Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6220166, 9 pages
https://doi.org/10.1155/2021/6220166

Research Article

Trustworthy Image Fusion with Deep Learning for Wireless Applications

Chao Zhang,[1] Haojin Hu,[1] Yonghang Tai,[1] Lijun Yun,[2] and Jun Zhang[1]

[1]*School of Physics and Electronic Information, Yunnan Normal University, Kunming, China*
[2]*School of Information Science and Technology, Yunnan Normal University, Kunming, China*

Correspondence should be addressed to Lijun Yun; yunlj@163.com and Jun Zhang; junzhang@ynnu.edu.cn

Received 21 April 2021; Revised 28 May 2021; Accepted 9 June 2021; Published 1 July 2021

Academic Editor: Weizhi Meng

To fuse infrared and visible images in wireless applications, the extraction and transmission of characteristic information security is an important task. The fused image quality depends on the effectiveness of feature extraction and the transmission of image pair characteristics. However, most fusion approaches based on deep learning do not make effective use of the features for image fusion, which results in missing semantic content in the fused image. In this paper, a novel trustworthy image fusion method is proposed to address these issues, which applies convolutional neural networks for feature extraction and blockchain technology to protect sensitive information. The new method can effectively reduce the loss of feature information by making the output of the feature extraction network in each convolutional layer to be fed to the next layer along with the production of the previous layer, and in order to ensure the similarity between the fused image and the original image, the original input image feature map is used as the input of the reconstruction network in the image reconstruction network. Compared to other methods, the experimental results show that our proposed method can achieve better quality and satisfy human perception.

1. Introduction

It is a big research challenge to fuse infrared and visible images to provide high-quality images for wireless applications, such as target recognition, visual enhancement, and cyber surveillance. Infrared images are mapped by infrared sensors capturing thermal radiation as a grayscale image and can emphasise thermal targets in low-light situations, but infrared images have a low resolution and do not show more detail in the scene. In contrast, the visible light sensor collects visible images to represent rich texture details, usually with higher resolution. Still, it is easily affected by imaging conditions (such as weather conditions, and lighting) [1]. The thermal radiation information of the infrared image and the texture information of the visible image can be fused to obtain an image with better visual quality and more information, which is the primary purpose of the fusion of infrared and visible images. Device can analyse the image which are been fused with computer vision and processing.

In the last few decades, many algorithms have been designed to implement the fusion of infrared and visible images, which get good fusion result. Fusion algorithms for infrared and visual images can be divided into general methods and deep learning-based methods. Various image processing techniques are used for feature extraction in available image fusion methods [2]. Different fusion rules are designed for multimodal images, making the design complex and the generalization of the fusion poor. Along with the continuous development of deep learning, numerous scholars have developed image fusion models based on deep learning models [3]. Liu et al. first proposed a convolutional neural network- (CNN-) based fusion algorithm for infrared and visible images [4], which provides better fusion results than traditional methods. Liu et al. [5] used CNN as a feature extraction model to achieve the fusion of multifocused images by rule-based fusion. Li and Wu [6] proposed an auto-encoder-based method for fusing infrared and visible images, which can use feature maps to obtain fused images eventually. The deep learning-based fusion method of infrared and visible images has the following drawbacks: (1) the method based on deep learning still cannot get rid of manual rule design, and the deep learning frames just as part of

the fusion architecture; (2) the fusion strategy cannot achieve the fusion of infrared images in the item. The information is balanced with the visible image, and the fusion image is only similar to the source image; (3) the extracted compelling features were largely lost in the transmission process, and the feature information used for the fusion image is reconstructed with only a small amount of feature information.

We proposed a framework for fusing infrared images with visible images based on a deep learning model to solve the above issues. Our model is composed of three parts: a feature extraction network, a fusion network, and a reconstruction network. To ensure effective extraction of feature information, the output of features extracted by the feature extraction network in each convolutional layer will be fed to the next layer together with the output of the previous layer; short direct connections are built between each layer and all layers in a feed-forward fashion, thus effectively reducing the loss of valid information. In the feature fusion process, we use point-to-point approach to merge the feature maps of different channels to obtain the fused feature maps. In reconstruction network, the fused feature maps are the input, and the source image pair also used for reconstruction of fusion image. Considering trustworthy is a critical issue in the real-world applications of image fusion [7–10], we also propose to apply blockchain technology to protect sensitive information.

