

## Research Article

# FLGC-Fusion GAN: An Enhanced Fusion GAN Model by Importing Fully Learnable Group Convolution

C. Yuan , C. Q. Sun , X. Y. Tang , and R. F. Liu 

*School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430023, China*

Correspondence should be addressed to R. F. Liu; [claye.liu@whpu.edu.cn](mailto:claye.liu@whpu.edu.cn)

Received 4 August 2020; Revised 20 August 2020; Accepted 21 August 2020; Published 23 October 2020

Guest Editor: Xiaojie Guo

Copyright © 2020 C. Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of image fusion is to combine the source images of the same scene into a single composite image with more useful information and better visual effects. Fusion GAN has made a breakthrough in this field by proposing to use the generative adversarial network to fuse images. In some cases, considering retain infrared radiation information and gradient information at the same time, the existing fusion methods ignore the image contrast and other elements. To this end, we propose a new end-to-end network structure based on generative adversarial networks (GANs), termed as FLGC-Fusion GAN. In the generator, using the learnable grouping convolution can improve the efficiency of the model and save computing resources. Therefore, we can have a better trade-off between the accuracy and speed of the model. Besides, we take the residual dense block as the basic network building unit and use the perception characteristics of the inactive as content loss characteristics of input, achieving the effect of deep network supervision. Experimental results on two public datasets show that the proposed method performs well in subjective visual performance and objective criteria and has obvious advantages over other current typical methods.

## 1. Introduction

Image fusion is an enhancement technique that aims to combine images obtained by different kinds of sensors to generate a robust or informative image [1]. This research sheds new light on military reconnaissance, remote sensing detection, medical health, computer vision, target recognition, etc. Mainly, multisensor data such as thermal infrared and visible images have been used to enhance the performance in terms of human visual perception, object detection, and target recognition [2]. The infrared images can highlight thermal infrared target area characteristics and is least affected by illumination change and artifact. However, images often lack detailed information and have low image contrast, which rarely affects the heat emitted by objects. The visible image is captured and used to record the spectral information of different reflected objects, which can provide better human visual characteristics. However, the targets in visible images may not be easily observed due to the influence of the external environments, such as nighttime

conditions, disguises, objects hidden in smoke, and cluttered background. Therefore, fusion technology is to combine the advantages of infrared and visible images that have abundant detailed information from the visible image and useful target areas from the infrared image [3].

According to different image fusion processing domains, image fusion can be roughly divided into two categories: the spatial domain and the transform domain. The focus of the fusion method is to extract relevant information from the source image and merge it. To this end, the researchers proposed a variety of feature extraction strategies and fusion rules. They can be divided into seven categories, namely, multiscale transform [4], sparse representation [5], neural network [6], subspace [7], saliency [8], hybrid models [9], and deep learning [10]. In general, the current fusion methods involve three crucial challenges, that is, image transform, activity-level measurement, and fusion rule designing [11]. The three constraints have become increasingly complex, especially for designing fusion rules in an old-fashioned way, which sharply limits the development of the fusion methods.

At present, the existing methods are usually the same significant characteristics to select the source image to be integrated into the fusion image, and the fusion image contains numerous kinds of information. However, the feature of infrared thermal radiation information is pixel intensity, while the feature of visible image texture details is edge and gradient. For these two different situations, the previous method cannot be chosen to express in the same way. In order to overcome this problem, Ma et al. proposed a new method called Fusion GAN [11], which was used for the first time to fuse the information of the infrared image and visible image by generative adversarial networks (GANs). Compared with previous studies, Fusion GAN was proposed in a subversive way. It could avoid the use of manually designed activity-level measurement and fusion rules like the traditional method, and the fusion effect was also greatly improved. On the other hand, for Fusion GAN using a generative adversarial network to image fusion as the pioneer of the field, there are still many things worth learning and innovating. Although this method has made an apparent breakthrough in vision compared with previous methods, many current methods generally reduce the brightness of fused images in the process of retaining the infrared radiation information and the gradient information of visible images. In the case of complex model training, the amount of computation and parameters will be increased, which will only bring high computing and memory costs and reduce the efficiency of computing units (e.g., CPU and GPU). We hope that the proposed method can be increasingly used on low-computational-cost devices. Considering the effect and the importance of including the image information fusion, we proposed a fusion method based on Fusion GAN improvement. This method can save the network effect under the premise that the fusion results retained sufficient image information; the method can also reflect in the final result more vivid image contrast. Figure 1 represents the contrast between thermal targets and the background. To better compare the differences between the methods, we framed a part of the regions in the fusion results of the Fusion GAN and our method to form a more visual comparison. In our results, it has a higher contrast and can better recognize the target. At the same time, more details in the visible image are more explicit (e.g., the lights, the bushes, and the shelves).

The contributions of this paper can be summarized as follows:

- (1) We propose a new end-to-end GAN framework, which can enhance the brightness and contrast of fused images while maintaining the information of infrared thermal radiation and visible gradient
- (2) We replace all the convolutional layers of the generator with learnable group convolution to reduce the high computational and memory costs of deep networks
- (3) We exploit the multilevel residual network as the fundamental network building unit of the generator to make the network capacity larger and easier to train
- (4) We use the inactive perception feature as the feature input of content loss to strongly constrain the overall information of the source image

The rest of the paper is organized as follows. Section 2 introduces the background and related work on GAN and learnable group convolution. Section 3 describes our proposed method. Section 4 validates the superiority of our model, qualitative and quantitative comparisons are made on the corresponding datasets, and the running time of various methods is also compared. Section 5 presents the conclusion of the paper.

## 2. Related Work

In this section, several methods for the fusion of visible and infrared images are briefly introduced along with GANs.

*2.1. Infrared and Visible Image Fusion.* In the past few years, many image fusion methods based on deep learning have emerged. They can be simply divided into seven categories including multiscale transform [12, 13], sparse representation [14], neural network [6, 15], subspace [16], and saliency-based [17] methods, hybrid models [18], and other methods [19]. Multiscale transform-based methods [19], the most actively used for fusion, assume that a source image can decompose original images into components of different scales. A final target fused image can be obtained by fusing its layers based on certain particular fusion rules. The most popular transforms used for decomposition and reconstruction are the wavelet [20], pyramid [4], curvelet [21], and their variants. Sparse representation image fusion methods [22] aim to learn an overcomplete dictionary from a large number of high-quality natural images. It has been found that an image can be represented by a sparse basis linear combination, which is a key factor to ensure the good performance of this method. Neural network-based methods [6] can imitate the perceptual behavior system of the human brain when dealing with neural information, and the neural network has the advantages of strong adaptability and fault tolerance and antinoise capabilities [16]. Subspace-based methods [23, 24] aim to project high-dimensional input images into low-dimensional subspaces. For most natural images, redundant information exists, and low-dimensional subspaces can help capture the intrinsic structures of the original images. Saliency-based methods [25, 26] are based on the fact that the attention is often captured by objects or pixels that are more significant than their neighbors, and saliency-based fusion methods can maintain the integrity of the salient object region and improve the visual quality of the fused image. Hybrid models [27] combine the advantages of the different methods and thus further enhance the image fusion performance.

