

Infrared and visible image fusion based on Laplacian pyramid and generative adversarial network

Juan Wang^{1,2}, Cong Ke^{1,2}, Minghu Wu^{1,2*}, Min Liu^{1,2} and Chunyan Zeng^{1,2}

¹Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan, 430068, P. R. China
[e-mail: happywj@hbut.edu.cn]

²Hubei Power Grid Intelligent Control and Equipment Engineering Technology Research Center, Wuhan, 430068, P. R. China
[e-mail: wuxx1005@mail.hbut.edu.cn]
*Corresponding author: Minghu Wu

*Received May 27, 2020; revised April 3, 2021; accepted May 11, 2021;
published May 31, 2021*

Abstract

An image with infrared features and visible details is obtained by processing infrared and visible images. In this paper, a fusion method based on Laplacian pyramid and generative adversarial network is proposed to obtain high quality fusion images, termed as Laplacian-GAN. Firstly, the base and detail layers are obtained by decomposing the source images. Secondly, we utilize the Laplacian pyramid-based method to fuse these base layers to obtain more information of the base layer. Thirdly, the detail part is fused by a generative adversarial network. In addition, generative adversarial network avoids the manual design complicated fusion rules. Finally, the fused base layer and fused detail layer are reconstructed to obtain the fused image. Experimental results demonstrate that the proposed method can obtain state-of-the-art fusion performance in both visual quality and objective assessment. In terms of visual observation, the fusion image obtained by Laplacian-GAN algorithm in this paper is clearer in detail. At the same time, in the six metrics of MI, AG, EI, MS_SSIM, Q_{abf} and SCD, the algorithm presented in this paper has improved by 0.62%, 7.10%, 14.53%, 12.18%, 34.33% and 12.23%, respectively, compared with the best of the other three algorithms.

Keywords: Laplacian pyramid, generative adversarial network, base layer, detail layer, fusion

This research was supported in part by the National Natural Science Foundation of China under Grants 62006073, and in part by the Major Technological Innovation Projects of Hubei under Grant No. 2018AAA028.

1. Introduction

With increasing sensor technology, different type images of fusion are a paramount issue in image processing field. Image fusion is to generate an image containing the complete details of the same scene [1]. In recent years, two different types of image fusion have become a crucial part of the image fusion field, such as infrared and visible images. It has been widely used in object detection, military actions, video surveillance [2] [3] and remote sensing [4] [5]. For example, the United Kingdom's Waterfall Solution Company developed an image fusion system. This system is applied to police helicopters [6]. Its function is to fuse infrared images with color visible images, and the image obtained has both infrared features and color image effect. In addition, William from the United States proposed an image fusion algorithm based on discrete Haar wavelet transform [7] [8], which was used in the vehicle-mounted system of night driving. Therefore, it is indispensable to further explore the fusion method of these two images.

Up to now, a host of fusion methods have been presented. Multi-scale transform method [9] is the most popular method in the early years. Classic multi-scale fusion methods include the Laplacian pyramid (LP)-based method [10], the morphological pyramid (MP)-based method [11], the discrete wavelet transform (DWT)-based method [12], the dual-tree complex wavelet transform (DTCWT)-based method [13] and the non-subsampled contourlet transform (NSCT)-based method [14]. These methods obtained the fused image by decomposition, fusion, and reconstruction [15]. However, the fused images obtained by these methods would lose a slice of detail information, which would influence the visual quality of the images. Moreover, these traditional fusion methods need to design complex fusion rules and perform poorly in extracting features.

Due to the advantages of deep learning in image feature extraction, some fusion methods based on deep learning had also attracted great attention. Yu Liu [16] proposed a deep convolutional neural network (CNN)-based multi-focus image fusion method. It was the first time to apply CNN to image fusion. However, this method only used the results of the last layer of the network model. Hui Li [17] presented a deep learning framework for infrared and visible image fusion. However, this method only used the average strategy to fuse the base layer, which would lead to the loss of some details. Jiayi Ma [18] introduced generative adversarial network (GAN) for infrared and visible image fusion. However, this method only obtains detailed information by distinguishing the generated image from the visible image, resulting in less infrared features in the generated image. Therefore, the network models of these deep learning fusion methods need to be improved.