2. Related Work

In this section, we briefly describe the infrared and visible image fusion methods based on general and deep learning that have been developed in recent years, in particular for wireless applications. Initially, signal processing algorithms were widely used in image fusion [11], using mean and median filtering to extract the fundamental and detail layers of features before using dominant features to obtain a weight map and then combining these three components to obtain a fused image. The existing traditional methods of image fusion mainly consist of multiscale transform-based methods and sparse representation-based fusion methods. The original input image is decomposed into scale components of different scales in a multiscale transform-based approach [12], and each scale component is then fused according to specific rules, and finally, the combined image is obtained by the corresponding inverse scale transform. The main multiscale transforms are the pyramid transform [13], the wavelet transform [14], and the nondown sampled contour wavelet transform [15]. Sparse representation-based fusion methods learn dictionaries from high-quality images and then use the learned dictionaries to sparse representations of the source images. The method first decomposes the source images into overlapping blocks by a sliding window strategy and learns dictionaries from high-quality images, using the dictionaries to encode each image path sparsely. The sparse representation coefficients are then fused according to the fusion rules, and finally, the fusion coefficients of the fused images are reconstructed using the dictionaries, such as the joint sparse representation [16], the directional gradient histogram-based fusion method [17], and the cosparse representation [18]. Traditional methods require the manual design of feature extraction rules, feature fusion rules, and image reconstruction rules, resulting in computationally intensive and challenging designs.

Deep learning in the field of digital image processing has shown advanced performance in recent years; for the complex relationship between data, it can model the context knowledge and automatically extraction the perform feature without human intervention. Liu et al. [4] designed a sparse convolutional representation- (CSR-) based image fusion method to overcome the cumbersome rules in manual design. In 2017, Liu et al. [19] proposed the fusion of medical images using convolutional neural networks, which uses convolutional neural networks to generate pixel weight maps, but the method did not achieve total neural network fusion but rather multiscale transform fusion using image pyramids. Masi et al. [20] propose a fusion method, which are entirely based on deep learning; the method based on deep learning can extract the feature from image and reconstruct the fused image. In ICCV2017, the unsupervised learning framework was used for multiexposure image fusion by Prabhakar et al. [21], which has an extraordinary fusion loss function. Li and Wu [6] add dense block fast to this structure and design a separate fusion strategy in the fusion layer. Xu et al. [22] proposed an unsupervised and unified densely connected network for different types of image fusion tasks. Mustafa et al. [23] use multilevel dense network multifocus image fusion. Ma et al. proposed FusionGan [24], which uses adversarial networks for image fusion, using discriminators to distinguish differences in the fused image from the original image. A dual-discriminator conditional generative adversarial network called DDcGAN [25] proposed by Ma et al. used to fusion multimodality medical images of different resolutions.

3. Materials and Methods

We proposed a deep neural network for infrared image, and visible image fusion is described in detail in this section. With the consideration of zero trust security model, blockchain is used to protect feature information. A private blockchain is implemented to store, share, and transmit feature data. The network consists of three main parts: a feature extraction network, a feature map fusion network, and a reconstruction network; above description is shown in Figure 1.

3.1. Feature Extraction Network. Extracting useful feature information from images of different modalities is a critical process in image fusion, and a good feature extraction strategy can reduce redundant feature information and provide more complex scene clues for subsequent processing. Therefore, the way of the feature extraction network is designed directly determines the effectiveness of the fusion. The feature extraction network proposed in this paper consists of 5 convolutional layers; each convolutional layers can obtain 48 feature maps by 3×3 filters. The first convolutional layer will extract the details and global information of the source image, and the subsequent convolutional layers are used for

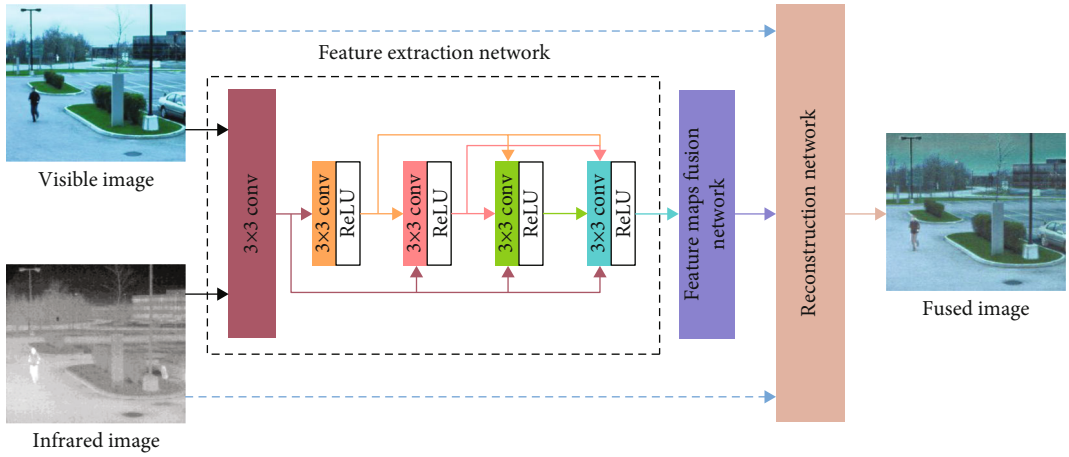

Figure 1: The overall architecture of the proposed network.

