.724**	**12.523**	**0.877**	**3.952**	**1.762**	**0.611**

Table 2 shows the average objective performance indicators of the fusion results obtained by 10 pairs of test images in the three experiments and the algorithms in this chapter. Seven evaluation indexes were used in quantitative analysis. In the table, the best values are shown in red bold and the next best values are shown in blue italics. It can be observed that the algorithm in this chapter obtains optimal values on all 7 indexes. The results obtained by the algorithm in this chapter are better than the other three experiments.

Through subjective and objective evaluation indicators, it can be seen that the complete algorithm has a better effect than the other three ablation experimental algorithms, and the proposed dense block and pyramid block and decomposition network are effective.

3.3. Contrast Test

In order to verify the superiority and feasibility of the algorithm in this chapter, the algorithm in this chapter is compared with other 6 representative algorithms from two aspects of qualitative and quantitative analysis. The comparison algorithms include: DenseFuse, IFCNN, U2Fusion, FusionGan, GANMcC and SDNet. In the image the red rectangle box is used to mark the infrared target and the green rectangle box is used to mark the detailed texture and is enlarged in the lower right corner.

As in the experiment in the previous chapter, the RGB image is converted to the YCbCr image, and the three channels Y, Cb and Cr are separated. The infrared image and visible image of the Y channel are combined with the Cb and Cr channels and finally converted to RGB images.

