Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(9): 16060–16082.
DOI: 10.3934/mbe.2023717
Received: 22 May 2023
Revised: 11 July 2023
Accepted: 20 July 2023
Published: 08 August 2023

Research article

FECFusion: Infrared and visible image fusion network based on fast edge convolution

Zhaoyu Chen[1], Hongbo Fan[2,*], Meiyan Ma[1] and Dangguo Shao[1,3]

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

[2] Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China

[3] Yunnan Province Key Laboratory of Computer, Kunming University of Science and Technology, Kunming 650500, China

* **Correspondence:** Email: 20212204010@stu.kust.edu.cn.

Abstract: The purpose of infrared and visible image fusion is to integrate the complementary information from heterogeneous images in order to enhance their detailed scene information. However, existing deep learning fusion methods suffer from an imbalance between fusion performance and computational resource consumption. Additionally, fusion layers or fusion rules fail to effectively combine heteromodal feature information. To address these challenges, this paper presents a novel algorithm called infrared and visible image fusion network base on fast edge convolution (FECFusion). During the training phase, the proposed algorithm enhances the extraction of texture features in the source image through the utilization of structural re-parameterization edge convolution (RECB) with embedded edge operators. Subsequently, the attention fusion module (AFM) is employed to sufficiently fuze both unique and public information from the heteromodal features. In the inference stage, we further optimize the training network using the structural reparameterization technique, resulting in a VGG-like network architecture. This optimization improves the fusion speed while maintaining the fusion performance. To evaluate the performance of the proposed FECFusion algorithm, qualitative and quantitative experiments are conducted. Seven advanced fusion algorithms are compared using MSRS, TNO, and M3FD datasets. The results demonstrate that the fusion algorithm presented in this paper achieves superior performance in multiple evaluation metrics, while consuming fewer computational resources. Consequently, the proposed algorithm yields better visual results and provides richer scene detail information.

Keywords: image fusion; edge operator; structural re-parameterization; infrared and visible images; deep learning

1. Introduction

As an image enhancement technology, image fusion is utilized to combine images captured by different types of sensors or under distinct shooting settings, aiming to obtain images with more comprehensive scene representation information [1]. Among various applications of image fusion, infrared and visible image fusion serves as a typical example. Infrared images are captured using infrared sensors, relying on thermal radiation. They are characterized by prominent targets, minimal environmental influence, high imaging noise and blurred details [2]. On the other hand, visible images exhibit rich texture details, high resolution and sensitivity to lighting conditions [3]. Image fusion enables the integration of unique and shared information from both modalities to generate fused images with enhanced texture and salient targets. These fused images play a crucial role in subsequent high-level vision tasks such as semantic segmentation [4] and nighttime vehicle target detection [5].

The research on infrared and visible image fusion has led to the development of various traditional methods that have been proposed [6–10]. These methods are often highly interpretable but rely on hand-designed fusion rules, which can limit their performance when dealing with more complex scene fusion tasks. However, with the advancements in deep learning, an increasing number of deep learning methods are being applied to image fusion tasks [11–15]. Deep networks exhibit strong capabilities in characterizing image features, surpassing the limitations of traditional feature extraction methods. They adopt a data-driven approach, enabling end-to-end generation of fused images.

In order to enhance the performance metrics of fusion, many existing deep learning-based fusion methods incorporate complex network modules that require more storage and computational resources to achieve improved performance metrics. For instance, Long et al. [16] proposed a network that aggregates residual dense blocks, combining dense connected blocks with residual connected blocks. Pu et al. [17] introduced a complex contextual information perceptual module for image reconstruction. Xu et al. [18] employed dissociative representation learning in an auto-encoder-based approach. These methods have demonstrated performance improvements in fusion results; however, they also introduce greater computational complexity due to the inclusion of complex modules in the network.

Furthermore, existing fusion algorithms often employ fusion layers that incorporate intricate fusion modules or fusion rules, with the primary aim of improving evaluation metrics. However, these algorithms often overlook the characteristics of different modalities. Notably, auto-encoder-based methods [18–21] utilize hand-designed fusion strategies for combining depth features. The use of such hand-designed fusion strategies may not assign proper weights to the depth features, leading to limitations in the performance of the fusion methods.

Currently, there is a lack of research on lightweight fusion models, which aim to reduce model parameters and convolutional depth channels. One example is PMGI [22], which performs information extraction through gradient and intensity scale preservation. It achieves this by reusing and fusing features extracted with fewer convolutional layers. Another lightweight model, FLFuse [23], generates fused images using a weight sharing encoder and feature swapping training strategy to ensure efficiency. However, FLFuse fails to fully extract and fuse image features due to its shallow network channel dimension and simplistic implicit fusion strategy, resulting in subpar visual effects and performance metrics.

We focus on exploring lightweight fusion methods based on structural re-parameterization. Existing structural re-parameterization methods have demonstrated high performance in training and fast inference speeds, making them effective for advanced vision tasks [24–27]. They are likely to be crucial in addressing

the imbalance between fusion performance and computational resource consumption. However, directly applying these structural re-parameterization blocks designed for high-level vision tasks provides limited improvement for infrared and visible image fusion. Specific structural re-parameterization blocks tailored for fusion tasks are required to efficiently extract richer information from different modal features.

To address the limitations of existing image fusion methods, this paper proposes a novel approach that combines edge operators with structural re-parameterization. This approach enables the rapid generation of fused images with enhanced edge texture information and prominent targets, effectively addressing the imbalance between fusion performance and computational resource consumption. The major contributions of this paper are outlined as follows:

- A fast edge convolution fusion network (FECFusion) for infrared and visible images is proposed, which combines edge operations with structural re-parameterization for the first time to rapidly generate fused images with rich edge texture information and salient target, solving the problem of imbalance between fusion performance and computational resource consumption.
- A structural re-parameterization edge convolution block (RECB) is proposed, which can deeply mine the edge information in the source images and improve the performance of the fusion model without introducing additional inference burden.
- An attention fusion module (AFM) is designed to sufficiently fuze the unique and common information of different modal features to effectively integrate the feature information of the source images with less computational effort.

2. Related works

2.1. Infrared and visible image fusion

In the current literature, there are numerous works that primarily focus on preserving texture details in images [28–30]. In contrast, our work aims to address the challenge of balancing lightweight design and performance in image fusion networks. One approach is IFCNN [31], where both the encoder and decoder components employ only two convolutional layers for feature extraction and image reconstruction. Additionally, the fusion rules are adjusted based on the source image type, resulting in a unified network capable of handling various fusion tasks. Another method, SDNet [32], tackles the fusion task by incorporating the generated fused image reconstruction into a squeezed network structure of the source image. This forces the fused image to contain more information from the source images. SeAFusion [33] utilizes dense blocks with gradient residuals for feature extraction and employs the semantic segmentation task loss to guide the training of the fusion network. Recently, FLFuse [23] achieves feature extraction implicitly through a weight sharing encoder and feature swapping training strategy, enabling the generation of fused images in a lightweight and fast manner.

However, existing methods for infrared and visible image fusion only reduce the network model's parameters through conventional lightweight network design approaches, which can lead to a degradation in fusion performance.

2.2. Structural re-parameterization

Many existing methods for infrared and visible image fusion rely on attention mechanisms or multi-scale feature extraction to enhance network performance, but these approaches often come at the cost of increased

computational complexity. Finding networks that effectively extract image features while maintaining high computational efficiency is challenging. In ACNet [34], Ding et al. proposed a method to convert multi-branch structures into a single-branch structure, thereby improving the performance of convolutional networks. In another work by Ding et al. [35], a concise VGG-like backbone network called RepVGG was introduced. RepVGG utilizes structural re-parameterization as its core technique, enabling efficient feature extraction while reducing network computation. RepVGG has demonstrated excellent performance in target detection tasks. Building upon this work, Ding et al. [36] proposed six network structures that can be structurally re-parameterized. The authors also explained the underlying reasons why the structural re-parameterization method is effective. Given the success of structural re-parameterization in various vision tasks, this approach holds promise for addressing the challenge of balancing fusion performance and computational resource consumption in infrared and visible image fusion tasks.

Unfortunately, the direct use of those structural re-parameterization blocks designed for advanced vision tasks provides little improvement for infrared and visible image fusion tasks. They still require the structural re-parameterization block specifically designed for the image fusion task to quickly extract the full wealth of information from the different modal features.

Therefore, we propose a new image fusion method, FECFusion, which substantially reduces computational resource consumption while maintaining high fusion performance through a well-designed structural re-parameterization technique.

