IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 27, 2021, accepted September 8, 2021, date of publication September 10, 2021,
date of current version September 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3111905

MIFFuse: A Multi-Level Feature Fusion Network for Infrared and Visible Images

DEPENG ZHU, WEIDA ZHAN, YICHUN JIANG, XIAOYU XU, AND RENZHONG GUO
National Demonstration Center for Experimental Electrical and Electronic Technology, Changchun University of Science and Technology, Changchun, Jilin 130022, China

Corresponding author: Weida Zhan (zhanweida@cust.edu.cn)

This work was supported by the "13th Five Year Plan" Scientific Research Program of Jilin Provincial Department of Education under Grant JJKH20200783KJ.

ABSTRACT Image fusion operation is beneficial to many applications and is also one of the most common and critical computer vision challenges. The perfect infrared and visible image fusion results should include the important infrared targets while preserving visible textural detail information as much as possible. A novel infrared and visible image fusion framework is proposed for this purpose. In this paper, the proposed fusion network (MIFFuse) is an end-to-end, multi-level-based fusion network for infrared and visible images. The presented approach makes effective use of the intermediate convolution layer's output features to preserve the primary image fusion information. We also build a cat_block to swap information between two paths to gain more sufficient information during the convolution steps. To reduce the model's running time even further, the proposed method that reduces the number of feature channels while maintaining the accuracy of the fusion performance. Extensive experiments on the TNO and CVC-14 image fusion datasets show that our MIFFuse outperforms the other methods in terms of both subjective visual effects and quantitative metrics. Furthermore, MIFFuse is approximately twice as fast as the most recent state-of-the-art methods. Our code and models can be found at https://github.com/depeng6/MIFFuse.

INDEX TERMS End-to-end framework, multi-level features, image fusion, concatenation block.

I. INTRODUCTION

More information about a target could not be obtained from a single sensor. The task of image fusion is to fuse multi-source information from multiple images into one image, which is convenient for people to view and post-process [1]. The uses of image fusion are mainly divided into four categories, including medical image fusion [2], [3], multi-focus image fusion [4]–[6], remote sensing image fusion [7], [8], infrared and visual image fusion [9], [10], which are examples of image fusion. The far more common image fusion scenario is infrared and visible image fusion [11]. In terms of target detection and surveillance camera tracking, infrared and visible image fusion technology is widely used. Thermal radiation released from surfaces is captured in infrared images, which can easily illuminate targets but lack texture information. Visible image, on the other hand, usually provide a lot of structural detail but are influenced by the background and lose goals. As a result, the fusion method's key emphasis shifts to how to efficiently merge complementary information.

Many conventional infrared and visible image fusion approaches were proposed after several years of development, and they can be roughly divided into four classes: multi-scale decomposition-based (MSD) methods [12], saliency-based methods [13], [14], sparse repre-sentation-based methods [15], [16], and hybrid-based methods [17]–[19]. Many scholars are now studying and using these approaches. These traditional image fusion approaches usually have multiple key elements, such as image transform, activity level measurements, fusion rules, and so on [11], [20]. Multi-scale decomposition, sparse representation methods, and non-downsampling methods are all examples of image transformation. The aim of activity level calculation is to collect quantitative data from various sources in order to distribute weights [21]. The weight distribution of each pixel is the core of the fusion rule [22]. Traditional fusion approaches have achieved strong fusion efficiency but finding the right activity level measurements and fusion rules to produce well-fused images remains difficult. In the meantime,

The associate editor coordinating the review of this manuscript and approving it for publication was Ravibabu Mulaveesala.

those modules are all made by hand, and the complexity of implementation and computing cost have become major issues.

Deep learning-based approaches [23], [24] have emerged as the most exciting and appealing path for image fusion in recent years. In the field of image fusion, convolutional neural networks [25] mainly fuse infrared and visible images by obtaining image features and recreating fused images or constructing different deep learning frameworks. For the multi-focus image fusion challenge, Liu *et al.* [24] proposed a convolutional neural network (CNN) [26]. Densefuse [27] uses the manually designed fusion strategy in the fusion of infrared and visible images, which does not have wide applicability. In addition, using the dense block alone in the convolution phase also results in a loss of detail. To recreate the fused image, various fusion techniques have been proposed. The fusion model designs two methods for combining intermediate features and compensation features, according to Jian *et al.* [28]. A general end-to-end image fusion network (IFCNN) was introduced by Zhang *et al.* [29], which is a simple but efficient fusion process. Two convolutional layers combine the fused deep features to create the fused image. Its framework, on the other hand, is too simple to extract powerful deep functions. Ma *et al.* [30] recently proposed a FusionGAN-based infrared and visible image fusion process. The training process of GAN (Generative Adversarial Networks) is unstable, and it is easy to cause the generation effect to be relatively poor [31]. Also, the most advanced approaches, such as FusinGANv2 [32] and DDcGAN [33], have problems preserving image detail. During the process of producing the fusion image, the GAN network generator altered part of the original image detail, resulting in a final generated image with no sense of truth.

To address the above issues, we propose a MIFFuse network for the fusion of infrared and visible images. This network without the need for manually design fusion rules. Then, to preserve more texture information during the convolution process, we use skip connections to migrate the feature map from the front convolutional layer to the back convolutional layer. Finally, as shown in Figure 1, we use cat_block in the two paths to exchange image information, making the image details output by the entire network clearer. The specific comparative experiment is in III-B5. Infrared and visible image characteristics are combined in the exchanged information.

Our paper makes three key contributions:

1) A new end-to-end network for infrared and visible image fusion is proposed, which does not include complex fusion rules or post-processing. Our proposed framework will quickly fuse infrared and visible images.
2) A cat_block is placed between two paths to efficiently collect and exchange infrared and visible image information, allowing for more infrared and visible image features to be preserved for fusion effects.
3) The number of parameters in our network is relatively small in our proposed framework, because we use the 1×1 convolution kernel to adjust the number of convolution output feature maps for each layer. Hence, the proposed MIFFuse can quickly finish the image fusion job.

The rest of this paper is organized as follows. In Section II, we explain the proposed method in depth. In Section III, we describe the datasets used and present quantitative and qualitative experimental findings. Then, in Section IV, we have discussion and followed by a conclusion in Section V.

II. METHODS

A. MULTI-LEVEL FEATURE FUSION NETWORK

By combining multi-source images from different sensors during the image fusion process, the aim is to obtain a more detailed and informative image [38]. Infrared(I_{ir}) and visible(I_{vi}) image fusion's aim is to acquire a fusion image(Y) and quantify the fusion image using the following function.

$$Y = F(I_{ir}, I_{vi}) \tag{1}$$

where $F(.,.)$ is a function that extracts more valuable information from visible and infrared images. For deep learning technology, the relevant parameters of this function can be obtained by training a deep learning network.

Specifically, the trained model extracts a sequence of attributes from two source images independently. The Multi-level Features Fusion Network is composed of *conv_block*, *res_block* and *cat_block* parts. These intermediate features generated by the first *conv_block* are referred to as $f_{m\_i}^{conv1} (i = 1, \ldots, 32)$. $m \in \{ir, vi\}$ represents the input source image. $m = ir$ or $m = vi$ represents infrared image or visible image, respectively. i represents the intermediate feature in the convolution process. *conv*1 represents the first *conv_block* layer feature map. Furthermore, the first *res_block* and *cat_block* layer feature map is defined as $f_{m\_i}^{res\_1} (i = 1, \ldots, 32)$ and $f_i^{cat\_1} (i = 1, \ldots, 32)$. We call these features multi-level features.

In addition, to integrate the characteristics of different levels, $f_i^{cat\_n}$ and $f_{m\_i}^{res\_n}$ can be represented as follows:

$$f_i^{cat\_n} = F_{cat\_block}(f_{ir\_i}^{res\_n}, f_{vi\_i}^{res\_n}) \tag{2}$$

$$f_{m\_i}^{res\_n+1} = F_{res\_block}[Concat(f_i^{cat\_n}, f_{m\_i}^{res\_n})] \tag{3}$$

where F_{cat_block} and F_{res_block} represent the transfer functions for *cat_block* and *res_block*. n denotes the *nth* block.