Although the above methods have achieved success in image fusion, their shortcoming lies in that they only pay attention to one type of source image while ignoring the details of another source image in retaining source image information. To overcome these deficiencies, we present a fusion method based on Laplacian pyramid and generative adversarial network (Laplacian-GAN). The image is decomposed into two parts through a guide filter [19]. We use the fusion method based on Laplacian pyramid to obtain the fused base layer. To address the detail layer problem, a novel generative adversarial network is employed to obtain the fused detail layer. The fused image is reconstructed by fusing the base layer and the detail layer.

The rest of this paper is organized as follows. Section2 describes the related work of the proposed Laplacian-GAN. In Section3, the proposed Laplacian-GAN is full introduced. Section4 describes the fusion results and further analysis. Finally, summary of this paper in Section5.

2. Related Work

Image pyramid and generative adversarial network are related to the research of this paper. This section focuses on the two related methods.

2.1 Laplacian Pyramid

Laplacian pyramid is to process images at multiple scales and resolutions. It can decompose the edge and texture features of the image into different resolution tower layers according to different scales [10]. But the Laplacian pyramid is based on the Gaussian pyramid [20].

The Gaussian pyramid is composed of a series of images, the bottom of which is the original image, and each upper layer of the image is obtained by the previous layer through low-pass filtering, so this operation is called subtraction operation, and is represented by the *REDUCE* function. The bottom layer of the pyramid is G_0 and the l layer is G_l , which is defined as follows.

$$G_l = REDUCE(G_{l-1}) = \sum_{m=-2}^2 \sum_{n=-2}^2 \omega(m,n) G_{l-1}(2i+m, 2j+n) \quad (1)$$

$$\omega = \frac{1}{256} * \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (2)$$

where ω is a 5×5 filter used for subsampled G_0 . i and j represent the rows and columns of the image. Then ω is convolved with the image of layer l to get the same size as the image of layer $l-1$.

The Laplacian pyramid consists of a series of error images L_0, L_1, \dots, L_N , each of which is obtained by subtracting the corresponding two layers of the Gaussian pyramid. As each layer of the Gaussian pyramid is obtained from the previous layer by subsampling, the image resolution of each layer is not the same. Therefore, the image of the current layer needs to be up-sampled to obtain the image with the same resolution size. Therefore, the *EXPAND* function is introduced to insert the up-sampled image to obtain the image with the same resolution as the upper layer. The formula is defined as follows.

$$G_l^* = EXPAND(G_l) = 4 \sum_{m=-2}^2 \sum_{n=-2}^2 \omega(m,n) G_l\left(\frac{i+m}{2}, \frac{j+n}{2}\right) \quad (3)$$

$G_l\left(\frac{i+m}{2}, \frac{j+n}{2}\right)$ has a corresponding value when $(i+m)/2$ and $(j+n)/2$ are integers, otherwise it is 0. The image of each layer of the Laplacian pyramid can be expressed by the following formula:

$$L_l = G_l - EXPAND(G_{l+1}) \quad (4)$$

where l represents the level of the pyramid, ranging from 0 to N , and set $L_N = G_N$. Fig. 1 is a diagram of the decomposition of the Laplacian pyramid.

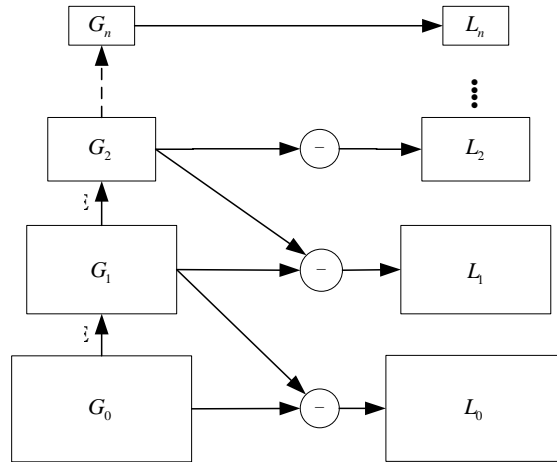


Fig. 1. Decomposition of the Laplacian pyramid

2.1 Generative Adversarial Network

Generative adversarial network [21] uses an adversarial model to estimate sample distribution and generate new data. The generator and discriminator form an adversarial network, represented by G and D , respectively. The generator G generates data similar to the sample data by analyzing the distribution of the sample data. It is the task of the discriminator D to calculate the probability that the samples come from real data instead of generating data. The probability that D output is large represents real data, otherwise it represents generated data. The schematic diagram of the generative adversarial network is shown in Fig. 2.