3. Methods

3.1. Network architecture

In this paper, FECFusion utilizes end-to-end convolutional neural networks to perform feature extraction, feature fusion and image reconstruction, enabling efficient and straightforward fusion tasks. The network architecture, as depicted in Figure 1, consists of three main components: an encoder, a fusion layer and a decoder. In the encoder, a two-branch structure is employed, comprising one convolutional layer and two structural re-parameterization edge convolution blocks. This setup allows for the extraction of depth features from both the infrared and visible images. The fusion layer combines these extracted features, leveraging the complementary information present in the two modalities. Subsequently, the decoder, consisting of three structural re-parameterization edge convolution blocks, reconstructs the hybrid features obtained from the fusion layer to generate the final fused image. Overall, FECFusion offers a simple and efficient solution for infrared and visible image fusion, utilizing end-to-end convolutional neural networks for image fusion of the two modalities.

To ensure that the fused images better retain the edge feature information of the source images, FECFusion has designed a structural re-parameterization edge convolution block(RECB) for improving the performance in infrared and visible image fusion tasks. In addition, we use an attention fusion module (AFM) to better fuse the feature information of different modal images extracted from different branches.

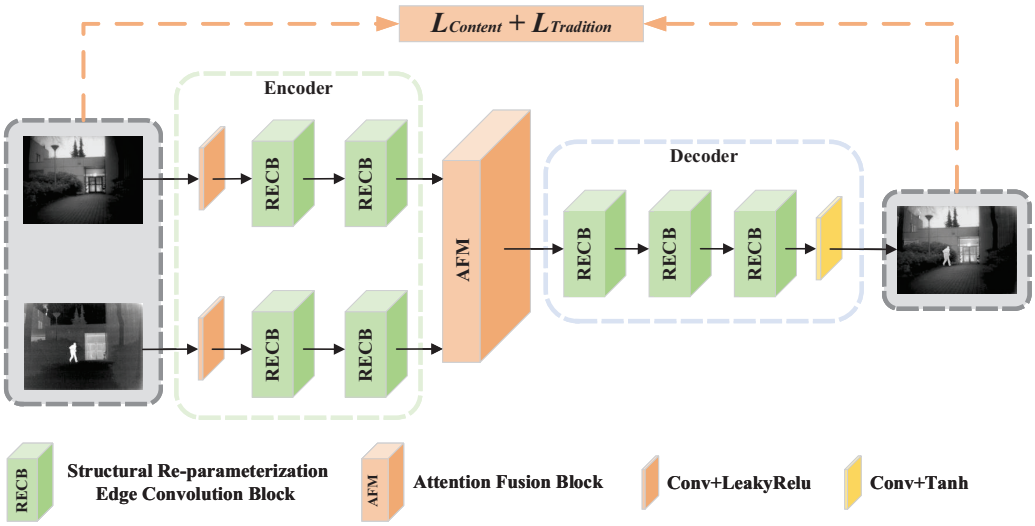


Figure 1. The overall structure of FECFusion consists of an encoder, a fusion layer (AFM), and a decoder. The infrared I_{ir} and visible images I_{vi} are simultaneously passed to a two-branch encoder to extract depth features, and a fusion layer to fuse common and unique features, finally reconstructed by a decoder to obtain the fused image I_f. The whole process is guided by both content loss $L_{content}$ and traditional loss $L_{tradition}$ to generate the fused image.

3.1.1. Structural re-parameterization edge convolution block (RECB)

Although it has some effects to use standard convolution to extract infrared and visible image feature information for fusion networks, it is inferior to complex models in terms of fusion performance. However, replacing standard convolution with complex blocks would make the network consume more computational resources. Therefore, the structural reparameterization technique is introduced in this paper to enrich the characterization capability of the network without increasing the computational resource consumption of the network in the inference stage.

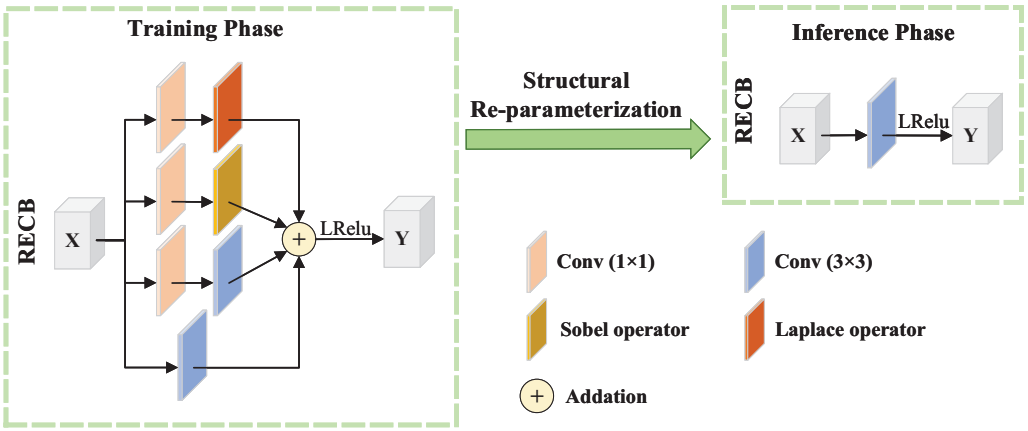


Figure 2. The specific devise of the structural re-parameterization edge convolution block(RECB). The RECB extracts fine-grained detail information of feature maps.

To ensure that the fused images are better able to retain the edge feature information of the source images, we design a structural re-parameterization edge convolution block(RECB) for improving the performance in infrared and visible image fusion task. The specific structure of RECB is shown in Figure 2.

In particular, the RECB consists of four elaborated operators as follows.

1) The branch of standard 3×3 convolution

To guarantee the basic performance of the module, we use a standard 3×3 convolution. This convolution is represented as:

$$F_n = K_n * X + B_n, \tag{3.1}$$

where F_n represents the output feature of 3×3 standard convolution. K_n represents the convolution kernel weight of 3×3 standard convolution. X represents the input feature. B_n represents the offset of 3×3 standard convolution.

2) The branch of feature expansion convolution

The representational power of the fusion task is improved by expanding the channels of the features, which helps to improve the extraction of more feature information. Specifically, the branch uses 1×1 convolution to expand the channel dimension of the features and 3×3 convolution to extract the feature information, which is expressed as:

$$F_e = K_n * (K_e * X + B_e) + B_n, \tag{3.2}$$

where F_e represents the output feature of the feature expansion convolution branch. K_e represents the convolution kernel of 1×1 convolution. B_e represents the offset of 1×1 convolution.

3) The branch of Sobel filter

Edge information is tremendously helpful for the performance improvement of the fusion task. Since it is usually difficult for the network model to learn the weights of the edge detection filters through training, a pre-defined Sobel edge filter is embedded in this branch for extracting the first-order spatial derivatives and learning the scaling factors of the filters. Specifically, the input features are firstly scaled by 1×1 convolution, then the edge information is extracted by horizontal and vertical Sobel filters, which are processed as follows:

$$D_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} \quad and \quad D_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \tag{3.3}$$

$$\begin{aligned} F_{D_x} &= (S_{D_x} \cdot D_x) \otimes (K_x * X + B_x) + B_{D_x}, \\ F_{D_y} &= (S_{D_y} \cdot D_y) \otimes (K_y * X + B_y) + B_{D_y}, \end{aligned} \tag{3.4}$$

$$F_{sobel} = F_{D_x} + F_{D_y}. \tag{3.5}$$

where K_x, B_x and K_y, B_y are the weights, bias of the 1×1 convolution in the horizontal and vertical directions. S_{D_x}, B_{D_x} and S_{D_y}, B_{D_y} are the scaling parameters and bias with the shape of $C \times 1 \times 1 \times 1$. $\otimes$ and $*$ represent DWConv and normal convolution. $(S_{D_x} \cdot D_x)$, $(S_{D_y} \cdot D_y)$ are in the shape of $C \times 1 \times 3 \times 3$.

4) The branch of Laplacian filter

In addition to the Sobel operator for extracting the first-order spatial derivatives, this branch employs a more stable Laplacian edge filter that is more robust to noise to extract the second-order spatial

derivatives of the image edge information. Similarly, this branch also uses 1×1 convolution for scaling and then uses the Laplacian operator to extract the edge information, processed as:

$$D_{lap} = \begin{bmatrix} 0 & +1 & 0 \\ +1 & -4 & +1 \\ 0 & +1 & 0 \end{bmatrix}, \tag{3.6}$$

$$F_{lap} = (S_{lap} \cdot D_{lap}) \otimes (K_{lap} * K + B_{lap}) + B_{lap}. \tag{3.7}$$

where K_{lap}, B_{lap} are the weights, bias of the 1×1 convolution. S_{lap}, B_{lap} are scaling factors and bias of DWConv, respectively.