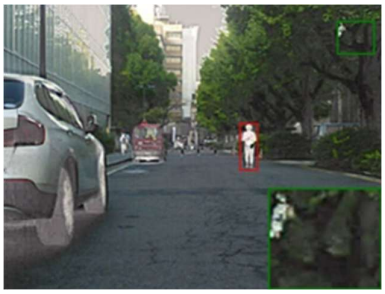
(a) Infrared Imagery

(b) Visible Image

(c) DenseFuse

(d) FusionGan

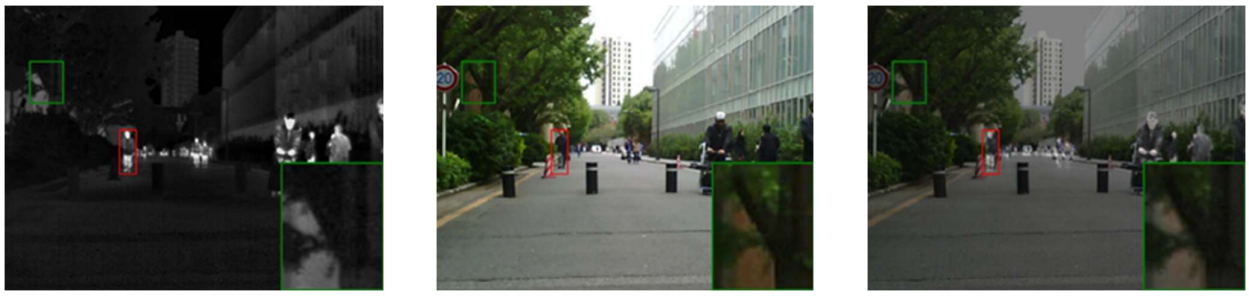
(e) GANMcC



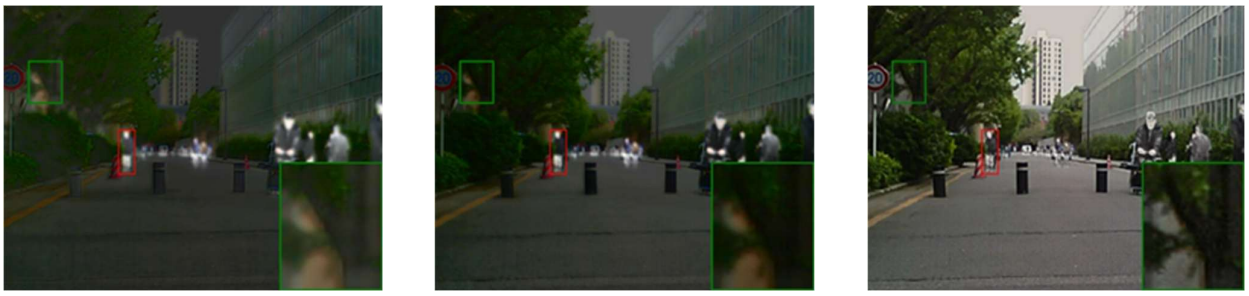
(f) IFCNN

(g) SDNet (h) U2Fusion (i) Algorithm in This Chapter

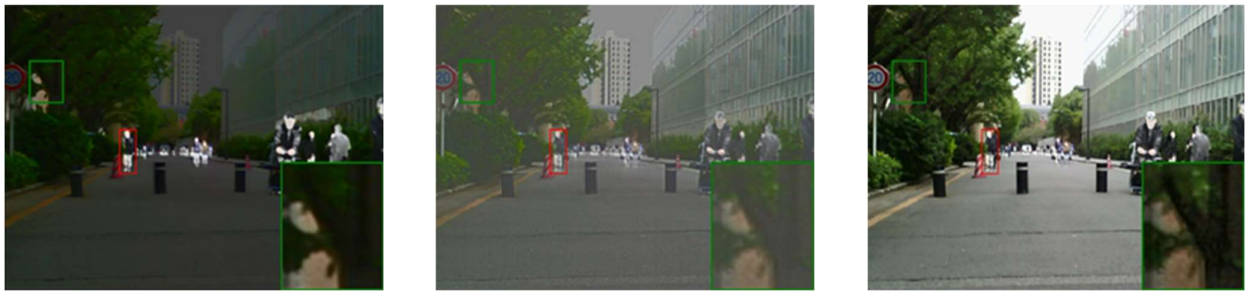
Fig. 6. Experimental Result 1 of Qualitative Analysis of the Algorithm in This Chapter



(a) Infrared Imagery (b) Visible Image (c) DenseFuse

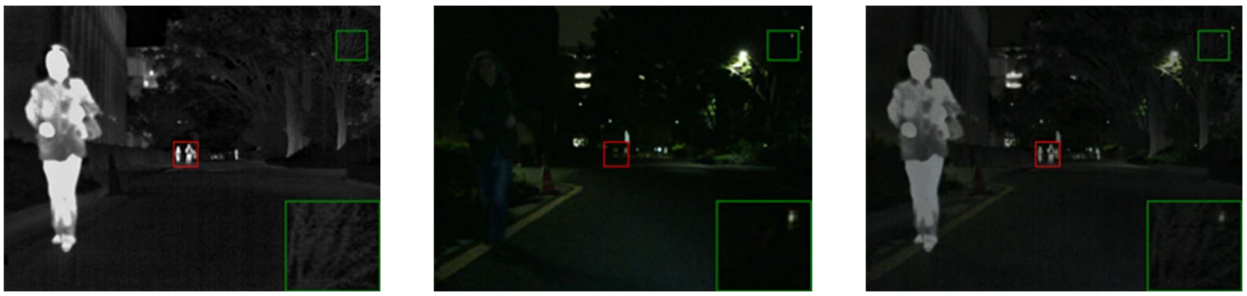


(d) FusionGan (e) GANMcC (f) IFCNN

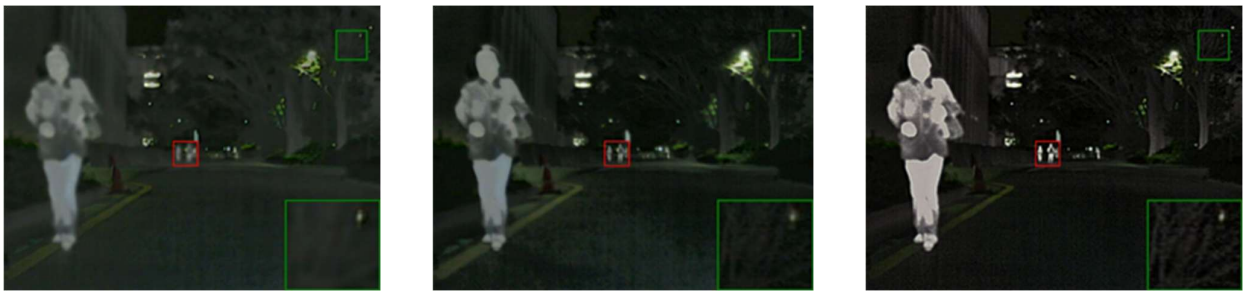


(g) SDNet (h) U2Fusion (i) Algorithm in This Chapter

Fig. 7. Experimental Result 2 of Qualitative Analysis of the Algorithm in This Chapter