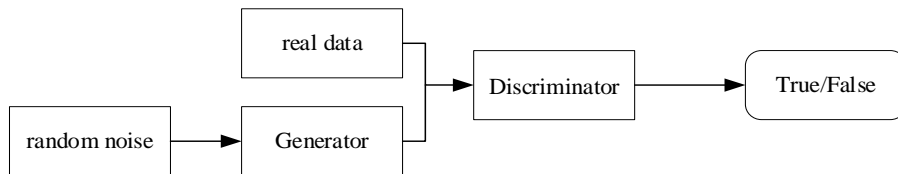


Fig. 2. Schematic diagram of generative adversarial network

The loss function of generative adversarial network is an important part used to predict whether the model is good or not. Its purpose is to make the data generated by the generator closer to the real data, that is, the difference between the two is getting smaller, and the discriminator cannot distinguish them. Therefore, the loss function of generative adversarial network is expressed as follows.

$$\min_G \max_D V_{GAN}(G, D) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (5)$$

where x represents the input sample, z denotes the noise of the input G . $P_{data}(x)$ represents real data distribution and $P_z(z)$ denotes noise distribution. $D(x)$ represents the probability that the discriminator determines whether the real data is real or not, and $D(G(z))$ denotes the probability that the discriminator determines whether the generated data is true or not.

3. Proposed Fusion Method

3.1 Image Decomposition

Each image can be decomposed two parts: the base layer and the detail layer. The base layer is a description of the change in the intensity value of an image. The detail layer reflects the changes in detail on an image. In our paper, we choose two preregistered images as input images. The input images are represented by I_k , $k \in \{1, 2\}$.

Nowadays, decomposition methods such as wavelet decomposition have achieved great results on images. But compared to those methods, the guide filter [19] is more efficient and time-saving. So this method is used to decompose the input images into two dimensions. For each input image I_k , the decomposed base layer and detail layer are represented by I_k^b and I_k^d , respectively. By solving the optimization problem in (6), the base layer can be gained.

$$I_k^b = \arg \min_{I_k^b} \|I_k - I_k^b\|_F^2 + \lambda (\|g_x * I_k^b\|_F^2 + \|g_y * I_k^b\|_F^2) \quad (6)$$

where $g_x = [-1 \ 1]$ represents the horizontal gradient operator, $g_y = [-1 \ 1]^T$ denotes the vertical gradient operator. In the above formula, the parameter λ is set to 5. The detail layer is the base layer obtained by subtracting the source image. The equation is as follows.

$$I_k^d = I_k - I_k^b \quad (7)$$

The method framework of this paper is shown in Fig. 3. Firstly, the input images obtain the base layers and detail layers by (6) and (7). Then the fused base layers are obtained by the Laplacian pyramid transform and the detail layers are fused by a novel generative adversarial network. Finally, the two parts are reconstructed to obtain the fusion image.