In addition, the BN layer is not used in the RECB, unlike the structural re-parameterization block designed for advanced vision tasks, because the BN layer would hinder the performance of the fusion network. Finally, the output features of these four branches are summed and mapped to the nonlinear activation layer:

$$F = F_n + F_e + F_{sobel} + F_{lap}. \tag{3.8}$$

where F is the output characteristic of RECB. The nonlinear activation layer used in this experiment is LeakyRelu.

The above RECB is the structure of the training phase. After the training is completed, the parameters of the four branch structures are equivalent to a 3×3 convolution parameter through the structure re-parameterization technique, so that the same effect can be obtained only through the 3×3 convolution processing after the structure re-parameterization in the inference phase.

3.1.2. Attention fusion module (AFM)

Since these two features come from source images of different modalities, they have object focus of different scenes, with complementary information and public information of each other. Thus, the fusion module should have to focus on the fusion of complementary information and public information of different modalities.

In Figure 3, it is evident that detecting pedestrians in visible images at night can be challenging due to inadequate lighting conditions. However, in infrared thermal images, pedestrians are clearly highlighted. Therefore, the key challenge lies in fusing these two features by leveraging the complementary information that exists in only one of the modalities. In the case of a well-illuminated thermal target, such as the vehicle in Figure 3, both cameras are capable of sensing it. During the fusion process, it is important to enhance both features simultaneously. If a method is used that focuses solely on processing the complementary information, there is a risk of weakening one of the features. In order to effectively fuse the infrared and visible features, it is crucial to devise a fusion approach that preserves and enhances the salient information from both modalities. This will ensure that both the infrared highlights and the visible details are effectively integrated, leading to improved detection results.

(a) Visible image (b) Infrared image

Figure 3. Illustrations of registered infrared and visible images, where infrared and visible images have unique (the red box) and common (the green box) information.

In order to better solve the problem of fusion of different modal information, the element-by-element addition method for extracting complementary information of heteromodal images and the element-by-element multiplication method for extracting common information of heteromodal images are employed here. The element-by-element addition and element-by-element multiplication methods are expressed as:

$$\begin{aligned} X_{add} &= X_{vi} + X_{ir}, \\ X_{mul} &= X_{vi} * X_{ir}. \end{aligned} \tag{3.9}$$

where X_{vi} and X_{ir}represent the depth features of infrared and visible images extracted by the encoder, respectively. Element-by-element addition X_{add} represents the addition of elements to visible image features and thermal target features to accumulate complementary information of different modes, while element-by-element multiplication X_{mul} represents the multiplication of elements to visible image features and thermal target features to enhance common information of different modes.

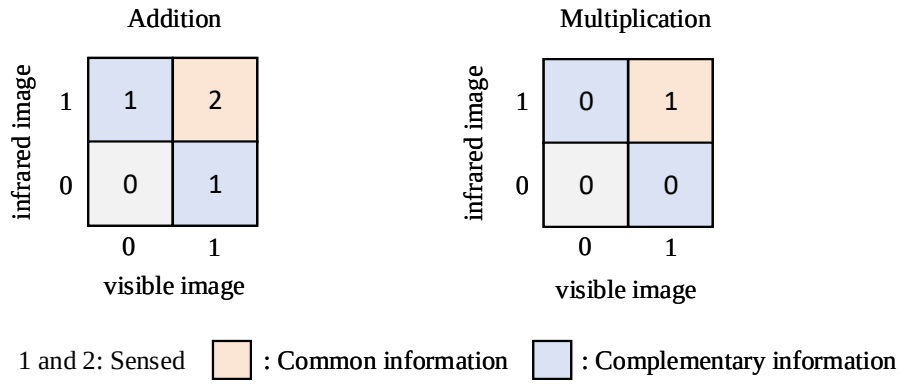


Figure 4. The simple illustration for the element-by-element addition of the complementary information and the element-by-element multiplication of the common information that are extracted from the infrared and visible images.

As in the simple example in Figure 4, 0 represents that the target is not sensed by the sensor and 1 represents that the target is sensed by the sensor. Suppose there are two cases of sensed and un-sensed target for infrared and visible sensors, respectively, which are illustrated by 0 and 1. Therefore, four

cases are generated, and the goal is to retain all the information as long as possible. Element-by-element addition preserves the target information to the maximum extent possible, the left in Figure 4, with the target being sensed by at least one of the sensors. Element-by-element multiplication allows filtering out the common information, the right image in Figure 4, with the target needing to be sensed by two sensors at the same time. Once the two types of information are obtained separately, they are preserved by feature concatenation, combining the unique information from both modalities. This approach allows for the retention of crucial information while effectively combining the features extracted from each sensor.

Therefore, the attention fusion module (AFM) based on element-by-element addition and element-by-element multiplication is designed to better fuse the information of these two depth features. The specific structure of the AFM is shown in Figure 5.

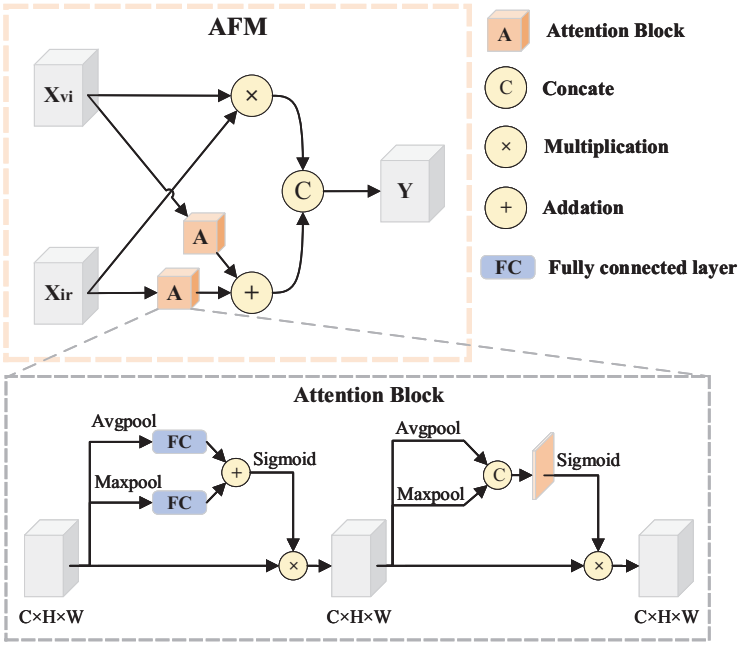


Figure 5. The specific devise of the attention fusion module (AFM), which is used to fuse unique and common information from different modal images.

The upper branch in AFM is used to enhance the common information of different modal features, while the bottom branch enhances the feature information of different modal features through the attention module and aggregates the complementary information of different modal features by feature summation, and then cascades the common and complementary information to allow both feature information to be retained as much as possible.AFM is processed as follows:

$$Y = Cat(X_{vi} * X_{ir}, A(X_{vi}) + A(X_{ir})). \tag{3.10}$$

where Cat represents concatenation; A represents the attention module, and CBAM is used in this article.

3.2. Loss function

The loss function is the key to guide the training of deep neural networks to achieve the desired results. The loss of structural similarity is often used to maintain a clear intensity distribution of the fused image. However, in the fusion task, the fused image needs to be similar to the two source images at the same time, and there is more complementary information between the two source images, which will weaken the complementary information region and lead to a decrease in fusion performance.

In order to better promote the recovery of texture details, this paper uses content loss $L_{content}$ and traditional loss $L_{tradition}$ to jointly constrain the training of the network. The total loss L_{fusion} formula is as follows:

$$L_{fusion} = \lambda L_{tradition} + L_{content}. \tag{3.11}$$

where λ is the weight coefficient to balance these two losses.

3.2.1. Traditional loss

The design of traditional loss function can enhance the similarity between the fused image and the two types of source images, guide the network to generate the fusion result with complete information faster, and avoid the single and incomplete information in the fusion result. The traditional loss $L_{tradition}$ calculation formula is :

$$L_{tradition} = \frac{1}{HW}\left\|I_f - 0.5 * (I_{ir} + I_{vi})\right\|_1 . \tag{3.12}$$

where I_f represents the fused image. I_{ir} and I_{vi} represent the infrared and visible images. $\|\cdot\|$ refers to the l_1-norm.