At present, there are many methods of infrared image fusion and visible image fusion [28–34], which are well adapted to the examples and verified. We analyzed several of these methods. In recent years, Ma et al. proposed the gradient transfer fusion (GTF) method for image fusion [19] to retain the main intensity distribution in infrared images



FIGURE 1: Schematic illustration of image fusion. From left to right: (a) infrared image, (b) visible image, (c) Fusion GAN [11], and (d) FLGC-Fusion GAN.

and gradient changes in visible images. The details of the image are more reflected in the texture, which is the gradient of the image. The visible image has more texture gradient information than the infrared image, so the gradient of the fusion image can be inherited from the gradient of the visible image. In fact, gradient changes do not adequately retain the useful details contained in the visible image. To address this issue, Ma et al. proposed Fusion GAN [11] method which establishes an adversarial game between a generator and a discriminator, where the generator aims to generate a fused image with primary infrared intensities together with additional visible gradients, and the discriminator aims to force the fused image to have more details existing in visible images. This enables the final fused image to simultaneously keep the thermal radiation in an infrared image and the textures in a visible image. However, only relying on confrontation training leads to the loss of detailed information. The content loss of Fusion GAN only attaches importance to the edge information of visible images. It ignores some information on infrared images, which leads to the blurring of the target edge of fusion images. Recently, Ma et al. proposed the latest method [35] to alleviate the problem of antagonistic learning based on retaining details. This method adds two new loss functions: detail loss and target edge-enhancement loss. The fusion image is provided with rich details of visible images and the effective target area of infrared images. Its advantage is to retain some infrared radiation information and gradient information, and its disadvantage is usually to reduce the brightness of image fusion.

**2.2. Generative Adversarial Network.** GAN was first proposed by Goodfellow et al. [36]. The main idea of GAN is to build a minimax two-player game between the learning of a generator and a discriminator. The generator takes the input information through various convolutional layers and attempts to transform this input into a more realistic image sample. Then, we input  $I_f$  with  $I_v$  into discriminator  $D$  to judge which of the samples is from the source data. GAN establishes an adversarial relationship between the discriminator and the generator. The discriminator determines whether the example comes from the model distribution or data distribution. The generator uses the adversarial relationship to generate samples that the discriminator cannot distinguish. The confrontation relationship between  $G$  and  $D$  is given by

$$\min_G \max_D V_{\text{GAN}}(G, D) = \mathbb{E}_{x \sim P_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

To keep the generator and discriminator in proper synchronization during training, until the two reach the Nash equilibrium, at present, many methods are to use the variant of GAN to achieve the purpose. The most widely used option of GAN is the conditional GAN [37], which applies GANs in the limited setting and forces the output to be conditioned on the input. Our method is also based on GAN variants to achieve image fusion.

**2.3. Fully Learnable Group Convolution.** In recent years, many fusion methods based on deep learning have been proposed and applied to many image-related fields. Liu et al. [38], adopted CNN to generate a weight map while the overall process is based on pyramids. In Li et al. [39], source images are decomposed into base parts and detail content, and deep learning [40] is used in the detail content to extract features. In some methods, deep learning is also used for reconstruction. Ram et al. [34] solved multiexposure fusion by utilizing a novel CNN. With the development of CNN, the increasing models are pursuing deeper and larger convolutional neural networks to improve network performance. In Condensenet [41], a learnable group convolution was proposed to select the input channels for each group automatically. However, the filters used for group convolution in each group are predefined and fixed, and this hard assignment hinders its representation capability even with random permutation after group convolution. To deal with all the above limitations once for all, in this work, Wang et al. [42] proposed a fully learnable group convolution (FLGC) method; the grouping structure including the input channels and filters in one group is dynamically optimized. And in this FLGC, the input channels and filters in one group (i.e., the group structure) are both dynamically determined and updated according to the gradient of the overall loss of the network through backpropagation, and it can be optimized in an end-to-end manner [42].

In a deep network, the convolution layer is computed as convolving the input feature maps with filters. Taking the  $K^{\text{th}}$  layer for an example, the input of the  $K^{\text{th}}$  layer can be denoted as  $X^k = \{x_1^k, x_2^k, \dots, x_C^k\}$ , where  $C$  is the number of channels and  $x_i^k$  is the  $i^{\text{th}}$  feature map. The filters of the  $K^{\text{th}}$

layer are denoted as  $W^k = \{w_1^k, w_2^k, \dots, w_N^k\}$ , where  $N$  denotes the number of filters, that is, the number of output channels, and  $w_i^k$  is the  $i^{\text{th}}$  3D convolutional filter. The output of this convolution layer is calculated as follows:

$$X^{k+1} = W^k \otimes X^k = \{w_1^k * X^k, w_2^k * X^k, \dots, w_N^k * X^k\}, \quad (2)$$

where  $\otimes$  in this work denotes the convolution between two sets and  $*$  denotes the convolution operation between a filter and the input feature maps.

Firstly, we formulate the grouping structure in the  $K^{\text{th}}$  layer as two binary selection matrices for input channels and filters, respectively, denoted as  $S^k$  and  $T^k$ . The  $S^k$  is a matrix for channel selection in shape of  $C \times G$ , with each element defined as

$$S^k(i, j) = \begin{cases} 1, & \text{if } x_i^k \in X_j^k, \\ 0, & \text{if } x_i^k \notin X_j^k, \end{cases} \quad i = [1, C]; j \in [1, G], \quad (3)$$

in which  $S^k(i, j) = 1$  means the  $i^{\text{th}}$  input channel is selected into the  $j^{\text{th}}$  group. As can be seen, the  $j^{\text{th}}$  column of  $S^k$  indicates which input channels belong to the  $j^{\text{th}}$  group. Then,  $X_j^k$  can be simply represented as follows:

$$X_j^k = X^k \odot S^k(:, j)^{\mathcal{T}}, \quad j \in [1, G], \quad (4)$$

where  $\odot$  denotes the elementwise selection operator (the element here means  $\forall x_i^k \in X^k$ ) and  $T$  denotes the transpose of a vector. For filter selection, we define a matrix  $T^k$  in shape of  $N \times G$ , with each element defined way similar to  $S^k(i, j)$ . As a result, the structure of group convolution is parameterized by two binary selection matrices  $S^k$  and  $T^k$ . Therefore, this parameterized group convolution can be embedded in any existing deep network with the objective as

$$\min_{W^k, S^k, T^k} \frac{1}{n} \sum_{i=1}^n L(Y_i; \hat{Y}_i | X_i, W^k, S^k, T^k), \quad (5)$$

in which  $X_i$  denotes the  $i^{\text{th}}$  input sample,  $n$  indicates the amount of training data,  $Y_i$  indicates the  $i^{\text{th}}$  sample's true category label,  $K$  is the number of layers, and  $\hat{Y}$  is the label predicted from a network with our group convolution parameterized by  $W^k$ ,  $S^k$ , and  $T^k$ .  $L(\cdot)$  denotes the loss function (e.g., cross-entropy loss) for classification or detection, etc. In the above objective, the filters  $W^k$ , the group structure including  $S^k$  and  $T^k$  can be all automatically optimized according to the overall objective function.