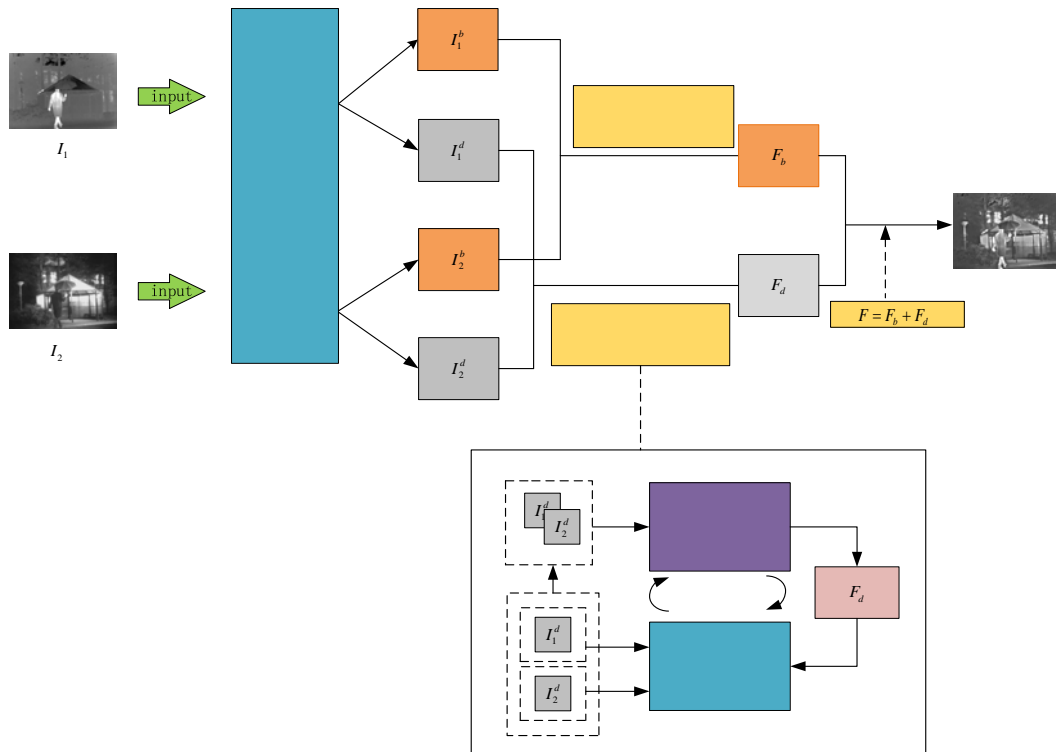


Fig. 3. Framework of the method

3.2 Fusion of Base Layer

Conventionally, most multi-scale fusion methods adopt average strategy for processing low-frequency information in images. However, the average strategy have some defects, which cannot well to keep the low-frequency information about the source images [22]. The image pyramid fusion method aims to decompose the input image into different frequencies and design corresponding fusion rules for each layer to obtain the fused pyramid layer. This method can effectively retain the intensity information of the base part. So in our paper, we utilize the Laplacian pyramid-based method to fuse these base layer. The procedure is shown in **Fig. 4**.

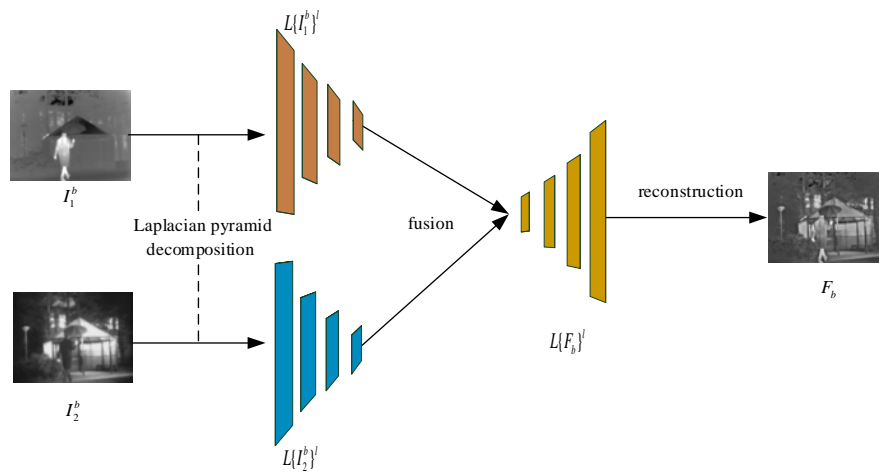


Fig. 4. The procedure of base layer fusion

I_1^b and I_2^b are decomposed into four levels [23] through the Laplacian pyramid, respectively. $L\{I_1^b\}^l$ and $L\{I_2^b\}^l$ represent the pyramid decomposition of layer L of I_1^b and I_2^b . First of all, for the fusion of the top-level image $L\{I_1^b\}^4$ and $L\{I_2^b\}^4$, it is necessary to calculate the regional average gradient $m*n$ ($m \geq 3$, $n \geq 3$, m and n are both odd) with each pixel as the center in the image. The calculation formula is as follows.

$$G = \frac{1}{(m-1)(n-1)} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{(\Delta I_x^2 + \Delta I_y^2) / 2} \quad (8)$$

where ΔI_x and ΔI_y represent the first-order difference of pixel point $f(x, y)$ on the x -axis and y -axis, and the formula is defined as follows.