3.2.2. Content loss

In order to promote the model to fuse more meaningful information, retain the saliency in the infrared image and the edge texture information of the source image, the content loss $L_{content}$ with bilateral filtering is designed in this paper. The content loss consists of two parts: the intensity loss L_{in} and the edge gradient loss with bilateral filtering L_{grad} . The formula is as follows:

$$L_{content} = \mu_1 L_{in} + \mu_2 L_{grad}. \tag{3.13}$$

where μ_1, μ_2 are the weighting coefficients for balancing these two losses.

Among them, the intensity loss L_{in} constrains the overall apparent intensity of the fused image. In order to better retain the salient target, the pixel intensity of the fused image should be biased towards the maximum intensity of the infrared and visible images. The formula of strength loss is as follows:

$$L_{in} = \frac{1}{HW}\left\|I_f - Max(I_{ir}, I_{vi})\right\|_1 . \tag{3.14}$$

where $Max(\cdot)$ stands of the element-wise maximum calculation.

In addition, in order to make the network model better preserve the edge texture details of the fused image, the existing methods use the maximum edge gradient of the source image to constrain the training of the network, but this loss is easily affected by noise in the infrared image. To this end, this paper uses a bilateral filter that preserves the edge gradient to denoise the infrared image, thereby reducing the

noise of the fused image. The following is the calculation formula of edge gradient loss of bilateral filtering:

$$L_{grad} = \frac{1}{HW}\left\|\left|\nabla I_f\right| - Max(|\nabla Bila(I_{ir})|, |\nabla I_{vi}|)\right\|_1. \tag{3.15}$$

where ∇ is the gradient operator for measuring image texture information, and the *Sobel* operator is used to calculate the gradient in this paper. $|\cdot|$ indicates the absolute operation. *Bila* represents a bilateral filter.

4. Experimental analysis

4.1. Experimental details

In this paper, FECFusion is trained with the MSRS dataset [37]. Since the existing infrared and visible image fusion dataset is small, the MSRS training set part of the infrared and visible images common dataset is expanded from 1083 to 26,112 pairs of images, and the size of the training set image pairs after data enhancement is 64 pixel × 64 pixel, which can basically meet the training requirements. In order to evaluate the effectiveness of FECFusion, it is tested on the test set of the MSRS dataset and picks 361 pairs of images as the test subject. In addition, to more comprehensively evaluate the generalization performance of FECFusion, it selected 42 and 300 pairs of images on the TNO [38] and M3FD [39] datasets, respectively, for generalization comparison experiments.

Since none of the current public datasets for infrared and visible image fusion have reference images, the quality of the fusion result images cannot be directly evaluated by ground truth, therefore, we evaluate the visualization image effects of different algorithms by human subjective visual perception as a qualitative assessment, and by objective generic image quality evaluation index results as a quantitative assessment.

In this paper, standard deviation (SD) [40], mutual information (MI) [41], visual information fidelity (VIF) [42], sum of correlation differences (SCD) [43], entropy (EN) [44] and Q_{abf} [45] are used. SD evaluates the contrast and distribution of the fused images from a statistical point of view. MI measures the amount of information from the source image to the fused image. VIF reflects the fidelity of the fused information from a human visual point of view. SCD measures the difference between the source image and the fused image. EN measures the amount of information contained in the image. Q_{abf} evaluates the amount of fused edge information from the source image. All the above metrics are positive metrics, and higher values mean better fusion results.

FECFusion is compared with seven fusion algorithms, including DenseFuse [46], FusionGAN [47], IFCNN [31], SDNet [32], U2Fusion [48], FLFuse [23] and PIAFusion [37]. All the compared algorithms are experimented in public code, where the relevant settings of the experiments are kept constant. In the superparameter settings of the proposed network, the network optimizer uses Adam, epoch = 10, batch size = 64, learning rate is 1×10^{-4}, loss function parameters are $\lambda = 10$, $\mu_1 = 12$, $\mu_2 = 45$. The parameters of bilateral filtering are $\sigma_d = 0.05$, $\sigma_r = 8.0$, and window size is 11×11. The training process of FECFusion is summarized in Algorithm 1.

Besides the comparative and generalization experiments, the effectiveness of RECB and AFM is verified by ablation experiments in this paper. In addition, FECFusion is verified to be helpful for the advanced vision task through segmentation experiments. Finally, we have compared the operational efficiency of FECFusion with other methods and compare the computational resource consumption with and without structural re-parameterization. Our experiments are all conducted on a GeForce RTX 2080Ti 11GB and an Intel Core i5-12600KF, with PyTorch of a deep learning framework.

Algorithm 1: Training procedure

Input: Infrared images I_{ir} and visible images I_{vi}
Output: Fused images I_f

for *M epochs* **do**
  for *p steps* **do**
    Select n infrared images $\{I_{ir}^1, I_{ir}^2, ..., I_{ir}^n\}$;
    Select n visible images $\{I_{vi}^1, I_{vi}^2, ..., I_{vi}^n\}$;
    synthesize n fused images $\{I_f^1, I_f^2, ..., I_f^n\}$ with our FECFusion;
    Calculate the total loss L_{fusion} according to Eq (3.11);
    Update the parameters of the FECFusion by Adam Optimizer;
  end
  Save the network model weights for the Mth round;
end
for *q steps* **do**
  Reads the current network module and parameters;
  if *The module has "switch_to_deploy" method* **then**
    Compute the parameter of the convolution kernel and bias after re-parameterization;
    Assign parameters to the newly created convolution block;
    Delete the original network structure;
  end
end
Save the model and weights of the network after structural re-parameterisation.

4.2. Comparative experiment

4.2.1. Qualitative results

It is an important challenge for the image fusion algorithm to generalize the performance of different scenes. In the MSRS dataset, we have chosen two daytime and two nighttime images to evaluate the subjective visualization performance, and the comparison results are shown in Figures 6 and 7. We mark the texture detail information with green boxes and the highlighted target information with red boxes.

In the daytime scene depicted in Figure 6, we can observe the performance of different fusion methods. DenseFuse, SDNet and U2Fusion fail to effectively highlight the infrared target and do not fully utilize the background information present in the visible image. FusionGAN manages to highlight the salient target to some extent, but it can be seen from the green box that it blurs the background. In contrast, IFCNN and FLFuse weaken the texture details of the background, as evident from the green box. Only PIAFusion and the method proposed in this paper successfully integrate the relevant information, effectively preserving both the infrared target and the background texture details. Therefore, it is evident that the method proposed in this paper exhibits superior performance in the daytime scene, achieving a balanced fusion result that highlights the target while retaining the background information.

In the night scene depicted in Figure 7, the visible image contains limited texture information, while the infrared image contains both background texture details and a salient target. Many existing fusion methods tend to overemphasize the information from one modality, making it challenging to achieve

satisfactory results across different scenes.

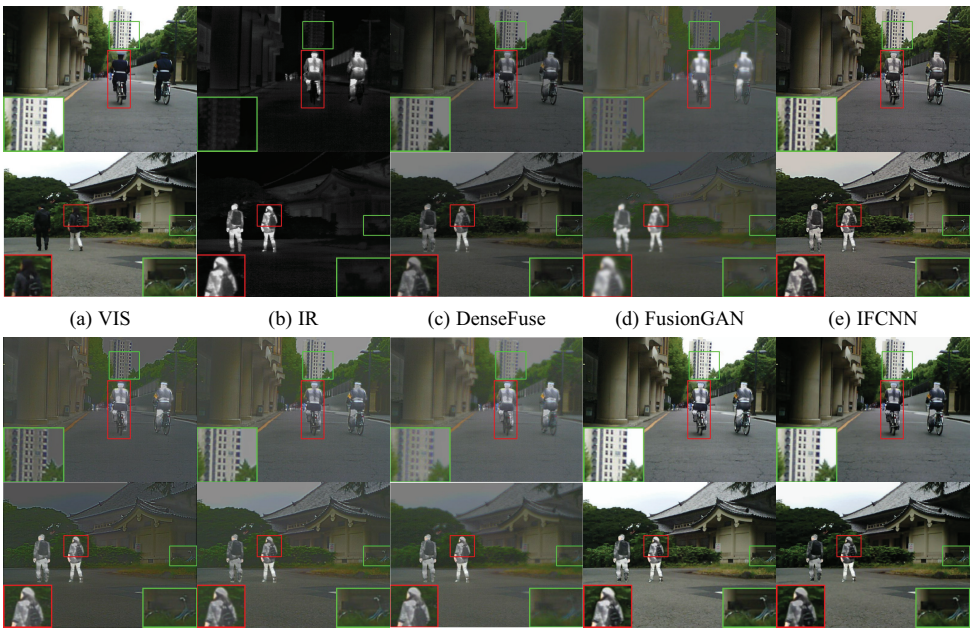


Figure 6. Qualitative comparison of FECFusion with 7 advanced algorithms on the daytime scene (00537D and 00633D) from the MSRS dataset. For a clear view of comparative detail, we have selected a textured region (the green box) and a salient region (the red box) in each image.