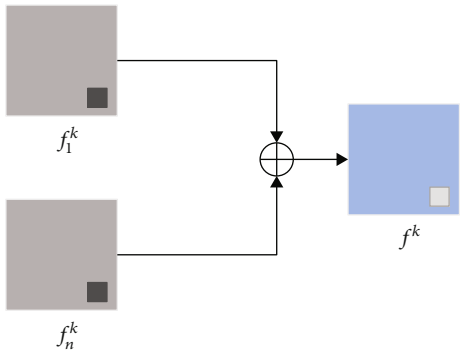

Figure 2: The architecture of combine with feature maps.

abstract feature generation. During the convolution process, the sequential sampling of the image makes the feature map gradually shrink, and a large number of valid information are lost. It cannot be repaired during the sampling process, this resulting in the disappearance of a large number of original features in the fused image. Therefore, we do not use pooling operations between the individual convolutional layers, but instead use the output of each layer, along with the output of the previous layer, as input to the next layer, model allowing valid information features to be passed throughout the convolutional network. Following Li and Wu [6], when the image of input is three-channel (RGB), each pair of the channel will be the input of the feature extraction network. To speed up convergence and avoid gradient sparsity, we use the ReLU activation function after each convolutional layer of the encoder.

3.2. Features Fusion. In DeepFuse [21], CNN is used to implement the fusion of exposure image pair; the feature maps obtained by the CNN are subjected to a point-to-point summation operation to get the final fused feature map; the same strategy was used in DenseFuse [6] by Li and Wu. Achieving accurate fusion is a difficult task, because the infrared and visible images are both come from different sensors. In this paper, following DeepFuse and DenseFuse, we implement the pixel-level's point-to-point merging of the feature maps from the feature extraction network by using an addition strategy, which is shown in Figure 2.

The input image is extracted by the feature network to form a feature map; $f_n^k(x, y)$ is the set composed of all feature maps, and $f^k(x, y)$ represents the merged feature map. The merging strategy is shown in Equation (1). (x, y) is the corresponding position coordinates of the feature map and the fused feature map. The merged feature map will be used as the input to the reconstruction network reconstruct the fused image.

$$f^k(x, y) = \sum_{1}^{n} f_n^k(x, y). \tag{1}$$

3.3. Reconstruction Network. Image reconstruction is also an essential task for networks, and deconvolution is used typically to reconstruct images. In our network, we replace the deconvolution layers of the reconstructed network with regular convolution. The reconstructed network consists of four Conv layers, using a ReLU layer of 3×3 kernel size. To feed the reconstruction network with more information, we use the input image as input to the reconstruction network, and the feature map and the original image were both used to reconstruct the fused image. The architecture of reconstruction network is shown in Figure 3. When the feature maps are calculated by feature extraction network and feature fusion layer, the source image pair and the fused maps are used for image reconstruction, the following equation defines this task:

$$\text{Fused}(x, y) = \sum_{i=1}^{n} f^i(x, y)\text{Source}_i(x, y), \tag{2}$$

where the $\text{Fused}(x, y)$ is the fused image, $f^i(x, y)$ represented the fused feature maps from feature maps fusion network, $\text{Source}_i(x, y)$ is the source image pair, and (x, y) represents the corresponding pixel point.

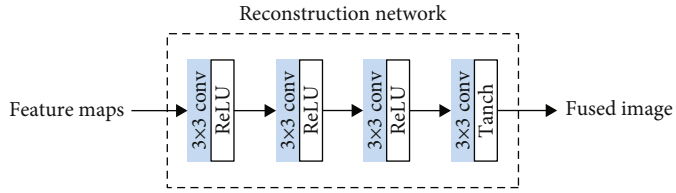

Figure 3: The architecture of combine with feature maps.

Figure 4: Source images of the VIFB dataset tested in our experiments, with the first row containing visible images and the second row containing infrared images.

4. Experimental Results and Discussion

In this section, first, we will describe the source images and the experimental environment. Secondly, we evaluate our fused images using subjective vision. Finally, the proposed algorithm is quantitatively assessed by using a variety of metrics. In order to validate the effectiveness of the deep learning model, we divided the comparison algorithms into general and deep learning-based methods in our experiments.