B. NETWORK ARCHITECTURE

Our framework is modular and can incorporate features information at different levels. We divide the fusion network into two paths because it can extract the feature information of both infrared and visible images at the same time, and the MIFFuse network's design also refers to the pseudo-Siamese network [34]. This two branch network is very suitable for image fusion, because infrared and visible image pairs have separate features at the corresponding pixel positions. In both

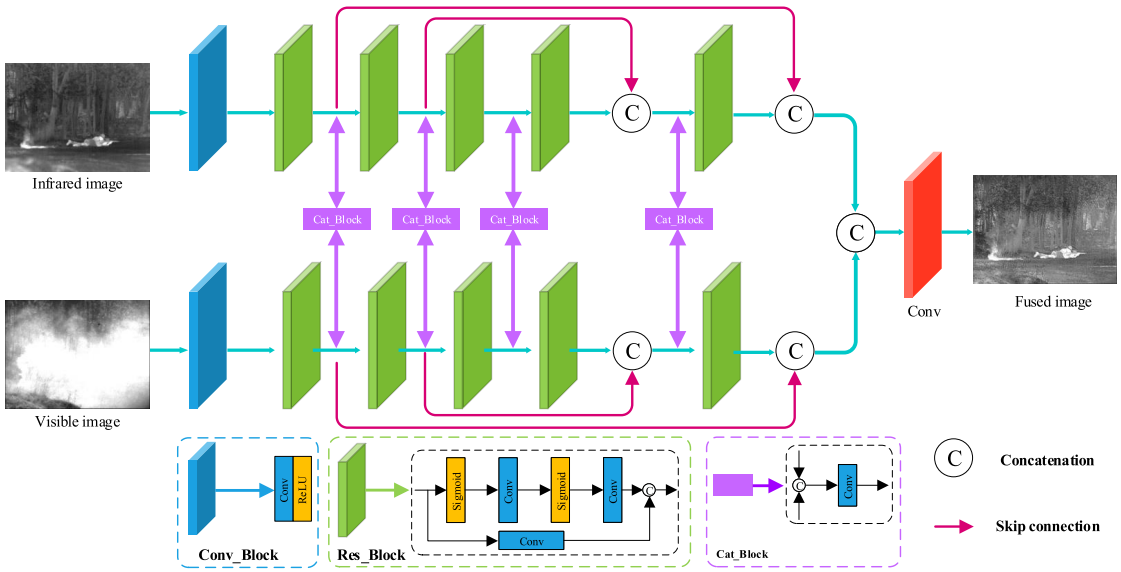

FIGURE 1. The proposed fusion framework (MIFFuse).

path, there are two conv_block and five res_block to extract the features. They all use the 3 × 3 convolution kernel. Figure 1 shows the MIFFuse fusion framework.

1) **Conv_block**. One convolutional layer and an activation function (ReLU) make up our conv_block. The size of the input training data will be anything you want. Convolution operations serve as a feature extractor, keeping all of the edge and texture detail from the infrared and visible images. In order to balance the calculation time of the model and the memory size of the GPU, the number of feature maps output by conv_block is 32, and the size of the feature maps is consistent with the original image.
2) **Res_block**. As shown in Figure 1, there are two main modifications to the res_block. Firstly, discard the BatchNorm layer in the traditional residual_block. Because image fusion is an image-to-image task, the absolute difference of images is very important, especially in image fusion and super-resolution. Secondly, use the Sigmoid as the activation function. The choice of the activation function will be determined by the experimental performance. Finally, we add five res_block for each path, with the aim to ensure the optimal training convergence in the deep network and extract more representative intermediate features.
3) **Cat_block**. We swap information between the two paths to acquire more sufficient information during the convolution step. To be more precise, the exchanged information is generated using the concatenating and convolution methods. The convolution layer's kernel size is 1 × 1. Then concatenated the exchanged information with the output of the previous convolutional layer as the input of the next convolutional layer.
4) **Skip connections**. In an investigation into skip connections [35], we discovered that the deep convolution operation keeps the high-level features of the original image but loses the texture information. Therefore, we use skip connections to transfer the feature information output by the previous convolutional layer to the back. More training details will be discussed in III-B1.

Finally, we merge the outputs of two paths and use a convolutional layer to create the fused image. The convolution layer's kernel size is 3 × 3 and it does not use the activation function.

C. LOSS FUNCTION

As we all know, the design of loss function is particularly important in image processing tasks based on deep learning. The goal of the image fusion task is to calculate an information fusion image from two images. Most image fusion algorithms will use L_2 as a part of the loss function, but we have found through experiments that adding L_2 to the loss function will reduce noise, but it produces visible splotchy artifacts.The specific experimental comparison is in III-B2. The loss function L_1 is effective in eliminating artifacts and enhancing image brightness. In the training process, it is found that the loss function *SSIM* (Structural Similarity) is not sensitive to uniform errors, which will cause the image to darken, and the L_1 loss function can just make up for this defect. The loss function L_1 and L_2 are calculated as

$$L_1 = \frac{1}{MN} \sum_{i \in M, j \in N} |y_{gt(i,j)} - y_{out(i,j)}| \tag{4}$$

$$L_2 = \frac{1}{MN} \sum_{i \in M, j \in N} (y_{gt(i,j)} - y_{out(i,j)})^2 \tag{5}$$

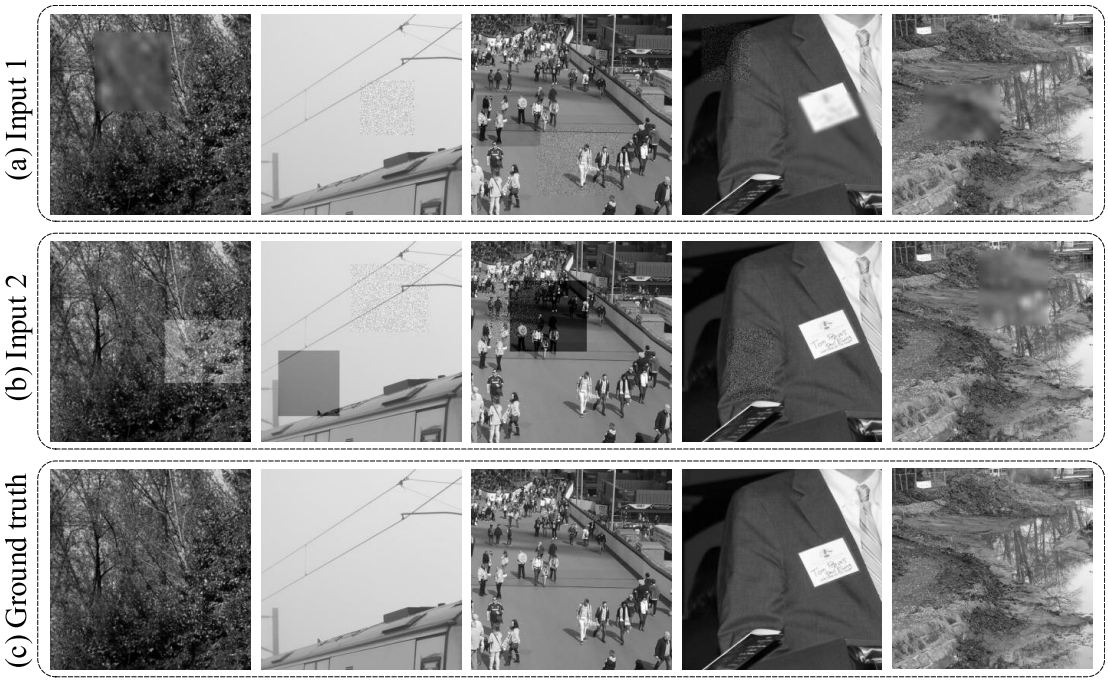

FIGURE 2. Shows the sample images in the Flickr2K datasets. (a) is the input1 image, (b) is the input2 image, (c) is the ground truth image.

where M and N represent the length and height of the image, (i, j) represents the pixel location. y_{gt} and y_{out} refer to the ground truth image and MIFFuse network output image.