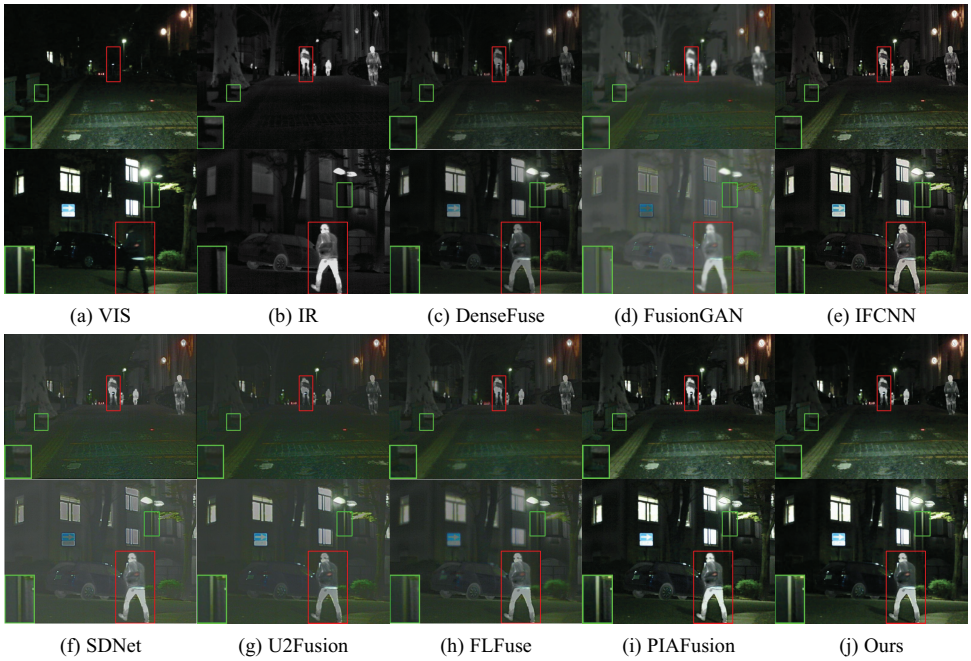


Figure 7. Qualitative comparison of FECFusion with 7 advanced algorithms on the nighttime scene (01023N and 01042N) from the MSRS dataset. For a clear view of comparative detail, we have selected a textured region (the green box) and a salient region (the red box) in each image.

Among the fusion methods examined, DenseFuse, SDNet and U2Fusion exhibit a bias towards infrared images, and weaken infrared targets. FusionGAN introduces artifacts into the fused images. Only IFCNN, PIAFusion, and the method proposed in this paper are capable of generating fused images with higher contrast in black night scenes. FLFuse, which is also a lightweight method, performs poorly in this scenario. It fails to fully leverage the characteristics of both modal images, leading to a degradation in both the background and the infrared target. Therefore, the method proposed in this paper demonstrates good performance in night scenes as well. It effectively captures the characteristics of both modal images and achieves better contrast, thereby preserving the infrared target while retaining background details.

4.2.2. Quantitative results

In this section, we perform quantitative evaluation on the MARS dataset and select six metrics for evaluation. The comparison of the metrics of different methods is shown in Table 1, where red represents the best result and blue represents the second one.

Table 1. Quantitative comparisons of the six metrics, i.e., SD, MI, VIF, SCD, EN and Q_{abf}, on image pairs from the MSRS dataset. Bold indicates the best result and underline represents the second best result.

Algorithm	Evaluation method					
	SD	MI	VIF	SCD	EN	Qabf
DenseFuse	7.0692	2.5409	0.6752	1.3296	5.8397	0.3552
FusionGAN	5.4694	1.9155	0.4253	0.8015	5.2260	0.1208
IFCNN	7.5947	2.7399	0.8283	1.6658	6.3109	0.5540
SDNet	5.3258	1.7398	0.3758	0.8364	4.8891	0.2944
U2Fusion	5.6231	1.8953	0.3967	1.0034	4.7525	0.2908
FLFuse	6.4790	2.0697	0.4860	1.1189	5.5157	0.3198
PIAFusion	*7.9268*	**4.1774**	*0.9072*	*1.7395*	*6.4304*	**0.6324**
Ours	**8.1413**	*3.6805*	**0.9282**	**1.8153**	**6.5104**	*0.5619*

From Table 1, it is clear that our method shows significant advantages in four metrics, SD, VIF, SCD and EN, while its performance in MI and Qabf is second only to PIAFusion. The value of SD is the best indicating that the fusion result of this paper method Shencheng achieves high contrast between infrared target and background; the highest value of VIF indicates that the fused image generated by this paper method is more in line with The highest value of VIF indicates that the fused images generated by this method are more consistent with the human visual system; the highest values of SCD and EN indicate that this method can generate fused images with more edge details and contain more realistic results than other methods. In conclusion, the quantitative experimental results show that this method can generate fused images with more information while reducing the computational effort.

4.3. Generalization experiment

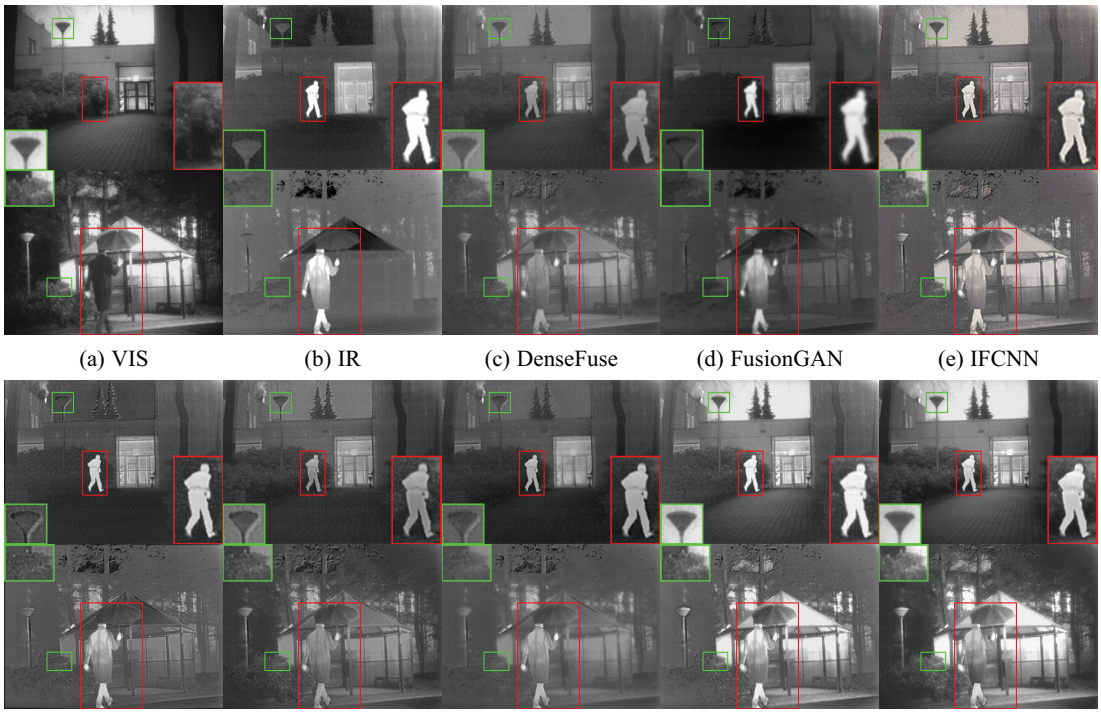
In the fusion task, it is required that the fusion model has a stronger generalization capability, which is applicable in different scenes. Therefore, we selected 20 and 300 pairs of images in the TNO and M3FD datasets, respectively, to evaluate the generalization ability of FECFusion. A qualitative comparison of the different algorithms on the TNO and M3FD datasets is presented in Figures 8 and 9.