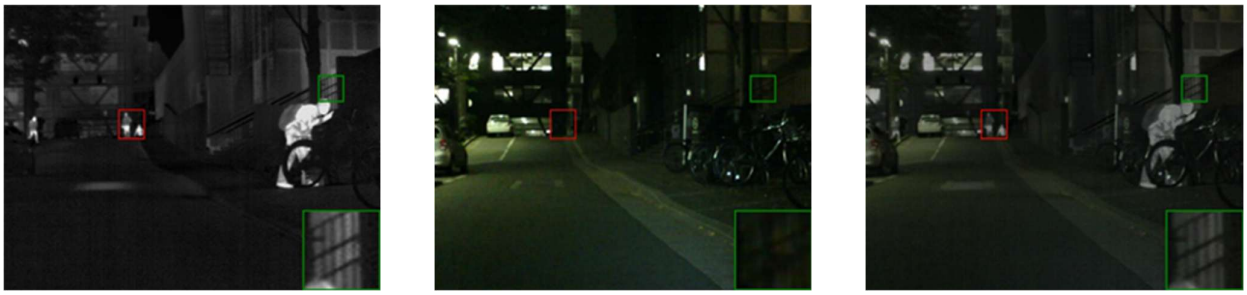
(a) Infrared Imagery (b) Visible Image (c) DenseFuse



(d) FusionGan (e) GANMcC (f) IFCNN

(g) SDNet (h) U2Fusion (i) Algorithm in This Chapter

Fig. 8. Experimental Result 3 of Qualitative Analysis of the Algorithm in This Chapter



(a) Infrared Imagery (b) Visible Image (c) DenseFuse

(d) FusionGan (e) GANMcC (f) IFCNN

(g) SDNet (h) U2Fusion (i) Algorithm in This Chapter

Fig. 9. Experimental Result 4 of Qualitative Analysis of the Algorithm in This Chapter

Figure 6, Figure 7, Figure 8 and Figure 9 show the subjective comparison results of the proposed algorithm with DenseFuse, IFCNN, U2Fusion, FusionGan, GANMcC and SDNet, which are based on deep learning. The green box marks the details and is enlarged in the lower right corner.

Figure 6 and Figure 7 are the fusion results of daytime images. It can be seen that road pedestrians are not obvious in visible images, while details and textures such as tree trunks and branches are not obvious in infrared images. The infrared information of DenseFuse, FusionGan and GANMcC is not obvious enough to highlight the outline of pedestrians on the road in the infrared image, and the texture information and overall clarity of the image are poor, and there is more noise information in the image with low brightness. The fusion result of FusionGan also has the problem of color distortion, and the trunk of the vehicle has obvious artifacts. The fusion results of IFCNN, SDNet and U2Fusion have a good retention effect on the infrared information, but the description of the information in the green box is slightly insufficient. The overall brightness of the image is low, the loss of detailed texture information is serious, and the details of the tree trunk in the shadow cannot be clearly displayed. It shows that these two algorithms have poor effect on information retention in visible image. However, the infrared information of pedestrians on the road can be prominently observed in the image obtained by the algorithm in this paper, and the texture of the details in the visible image has a better retention effect.