4.1. Experimental Settings. We train our proposed method with 5000 input images that we choose from MS-COCO dataset [26]; the learning rate is set to 10^{-4}; the batch size is set 24. Because there are no fully connected layers in our method, any same-scale infrared and visible image pairs can be fused using our model. Our experiment compares our model with even state-of-art image fusion methods in VIFB [27] with the particular consideration of wireless applications including object recognition and cyber resilience. VIFB is a visible and infrared image fusion benchmark, which consist 21 image pairs, and the size of each image pair is different. Four examples of VIFB are shown in Figure 4. The fusion methods used in this paper fall into two categories: general methods and methods based on deep learning. General methods include ADF [28], guided filter algorithm (GFF) [29], cross bilateral filter (CBF) [30], and VSMWLS [31]; methods based on deep learning include DenseFuse [6], CNN [5], ResNet [32], and our method. DenseFuse, CNN, ResNet, and our model are implemented with Pytorch and trained with double Tesla V100, 16GB RAM GPUs. Other methods are implemented with MATLAB 2016B.

4.2. Subjective Visual Evaluation. In this section, subjective visual evaluations are used to assess the performance of various infrared and visible image fusion algorithms, which is based on the way of the human visual system. In order to validate the effectiveness of deep learning models, we classify the current fusion methods into categories: general and deep learning. We chose four images for the night environment and the daytime environment. In the daytime environment, the first image is darker in the evening, and the second chapter is better lit; in the dark night environment, the light is weaker in the first image than in the second image. All four images we selected contained thermal targets for verifying the algorithm's performance in highlighting thermal targets. The fusion results obtained by the different fusion algorithms are presented in Figure 5.

As shown in Figure 5, we use the red dashed line to divide the images into three groups. The first group shows the original input visible image with infrared image, the second group shows the image fusion results using the general methods, from top to bottom, ADF, GTF, CBF, and VSMWLS, respectively, and the third group offers the fused images based on deep learning methods, from top to bottom,

DenseFuse, CNN, ResNet, and our proposed methods. Among the general techniques, ADF and VSMMLS work better; the fused images obtained by GTF produce more significant artefacts, and the fused images contain more information about the infrared than the visible images; the CBF method achieves fusion, but a large number of blurred areas appear in the fused images. Deep learning-based methods achieve good image fusion with minimal visual discrepancies; CNN methods show coloured streaks when fusing images in daylight. DenseFuse and ResNet achieve better fusion results, and these methods achieve fused image images that contain more information about the original. Our fused images have three main advantages over other methods. Firstly, our results for hot tar (e.g., human portraits) have high contrast. Secondly, the images we obtain contain rich textural detail and more detailed information in the background. Thirdly, our method produces images that better balance the modalities of infrared and visible images and have a better visual perception, resulting in a more natural fusion.

4.3. Quantitative Evaluation. This section compares our approach with general methods and the approach base on the deep learning carried out in VIFB 21 for the quantitative analysis of images. We use ten metrics such as average gradient (AG) [33], correlation coefficient (CC), peak signal-to-noise ratio (PSNR) [34], information entropy (EN) [35], structural similarity of images (SSIM) [36], mutual information (MI) [37], image similarity metric based on edge information (Qabf) [38], pixel feature mutual information (FMI_pixel) [39], discrete cosine characteristic mutual information(FMI_dct) [39], and wavelet features mutual information(FMI_w) [40] for evaluation.

(i) *Average Gradient (AG).* This evaluation indicator reflects the sharpness of the image. The average gradient is calculated only necessary to consider the fused image, an evaluation metric that reflects the sharpness of the image and is defined by the following equation:

$$\mathrm{AG}(F) = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\frac{(I(i+1,j) - I(i,j))^2 + (I(i,j+1) - I(i,j))^2}{2}}, \tag{3}$$

where M and N are the fused image's width and height and $I(x, y)$ is the pixel value of the image at that spot.