As the total loss function, we used the L_1 and $SSIM$. The L_{total} is written as follows:

$$L_{total} = \lambda L_1 + (1 - \lambda) L_{ssim} \tag{6}$$

where L_{total}, L_1 and L_{ssim} represent the total loss, L_1 and $SSIM$, respectively. Two images measure the structural similarity through the SSIM. The λ represents the weight. Furthermore, the loss terms' parameters are set to $\lambda = 0.16$ [36].

The $SSIM$ is obtained by

$$L_{ssim} = 1 - SSIM(y_{gt}, y_{out}) \tag{7}$$

where $SSIM(\cdot)$ represents the structural similarity function [37], the function can be formulated as

$$SSIM(x, y) = [l(x, y)]^{\alpha} \cdot [c(x, y)]^{\beta} \cdot [s(x, y)]^{\gamma} \tag{8}$$

where $l(x, y)$ represents brightness contrast function, $c(x, y)$ represents the contrast comparison function, $s(x, y)$ represents structure comparison function. The α, β, and γ was set as 1. The $l(x, y)$, $c(x, y)$ and $s(x, y)$ are calculated as

$$l(x, y) = \frac{2\mu_{x_i}\mu_{y_i} + C_1}{\mu_{x_i}^2 + \mu_{y_i}^2 + C_1} \tag{9}$$

$$c(x, y) = \frac{2\sigma_{x_i}\sigma_{y_i} + C_2}{\sigma_{x_i}^2 + \sigma_{y_i}^2 + C_2} \tag{10}$$

$$s(x, y) = \frac{\sigma_{x_i y_i} + C_3}{\sigma_{x_i}\sigma_{y_i} + C_3} \tag{11}$$

where μ denotes the mean value, σ represents the standard deviation/covariance, and C_1, C_2, and C_3 are the parameters to make the metric stable.

Our proposed framework uses skip connections, and the loss function is designed to be differentiable. Therefore, we can use the backpropagation algorithm to update the model parameters. In the end, the model parameters are optimized, and the fusion result is the best.

III. EXPERIMENTAL RESULTS

We evaluate MIFFuse on a publicly accessible dataset (TNO [38] and CVC-14 [39]) and compare it to other deep learning-based methods in this section. Those methods including CSR [40], DenseFuse [27], FusionGAN [30], IFCNN [29], SEDRFuse [28]. First, we briefly introduce the experimental setup and implementation details. Then we do an ablation study to see how the connections, loss functions and blocks affect image fusion. Following that, we show quantitative and qualitative results for the two different datasets. In addition, the performance analysis of running speed shows that the proposed method is faster than other methods.

A. EXPERIMENTS

1) TRAINING DETAIL

To successfully repeat the results of the entire paper, we will describe the training details as follows.

First, we need to prepare training dataset. The Flickr2K [41] dataset was used for training in this study. The Flickr2K dataset consist of 2650 high-definition color images

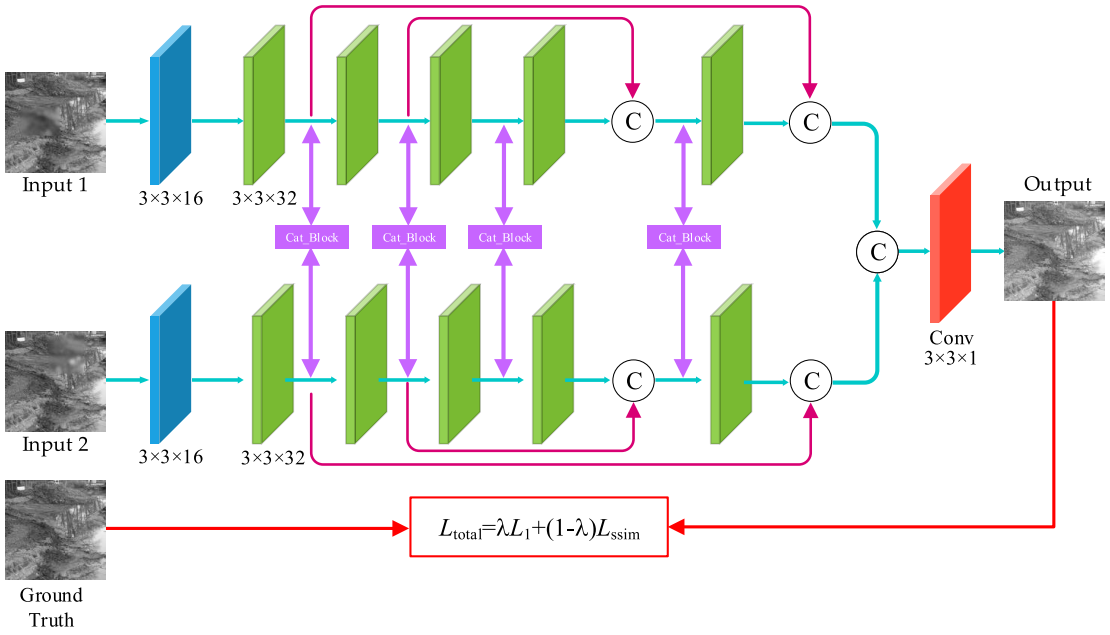

FIGURE 3. The framework of training process.

(2K resolution). To be more precise, we convert the images from RGB to YCbCr color space first. Since structural detail and brightness variance can be represented in the Y channel (luminance channel). Second, we cropped each original image to size 512 × 512, and finally obtained a total of 16335 training images. We add random size and position Gaussian noise and gamma transformation to the cropped image. Because the Gaussian blur of random size and position can simulate the loss of confidence in the camera system's signal acquisition process. The maximum Gaussian blur area is 300 × 300. The value range of gamma in the gamma transformation is $\gamma \in [0.3, 1.3]$. Third, we adopted a supervised method to train the MIFFuse network. The original images are used as the ground truth, we make a group of each original image and the corresponding two processed images. The MIFFuse network is trained using a supervised method. The sample images are seen in Figure 2.

The detailed framework throughout the training process is shown in the Figure 3. This supervised training strategy is an innovation in the field of fusion, and the experimental results prove that the method does have certain advantages.

All the experiments in this study were carried out on a desktop with two NVIDIA GTX Titan XP GPUs and an Intel i7-7820X CPU. Pytorch was used to program the network framework.

A total of 60 epochs were used in the training. The learning rate was set to 1×10^{-4}. AdamOptimizer updates the parameters in our MIFFuse.

Figure 4 depicts the training total loss curve. A cumulative loss value is calculated after 100 iterations. 7041 iterations are needed for each epoch. The parameters of the entire network begin to stabilize when the network is trained to the 40th epochs, as shown in the curve.

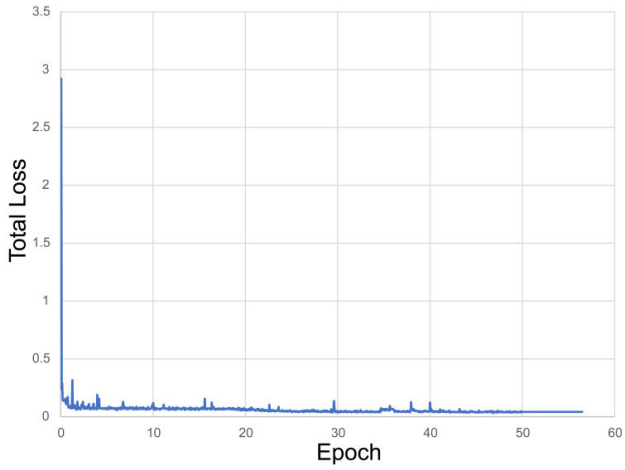

FIGURE 4. The total loss value curve in training stage.

It indicates that the learned model has found its optimum configuration.

2) TESTING DETAILS

In our experiments, the TNO and CVC-14 datasets was used to assess the MIFFuse performance. Twenty-five pairs of images were selected in the two datasets for testing. Figure 5 shows some of the dataset's examples.