As can be seen from equation (5), the group structure in this method is automatically optimized rather than manually defined. Furthermore, different from those methods only considering the magnitude and impact of the connection in one or two layers, the group structure is determined according to the objective loss of the whole network. Therefore, the group structures of all layers in our method are jointly optimized.

### 3. Method

This section introduces the method proposed in this paper. First, the structural framework of the model is described, and then we will introduce the practical methods of our application in detail in conjunction with the model framework.

**3.1. Problem Formulation.** Considering that the complex model can reduce the amount of calculation during the training process, it can efficiently maintain the thermal radiation information of the infrared image and the rich texture information of the visible image, while also improving the overall brightness and contrast of the fused image. In order to keep the thermal radiation information of the infrared image and rich texture information of the visible image, the overall brightness and contrast of the fused image can be improved. We have made further improvements to our latest fusion strategy. On the whole, GAN is still used to transform the fusion problem of infrared image and visible image into the generation confrontation problem, as shown in Figure 2. First, we connect infrared image  $I_i$  and visible image  $I_v$  to the channel dimensions. During the training phase, two images are stacked on the channel dimensions and then entered into generator  $G$ , a multilevel residual network structure. Under the guidance of the loss function, generator  $G$  can generate the original fused image  $I_f$ . In the fusion problem, there is no exact fusion image in the dataset to provide a reference for the discriminator, so the next step is to input the original fused image  $I_f$  and the original visible image  $I_v$  into the discriminator  $D$ . The structure of the discriminator  $D$  is similar to VGG-Net [43] used to determine which sample came from the sample data source and give the corresponding probability. The confrontation training was carried out between  $G$  and  $D$  until they reached Nash equilibrium, and the discriminator cannot distinguish the fused image from the visible image. In this way, the generator has a strong ability to generate fusion images; the generator and discriminator of the model gradually mature in the whole training stage. The resulting fused image has a more realistic image effect and other features of the fused image (e.g., pixel intensity contrast, saturation, and brightness) are significantly improved. In the test phase, we input the visible image and infrared image concatenate into the trained generator to get the final fusion image.

**3.2. Network Architecture.** The network architecture of this model is mainly realized based on the generation countermeasure network, which is mainly composed of two parts, namely, generator and discriminator. The architecture of each part is based on a learnable grouping convolutional neural network.

**3.2.1. Network Architecture of the Generator.** Our network architecture of the generator  $G_{\theta_G}$  is presented in Figure 3. As shown,  $G_{\theta_G}$  is based on the multilevel residual network design. First, we replace all convolutional layers in the

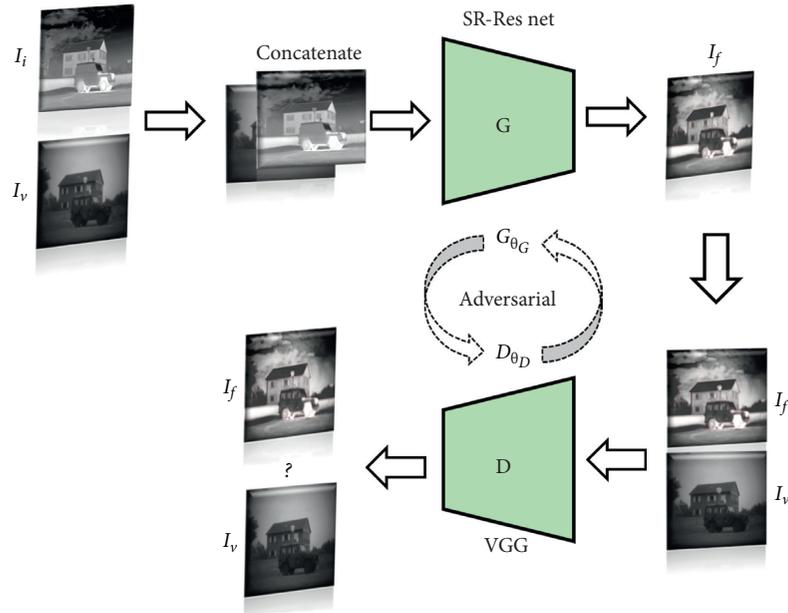


FIGURE 2: General architecture of the infrared and visible image fusion model.

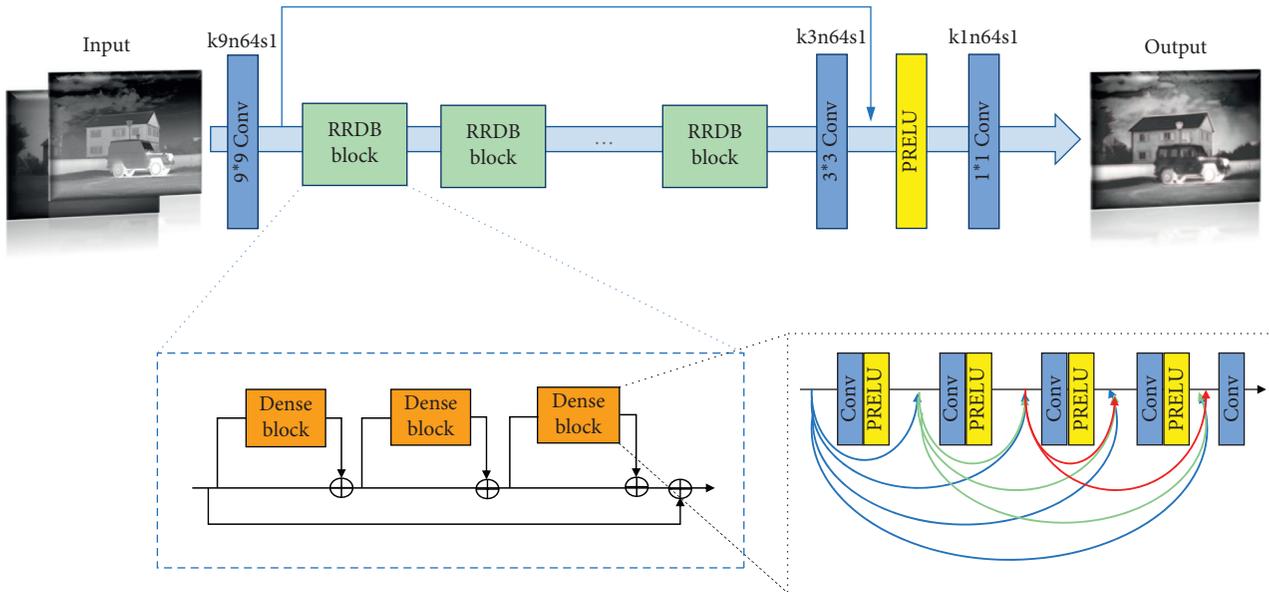


FIGURE 3: Generator network framework: the internal information of each RRDB block and the internal structure of the dense internal blocks included therein. In this structure, the dense block of RRDB consists of the PReLU activation layer and a convolutional layer without the batch normalization layer.