From the figures, it is evident that DenseFuse, SDNet, U2Fusion and FLFuse tend to blend the background and the target together, making it difficult to distinguish the salient infrared target. FusionGAN, on the other hand, exhibits high overlap with the infrared image and lacks the inclusion of background information. In comparison, IFCNN and PIAFusion show performance similar to the proposed method in this paper. However, it is important to note that these methods may not match the inference speed of the proposed method, which offers faster processing capabilities. Therefore, based on objective evaluation and considering the faster inference speed, the proposed method in this paper demonstrates competitive performance and provides a promising solution for infrared and visible image fusion tasks.

The results of quantitative metrics for the generalization experiments are shown in Table 2. The metrics performance of our method is the best or the second best on both datasets, which indicates that our method can both preserve the texture details of the source image and improve the contrast of the target. In conclusion, the qualitative and quantitative results show that FECFusion performs excellently in generalization. In addition, the method in this paper effectively maintains the intensity distribution of the target region and preserves the texture details of the background region, benefiting from the proposed RECB and AFM.

Table 2. Quantitative comparisons of the six metrics, i.e., SD, MI, VIF, SCD, EN and Q_{abf}, from the TNO and M3FD datasets. Bold indicates the best result and underline represents the second best result.

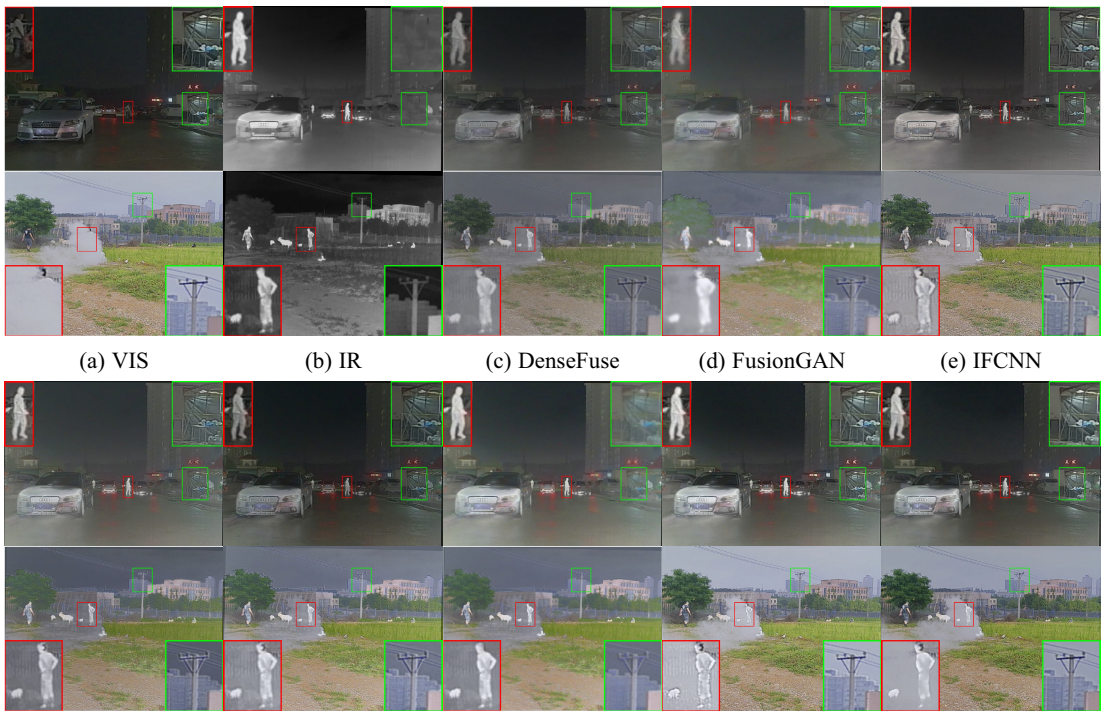
Dataset	Algorithm	Evaluation method					
		SD	MI	VIF	SCD	EN	Qabf
TNO	DenseFuse	8.5765	2.1987	0.6704	1.5916	6.3422	0.3427
	FusionGAN	8.6703	2.3353	0.6541	1.3788	6.5578	0.2339
	IFCNN	9.0058	2.4154	0.7996	1.6850	6.7413	0.5066
	SDNet	9.0679	2.2606	0.7592	1.5587	6.6947	0.4290
	U2Fusion	8.8553	1.8730	0.6787	1.5862	6.4230	0.4245
	FLFuse	*9.2628*	2.1925	0.8084	*1.7308*	6.3658	0.4177
	PIAFusion	9.1093	*3.2464*	*0.8835*	1.6540	*6.8937*	**0.5556**
	Ours	**9.2721**	**3.7136**	**0.9496**	**1.7312**	**6.9856**	*0.5311*
M3FD	DenseFuse	8.6130	2.8911	0.6694	1.5051	6.4264	0.3709
	FusionGAN	8.8571	2.9921	0.5176	1.1292	6.4750	0.2530
	IFCNN	9.2815	2.9560	0.7738	1.5353	6.6966	0.6053
	SDNet	8.8855	3.1798	0.6329	1.3914	6.6102	0.5005
	U2Fusion	9.0141	2.7531	0.7061	*1.5488*	6.6285	0.5303
	FLFuse	8.7580	3.2425	0.6986	1.4975	6.5744	0.2640
	PIAFusion	**10.1639**	**4.6942**	*0.9300*	1.3363	**6.8036**	*0.6348*
	Ours	*9.9899*	*4.3123*	**0.9350**	**1.5502**	*6.7685*	**0.6440**



(a) VIS (b) IR (c) DenseFuse (d) FusionGAN (e) IFCNN

(f) SDNet (g) U2Fusion (h) FLFuse (i) PIAFusion (j) Ours

Figure 8. The visualisation results of FECFusion with 7 advanced algorithms on the TNO dataset. For a clear view of comparative detail, we selected a textured region (the green box) and a salient region (the red box) in each image.



(a) VIS (b) IR (c) DenseFuse (d) FusionGAN (e) IFCNN

(f) SDNet (g) U2Fusion (h) FLFuse (i) PIAFusion (j) Ours

Figure 9. The visualisation results of FECFusion with 7 advanced algorithms on the M3FD dataset. For a clear view of comparative detail, we selected a textured region (the green box) and a salient region (the red box) in each image.

4.4. Ablation experiment

In order to verify the effectiveness of adding RECB and AFM to our FECFusion, ablation experiments are designed in this section to further analyze the role of these two proposed modules in the network model. First, RECB is a structure that equates multiple branch structures into a single branch structure by structural re-parameterization; therefore, in the ablation experiments, the RECB part is directly replaced with the structure after structural re-parameterization for training, i.e., the structure of a single ordinary convolution. For the ablation experiments of AFM, the network is trained with a direct feature cascade instead of AFM. This experiment is performed on the MSRS dataset, and the experimental results are shown in the Figure 10, where the background texture is marked with a green solid box and the infrared salient targets are marked with a red solid box.

From the experimental results, it can be seen that without RECB, the fused images are blurred at the edges to some extent, which proves that the module contributes to maintaining the edge information of the fusion results. Without AFM, the saliency of the fusion results decreases. If both RECB and AFM are not available, the fusion results have a decreased target significance and blurred edge texture. The evaluation metrics for this ablation experiment are shown in Table 3. We observe that the absence of both RECB and AFM leads to a decrease in the evaluation metric values to different degrees, proving the effectiveness of each part of our FECFusion.

(a) VIS (b) IR (c) Ours (d) W/O RECB (e) W/O AFM (f) W/O both RECB and AFM

Figure 10. Visualized results of ablation on the MSRS dataset. From left to right: visible images, infrared images, fused results of FECFusion, FECFusion without RECB, FECFusion without AFM and FECFusion without both RECB and AFM.

Table 3. The results of ablation study for RECB and AFM on the MSRS dataset. The bolded values indicate the best results.

RECB	AFM	Evaluation method					
		SD	MI	VIF	SCD	EN	Qabf
✓	✓	**8.1413**	**3.6805**	**0.9282**	**1.8153**	**6.5104**	**0.5619**
✓	✗	7.4551	3.0648	0.7234	1.6117	6.0100	0.4488
✗	✓	7.7954	3.0922	0.8355	1.8060	6.2925	0.5098
✗	✗	6.6285	2.6117	0.5500	1.2793	5.6158	0.4362

4.5. Efficiency comparison experiment

To verify the execution efficiency of the proposed algorithm, the average processing time of forward propagation of each fusion method is tested on the MSRS dataset in this paper, and the comparison results are shown in Table 4 where red represents the best and blue represents the second best. It can be seen that our method is more efficient than most methods, while FLFuse is faster than our method, the method in this paper works better, so these differences in running efficiency are acceptable.