$$\Delta I_x = f(x, y) - f(x-1, y) \quad (9)$$

$$\Delta I_y = f(x, y) - f(x, y-1) \quad (10)$$

The average regional gradient of each pixel in the top-level image $L\{I_1^b\}^4$ and $L\{I_2^b\}^4$ is represented by G_1 and G_2 respectively. Since the average gradient represents the texture feature changes and the richness of detail information of the image, the top-level image fusion can be expressed as follows.

$$L\{F_b\}^4(i, j) = \begin{cases} L\{I_1^b\}^4(i, j) & G_1(i, j) \geq G_2(i, j) \\ L\{I_2^b\}^4(i, j) & G_1(i, j) < G_2(i, j) \end{cases} \quad (11)$$

Next is the fusion of other layers of images. When $0 < l < 4$, the regional energy of these layers needs to be calculated, and the calculation formula is as follows.

$$ARE(i, j) = \sum_{p=-p}^p \sum_{q=-q}^q \varpi(p, q) |L\{I_1^b\}^l(i+p, j+q)| \quad (12)$$

$$BRE(i, j) = \sum_{p=-p}^p \sum_{q=-q}^q \varpi(p, q) |L\{I_2^b\}^l(i+p, j+q)| \quad (13)$$

where ARE and BRE represent the regional energy of the two source images in the corresponding pyramid layer respectively. p and q are set as 1. ϖ is a 3×3 matrix, represented as follows.

$$\varpi = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (14)$$

When $0 < l < 4$, the fusion result of the Laplacian image pyramid of the L layer is shown as follows.

$$L\{F_b\}^l(i, j) = \begin{cases} L\{I_1^b\}^l(i, j) & ARE(i, j) \geq BRE(i, j) \\ L\{I_2^b\}^l(i, j) & ARE(i, j) < BRE(i, j) \end{cases} \quad (15)$$

The above formula can be used to obtain the fusion image of each layer of the Laplacian pyramid, namely, $L\{F_b\}^1$, $L\{F_b\}^2$, $L\{F_b\}^3$ and $L\{F_b\}^4$. Then, the fused Laplace pyramid layer is reconstructed to obtain the corresponding Gaussian pyramid, and the final fused image G_1 is obtained. The reconstruction formula is shown below.

$$\begin{cases} G_4 = L\{F_b\}^4 \\ G_l = L\{F_b\}^l + EXPAND(G_{l+1}) \end{cases} \quad (16)$$

3.3 Fusion of Detail Layer

In terms of detail layer I_1^d and I_2^d , a fusion strategy based on generative adversarial network is presented in this paper. Because the detail layer contains the change of the detail information of the image, the generative adversarial network can better retain the detail information of the source image. This framework is shown in Fig. 5.

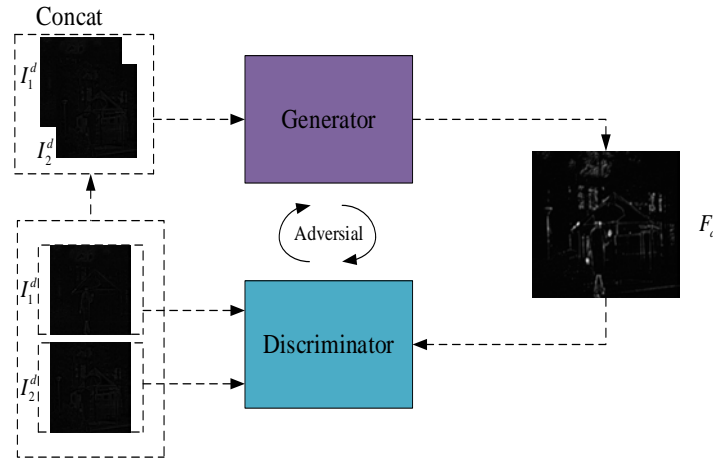


Fig. 5. Framework of detail layer fusion

In Fig. 5, we employ the generator to generate a fused image. Then employ the discriminator to distinguish the generated image and two detail layers' images to get a fused detail image.

3.3.1 Network Structure of the GAN

The network structure of GAN in this paper includes two parts: generator model G and discriminator model D . To ensure that the image information is not subject to large loss in network propagation, both the generator and the discriminator use a convolutional network.