(ii) *Correlation Coefficient (CC).* Correlation coefficient reflects the degree of correlation among the IR image and the visible image as well as the fused image. We calculated the correlation coefficients $CC(I, F)$ and $CC(V, F)$ for the infrared and visible images and the fused image, respectively, and finally obtained the overall correlation coefficient, which is defined by the following equation:

$$\mathrm{CC}(I, V, F) = \frac{1}{2}\left(\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{i,j} - \bar{I}\right)\left(F_{i,j} - \bar{F}\right)}{\sqrt{\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{i,j} - \bar{I}\right)^2\left(F_{i,j} - \bar{F}\right)^2\right)}} + \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(V_{i,j} - \bar{V}\right)\left(F_{i,j} - \bar{F}\right)}{\sqrt{\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(V_{i,j} - \bar{V}\right)^2\left(F_{i,j} - \bar{F}\right)^2\right)}}\right), \tag{4}$$

where I and V represent the infrared image and the visible image; F represents the fused image; $I(i, j)$, $V(i, j)$, and $F(i, j)$ are the pixels corresponding to the pixel value of the pixel point; and $\bar{I}$, $\bar{V}$, and $\bar{F}$ are the mean values.

(iii) *Peak Signal-to-Noise Ratio (PSNR).* This assessment measures whether the image is distorted or not. Its value is the ratio of valid information to noisy information in the image. Its formula is as follows:

$$\mathrm{PSNR}(I, V, F) = \frac{1}{2}\left(10\log_{10}\left(\frac{L^2}{\mathrm{MSE}(I, F)}\right) + 10\log_{10}\left(\frac{L^2}{\mathrm{MSE}(V, F)}\right)\right). \tag{5}$$

MSE represents the mean squared error, $\mathrm{MSE}(x, y) = (1/mn)\sum_{i=0}^{m}\sum_{i=0}^{n}\|x(i, j) - y(i, j)\|^2$, and $x(x, j)$ and $y(i, j)$ are the pixels at the corresponding positions. When the peak signal-to-noise ratio is higher, the difference between the fused image and the original image is more minor.

(iv) The information entropy (EN) can represent the average amount of information in an image, a metric does not need to take into account the input

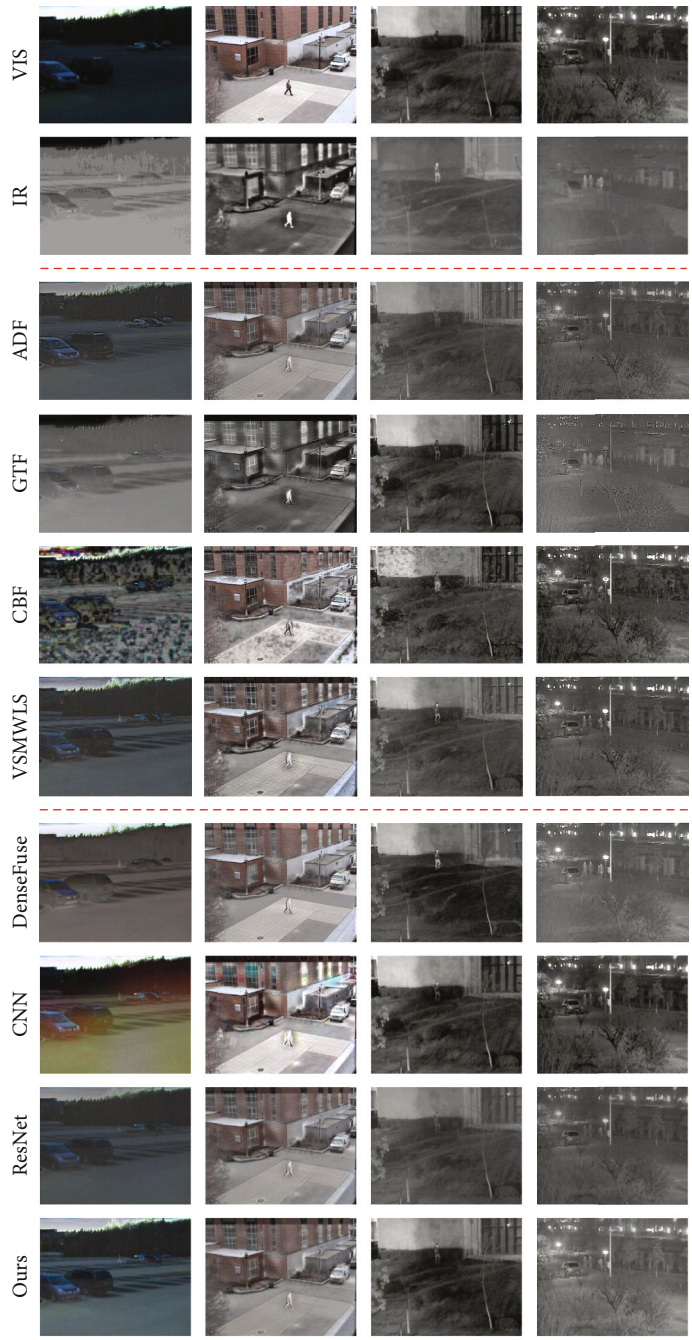

Figure 5: Visual fusion results of images from different scenes in the VIFB dataset. From top to bottom: infrared images, visible images, ADF, GTF, CBF, VSMWLS, DenseFuse, CNN, fusion results from ResNet, and our proposed method; the red dashed line divides the images into three parts: original images, fused images from the general method, and fused images based on the deep learning method.