network with learnable packet convolution, as shown in Figure 4(a). They can automatically select input channels for each group, and the filters used in each group of channels are learnable. In actual use, the filters would need to be reordered with an index-Recording layer to align them in a group. Secondly, considering that the input channel is also the output channel of the previous layer, we combine the index of the output channel of the previous layer and the index of the input channel in this layer into a single index to obtain the correct input channel order as shown in Figure 4(b). We deepened the generator's network with

dense connections [44] and multilevel residuals, and the original residuals were replaced by residuals dense blocks (RRDB). Each RRDB block is composed of a 3-layer dense block, and the internal structure of the dense block is connected by a learnable multilayer grouping convolution in the form of multilevel residuals. Additionally, there is no BN layer in the Dense block to prevent the calculation of lag in the deeper network training. Instead, we make the initial parameter variance smaller so that the residual network is more comfortable to train. Activation functions in multilevel residuals in each layer use parameterized modified linear

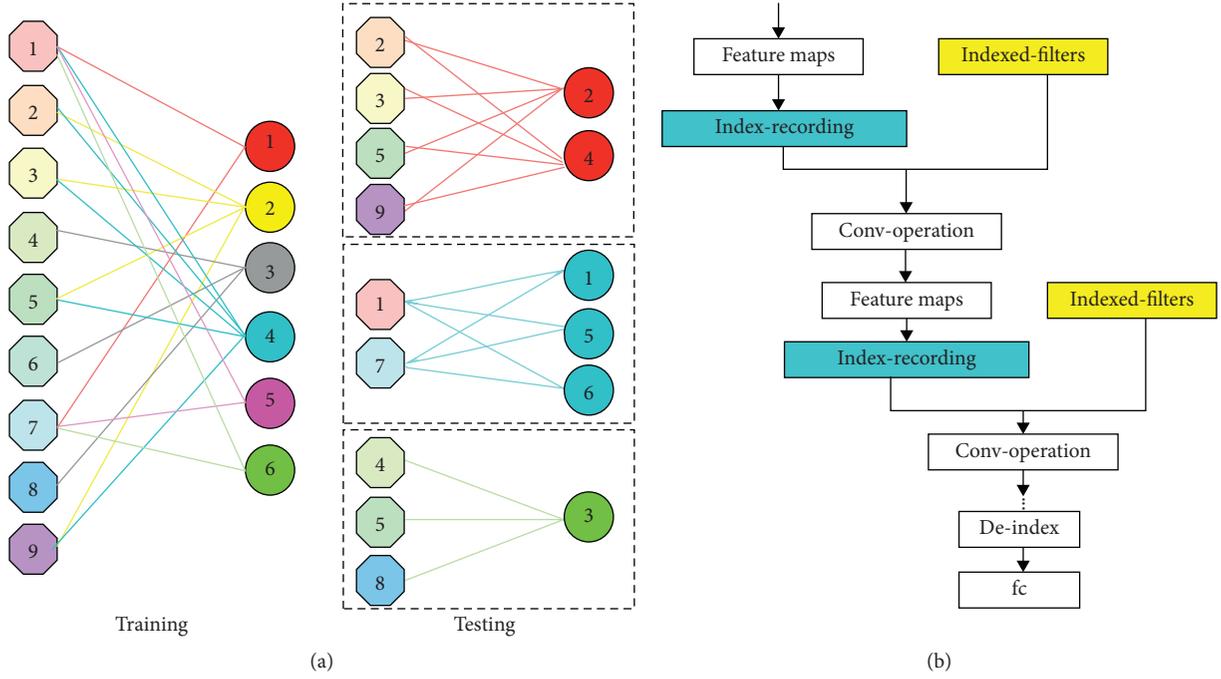


FIGURE 4: A block convolutional layer example used in  $G_{\theta_G}$ . (a) The FLGC convolution, in which grouping structures including input channels and filters are dynamically learned. The octagon represents the input channel, and the circle represents the filter. (b) The structure drawing for the index-recording layer to rearrange the channels in a practical application for grouping convolution.

units (PReLU) rather than the typical ReLU, which can be seen as a variation of leaky ReLU, some of which are common with adding a linear term to negative input. The key difference between the two is that the slope of this linear term is learned in model training. We use the features of the first convolutional layer as the input of each layer. In other words, the features before activation are used as the feature input of content loss, so as to constrain some possible factors; the pixel intensity of thermal radiation information of infrared images and texture information of visible images are strongly monitored. As dense  $1 \times 1$  convolution takes up a lot of computing cost, leaving much space for further acceleration, we use the FLGC layer to replace the convolutional layer  $1 \times 1$  with the number of filters greater than 96 to construct a complete convolutional network. Also, we only need to double the span of the first layer and add an FC layer.

**3.2.2. Network Architecture of the Discriminator.** The design of the discriminator  $D_{\theta_D}$  is based on the VGG network [43], as shown in Figure 5. Five convolution layers and five max-pooling layers are used in VGG11. However, for this network, there is a batch normalization layer behind each convolutional layer, which is used to improve the network training speed and performance. In addition, we use the leaky ReLU activation function to replace the general PReLU in the first five layers to adjust the degree of leakiness during backpropagation. Unlike the VGG network, we add another convolutional layer at the end of the network to reduce the size. The final linear layer is used mainly for classification.

**3.3. Loss Function.** The loss function of our FLGC-Fusion GAN consists of two parts, that is, the loss function of generator  $G_{\theta_G}$  and the loss function of discriminator  $D_{\theta_D}$ . In the following, we will introduce them separately.