3) EVALUATING METRICS

The quantitative and qualitative evaluations are performed on five approaches chosen to understand their performance objectively. More details will be discussed in III-C and III-D. The subjective sensory experience of humans is

FIGURE 5. The examples of infrared and visible images in the TNO and CVC-14 datasets. All images have been registered. (a) and (b) is the infrared and visible images examples of TNO dataset; (c) and (d) is the infrared and visible images examples of CVC-14 dataset.

used in qualitative assessment. A successful fused outcome keeps the source images' sharpness while preserving the data. Quantitative assessment refers to the use of common quality metrics to assess the value of fused images. To calculate the fusion results, eight common statistics are chosed as objective metrics, such as standard deviation(SD) [42], entropy(EN) [43], spatial frequency(SF) [44], edge intensity(EI) [45], contrast(CON) [46], average gradient(AG) [47], structural similarity fusion metric($SSIM_f$) [37], $Q^{AB/F}$ [48].

$$SSIM_f = (SSIM(F, I_{ir}) + SSIM(F, I_{vi}))/2 \quad (12)$$
$$Q^{AB/F} = (Q^{AB/F}(F, I_{ir}) + Q^{AB/F}(F, I_{vi}))/2 \quad (13)$$

where $SSIM(\cdot)$ and $Q^{AB/F}$ denote the SSIM function, $Q^{AB/F}$ function, respectively. I_{ir} and I_{vi} are infrared and visible images. F is the fused image, which serves as reference image.

$Q^{AB/F}$: this metric measures the amount of edge information that is transferred from source images to the fused image. $Q^{AB/F}$ is defined as:

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(w^A(i,j) + w^B(i,j)\right)} \quad (14)$$

where $Q^{XF}(i,j) = Q_g^{XF}(i,j)Q_a^{XF}(i,j)$, $Q_g^{XF}(i,j)$ and $Q_a^{XF}(i,j)$ indicate the edge strength and orientation values at location (i,j), respectively. w^X is the weight that expresses the importance of each source image to the fused image. A large $Q^{AB/F}$ means that considerable edge information is transferred to the fused image.

B. ABLATION STUDY

1) EFFECT OF SKIP CONNECTIONS

The impact of with and without skip connections is discussed in this section. In Figure 6, We randomly selected two sets of experimental results. The clarity of the fused effects is insufficient where there are no skip connections, as in the case of window and soldier (red box) in Figure 6. With skip connections, the infrared image's contour detail of windows and humans will be clearly preserved, and the show effect will be enhanced. Without skip connections, this effect cannot be achieved. In table 1, We show the results of quantitative experiments with or without skip connections, and the results show that fusion with skip connections has a general superiority. In the proposed framework, skip connection can fuse shallow features with deep features to obtain a result with more detailed information.

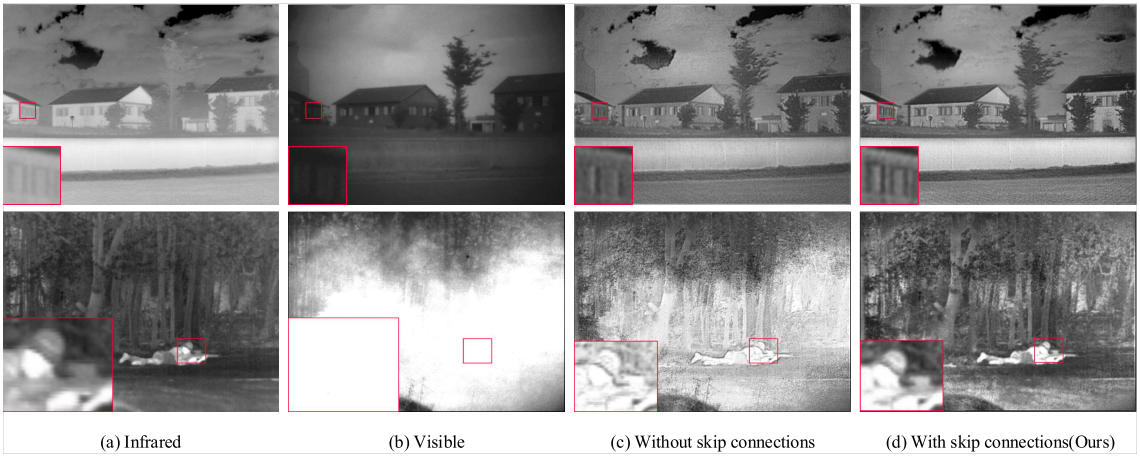

FIGURE 6. Ablation experiment of skip connections. The overall sharpness and defogging ability of images with skip connections are the best. (a) infrared image; (b) visible image; (c) without skip connections; (d) with skip connections.

TABLE 1. Average evaluation metric values of with and without skip connections. The best values in each metric are denoted in bold.

Fusion strategies	SF	EI	CON	EN	AG	SD	SSIM	$Q^{[AB/F]}$
Without skip connection	6.4838	54.0815	73.6979	**6.9391**	5.2219	**37.1849**	**1.2943**	**0.3751**
With skip connection (Ours)	**6.6187**	**58.2481**	**75.0003**	6.8072	**5.5286**	32.4702	1.2658	0.3656

2) USAGE OF LOSS FUNCTIONS

In our method, the fusion results of $SSIM + L_1$ and $SSIM + L_2$ were investigated as loss functions. Fusion results of two scenes are shown in Figure 7. The observations of the texture of the ground and window areas (red box) show that the $SSIM + L_1$ loss function preserved more texture information than $SSIM + L_2$. At the same time, in Table 2, the EI and CON of $SSIM + L_1$ are much higher than those of $SSIM + L_2$, which also proves the superiority of $SSIM + L_1$. As a result, the $SSIM + L_1$ is better for image fusion.

3) ACTIVATION FUNCTION FOR RESIDUAL BLOCK

For residual block, the ReLU (Rectified Linear Unit) and Sigmoid activation functions were investigated. Figure 8 shows the effects of fusing two image pairs with two different activation functions. According to the visual assessment, the Sigmoid activation function does not cause halos and lose details than the ReLU. Particularly, the roof and human clothes (red box) fused with Sigmoid activation function are clearer than ReLU. Then, we use the TNO dataset and the chosen fusion metrics to evaluate the two activation functions. Table 3 shows that most of the fusion metrics have reached the best amount when using the Sigmoid activation function.

4) FUSION STRATEGY FOR SKIP CONNECTION

For skip connection feature fusion, addition, choose-max, and concatenation were investigated. Figure 9 shows the effects of fusing two image pairs using three different strategies. The observation of streetlamp and display window (red box) showed that the concatenation fusion strategy makes the target clearer than those with other two strategies. Besides, to evaluate concatenation fusion strategy more comprehensively, we also provided comparative experiments about three fusion strategies, as show in Table 4. In terms of fusion metrics, the concatenation fusion approach yields the best performance. As a result, the proposed framework employs the concatenation fusion strategy.

5) EFFECT OF CAT_CLOCK

The impact of with and without cat_block is discussed in the section. In Figure 10, we have compared the area in the two sets of images. In the first group of comparisons, it is clear that the branches at the eaves are clearer (with cat_block) than another set of images (without cat_block). In the second set of comparison images, with cat_block, the texture of the infrared and visible images is preserved in the fusion result, and the clarity of the image is improved, especially in the red box area. Without cat_block, the whole image is blurry, and the details of the house in the distance are seriously lost. In table 5, we show the results of quantitative experiments with or without cat_block, and the results show that fusion with cat_block has a general superiority. In the proposed framework, cat_block can fuse shallow features with deep features to obtain a result with more detailed information. The cat_block can fully integrate the characteristic information in the infrared and visible light images, and the final fused image has rich details and sharp contours.