The generator network structure is shown in Fig. 6. The purpose is to extract more detailed information in the source images and generate a fused image with rich details. As shown, a five-layer convolutional network constitutes G , where the first two layers use 5×5 filters, the middle two layers use 3×3 filters, the last layer uses a 1×1 filter and stride of each layer is set to 1. Besides, to solve problems such as vanishing gradient, we use BatchNorm [24] to normalize the data at each layer, then employ LeakyRelu [25] activation function to enhance the non-linear degree of the network.

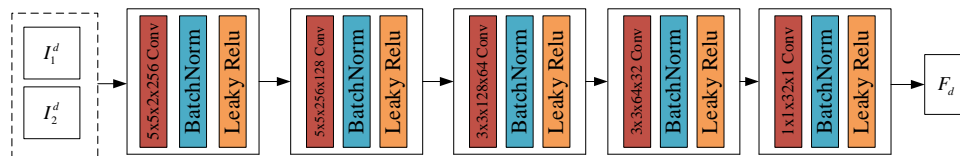


Fig. 6. Generator network structure

The LeakyRelu activation function is a variation of the Relu activation function. Relu is a rectified line unit in the neural network. When its input is x and x is greater than 0, its gradient

is not 0. In this case, Relu can be used to update the weight. When x is less than 0, then its gradient is 0, and Relu cannot be used to update weights. And the advantage of LeakyRelu is that when x is less than zero, LeakyRelu preserves a very small constant. In this way, when the input is less than 0, the information cannot be completely lost, and the corresponding information is retained. In other words, the LeakyRelu still has a small gradient for updating the weight in the part where x is less than zero. Schematic diagrams of the two functions are shown in Fig. 7.

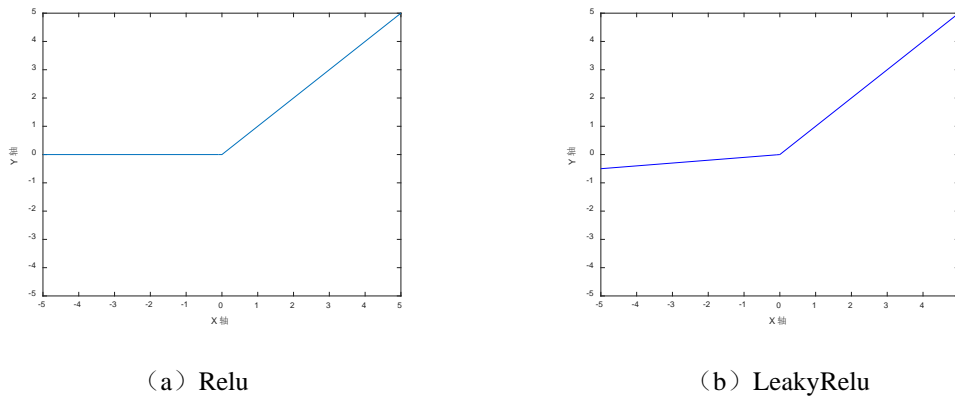


Fig. 7. Schematic diagram of two activation functions

The discriminator network structure is shown in Fig. 8. The purpose is to correctly distinguish between the generated image and the source image. The discriminator D is an eight-layer convolutional network in Fig. 8. Each layer uses 3×3 filters in the discriminator. The second to fourth layers of the stride are set to 2, and the remaining layers are set to 1.

To avoid introducing noise, the pooling layer is replaced by a convolutional layer with stride is set to 2, so that the classification effect of the discriminator is better. Similar to the generator, the first layer to the seventh layer employ BatchNorm to normalize the data, then use LeakyRelu as the activation function, and the last layer is the linear layer for classification.