The generator loss function of this method is composed of four types and is given by the following equation, including content loss, detail loss, target edge-enhancement loss, and adversarial loss:

$$\text{loss}_{\text{total}} = \underbrace{L_{\text{image}} + \alpha L_{\text{gradient}}}_{\text{content loss}} + \beta L_{\text{detail}} + \delta L_{\text{tee}} + \gamma L_{\text{adversarial}}, \quad (6)$$

where the content loss constrains the fused image to one with similar pixel intensities as those of the infrared image and similar gradient variation as that of the visible image. The detail loss  $L_{\text{detail}}$  and adversarial loss  $L_{\text{adversarial}}$  aim at adding more abundant detailed information to the fused image. The target edge-enhancement loss  $L_{\text{tee}}$  is for sharpening the edges of highlighted objectives in the fusion image. During this process, the accuracy of both models is improved. Then, we use the weight parameters of  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  to trade off among different items in the generator loss.

The pixelwise image loss  $L_{\text{image}}$  is given by the following formula; the image loss in terms of pixel intensity distribution makes the fused image consistent with the infrared image:

$$L_{\text{image}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \left( I_{r(x,y)} - I_{f(x,y)} \right)^2, \quad (7)$$

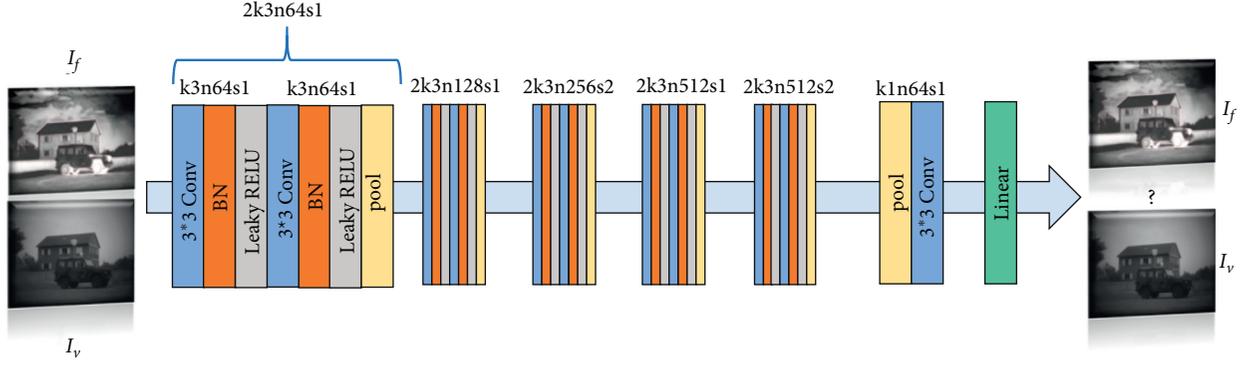


FIGURE 5: Discriminator network framework.

where  $I_r$  is the original infrared image,  $I_f$  is the final output of the generator, and  $W$  and  $H$  denote the width and height of the image.

The gradient loss is given by formula (8). In the fusion problem, we hope that the fusion image can fuse more texture information in the visible image:

$$L_{\text{gradient}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \left( D_{v(x,y)} - D_{f(x,y)} \right)^2, \quad (8)$$

where  $D_{v(x,y)}$  denotes the gradient of the visible image and  $D_{f(x,y)}$  means the gradient of the fused image.

We define the difference of the discriminator feature map between the fused image and the visible image as the detail loss as follows:

$$L_{\text{detail}} = \sum_{i=1}^N \sum_{j=1}^M \left( \phi_{v(i,j)} - \phi_{f(i,j)} \right)^2, \quad (9)$$

where  $\phi$  depicts the feature map obtained by the convolution within the discriminator,  $\phi_v$  and  $\phi_f$  denote the feature representations of the visible and fused images, and  $N$  and  $M$  denote the width and height of the result, which is the input image computed by conventional feature maps.

The target edge-enhancement loss is given by the following formula:

$$L_{\text{tee}} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \left( I_{r(x,y)} - I_{f(x,y)} \right)^2 \cdot G(x, y), \quad (10)$$

in order to make target boundaries more sharpened, a weight map  $G$  is designed to pay more attention to the target boundary area and multiplied to  $L_{\text{image}}$ .

The adversarial loss is adopted for our generator network with a discriminator to generate better-fused images. The adversarial loss is defined based on the probabilities of the discriminator  $\log D_{\theta_D}(G_{\theta_G}(I^{\text{mix}}))$  over all the training samples given by the following formula:

$$L_{\text{adversarial}} = \sum_{n=1}^N \left( 1 - \log D_{\theta_D}(G_{\theta_G}(I^{\text{mix}})) \right), \quad (11)$$

where  $I^{\text{mix}}$  is the stack of the infrared and visible images,  $\log D_{\theta_D}(G_{\theta_G}(I^{\text{mix}}))$  is the probability of the fusion image like a visible image, and  $N$  is the size of the batch.

The discriminators are trained to discriminate between the real data and the generated data. The adversarial losses of discriminators can calculate the JS divergence between distributions to determine whether the pixel intensity or texture information is true. The adversarial losses of the discriminators are defined as follows:

$$L_D = \frac{1}{N} \sum_{n=1}^N \left( D_{\theta_D}(I_v) - b \right)^2 + \frac{1}{N} \sum_{n=1}^N \left( D_{\theta_D}(I_f) - a \right)^2, \quad (12)$$

where  $a$  and  $b$  denote the labels of fused image  $I_f$  and visible image  $I_v$ , respectively, and  $D_{\theta_D}(I_v)$  and  $D_{\theta_D}(I_f)$  denote the classification results of the visible and fused images, respectively.

## 4. Experiments

In this section, we first briefly introduce the fusion metrics used in this paper, then demonstrate the efficacy of the proposed method on public datasets, and compare it with eight state-of-the-art fusion methods including curvelet transform (CVT) [18], dual-tree complex wavelet transform (DTCWT) [45], guided filtering-based fusion (GFF) [4], gradient transfer and total variation minimization (GTF) [19], and Fusion GAN [11]. In Section 4.1, several common image fusion evaluation indexes are introduced. In Section 4.2, the fusion performance of this method is experimentally verified. The first part introduces the dataset and the corresponding training settings since this experiment. The second and third parts, respectively, introduce the experimental results based on the TNO dataset and the RGB-NIR Scene dataset. On the basis of the results of the experiment, qualitative and quantitative performance comparisons are made, respectively.

**4.1. Fusion Metrics.** It is often difficult to judge the fusion performance by only subjective evaluation. Thus, we also use objective fusion indicators for objective evaluation. In this paper, we choose six objective evaluation indexes including spatial frequency (SF) [46], cosine similarity (COSIN) [47], information entropy (EN) [48], structural similarity index measure (SSIM) [49], standard deviation (SD) [50], and the correlation coefficient (CC) [51].