image, is determined only from the fused image and is defined by the following equation:

$$\mathrm{EN}(x) = \sum_{i=0}^{l-1} p(x_i) \log_b p(x_i), \tag{6}$$

where $p(x_i)$ is the percentage of pixels within the grayscale image x with grayscale i and l is taken to be 256 and is the grayscale level; this equation is a 256 element entropy function; each element can be obtained with equal probability of occurrence as the maximum value; when the value of EN is larger, it means that there is more information in the image.

(v) *Structural Similarity of Images (SSIM)*. The structural similarity of an image can be measured in terms of luminance, contrast, and structure, where the mean, standard deviation, and covariance are

used as estimates of the illumination, contrast, and structural similarity phases, which given by the following formula:

$$\mathrm{SSIM}(I,V,F)=\frac{1}{2}\left(\frac{(2\mu_I\mu_F+c_1)(2\sigma_{\mathrm{IF}}+c_2)}{(\mu_I^2+\mu_F^2+c_1)(\sigma_I^2+\sigma_F^2+c_2)}+\frac{(2\mu_V\mu_F+c_1)(2\sigma_{VF}+c_2)}{(\mu_V^2+\mu_F^2+c_1)(\sigma_V^2+\sigma_F^2+c_2)}\right), \tag{7}$$

where μ_I, μ_V, and μ_F are the image mean; σ_I, σ_V, and σ_F are the standard deviation; σ_{IF} and σ_{VF} are the covariance; and $c_1=(k_1L^2)$ and $c_2=(k_2L^2)$,where k_1=0.01, k_2=0.03, and $L=255$.

(vi) *Mutual Information (MI)*. Mutual Information measures the dependence between two domain variables. It measures the similarity between the fused image and the source image based on the amount of information retained by the combined image in the source image and is calculated as follows:

$$\mathrm{MI}(I,V,F)=\sum_{i,j}p_{\mathrm{IF}}(i,j)\log_2\frac{p_{\mathrm{IF}}(i,j)}{p_I(i)p_F(j)}+\sum_{i,j}p_{VF}(i,j)\log_2\frac{p_{VF}(i,j)}{p_V(i)p_F(j)}. \tag{8}$$

(vii) *Image Similarity Metric Based on Edge Information (Q_abf)*. Xydeas et al. [34] argue that image quality is closely related to the integrity and sharpness of the edges and that the similarity between the fused image and the source image is measured from the edge perspective

(viii) *Feature Mutual Information (FMI)*. FMI measures the quality of an image by calculating the mutual information of image features, and a higher value of FMI indicates better fusion quality:

$$\mathrm{FMI}(I,V,F)=\frac{1}{2}(T(I;F)+T(V;F)), \tag{9}$$

where $T(I;F)=(2/n)\sum_{i=1}^{n}(T_i(I;F)/(H_i(I)+H_i(F)))$ and $T(V;F)=(2/n)\sum_{i=1}^{n}(T_i(V;F)/(H_i(V)+H_i(F)))$, where $H_i(I)$, $H_i(V)$, and $H_i(F)$ are the entropy of the corresponding windows from the three images; $n=M\times N$, n is the size of the image, and a more significant value of FMI indicates better image fusion performance. In the paper, we will calculate the pixel feature mutual information (FMI_pix) and discrete cosine feature mutual information (FMI_dct) and wavelet feature mutual information (FMI_w) to evaluate our fusion performance.

The results of our quantitative analysis are shown in Figure 6, where the values are the average values of the different evaluation metrics for the 21 pairs of images by the other algorithms. Overall, image fusion methods based on general methods achieved the best values on 3 metrics and fusion methods based on deep learning achieved the best values on 7 metrics. In the general method, ADF, CBF, and GTF