C. RESULTS ON THE TNO DATASET

1) QUALITATIVE EXPERIMENTS

On the fusion results of each method, qualitative and quantitative analysis were performed to validate the proposed

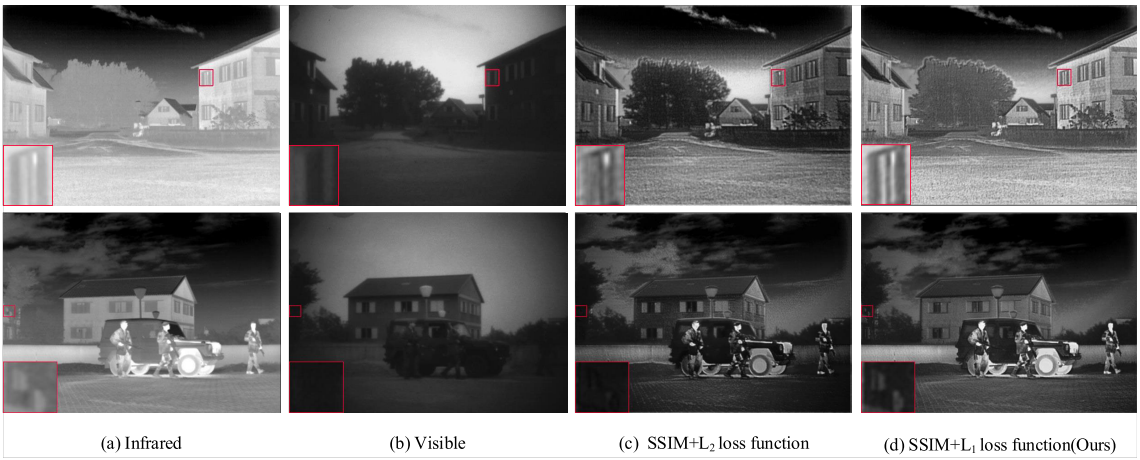

FIGURE 7. Ablation experiment of loss functions. The ability of the $SSIM + L_1$ loss function to maintain texture and dark light details is the best. (a) infrared image; (b) visible image; (c) the result of $SSIM + L_2$ loss function; (d) the result of $SSIM + L_1$ loss function.

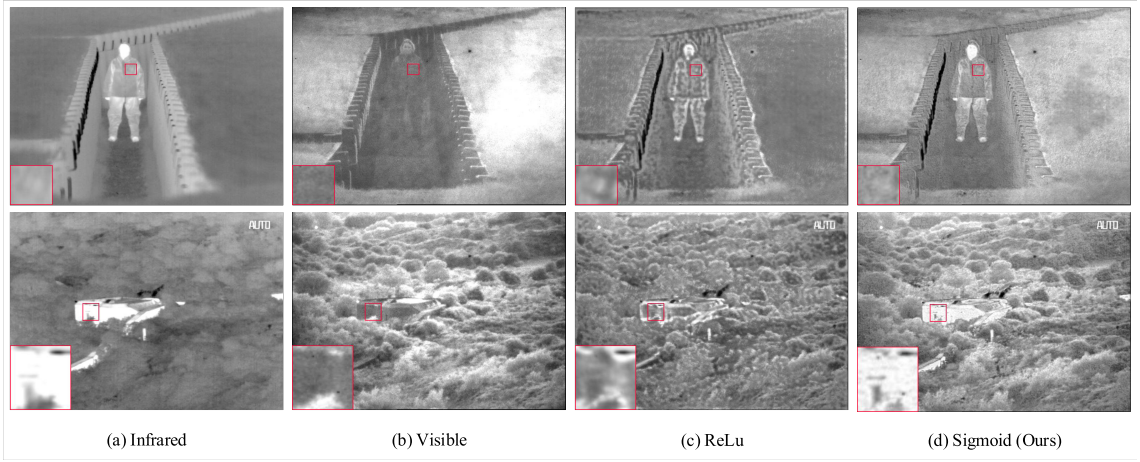

FIGURE 8. Ablation experiment of activation functions. The Sigmoid activation function will not cause a large area of halo and cause loss of image information. (a) infrared image; (b) visible image; (c) the result of ReLu activation function; (d) the result of Sigmoid activation function.

TABLE 2. Average evaluation metric values of with different loss functions. The best values in each metric are denoted in bold.

Fusion strategies	SF	EI	CON	EN	AG	SD	SSIM	$Q^{[AB/F]}$
$SSIM + L_2$	6.3811	50.0391	71.6551	**6.8356**	4.8801	**34.3849**	**1.3137**	0.3647
$SSIM + L_1(Ours)$	**6.6187**	**58.2481**	**75.0003**	6.8072	**5.5286**	32.4702	1.2658	**0.3656**

TABLE 3. Average evaluation metric values of with different activation functions. The best values in each metric are denoted in bold.

Fusion strategies	SF	EI	CON	EN	AG	SD	SSIM	$Q^{[AB/F]}$
ReLU	4.2157	33.7274	32.4461	**6.6455**	3.1495	**31.0051**	**1.3252**	0.3351
Sigmoid (Ours)	**6.6187**	**58.2481**	**75.0003**	6.8072	**5.5286**	32.4702	1.2658	**0.3656**

MIFfuse's good efficiency. In Figure 11, shows the fusion results of several different methods. Figure 11 (a1) and (a2) are the infrared images, Figure 11 (b1) and (b2) are the visible image. The effects of the six fusion methods (CSR, DenseFuse, Fusion-GAN, IFCNN, SEDRFuse, and the proposed method) are seen in subfigures (c) to (j).

As a consequence, we can see that our MIFFuse has distinct benefits over other approaches. First and foremost, our approach is capable of properly preserving the details of source images, including brightness information and contrast information (see Figure 11 (h1) and 11 (h2)). However, fusion results by the Fusion_GAN cannot reflect the

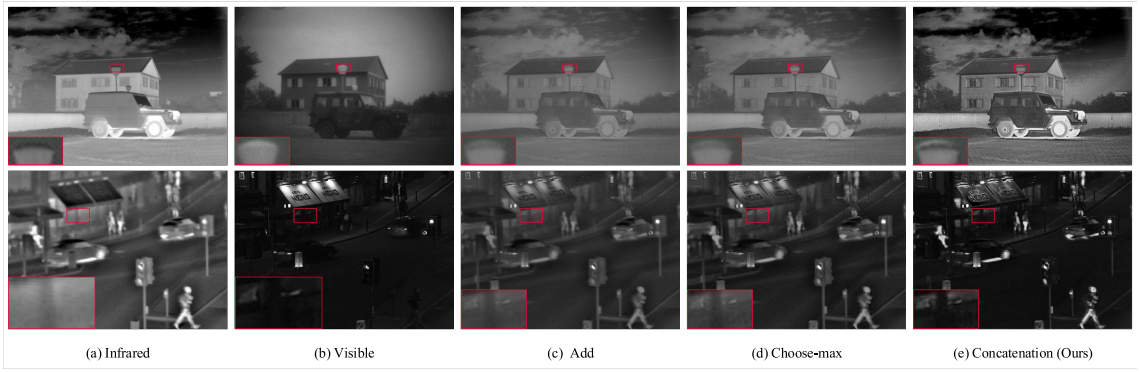

FIGURE 9. Ablation experiment of different fusion strategies for skip connection features. The fusion effect using the concatenation strategy has higher contrast and clarity. (a) infrared image; (b) visible image; (c) the result of add strategy; (d) the result of choose-max strategy;(e) the result of concatenation strategy.

TABLE 4. Average evaluation metric values of with different fusion strategies. The best values in each metric are denoted in bold.