EN is a statistical feature form that reflects the average amount of information in an image. EN is mathematically defined as follows:

$$EN = - \sum_{l=0}^{L-1} p_l \log_2 p_l, \quad (13)$$

where  $L$  represents the image's gray levels and  $p_l$  represents the proportion of pixels with gray value  $l$  in the total pixels. A larger EN means that an enormous amount of information exists in the fused images.

COSIN converts the corresponding image into a vector and calculates the cosine similarity between the vectors, and the mathematical formula of COSIN is defined as follows:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}. \quad (14)$$

This method has a large amount of calculation, while the result is more reliable than SSIM. The larger the COSIN value, the more similar the pixels in the image.

SF is based on the image gradient to reflect the details and texture of the image. First, the horizontal frequency RF and vertical frequency CF are calculated, and then the SF is calculated. The larger the value, the richer the edge texture information. The mathematical formula of SF is defined as follows:

$$RF = \sqrt{\sum_{i=1}^M \sum_{j=2}^N (x_{i,j} - x_{i,j-1})^2}, \quad (15)$$

$$CF = \sqrt{\sum_{i=2}^M \sum_{j=1}^N (x_{i,j} - x_{i-1,j})^2}, \quad (16)$$

$$SF = \sqrt{RF^2 + CF^2}. \quad (17)$$

The SSIM considers image distortion by comparing changes in image structure information, thereby obtaining an objective quality evaluation. The mathematical definition of SSIM is as follows:

$$SSIM(X, Y) = \left( \frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right)^\alpha \cdot \left( \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right)^\beta \cdot \left( \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3} \right)^\gamma, \quad (18)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y), \quad (19)$$

where  $x$  and  $y$  are the reference image and fused image, respectively;  $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ , and  $\sigma_{xy}$  represent the mean value and variance and covariance of images  $x$  and  $y$ , respectively; and  $c_1, c_2$ , and  $c_3$  are small normal numbers to avoid having

a denominator of zero. Parameters  $\alpha, \beta, \delta$ , and  $\gamma$  are used to adjust the proportions.

In SD, formula (20) is an objective evaluation index to measure the richness of image information. The larger the value is, the more information the image carries and the better the fusion image quality is. The mean value measurement formula (21) is an indicator that reflects the brightness information. If the mean value is moderate, the fusion image quality will be better. The mathematical definition of SD is as follows:

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=2}^N [H(i, j) - \bar{H}]^2}, \quad (20)$$

where  $\bar{H}$  represents the mean value and can also be used to evaluate fusion images:

$$\bar{H} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=2}^N H(i, j). \quad (21)$$

The CC measures the degree of linear correlation between a fused image and infrared and visible images and is mathematically defined as follows:

$$CC(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \quad (22)$$

where  $\text{Cov}(X, Y)$  represents the covariance of  $X$  and  $Y$  and  $\text{Var}(X)$  and  $\text{Var}(Y)$  represent the variance of  $X$  and  $Y$ , respectively. The larger the CC is, the higher the degree of correlation between the fused images and visible and infrared images is and the higher the similarity is.

**4.2. Experimental Validation of Fusion Performance.** In this experiment, we first focus on qualitative and quantitative comparisons on the fusion performance of different methods on the surveillance images from TNO Human Factors. We selected 45 pairs of infrared and visible images from the publicly available TNO image fusion dataset. Besides, we also test our method on the RGB-NIR Scene dataset. This dataset consists of 477 images in 9 categories captured in RGB and near-infrared (NIR). The images were captured using separate exposures from modified SLR cameras, using visible and NIR filters. The scene categories are country, field, forest, indoor, mountain, old building, street, urban, and water [52]. Our training parameters are set as follows: the batch image size is 64, the number of training iterations is 400, and the discriminator training step is 2. The parameters  $\alpha, \beta, \delta$ , and  $\gamma$  are set as follows:  $\alpha = 100, \beta = 0.2, \delta = 5$ , and  $\gamma = 0.005$ . The learning rate is set to  $10^{-5}$ . In the training phase,  $88 \times 88$  random crops are used as input to the original image pair in each iteration. The Adam solver is used to optimize the loss function, and the generator and discriminator update their parameters with each iteration. In the testing stage, we put the stacked images of the same proportion of infrared and visible images into the generator, and  $G_{\theta_G}$  generates a fusion image of the same size as the input. The experiments are



FIGURE 6: Qualitative fusion results on five typical infrared and visible image pairs from the TNO database. From left to right: Bunker, Kaptein\_1123, Kaptein\_1654, Marne\_07, and Tank. From top to bottom: (a) the visible image, (b) the infrared image, (c) CVT [18], (d) DTCWT [45], (e) GFF [4], (f) GTF [19], (g) Fusion GAN [11], and (h) FLGC-Fusion GAN.

TABLE 1: The average evaluation index value of all methods on the 20 images of the TNO dataset.

Methods	SF	EN	CC	COSIN	SD	SSIM
CVT	6.9696	6.2151	0.4023	0.9118	28.7006	0.394
DTCWT	6.9266	6.6617	0.3892	0.9549	32.6006	<b>0.4652</b>
GFF	7.0902	7.1238	0.3711	0.8683	43.2243	0.3763
GTF	7.2161	7.1464	0.3622	0.9739	44.3197	0.3858
Fusion GAN	7.4707	7.377	0.4473	0.9715	48.5708	0.4132
FLGC-Fusion GAN	<b>8.0068</b>	<b>7.5199</b>	<b>0.4861</b>	<b>0.9801</b>	<b>50.6215</b>	0.4541

The number in each column indicates the average value of the evaluation index.

conducted on a computer with 3.7 GHz Intel 10900x, GPU RTX 2080Ti with 11 GB memory.

*4.2.1. Results on TNO Database.* Considering giving some intuitive results on the fusion performance, we select five typical image pairs for qualitative evaluation: Bunker, Kaptein\_1123, Kaptein\_1654, Marne\_07, and Tank. The fusion results of the proposed method are compared with those of Fusion GAN and other methods, as shown in Figure 6. The first two rows present the original visible images and infrared images, the third to the sixth rows present the fusion images of other methods, the seventh row shows the fusion images of Fusion GAN, and the last row is the fusion results of our methods. From the results, we see that all methods can well fuse the information of the visible image and infrared image to some extent. For example, in the Tank, both the Fusion GAN and our approach incorporate the thermal radiation target well. However, the Fusion GAN effect does not seem to highlight the target. Besides, our method retains obvious heat source information and outlines obvious detail edge information in all five groups of images. In contrast, other methods have some mottled artifacts in the fusion results. By comparing Fusion GAN with our method, it can be found that the scene information in our method can be clearly seen, each part has a prominent contrast effect with the surrounding environment, and the brightness also improves significantly. Nevertheless, in our proposed method can appear extremely individual image edges slightly blurred. This is due to the fact that, in order to preserve the source information and image contrast of the image during the training process, the pixel intensities of some areas in the fused image may be changed during training. In order to verify the authenticity of this conclusion, we use six image fusion indicators to illustrate.