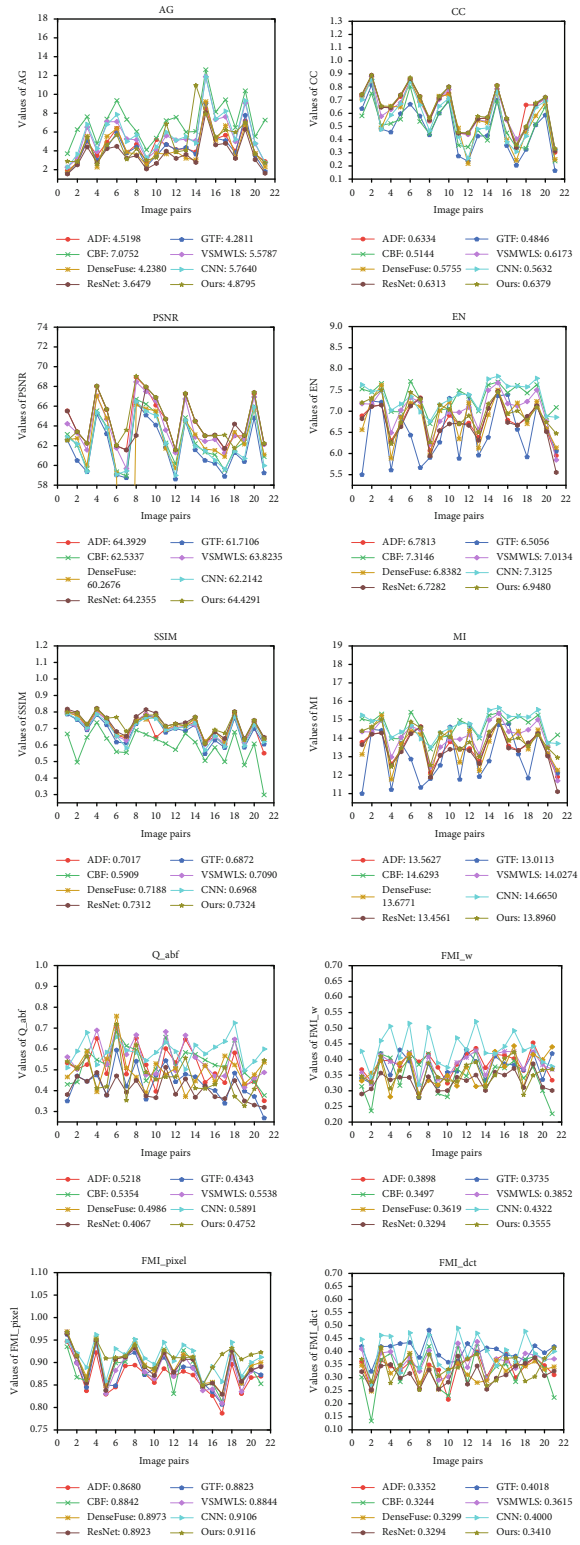

Figure 6: Comparative results of quantitative analysis of different fusion algorithms in ten fusion metrics.

Table 1: Average runtime comparison of different methods on 21 testing image pairs.

Method	ADF	GTF	CBF	VSMWLS
Running time	1.32	0.37	20.58	3.42
Method	DenseFuse	CNN	ResNet	Ours
Running time	9.85	33.25	4.53	0.95

achieved the best values for AG, EN, and FMI_dct, respectively. In the deep learning-based approach, CNN obtained the best values for MI, Qabf, and FMI_w. The fused images generated by our method achieved the best values on four metrics: CC, SSIM, PSNR, and FMI_pixel.

The average runtime of the 8 methods on the 21 testing image pairs is also reported in Table 1. It can be seen that the running times of the image fusion methods vary considerably. In our comparison method, our method is the fastest deep learning-based method; although the GPU is used to perform the computation, it still took an average of 0.95 seconds to fuse an image pair.

5. Conclusions

This proposes a novel and effective deep learning structure for wireless applications to implement the fusion of infrared and visible images. Our fusion structure consists of three main components: a feature extraction network, a feature map fusion network, and a reconstruction network. The feature output extracted by the feature extraction network of each convolutional layer will be fed to the next layer together with the previous layer output, and the original image is also involved in the reconstruction of the image, thus effectively reducing the loss of feature information. The images we obtain contain rich texture details and more background detail information, which can better balance the modality of infrared and visible images, have a better visual experience, and achieve a more natural fusion.

Data Availability

We use the open dataset MS-COCO that is publicly available on https://cocodataset.org/#home.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: a survey," *Information Fusion*, vol. 45, pp. 153–178, 2019.

[2] Y. Miao, C. Chen, L. Pan, Q. L. Han, J. Zhang, and Y. Xiang, "Machine learning based cyber attacks targeting on controlled information: a survey," *ACM Computing Survey*, 2021.

[3] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.

[4] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Information Fusion*, vol. 36, pp. 191–207, 2017.

[5] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 16, no. 3, pp. 1850018–1850018:20, 2018.