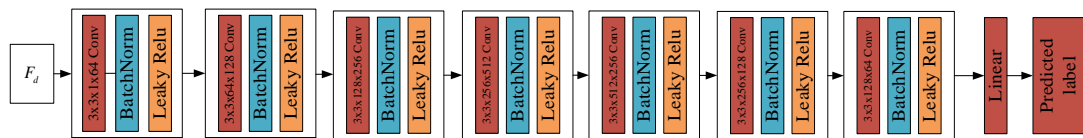


Fig. 8. Discriminator network structure

3.3.2 Loss Function

In this paper, the loss function of GAN is composed of two parts, namely L_G and L_D . The purpose is to minimize the loss function so as to obtain the best training model. Next, we will introduce the total loss function L_{GAN} , generator loss function L_G and discriminator loss function L_D , respectively. First, the total loss function of GAN is shown as follows.

$$L_{GAN} = \{\min(L_G), \min(L_D)\} \quad (17)$$

where L_{GAN} denotes the total loss, L_G denotes the generator loss and L_D denotes the discriminator loss. The generator loss function L_G is defined as follows.

$$L_G = V + \alpha L_{\text{content}} \quad (18)$$

where parameter α is used to balance V and L_{content} , V represents the adversarial loss between generator and discriminator, as shown in (19). The second term L_{content} denotes the content loss of image details during the generation process, as shown in (20).

$$V = \frac{1}{N} \sum_{n=1}^N (D(F_d^n) - a)^2 \quad (19)$$

$$L_{\text{content}} = \frac{1}{HW} [\beta (\|\nabla F_d - \nabla I_1^d\|_F^2 + \gamma \|\nabla F_d - \nabla I_2^d\|_F^2)] \quad (20)$$

where N denotes the number of fused image, F_d denotes the fused image, and $D(F_d^n)$ represents the result of classification. I_1^d and I_2^d represent the detail layer of the input two source images, respectively. a is the value of the generator generated fake data. In (20), H denotes the height of I_1^d or I_2^d , W denotes the width of I_1^d or I_2^d , ∇ denotes the gradient operator [26]. $\|\cdot\|_F$ represents the matrix Frobenius norm. L_{content} aims to retain more gradient detail information in both I_1^d and I_2^d . The parameters β and γ are designed to control the balance between them. To obtain better effect of the fused image, discriminator is introduced to distinguish the fused image. Specifically, the discriminator loss function L_D is shown as follows.

$$L_D = \frac{1}{N} \sum_{n=1}^N (D(I_1^d) - b)^2 + \frac{1}{N} \sum_{n=1}^N (D(I_2^d) - c)^2 + \frac{1}{N} \sum_{n=1}^N (D(F_d^n) - d)^2 \quad (21)$$

where b and c denote the labels of the detail layer I_1^d and I_2^d , with values ranging from 0.4 to 0.7 and 0.8 to 1.1, respectively. d denote the labels of the fused detail layer, whose value ranges from 0 to 0.3. $D(I_1^d)$, $D(I_2^d)$ and $D(F_d^n)$ represent the discriminator distinguish results of I_1^d , I_2^d and F_d , respectively.

3.3.3 Training Details

41 image pairs were selected from the TNO database as training for this experiment. Although 41 pairs of images were not adequate for GAN training, we crop these source images to increase the diversity of training data, and we design a sliding window with size of 120×120 and stride of 14 to crop the detail layer image. Thus, 58,658 pairs of infrared and visible patches were obtained. Each patch was padded to the size of 132×132 and then fed into the generator. The size of the fusion detail image generated is 120×120 . The generated fusion image is used as the input of discriminator, and Adam [27] is utilized to optimize the algorithm until the maximum number of training is reached.

3.3.4 Reconstruction

After obtaining the fused detail layer F_d , the fused image F is obtained by using F_b and F_d . The formula is as below.

$$F = F_b + F_d \quad (22)$$

4. Experimental Results and Analysis

4.1 Experimental Settings

To verify the performance of the proposed Laplacian-GAN, compare it with three fusion methods including FusionGAN [18], deep learning framework (DLF) [17], convolutional neural network (CNN) [16]. In order to more intuitively express the effectiveness of the Laplacian-GAN method, we used six fusion evaluation metrics. The server configuration for this experiment is as follows: Ubuntu server 18.04, CPU is Intel Xeon(R) Gold 5120 and GPU is Nvidia Tesla P100.

4.2 Fusion Metrics

Since the qualitative evaluation is hard to accurately evaluate the fused image, we also need metrics to quantitatively evaluate fused images. In recent years, various evaluation metrics have been proposed, but different metrics represent the performance of fused images in different aspects. Multiple metrics need to be selected to evaluate these fused images. So six metrics are used to quantify the performance of fusion images, *i.e.* mutual information (MI) [28], average gradient (AG) [29], edge intensity (EI) [30], multi-scale structural similarity (MS_SSIM) [31], quality of edge (Q_{abf}) [32], and the sum of the correlations of differences (SCD) [33]. These metrics are defined as follows.