We use some qualitative evaluation indexes to show the quality of fusion images objectively. We further use the above six fusion indicators to conduct a quantitative comparison between any ten pairs of infrared images and visible images in the TNO dataset, as shown in Table 1. Of the 20 image pairs tested, our method was the largest in terms of the average value of the evaluation index. Our method obtains the best SF, EN, CC, COSIN, SD, and SSIM for most image pairs, and the average values for six evaluation metrics, including CC, are the largest relative to the other methods. The largest SF indicates that our fused image has richer edges and textures. The largest EN demonstrates that our fused image has much more abundant information than the other

seven competitors. The largest CC shows that our fused image is strongly correlated with the two source images. The largest COSIN indicates that the distortion of the fused images is small, and the whole image is closer. Among them, SSIM is higher in our proposed method, but not as good as DTCWT. This indicates that our few fused images do not achieve the best-expected results with the source image in terms of brightness, contrast, and structure so that the total SSIM score is not the highest. This also verifies that the edges of people in our fusion image in the second column of Figure 6 are slightly blurred, but this does not affect the overall effect of our fusion method. Compared with the six fusion methods, there are two indicators SD, and SF is more prominent than the other four results. This shows that our fusion image has higher contrast and richer edge texture overall.

*4.2.2. Results on RGB-NIR Scene Database.* In order to verify the versatility of FLGC-Fusion GAN and further observe the effect of this method on different datasets. We test our method on the RGB-NIR Scene dataset, which is trained on the TNO dataset. In this way, we can verify the universality of FLGC-Fusion GAN. Figure 7 is a sample of six scenarios from the dataset (i.e., country, field, forest, old building, street, and water). Figure 7 illustrates that the fusion result of each scene has a clear and bright outline. The fusion image as a whole well fuses the thermal radiation information of the infrared image while maintaining consistency with the visible RGB image. The improvement of this paper is based on the Fusion GAN model. Compared with the previous methods, Fusion GAN has better evaluation results in every aspect. Therefore, we make a quantitative comparison between the Fusion results of datasets on this model and those on Fusion GAN. The Fusion indicators were still SF, COSIN, EN, SSIM, SD, and CC. The comparison results are fully reflected in Table 2. The results intuitively show that the average evaluation index of our method is the best regardless of the dataset. For some evaluation indexes, we, respectively, refer to the infrared image and visible image for evaluation, and the evaluation results are superior to the Fusion GAN. Therefore, our proposed method retains enough visible image texture information while keeping more infrared thermal radiation information, and the brightness of the fusion image is significantly improved compared to the previous way.

We also present the run-time comparison of the six methods in Table 3. Compared with the other five methods,

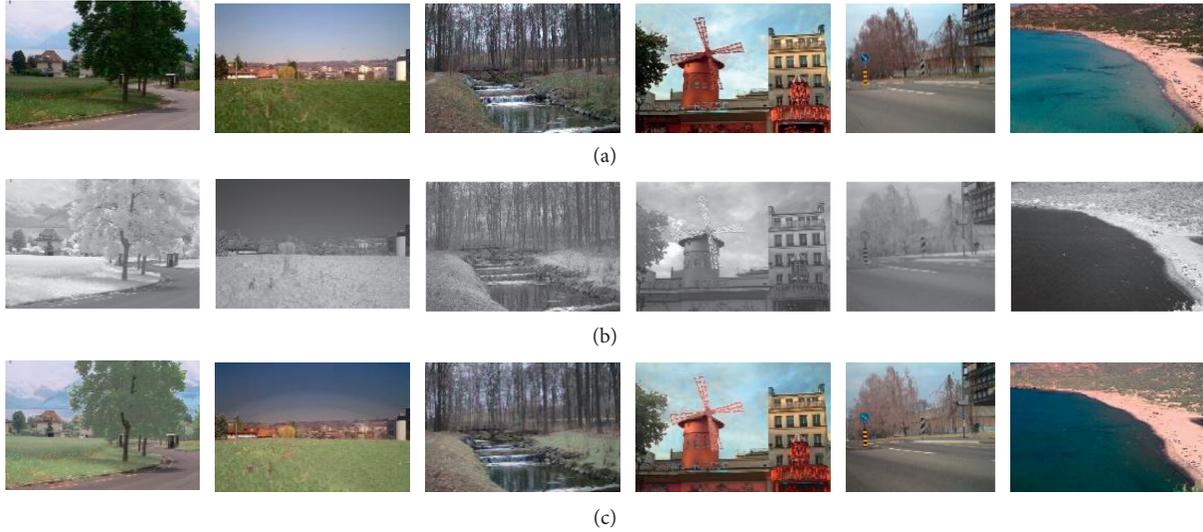


FIGURE 7: The fusion image is generated on the RGB-NIR Scene dataset using the proposed method. From left to right: country, field, forest, old building, street, and water. From top to bottom: (a) the RGB image, (b) the NIR image, and (c) fused results of our FLGC-Fusion GAN.

TABLE 2: The average evaluation index value of all methods on the 20 images of the RGB-NIR Scene dataset.

Methods	SF	EN	CC	COSIN	SD	SSIM
Fusion GAN	5.5027	7.0204	0.7481	0.9896	43.8332	0.722
FLGC-Fusion GAN	6.4792	7.3313	0.809	0.9935	45.5027	0.7814

The number in each column indicates the average value of the evaluation index.

TABLE 3: Run-time comparison of six methods on the TNO datasets.

Method	TNO
CVT	1.46
DTCWT	$3.32 \times 10^{-1}$
GFF	$3.18 \times 10^{-1}$
GTF	4.82
Fusion GAN	$4.61 \times 10^{-2}$
FLGC-Fusion GAN	$3.07 \times 10^{-2}$

The CVT, DTCWT, GFF, GTF, Fusion GAN, and FLGC-Fusion GAN are performed on GPU while all the others are performed on CPU. Each value denotes the mean of the run time of a certain method on a dataset (unit: second).

our method has achieved considerable efficiency by using a learnable grouping convolution.

## 5. Conclusion

In this paper, we propose a new method of infrared image and visible image fusion network based on GAN. This method is an end-to-end fusion model, the characteristic of the proposed method is to enhance the saliency of targets in the fused image, which can perform a variety of different image fusion tasks well. The main contribution of this paper is to design a new generator network, which adopts the grouping convolution to realize high efficiency and good

fusion effects. We use the learnable grouping convolution in the generator network to optimize the group structure in all layers at the same time in an end-to-end manner, accelerating the training of the model. The use of multilevel residual networks deepens the depth of the network and prevents the BN layer from having a certain influence on the stability in the deep network training process. Moreover, the inactive perception feature is used as the feature input of content loss, so as to constrain the pixel intensity of thermal radiation information of the infrared image and the texture information of the visible image. Through experiments on two standard datasets, FLGC-Fusion GAN can adapt to different image Fusion tasks, and the result can retain approximately the maximum amount of information in the source image. Qualitative and quantitative experiments show that compared with the existing methods, our FLGC-Fusion GAN has advantages in various image Fusion tasks and is faster than Fusion GAN.