Fusion strategies	SF	EI	CON	EN	AG	SD	SSIM	$Q^{[AB/F]}$
Add	3.1351	21.4274	13.5653	6.3183	2.0729	23.8741	**1.5573**	0.3445
Choose-max	3.8318	27.2236	22.2301	6.5576	2.6623	28.7663	1.5427	**0.4165**
Concatenation (Ours)	**6.6187**	**58.2481**	**75.0003**	**6.8072**	**5.5286**	**32.4702**	1.2658	0.3656

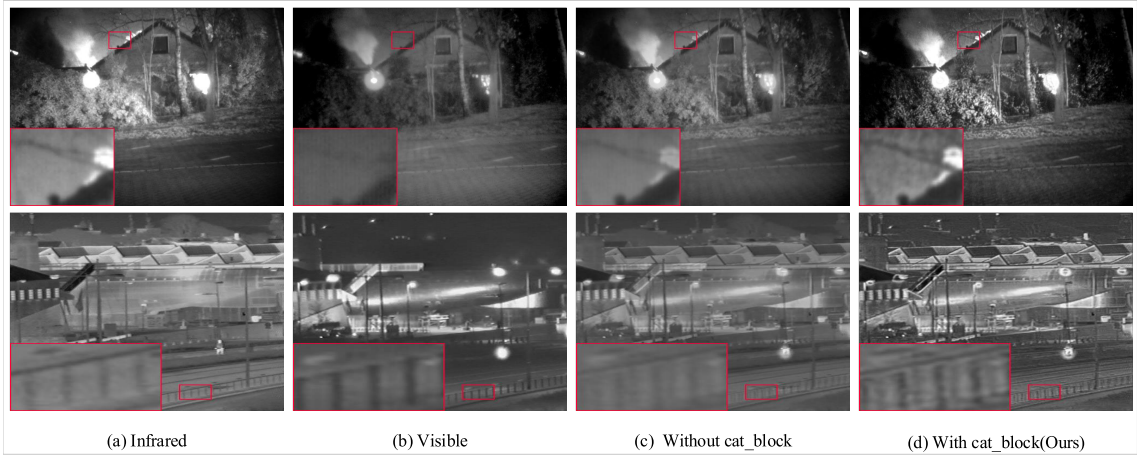

FIGURE 10. Ablation experiment of cat_block. The overall sharpness and defogging ability of images with cat_block is the best. (a) infrared image; (b) visible image; (c) without cat_block; (d) with cat_block.

TABLE 5. Average evaluation metric values of with and without cat_block. The best values in each metric are denoted in bold.

Fusion strategies	SF	EI	CON	EN	AG	SD	SSIM	$Q^{[AB/F]}$
Without cat_block	3.0575	21.2066	14.3239	6.3274	2.0534	24.2212	**1.5522**	0.3376
With cat_block	**6.6187**	**58.2481**	**75.0003**	**6.8072**	**5.5286**	**32.4702**	1.2658	**0.3656**

detail of source images clearly. In Figure 11 (e1), the structure of the house and the car edge, for example, are difficult to discern. Figure 11 (e2) depicts a similar result, which loses a lot of edge information of leaf. In addition, the CSR, DenseFuse and IFCNN methods have low contrast (see Figure 11 (c1), (d1) and (f1)) and the traffic sign is not obvious (see Figure 11 (c2), (d2) and (f2)). The fused results obtained by SEDRFuse method can achieve good performance to some extent. But it still shorts on the target clarity. For example, there is a lack of clarity in the position of windows and vehicles (see Figure 11 (g1) and (g2)).

2) QUANTITATIVE EXPERIMENTS

We conduct fusion on twenty-five pairs of source images of various scenes to quantitatively assess the performance of different fusion methods, and we list the average scores belonging to the eight metrics in Table 6. Values in boldface represent the best results. We can observe from Table 6,

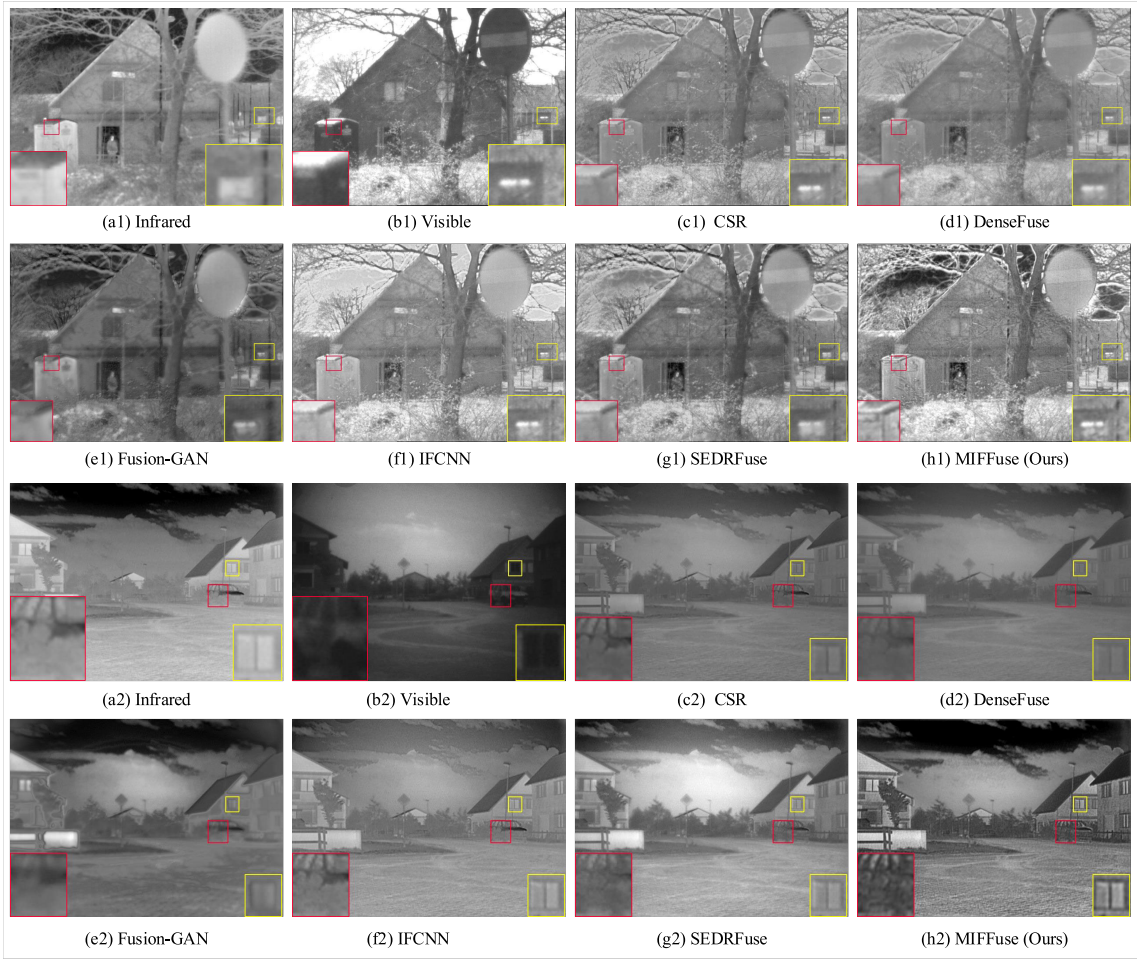

FIGURE 11. Fused results on the TNO dataset. (a) infrared image; (b) visible image; (c) CSR; (d) DenseFuse; (e) Fusion-GAN; (e) Fusion-GAN; (e) Fusion-GAN; (f) IFCNN; (g) SEDRFuse; (h) Our MIFFuse.

TABLE 6. The average values of eight metrics for 25 fused images. The best values in each metric are denoted in bold.

Methods	SF	EI	CON	EN	SD	AG	SSIM	$Q^{[AB/F]}$
CSR	4.1296	28.1597	34.1713	6.3846	25.2134	2.7586	1.5192	**0.5252**
DenseFuse	3.2661	21.2677	15.3294	6.3214	24.0391	2.0987	**1.5607**	0.3491
Fusion_GAN	3.3075	21.0156	14.0007	6.5561	29.5891	2.0468	1.3469	0.2303
IFCNN	5.5126	38.0881	54.3882	6.7143	32.7975	3.8218	1.4766	0.5129
SEDRFuse	5.2556	35.5028	41.8893	**6.9564**	**39.9277**	3.5073	1.4424	0.4797
MIFFuse(our)	**6.6187**	**58.2481**	**75.0003**	6.8072	32.4702	**5.5286**	1.2658	0.3656

it is obvious that the MIFFuse has four best average values (SF, EI, CON, AG), while the EN and SD measures are top two overall. The fused results have the highest SF and AG, indicating that they have a lot of structural details and a clear edge. The results of our method have clear textures and rich detail information, as evidenced by the largest EI and CON.