Table 4. Mean of the running times of all methods on the MSRS dataset (underline: second, bold indicates the best result and italic represents the second best result).

Algorithm	DenseFuse	FusionGAN	IFCNN	SDNet	U2Fusion	FLFuse	PIAFusion	Ours
Running time	0.374	0.082	0.019	0.014	0.155	**0.001**	0.081	*0.002*

In addition, with reference to the image size used in the MSRS data set, the data of $640 \times 480 \times 1$ is used as the input of the network forward propagation, and the parameters and weights of the network model are calculated by the TorchSummary library. The forward propagation time, running storage space, parameter quantity, weight size and the cumulative deviation of the fusion results pixel by pixel before and after the structural re-parameterization are compared experimentally. The comparison results are shown in Table 5.

Table 5. Model properties of FECFusion with and without the structural re-parameterisation.

Structural re-parameterisation	Forward time	Forward pass size	Params	Params size	Cumulative pixel deviation of results
W/O	0.0299 s	5451.57 MB	146,389	0.56 MB	/
W	0.0020 s	2601.57 MB	145,477	0.55 MB	1×10^{-4}

By comparing the results, it can be seen that there is almost no difference in the fusion results of the network before and after the structural re-parameterization, indicating that the structural re-parameterization can effectively reduce the running time, running storage space, parameter quantity and weight size in the case of very low deviation.

4.6. Segmentation experiment

Semantic segmentation algorithms are an important general-purpose computer vision method whose performance reflects well on the semantic information of the fused resultant image. To verify that the fused images can be helpful for subsequent vision tasks, DeepLabV3+ [49], a semantic segmentation model pre-trained on the Cityscapes dataset [50], is also used in this section to evaluate the performance of the fused images, and the semantic segmentation results are shown in the Figure 11.

From the experimental results, the semantic segmentation results of the fused result images are all a little better than the infrared and visible images, especially at night when the lighting conditions are poorer, the visible sensors have difficulty capturing enough information, and the semantic segmentation models often have difficulty detecting hot targets such as travelers, so to some extent, it can be shown that the image fusion has an enhanced effect on subsequent vision tasks.

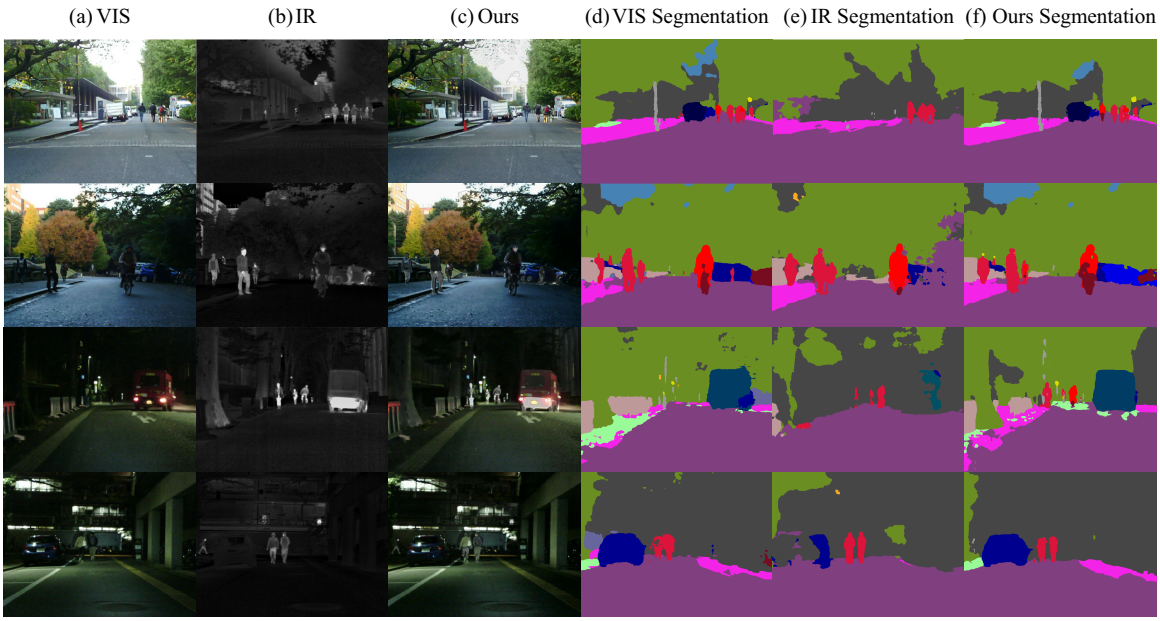


Figure 11. Segmentation results for infrared, visible, and fused images from the MSRS dataset. The segmentation model is Deeplabv3+, pre-trained on the Cityscapes dataset.

5. Conclusions

In this paper, we propose FECFusion, an infrared and visible image fusion network based on fast edge convolution. The network consists of several key components. First, the main part of the network employs the RECB to extract features, including detailed texture features and salient image features. These extracted features are then fused using the AFM, and the fused image is reconstructed. After the completion of training, the network undergoes a structural re-parameterization operation to optimize the inference speed and storage space required while preserving the original training effectiveness. Through subjective and objective experimental results, we demonstrate that FECFusion achieves superior fusion results compared to other algorithms. It offers better real-time performance and requires less inference memory footprint, making it more suitable for practical engineering applications that involve the design of custom hardware accelerated circuits. In future research, we will explore specific applications of FECFusion on mobile devices and further optimize its performance. This includes enhancing the network's ability to learn multi-scale image features and achieve better fusion results with lower computational resource consumption.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is supported by National Natural Science Foundation of China(NO.62266025).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. Chen, X. Li, L. Luo, J. Ma, Multi-focus image fusion based on multi-scale gradients and image matting, *Trans. Multimedia*, **24** (2021), 655–667. https://doi.org/10.1109/TMM.2021.3057493
2. S. Karim, G. Tong, J. Li, A. Qadir, U. Farooq, Y. Yu, Current advances and future perspectives of image fusion: A comprehensive review, *Inf. Fusion*, **90** (2023), 185–217. https://doi.org/10.1016/j.inffus.2022.09.019
3. H. Zhang, H. Xu, X. Tian, J. Jiang, J. Ma, Image fusion meets deep learning: A survey and perspective, *Inf. Fusion*, **76** (2021), 323–336. https://doi.org/10.1016/j.inffus.2021.06.008
4. H. Liu, F. Chen, Z. Zeng, X. Tan, AMFuse: Add–multiply-based cross-modal fusion network for multi-spectral semantic segmentation, *Remote Sens.*, **14** (2022), 3368. https://doi.org/10.3390/rs14143368
5. P. Gao, T. Tian, T. Zhao, L. Li, N. Zhang, J. Tian, GF-detection: Fusion with GAN of infrared and visible images for vehicle detection at nighttime, ***Remote Sens.***, **14** (2022), 2771. https://doi.org/10.3390/rs14122771
6. J. Chen, X. Li, L. Luo, X. Mei, J. Ma, Infrared and visible image fusion based on target-enhanced multiscale transform decomposition, *Inf. Sci.*, **508** (2020), 64–78. https://doi.org/10.1016/j.ins.2019.08.066
7. H. Tang, G. Liu, L. Tang, D. P. Bavirisetti, J. Wang, MdedFusion: A multi-level detail enhancement decomposition method for infrared and visible image fusion, *Infrared Phys. Technol.*, **127** (2022), 104435. https://doi.org/10.1016/j.infrared.2022.104435
8. Y. Li, G. Li, D. P. Bavirisetti, X. Gu, X. Zhou, Infrared-visible image fusion method based on sparse and prior joint saliency detection and LatLRR-FPDE, *Digital Signal Process.*, **134** (2023), 103910. https://doi.org/10.1016/j.dsp.2023.103910
9. J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, *Inf. Fusion*, **31** (2016), 100–109. https://doi.org/10.1016/j.inffus.2016.02.001
10. J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, *Infrared Phys. Technol.*, **82** (2017), 8–17. https://doi.org/10.1016/j.infrared.2017.02.005
11. L. Tang, Y. Deng, Y. Ma, J. Huang, J. Ma, SuperFusion: A versatile image registration and fusion network with semantic awareness, *IEEE/CAA J. Autom. Sin.*, **9** (2022), 2121–2137. https://doi.org/10.1109/JAS.2022.106082
12. J. Ma, L. Tang, M. Xu, H. Zhang, G. Xiao, STDFusionNet: An infrared and visible image fusion network based on salient target detection, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–13. https://doi.org/10.1109/TIM.2021.3075747
13. J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, Y. Ma, SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer, *IEEE/CAA J. Autom. Sin.*, **9** (2022), 1200–1217. https://doi.org/10.1109/JAS.2022.105686