## Data Availability

All data included in this study are available upon request to the corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

C. Yuan and C. Q. Sun contributed equally to this work and should be considered the co-first authors.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. F060609) and Natural Science Fund of Hubei Province (Grant no. 2019CFB250).

## References

- [1] A. Dogra, B. Goyal, and S. Agrawal, "From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications," *IEEE Access*, vol. 5, pp. 16040–16067, 2017.
- [2] A. Toet, "Image fusion by a ratio of low-pass pyramid," *Pattern Recognition Letters*, vol. 9, no. 4, pp. 245–253, 1989.
- [3] L. Xu, G. Cui, C. Zheng, and J. Zhao, "Visible infrared image fusion method based on multi-scale decomposition and saliency region extraction," *Laser & Optoelectronics Progress*, vol. 54, no. 11, pp. 111–120, 2017.
- [4] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 22, no. 7, pp. 2864–2875, 2013.
- [5] B. Yang and S. Li, "Visual attention guided image fusion with sparse representation," *Optik*, vol. 125, no. 17, pp. 4881–4888, 2014.
- [6] W. Kong, L. Zhang, and Y. Lei, "Novel fusion method for visible light and infrared images based on NSST-SF-PCNN," *Infrared Physics & Technology*, vol. 65, pp. 103–112, 2014.
- [7] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proceedings of the 20th International Conference on Information Fusion (Fusion)*, pp. 1–9, Xi'an, China, July 2017.
- [8] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infrared Physics & Technology*, vol. 76, pp. 52–64, 2016.
- [9] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147–164, 2015.
- [10] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: a survey of the state of the art," *Information Fusion*, vol. 33, pp. 100–112, 2017.
- [11] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: a generative adversarial network for infrared and visible image fusion," *Information Fusion*, vol. 48, pp. 11–26, 2019.
- [12] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Information Fusion*, vol. 12, no. 2, pp. 74–84, 2011.
- [13] G. Pajares and J. Manuel de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.
- [14] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3450–3459, 2012.
- [15] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 16, no. 3, Article ID 1850018, 2018.
- [16] T. Xiang, L. Yan, and R. Gao, "A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain," *Infrared Physics & Technology*, vol. 69, pp. 53–61, 2015.
- [17] J. Zhao, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Infrared image enhancement through saliency feature analysis based on multi-scale decomposition," *Infrared Physics & Technology*, vol. 62, pp. 86–93, 2014.
- [18] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017.
- [19] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Information Fusion*, vol. 31, pp. 100–109, 2016.
- [20] 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," *Information Fusion*, vol. 42, pp. 158–173, 2018.
- [21] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, 2003.
- [22] S. Rajkumar and P. V. S. S. R. C. Mouli, "Infrared and visible image fusion using entropy and neuro-fuzzy concepts," in *Proceedings of the ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol I*, pp. 93–100, Novotel Varun Beach, Visakhapatnam, December 2014.
- [23] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proceedings of the International Conference on Information Fusion*, pp. 1–9, Xi'an, China, July 2017.
- [24] X. Jin, Q. Jiang, S. Yao et al., "A survey of infrared and visual image fusion," *Infrared Physics & Technology*, vol. 85, pp. 478–501, 2017.
- [25] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.
- [26] H. Hermessi, O. Mourali, and E. Zagrouba, "Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain," *Neural Computing and Applications*, vol. 30, pp. 1–17, 2018.
- [27] X. Zhang, Y. Ma, F. Fan, Y. Zhang, and J. Huang, "Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition," *Journal of the Optical Society of America A*, vol. 34, no. 8, pp. 1400–1410, 2017.
- [28] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: a survey," *Information Fusion*, vol. 45, pp. 153–178, 2019.
- [29] H. Xu, P. Liang, W. Yu, J. Jiang, and J. Ma, "Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3954–3960, Beijing, China, August 2019.
- [30] K. Jiang, Z. Wang, and P. Yi, "Multi-scale progressive fusion network for single image deraining," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, November 2020.
- [31] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4980–4995, 2020.
- [32] J. Chen, X. Li, L. Luo, X. Mei, and J. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Information Sciences*, vol. 508, pp. 64–78, 2020.
- [33] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognition*, vol. 48, no. 3, pp. 772–784, 2015.
- [34] K. Ram, V. S. Srikar, and R. V. Babu, "Deepfuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4724–4732, Venice, Italy, October 2017.

- [35] J. Ma, P. Liang, W. Yu et al., "Infrared and visible image fusion via detail preserving adversarial learning," *Information Fusion*, vol. 54, pp. 85–98, 2020.
- [36] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 2, pp. 2672–2680, 2014.
- [37] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, <https://arxiv.org/abs/1411.1784v1>.
- [38] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Information Fusion*, vol. 36, pp. 191–207, 2017.
- [39] H. Li, X. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proceedings of the International Conference on Pattern Recognition 2018*, IEEE, Prague, Czech Republic, October 2018.
- [40] X. Xu, Y. Li, and C. Yuan, "Identity bracelets" for deep neural networks," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.
- [41] G. Huang, S. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Condensenet: an efficient densenet using learned group convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2752–2761, IEEE, Salt Lake City, UT, USA, June 2018.
- [42] X. Wang, M. Kan, and S. Shan, "Fully learnable group convolution for acceleration of deep neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556v6>.
- [44] Yi Peng, Z. Wang, K. Jiang, Z. Shao, and J. Ma, "Multi-temporal ultra dense memory network for video super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, 2019.
- [45] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2, pp. 119–130, 2007.
- [46] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.
- [47] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, and H. Jégou, "Word translation without parallel data," 2017, <https://arxiv.org/abs/1710.04087v3>.
- [48] J. W. Roberts, J. Van Aardt, and F. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *Journal of Applied Remote Sensing*, vol. 2, no. 1, pp. 1–28, 2008.
- [49] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [50] Y.-J. Rao, "In-fibre bragg grating sensors," *Measurement Science and Technology*, vol. 8, no. 4, p. 355, 1997.
- [51] M. Deshmukh and U. Bhosale, "Image fusion and image quality assessment of fused images," *International Journal of Image Processing (IJIP)*, vol. 4, no. 5, p. 484, 2010.
- [52] M. Brown and S. Susstrunk, "Multi-spectral SIFT for scene category recognition," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR11)*, Colorado SPGS, USA, June 2011.