Figure 8 and Figure 9 are the fusion results on the night image. The silhouette of pedestrians cannot be displayed in the visible image, and the illumination and brightness information of street lights and landmarks on the road are almost completely invisible in the infrared image. Among them, the results of DenseFuse, U2Fusion and FusionGan algorithms are not obvious enough for the display of the main pedestrian infrared information in the red box, and the details in the green box are also vague. There is a lot of noise information in DenseFuse, while the infrared characteristics of GANMcC and IFCNN are relatively obvious, but the infrared information in IFCNN fusion results has obvious color distortion compared with the original infrared image. Moreover, the infrared information in the results of GANMcC is dimmer than that of IFCNN, and the texture information in the results of these two algorithms has a weak ability to retain, and neither can clearly show the details of the texture such as

trees and Windows in the dark. The fusion effect of SDNet is better than that of the previous algorithms, but the overall image is dark, and the shadow of the part illuminated by the street lamp cannot be well displayed in the fused image. Although the algorithm has a good effect on retaining the ground information of infrared and visible images, the overall image quality is not as good as the algorithm in this paper. The results of the algorithm in this paper are better for the retention of details and textures such as pedestrians in the dark, branches and Windows, which indicates that the algorithm in this paper has less information loss, and the overall effect of the fusion result map of the algorithm in this paper is better, with higher resolution and good visual effects.

In order to further illustrate the effectiveness of the algorithm, the quality of the fused images of different algorithms will be analyzed from an objective point of view. Table 3 shows the average index of 15 images fused by different algorithms, where the red bold indicates the optimal value and the blue italic indicates the suboptimal value.

Table 3. Mean Values of Performance Indicators of Different Algorithms

	EN	*SD*	*SF*	*VIF*	*MI*	*SCD*	Q_{abf}
DenseFuse	**6.854**	30.466	6.729	0.659	*2.711*	1.322	0.363
FusionGan	5.895	25.789	5.329	0.411	2.144	0.878	0.164
GANMcC	6.204	33.097	6.424	0.578	2.482	*1.626*	0.299
IFCNN	*6.669*	42.04	**12.853**	*0.744*	2.642	1.432	*0.586*
SDNet	5.74	*36.199*	7.964	0.431	2.467	1.363	0.205
U2Fusion	5.351	23.582	7.021	0.485	2.044	0.971	0.303
Algorithms in This Chapter	6.31	**50.724**	*12.523*	**0.877**	**3.952**	**1.761**	**0.611**

As can be seen from the table, the proposed algorithm achieves the optimal value in five indicators and the sub-optimal value in one indicator, indicating that the fused image of the proposed algorithm has rich information and texture. The retention effect of the main thermal radiation information of infrared images is also better, and the information loss of the algorithm in this chapter is less than that of other algorithms.

4. Conclusion

In this paper, an infrared visible road image fusion algorithm based on dual branch and result decomposition is proposed. Due to the large modal difference between infrared image and visible image, using the same structure for feature extraction of the two images may not obtain good results. Therefore, a dual-branch structure is proposed to realize the feature extraction of the two kinds of images. In the visible light branch, three dense blocks are used to extract the feature of the visible light image. The dense connection structure in the dense block can realize the interconnection between the convolutional layers and reduce the loss of detail information in the visible image during feature extraction. In the infrared branch, three pyramid blocks are used to extract infrared image features. The pyramid block can divide the input features into grids of different sizes and carry out average pooling, which is conducive to obtaining global information. This structure has a good effect on the extraction of edge features in infrared images. In addition to this, an outcome decomposition network is designed. The output fused image is input into the result decomposition network, and the decomposed infrared and visible images are obtained. Since all the information in the fused image comes from the two source images, the infrared and visible images obtained by the decomposition of the fused image are compared with the source image, and the comparison results are input into the loss function, so that the network can continuously reduce the gap between the decomposed image and the source image in the training process, and the information loss in the fusion process will also be continuously reduced. Experimental results show that compared with the other six comparison algorithms, the proposed algorithm has better results in subjective and objective evaluation, and the

fused image contains more information.

References

[1] Ma, W., Wang, K., Li, J., Yang, S. X., Li, J., Song, L., & Li, Q. (2023). Infrared and visible image fusion technology and application: A review. Sensors, 23(2), 599.