[6] H. Li and X.-J. Wu, "Densefuse: a fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2018.

[7] J. Qiu, J. Zhang, W. Luo, L. Pan, S. Nepal, and Y. Xiang, "A survey of android malware detection with deep neural models," *ACM Computing Survey*, vol. 53, no. 6, pp. 1–36, 2021.

[8] M. Wang, T. Zhu, T. Zhang, J. Zhang, S. Yu, and W. Zhou, "Security and privacy in 6G networks: new areas and new challenges," *Digital Communications and Networks*, vol. 6, no. 3, pp. 281–291, 2020.

[9] G. Lin, S. Wen, Q.-L. Han, J. Zhang, and Y. Xiang, "Software vulnerability detection using deep neural networks: a survey," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1825–1848, 2020.

[10] R. Coulter, Q. Han, L. Pan, J. Zhang, and Y. Xiang, "Data-driven cyber security in perspective—intelligent traffic analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3081–3093, 2020.

[11] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infrared Physics & Technology*, vol. 76, pp. 52–64, 2016.

[12] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147–164, 2015.

[13] G. Liu, Z. Jing, S. Sun, J. Li, Z. Li, and H. Leung, "Image fusion based on expectation maximization algorithm and steerable pyramid," *Chinese Optics Letters*, vol. 2, no. 7, pp. 386–389, 2004.

[14] Y. Zou, X. Liang, and T. Wang, "Visible and infrared image fusion using the lifting wavelet," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, no. 11, pp. 6290–6295, 2013.

[15] F. Meng, M. Song, B. Guo, R. Shi, and D. Shan, "Image fusion based on object region detection and non-subsampled contourlet transform," *Computers & Electrical Engineering*, vol. 62, pp. 375–383, 2017.

[16] J.-j. Zong and T.-s. Qiu, "Medical image fusion based on sparse representation of classified image patches," *Biomedical Signal Processing and Control*, vol. 34, pp. 195–205, 2017.

[17] R. Gao, S. A. Vorobyov, and H. Zhao, "Image fusion with cosparse analysis operator," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 943–947, 2017.

[18] Q. Zhang, Y. Fu, H. Li, and J. Zou, "Dictionary learning method for joint sparse representation-based image fusion," *Optical Engineering*, vol. 52, no. 5, article 057006, 2013.

[19] N. Sun, J. Zhang, P. Rimba, S. Gao, L. Y. Zhang, and Y. Xiang, "Data-driven cybersecurity incident prediction: a survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 2, pp. 1744–1772, 2019.

[20] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–7, Xi'an, China, July 2017.

[21] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sensing*, vol. 8, no. 7, p. 594, 2016.

[22] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4724–4732, Venice, Italy, 2017.

[23] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A Unified Densely Connected Network for Image Fusion," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12484–12491, 2020.

[24] H. T. Mustafa, M. Zareapoor, and J. Yang, "MLDNet: multi-level dense network for multi-focus image fusion," *Signal Processing: Image Communication*, vol. 85, pp. 110923–115965, 2020.

[25] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: a generative adversarial network for infrared and visible image fusion," *Information Fusion*, vol. 48, pp. 11–26, 2019.

[26] 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," *IEEE Transactions on Image Processing*, vol. 29, pp. 4980–4995, 2020.

[27] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft coco: common objects in context," in *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., pp. 740–755, Springer, Cham, 2014.

[28] X. Zhang, P. Ye, and G. Xiao, "VIFB: a visible and infrared image fusion benchmark," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 104-105, Seattle, WA, USA, June 2020.

[29] D. P. Bavirisetti and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 203–209, 2015.

[30] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Information Fusion*, vol. 31, pp. 100–109, 2016.

[31] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1193–1204, 2015.

[32] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017.

[33] H. Li, X.-j. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infrared Physics & Technology*, vol. 102, article 103039, 2019.

[34] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Optics Communications*, vol. 341, pp. 199–209, 2015.

[35] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, pp. 133–142, 2015.

[36] J. W. Roberts, J. A. Van Aardt, and F. B. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *Journal of Applied Remote Sensing*, vol. 2, no. 1, article 023522, 2008.

[37] L. Liu, O. de Vel, Q. Han, J. Zhang, and Y. Xiang, "Detecting and preventing cyber insider threats: a survey," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 2, pp. 1397–1417, 2018.

[38] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38, no. 7, pp. 313–315, 2002.

[39] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308-309, 2000.

[40] M. Haghighat and M. A. Razian, "Fast-FMI: non-reference image fusion metric," in *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1–3, Astana, Kazakhstan, October 2014.