Nevertheless, the SSIM and $Q^{AB/F}$ need the source image as the reference image for calculation. Hence, the closer the fused image is compared to the infrared and visible image, the greater their value of SSIM and $Q^{AB/F}$. The image fused results need to retain the detail information and enhance the clarity of the target. This will lead to a large deviation between the fused result and the source image, so the above four metric values will be relatively low.

Our method has the best SF, EI, CON, and AG values, indicating that it has obvious advantages in terms of edge preservation of image content and infrared target clarity. The MIFFuse approach is better suited to target detection in foggy or snowy conditions.

D. RESULTS ON THE CVC-14 DATASET

1) QUALITATIVE EXPERIMENTS

Figure 12 is the fusion results of two couples of the CVC-14 dataset. Our method can clearly retain contour and

FIGURE 12. Fused results on the CVC-14 dataset. (a) infrared image; (b) visible image; (c) CSR; (d) DenseFuse; (e) Fusion-GAN; (e) Fusion-GAN; (e) Fusion-GAN; (f) IFCNN; (g) SEDRFuse; (h) Our MIFFuse.

TABLE 7. The average values of eight metrics for 25 fused images. The best values in each metric are denoted in bold.

Methods	SF	EI	CON	EN	SD	AG	SSIM	$Q^{[AB/F]}$
CSR	4.7067	26.8429	36.9844	6.5217	28.3533	2.8352	1.4795	**0.4759**
DenseFuse	4.0919	21.5795	20.4372	6.4667	27.3216	2.2988	**1.5349**	0.3327
Fusion_GAN	3.4802	17.9909	15.5682	6.0266	22.8447	1.9106	1.1075	0.0913
IFCNN	**6.8408**	39.9967	72.4956	6.9597	36.8734	4.3472	1.4132	0.4484
SEDRFuse	4.3638	27.1132	28.6216	6.8271	35.7883	2.6898	1.4535	0.4309
MIFFuse(our)	6.6475	**57.5336**	**82.0031**	**7.0143**	**37.7771**	**5.5116**	1.2636	0.3485

TABLE 8. Comparison results for different fusion framework in running time. The best values in each metric are denoted in bold.

Methods	CSR	DenseFuse	Fusion-GAN	IFCNN	SEDRFuse	MIFFuse(our)
Time(seconds per image pair)	112.3072	2.3331	1.6379	0.7842	0.9471	**0.3575**

detail information, as shown in Figure 12, while DenseFuse, Fusion_GAN and SEDRFuse cannot. Figure 12 (h1) shows partial enlargements of fusion results, and it is obvious that in contrast methods, ''the car edge'' is blurred, while the proposed fusion method has a relatively clear marginal structure. Furthermore, the results obtained using the six methods vary marginally in contrast and details. Our method yields an image Figure 12 (h2) with more texture details in the ''person'' area than the others. it can be found that the silhouette of a person is relatively fuzzy.

Additionally, the proposed method has advantages in enhancing the detail texture. while CSR, Fusion_GAN and SEDRFuse cannot. For example, in the experimental results of other methods, Figure 12 (c2)-(g2), it can be found that the outline of a person is relatively fuzzy, while is sharpened in the MIFFuse result. As a result, we may conclude that our

FIGURE 13. Fused results for infrared and visible (RGB) images. (a) Infrared image; (b) Visible image; (c) Our MIFFuse.

proposed approach outperforms the other five comparative algorithms in terms of fusion performance.

2) QUANTITATIVE EXPERIMENTS

Twenty-five pairs of images were selected in the CVC-14 dataset to verify the robustness of our method, and the results are shown in Table 7. The best results are indicated in bold. Our proposed MIFFuse method also earns the best score in five fusion metrics, including EI, CON, EN, SD, and AG, according to objective evaluation. For the metrics EI and CON, our method reaches the maximum value, and the image obtained by fusion contains rich information and contrast. The MIFFuse result has the highest value for EN and AG, indicating that the suggested method's results contain a lot of detail information and sharp edges. As a result, the proposed MIFFuse method outperforms all other approaches in terms of quantitative assessment.

E. RGB AND INFRARED IMAGE FUSION

We use the proposed method to test the fusion effect of RGB and infrared images. The images for the input are provided from [49]. The dataset includes 221 RGB visible and infrared

image pairs, which has been entirely registered. Figure 13 shows the fused results for RGB visible and infrared images. First, we convert the RGB image to YCbCr format and extract the Y channel separately. Then, we input the Y channel and the infrared image into the MIFFuse network to obtain the fusion result. Finally, we merge the fusion result with the Cb and Cr channels and get the fused color image.

F. RUNTIME COMPARSION

The averaged running time of various fusion methods is compared in Table 7. As can be shown, our method has the highest running performance, running almost twice as quickly as the other methods. In summary, our proposed fusion method can make faster and more effective with infrared and visible image fusion.

IV. DISCUSSION

Consider both the quantitative and the qualitative results, we can make some overall comments for the investigated methods. Firstly, when it comes to deep learning-based image fusion approaches, most methods are unsupervised because there is no ground truth. Secondly, there is no unified dataset in infrared and visible image fusion scopes to compare the performance of various algorithms. As a result, different images are utilized in research, making it difficult to compare the advantages and disadvantages of various algorithms and determine the future research direction. Third, there are many evaluation metrics for images, but none of them are specifically designed for the field of image fusion. This makes quantitative performances comparisons difficult. However, it is difficult to align the two band images, because the focal length and resolution of the visible light camera and the infrared camera are different. Therefore, the registration algorithm [50] of infrared and visible images hinders the development of this field to some extent.

In comparison to previous state-of-the-art methods, our proposed method incorporates several innovative features that could help to improve the quality of infrared and visible image fusion. To begin with, our approach is an end-to-end framework that speeds up the fusion. In contrast, reference [27], [28] proposed a two-stage deep learning method for fusing infrared and visible images. Furthermore, our proposed method can complete the fusion of RGB and infrared images compared with other fusion algorithms [28], [30], [40].

When it comes to image fusion, it is worth noting that the evaluation metrics aren't all the same. As a result, it is difficult for researchers to uniformly evaluate the advantages and disadvantages of the existing state-of-the-art methods. As we have already mentioned, MIFFuse has achieved four firsts (SF, EI, CON, AG) among the six non-reference metrics (SF, EI, CON, EN, SD, AG), which also shows the superiority of our method. Of course, another important research direction of image fusion is to accelerate the speed of image fusion while ensuring the quality of fusion, this will reserve more operating space for target detection and recognition. Although our method runs more than twice as fast as the other methods, it has not yet reached the level of real-time processing.

V. CONCLUSION

In this study we tackle the challenging problem of image fusion by employing a multi-level feature fusion framework based on CNN and residual block. The MIFFuse framework is an end-to-end framework while it does not require any preprocessing steps. This means that the user can omit the process of designing fusion strategies, something which usually requires much time and attention. The experiments on two datasets revealed the potentials of our method, these experiments indicated that the proposed fusion method could achieve high quality, preserve a large amount of complementary information, and sharp edge features while avoiding artifacts and ambiguities. Additionally, the proposed method provided fast inference times, two times faster than existing state-of-the-art methods. In the future, we plan to reduce the running time of our method and eventually achieve real-time computing performance. Finally, we intend to deploy the optimized model to an embedded computing platform, e.g., the NVIDIA Jetson AGX Xavier.[1]