[2] Jin, X., Jiang, Q., Yao, S., Zhou, D., Nie, R., Hai, J., & He, K. (2017). A survey of infrared and visual image fusion methods. Infrared Physics & Technology, 85, 478-501.

[3] Zhang, Y., Zhang, L., Bai, X., & Zhang, L. (2017). Infrared and visual image fusion through infrared feature extraction and visual information preservation. Infrared Physics & Technology, 83, 227-237.

[4] Zhang, X., Liu, G., Huang, L., Ren, Q., & Bavirisetti, D. P. (2023). IVOMFuse: An image fusion method based on infrared-to-visible object mapping. Digital Signal Processing, 137, 104032.

[5] Mo, Y., Kang, X., Duan, P., Sun, B., & Li, S. (2021). Attribute filter based infrared and visible image fusion. Information Fusion, 75, 41-54.

[6] Chen, L., Yang, X., Lu, L., Liu, K., Jeon, G., & Wu, W. (2019). An image fusion algorithm of infrared and visible imaging sensors for cyber-physical systems. Journal of Intelligent & Fuzzy Systems, 36(5), 4277-4291.

[7] Yang, K., Xiang, W., Chen, Z., Zhang, J., & Liu, Y. (2024). A review on infrared and visible image fusion algorithms based on neural networks. Journal of Visual Communication and Image Representation, 104179.

[8] Sun, C., Zhang, C., & Xiong, N. (2020). Infrared and visible image fusion techniques based on deep learning: A review. Electronics, 9(12), 2162.

[9] Zhao, Z., Xu, S., Zhang, J., Liang, C., Zhang, C., & Liu, J. (2021). Efficient and model-based infrared and visible image fusion via algorithm unrolling. IEEE Transactions on Circuits and Systems for Video Technology, 32(3), 1186-1196.

[10] Zhang, Y., Li, D., & Zhu, W. (2020). Infrared and visible image fusion with hybrid image filtering. Mathematical Problems in Engineering, 2020(1), 1757214.

[11] Chen, J., Li, X., Luo, L., Mei, X., & Ma, J. (2020). Infrared and visible image fusion based on target-enhanced multiscale transform decomposition. Information Sciences, 508, 64-78.

[12] An, W. B., & Wang, H. M. (2020). Infrared and visible image fusion with supervised convolutional neural network. Optik, 219, 165120.

[13] Ma, J., Zhou, Z., Wang, B., & Zong, H. (2017). Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Physics & Technology, 82, 8-17.

[14] Hao, S., He, T., An, B., Ma, X., Wen, H., & Wang, F. (2022). VDFEFuse: A novel fusion approach to infrared and visible images. Infrared Physics & Technology, 121, 104048.

[15] Rao, D., Xu, T., & Wu, X. J. (2023). TGFuse: An infrared and visible image fusion approach based on transformer and generative adversarial network. IEEE Transactions on Image Processing.

[16] Li, Z., Hu, H. M., Zhang, W., Pu, S., & Li, B. (2020). Spectrum characteristics preserved visible and near-infrared image fusion algorithm. IEEE Transactions on Multimedia, 23, 306-319.

[17] Zhao, L., Zhang, Y., Dong, L., & Zheng, F. (2022). Infrared and visible image fusion algorithm based on spatial domain and image features. Plos one, 17(12), e0278055.

[18] Ma, J., & Zhou, Y. (2020). Infrared and visible image fusion via gradientlet filter. Computer Vision and Image Understanding, 197, 103016.

[19] Li, H., & Wu, X. J. (2018). DenseFuse: A fusion approach to infrared and visible images. IEEE Transactions on Image Processing, 28(5), 2614-2623.

[20] Junwu, L., Li, B., & Jiang, Y. (2020). An infrared and visible image fusion algorithm based on LSWT-NSST. IEEE Access, 8, 179857-179880.

[21] Tang, L., Yuan, J., & Ma, J. (2022). Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. Information Fusion, 82, 28-42.

[22] Zhang, L., Li, H., Zhu, R., & Du, P. (2022). An infrared and visible image fusion algorithm based on ResNet-152. Multimedia Tools and Applications, 81(7), 9277-9287.