14. H. Li, Y. Cen, Y. Liu, X. Chen, Z. Yu, Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion, *IEEE Trans. Image Process.*, **30** (2021), 4070–4083. https://doi.org/10.1109/TIP.2021.3069339
15. H. Liu, M. Ma, M. Wang, Z. Chen, Y. Zhao, SCFusion: Infrared and visible fusion based on salient compensation, *Entropy*, **25** (2023), 985. https://doi.org/10.3390/e25070985
16. Y. Long, H. Jia, Y. Zhong, Y. Jiang, Y. Jia, RXDNFuse: A aggregated residual dense network for infrared and visible image fusion, *Inf. Fusion*, **69** (2021), 128–141. https://doi.org/10.1016/j.inffus.2020.11.009
17. Q. Pu, A. Chehri, G. Jeon, L. Zhang, X. Yang, DCFusion: Dual-headed fusion strategy and contextual information awareness for infrared and visible remote sensing image, *Remote Sens.*, **15** (2023), 144. https://doi.org/10.3390/rs15010144
18. H. Xu, X. Wang, J. Ma, DRF: Disentangled representation for visible and infrared image fusion, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–13. https://doi.org/10.1109/TIM.2021.3056645
19. H. Li, X. J. Wu, J. Kittler, RFN-Nest: An end-to-end residual fusion network for infrared and visible images, *Inf. Fusion*, **73** (2021), 72–86. https://doi.org/10.1016/j.inffus.2021.02.023
20. H. Xu, M. Gong, X. Tian, J. Huang, J. Ma, CUFD: An encoder–decoder network for visible and infrared image fusion based on common and unique feature decomposition, *Comput. Vision Image Understanding*, **218** (2022), 103407. https://doi.org/10.1016/j.cviu.2022.103407
21. H. Li, X. J. Wu, T. Durrani, NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models, *IEEE Trans. Instrum. Meas.*, **69** (2020), 9645–9656. https://doi.org/10.1109/TIM.2020.3005230
22. H. Zhang, H. Xu, Y. Xiao, X. Guo, J. Ma, Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **34** (2020), 12797–12804. https://doi.org/10.1609/aaai.v34i07.6975
23. W. Xue, A. Wang, L. Zhao, FLFuse-Net: A fast and lightweight infrared and visible image fusion network via feature flow and edge compensation for salient information, *Infrared Phys. Technol.*, **127** (2022), 104383. https://doi.org/10.1016/j.infrared.2022.104383
24. X. Zhang, H. Zeng, L. Zhang, Edge-oriented convolution block for real-time super resolution on mobile devices, in *Proceedings of the 29th ACM International Conference on Multimedia*, (2021), 4034–4043. https://doi.org/10.1145/3474085.3475291
25. P. K. A. Vasu, J. Gabriel, J. Zhu, O. Tuzel, A. Ranjan, MobileOne: An improved one millisecond mobile backbone, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2023), 7907–7917. https://doi.org/10.48550/arXiv.2206.04040
26. P. K. A. Vasu, J. Gabriel, J. Zhu, O. Tuzel, A. Ranjan, FastViT: A fast hybrid vision transformer using structural reparameterization, preprint, arXiv:2303.14189. https://doi.org/10.48550/arXiv.2303.14189
27. X. Ding, X. Zhang, J. Han, G. Ding, Scaling up your kernels to 31x31: Revisiting large kernel design in cnns, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 11963–11975. https://doi.org/10.1109/CVPR52688.2022.01166
28. X. Liao, J. Yin, M. Chen, Z. Qin, Adaptive payload distribution in multiple images steganography based on image texture features, *IEEE Trans. Dependable Secure Comput.*, **19** (2020), 897–911. https://doi.org/10.1109/TDSC.2020.3004708

29. X. Liao, Y. Yu, B. Li, Z. Li, Z. Qin, A new payload partition strategy in color image steganography, *IEEE Trans. Circuits Syst. Video Technol.*, **30** (2019), 685–696. https://doi.org/10.1109/TCSVT.2019.2896270

30. J. Tan, X. Liao, J. Liu, Y. Cao, H. Jiang, Channel attention image steganography with generative adversarial networks, *IEEE Trans. Network Sci. Eng.*, **9** (2021), 888–903. https://doi.org/10.1109/TNSE.2021.3139671

31. Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, L. Zhang, IFCNN: A general image fusion framework based on convolutional neural network, *Inf. Fusion*, **54** (2020), 99–118. https://doi.org/10.1016/j.inffus.2019.07.011

32. H. Zhang, J. Ma, SDNet: A versatile squeeze-and-decomposition network for real-time image fusion, *Int. J. Comput. Vision*, **129** (2021), 2761–2785. https://doi.org/10.1007/s11263-021-01501-8

33. L. Tang, J. Yuan, J. Ma, Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network, *Inf. Fusion*, **82** (2022), 28–42. https://doi.org/10.1016/j.inffus.2021.12.004

34. X. Ding, Y. Guo, G. Ding, J. Han, Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 1911–1920. https://doi.org/10.1109/ICCV.2019.00200

35. X. Ding, X. Zhang, N. Ma, et al., Repvgg: Making vgg-style convnets great again, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), 13733–13742. https://doi.org/10.1109/CVPR46437.2021.01352

36. X. Ding, X. Zhang, J. Han, G. Ding, Diverse branch block: Building a convolution as an inception-like unit, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), 10886–10895. https://doi.org/10.1109/CVPR46437.2021.01074

37. L. Tang, J. Yuan, H. Zhang, X. Jiang, J. Ma, PIAFusion: A progressive infrared and visible image fusion network based on illumination aware, *Inf. Fusion*, **83** (2022), 79–92. https://doi.org/10.1016/j.inffus.2022.03.007

38. A. Toet, TNO image fusion dataset, 2014. Available from: https://figshare.com/articles/dataset/TNO Image Fusion Dataset/1008029.

39. J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, et al., Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 5802–5811. https://doi.org/10.1109/CVPR52688.2022.00571

40. Y. J. Rao, In-fibre Bragg grating sensors, *Meas. Sci. Technol.*, **8** (1997), 355. https://doi.org/10.1088/0957-0233/8/4/002

41. G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, *Electron. Lett.*, **38** (2002), 1. https://doi.org/10.1049/el:20020212

42. Y. Han, Y. Cai, Y. Cao, X. Xu, A new image fusion performance metric based on visual information fidelity, *Inf. Fusion*, **14** (2013), 127–135. https://doi.org/10.1016/j.inffus.2011.08.002

43. V. Aslantas, E. Bendes, A new image quality metric for image fusion: The sum of the correlations of differences, *AEU-Int. J. Electron. Commun.*, **69** (2015), 1890–1896. https://doi.org/10.1016/j.aeue.2015.09.004

44. J. W. Roberts, J. A. V. Aardt, F. B. Ahmed, Assessment of image fusion procedures using entropy, image quality, and multispectral classification, *J. Appl. Remote Sens.*, **2** (2008), 023522. https://doi.org/10.1117/1.2945910

45. C. S. Xydeas, V. Petrovic, Objective image fusion performance measure, *Electron. Lett.*, **36** (2000), 308–309. https://doi.org/10.1049/el:20000267

46. H. Li, X. J. Wu, DenseFuse: A fusion approach to infrared and visible images, *IEEE Trans. Image Process.*, **28** (2018), 2614–2623. https://doi.org/10.1109/TIP.2018.2887342

47. J. Ma, W. Yu, P. Liang, C. Li, J. Jiang, FusionGAN: A generative adversarial network for infrared and visible image fusion, *Inf. Fusion*, **48** (2019), 11–26. https://doi.org/10.1016/j.inffus.2018.09.004

48. H. Xu, J. Ma, J. Jiang, X. Guo, H. Ling, U2Fusion: A unified unsupervised image fusion network, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2020), 502–518. https://doi.org/10.1109/TPAMI.2020.3012548

49. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 801–818. https://doi.org/10.1007/978-3-030-01234-2_49

50. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, et al., The cityscapes dataset for semantic urban scene understanding, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 3213–3223. https://doi.org/10.1109/CVPR.2016.350

AIMS Press