REFERENCES

[1] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253517307972
[2] X. Huang, B. Zhang, X. Zhang, M. Tang, Q. Miao, T. Li, and G. Jia, "Application of U-Net based multiparameter magnetic resonance image fusion in the diagnosis of prostate cancer," *IEEE Access*, vol. 9, pp. 33756–33768, 2021.
[3] B. Li, H. Peng, X. Luo, J. Wang, X. Song, M. J. Pérez-Jiménez, and A. Riscos-Núñez, "Medical image fusion method based on coupled neural P systems in nonsubsampled shearlet transform domain," *Int. J. Neural Syst.*, vol. 31, no. 1, Jan. 2021, Art. no. 2050050.
[4] H. Zhang, Z. Le, Z. Shao, H. Xu, and J. Ma, "MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion," *Inf. Fusion*, vol. 66, pp. 40–53, Feb. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253520303572
[5] B. Wei, X. Feng, and W. Wang, "3M: A multi-scale and multi-directional method for multi-focus image fusion," *IEEE Access*, vol. 9, pp. 48531–48543, 2021.
[6] Y. Liu, X. Chen, H. Peng, and Z. F. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253516302081
[7] P. Hartzell, C. Glennie, and S. Khan, "Terrestrial hyperspectral image shadow restoration through lidar fusion," *Remote Sens.*, vol. 9, no. 5, p. 421, Apr. 2017. [Online]. Available: https://www.mdpi.com/2072-4292/9/5/421
[8] H. Li, W. Ding, X. Cao, and C. Liu, "Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing," *Remote Sens.*, vol. 9, no. 5, p. 441, May 2017. [Online]. Available: https://www.mdpi.com/2072-4292/9/5/441
[9] F. Palsson, J. Sveinsson, and M. Ulfarsson, "Sentinel-2 image fusion using a deep residual network," *Remote Sens.*, vol. 10, no. 8, p. 1290, Aug. 2018.
[10] Y. Liu, L. Dong, Y. Chen, and W. Xu, "An efficient method for infrared and visual images fusion based on visual attention technique," *Remote Sens.*, vol. 12, no. 5, p. 781, Feb. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/5/781

[1] https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-agx-xavier/

[11] D. Xu, Y. Wang, S. Xu, K. Zhu, N. Zhang, and X. Zhang, "Infrared and visible image fusion with a generative adversarial network and a residual network," *Appl. Sci.*, vol. 10, no. 2, p. 554, Jan. 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/2/554

[12] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[13] C. Liu and W. Ding, "Variational model for infrared and visible light image fusion with saliency preservation," *J. Electron. Imag.*, vol. 28, no. 2, p. 1, Mar. 2019.

[14] J. Zhao, Q. Zhou, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition," *Infr. Phys. Technol.*, vol. 56, pp. 93–99, Jan. 2013.

[15] B. Yang and S. Li, "Visual attention guided image fusion with sparse representation," *Optik*, vol. 125, no. 17, pp. 4881–4888, Sep. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030402614004847

[16] J. Wang, J. Peng, X. Feng, G. He, and J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infr. Phys., Technol.*, vol. 67, pp. 477–489, Nov. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449514001984

[17] G. He, S. Xing, X. He, J. Wang, and J. Fan, "Image fusion method based on simultaneous sparse representation with non-subsampled contourlet transform," *IET Comput. Vis.*, vol. 13, no. 2, pp. 240–248, 2019.

[18] G.-Q. He, Q.-Q. Zhang, J.-Q. Ji, D.-D. Dong, H.-X. Zhang, and J. Wang, "An infrared and visible image fusion method based upon multi-scale and top-hat transforms," *Chin. Phys. B*, vol. 27, no. 11, Nov. 2018, Art. no. 118706.

[19] S. Zhang, F. Huang, H. Zhong, B. Liu, Y. Chen, and Z. Wang, "Multi-modal image fusion via sparse representation and multi-scale anisotropic guided measure," *IEEE Access*, vol. 8, pp. 35638–35649, 2020.

[20] C. Sun, C. Zhang, and N. Xiong, "Infrared and visible image fusion techniques based on deep learning: A review," *Electronics*, vol. 9, no. 12, p. 2162, Dec. 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/12/2162

[21] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jun. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253516300458

[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 *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4724–4732.

[23] H. Li, X. J. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infr. Phys. Technol.*, vol. 102, Mar. 2019, Art. no. 103039. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449519301525

[24] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253516302081

[25] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253517305936

[26] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2705–2710.

[27] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8580578/

[28] L. Jian, X. Yang, Z. Liu, G. Jeon, M. Gao, and D. Chisholm, "SEDRFuse: A symmetric encoder–decoder with residual block network for infrared and visible image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5002215. [Online]. Available: https://ieeexplore.ieee.org/document/9187663/

[29] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020. https://www.sciencedirect.com/science/article/pii/S1566253518305505 and [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1566253518305505

[30] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1566253518301143

[31] H. Li, X.-J. Wu, and J. Kittler, "RFN-nest: An end-to-end residual fusion network for infrared and visible images," *Inf. Fusion*, vol. 73, pp. 72–86, Sep. 2021.

[32] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, and J. Wu, "Infrared and visible image fusion via detail preserving adversarial learning," *Inf. Fusion*, vol. 54, pp. 85–98, Feb. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519300314

[33] 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 Trans. Image Process.*, vol. 29, pp. 4980–4995, 2020.

[34] L. H. Hughes, M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu, "Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 784–788, May 2018.

[35] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," 2016, *arXiv:1603.09056*. [Online]. Available: http://arxiv.org/abs/1603.09056

[36] H. Zhu and T. Kaneko, "Comparison of loss functions for training of deep neural networks in Shogi," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Nov. 2018, pp. 18–23. [Online]. Available: https://ieeexplore.ieee.org/document/8588470/

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[38] A. Toet, "The TNO multiband image data collection," *Data Brief*, vol. 15, pp. 249–251, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340917304699

[39] A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, J. Xu, and A. López, "Pedestrian detection at day/night time with visible and FIR cameras: A comparison," *Sensors*, vol. 16, no. 6, p. 820, Jun. 2016. [Online]. Available: https://www.mdpi.com/1424-8220/16/6/820

[40] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[41] A. Ignatov *et al.*, "PIRM challenge on perceptual image enhancement on smartphones: Report," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2019, pp. 315–333. [Online]. Available: https://arxiv.org/pdf/1810.01641.pdf

[42] Y.-J. Rao, "In-fibre Bragg grating sensors," *Meas. Sci. Technol.*, vol. 8, no. 4, p. 355, 1997.

[43] W. J. Roberts, J. A. A. Van, and F. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, pp. 1–28, 2008. [Online]. Available: https://doi.org/10.1117/1.2945910

[44] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.

[45] B. Rajalingam and R. Priya, "Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis," *Int. J. Eng. Sci. Invention*, vol. 2, pp. 52–60, Jan. 2018.

[46] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 6, pp. 460–473, Jun. 1978.

[47] 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," *Opt. Commun.*, vol. 341, pp. 199–209, Apr. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401814011833

[48] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[49] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A unified densely connected network for image fusion," in *Proc. 34th AAAI Conf. Artif. Intell. (AAAI)*, 2020, pp. 12484–12491.

[50] X. Chen, G. Tian, J. Wu, C. Tang, and K. Li, "Feature-based registration for 3D eddy current pulsed thermography," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6998–7004, Aug. 2019.

DEPENG ZHU was born in Wuwei, Gansu, China, in 1996. He received the B.S. degree in electronic and information engineering from Changchun University of Science and Technology, where he is currently pursuing the Ph.D. degree. His research interests include image fusion, image registration, and object detections.

WEIDA ZHAN is currently a Professor and a Supervisor of Ph.D. candidates with Changchun University of Science and Technology. His research interests include digital image processing, infrared imaging technology, and automatic target recognition technology.

YICHUN JIANG was born in Shaoguan, Guangdong Province. He received the B.S. degree in electrical engineering and its automation from Changchun University of Science and Technology, where he is currently pursuing the Ph.D. degree. His research interests include image super resolution and object detections.

XIAOYU XU was born in Yantai, Shandong, China, in 1997. He received the B.S. degree in automation from Changchun University of Science and Technology, where he is currently pursuing the master's degree. His research interest includes image fusion.

RENZHONG GUO was born in Siping, Jilin, China, in 1998. He received the B.S. degree in electronic and information engineering from Changchun Institute of Technology. He is currently pursuing the master's degree with Changchun University of Science and Technology. His research interests include image fusion and image registration.