(1) MI measures the preserved input images information about fused image. The MI formula is as follows.

$$M_{A,B} = \sum_{a,b} p_{A,B}(a,b) \log \frac{p_{A,B}(a,b)}{p_A(a)p_B(b)} \quad (23)$$

$$MI = M_{A,H} + M_{B,H} \quad (24)$$

where $p_A(a)$ denotes the normalized histogram of image A , $p_B(b)$ denotes the normalized histogram of image B . $p_{A,B}$ represents the jointly normalized histogram of the input two images. $M_{A,H}$ and $M_{B,H}$ represent the information preserved by the fused image from two input images, respectively.

(2) AG is an evaluation metric that represents the fused image gradient information. Its value denotes the details and texture of the fused image. The calculation formula is as follows:

$$AG = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{[H(i,j) - H(i+1,j)]^2 + [H(i,j) - H(i,j+1)]^2}{2}} \quad (25)$$

where M and N represent the height and width of the fused image. $H(i,j)$ denotes the value of the fused image at (i,j) .

(3) EI refers to the amplitude of the gradient of the image edge point, which is defined as follows:

$$EI = \sqrt{\nabla_x f(i,j)^2 + \nabla_y f(i,j)^2} \quad (26)$$

where ∇_x and ∇_y represent the first order difference of the fused image in the direction x and y at (i,j) , respectively. $f(i,j)$ represents the value of the fused image at (i,j) . These

formulas are defined as.

$$\nabla x = f(i, j) - f(i-1, j) \quad (27)$$

$$\nabla y = f(i, j) - f(i, j-1) \quad (28)$$

(4) MS_SSIM represents the structural similarity of two images at multiple scales.

$$MS_SSIM(Z, K) = [l_M(Z, K)]^{\alpha_M} \prod_{i=1}^M [s_i(Z, K)]^{\beta_i} [z_i(Z, K)]^{\gamma_i} \quad (29)$$

where M is the highest scale selected for the reference image, $l(Z, K)$, $s(Z, K)$ and $z(Z, K)$ denote brightness similarity, contrast similarity and structure similarity between image Z and K , respectively. α , β and γ are the weight of the three, which are generally set as 1.

$$l(Z, K) = \frac{2\mu_Z\mu_K + C_1}{\mu_Z^2 + \mu_K^2 + C_1} \quad (30)$$

$$s(Z, K) = \frac{2\sigma_Z\sigma_K + C_2}{\sigma_Z^2 + \sigma_K^2 + C_2} \quad (31)$$

$$z(Z, K) = \frac{2\sigma_{ZK} + C_3}{\sigma_Z\sigma_K + C_3} \quad (32)$$

where μ_Z and μ_K , σ_Z and σ_K are the mean and standard deviation of image Z and K , respectively. σ_{ZK} is the gray covariance between image Z and K . C_1 , C_2 and C_3 are all constants.

(5) Q_{abf} is to calculate the edge information transferred from the source images to the fused image to evaluate the quality of the fused image. The calculation is as follows:

$$Q_{abf} = \frac{\sum_{i=1}^M \sum_{j=1}^N Q_{AF}(i, j)w_A(i, j) + Q_{BF}(i, j)w_B(i, j)}{\sum_{i=1}^M \sum_{j=1}^N w_A(i, j) + w_B(i, j)} \quad (33)$$

where w_A and w_B are weighting coefficients for calculating the edge intensity value of the source image using Sobel operator. Q_{AF} and Q_{BF} represent the edge information preserved by the fused image F from image A and image B , respectively.

$$Q_{AF}(i, j) = Q_g(i, j)Q_\alpha(i, j) \quad (34)$$

where Q_g and Q_α are edge strength and direction similarity calculated by Sobel operator.

(6) SCD is the sum of the differential relations between the source images and the fused image, and its computational formula is defined as follows.

$$L_1 = F - I_1 \quad (35)$$

$$L_2 = F - I_2 \quad (36)$$

$$SCD = r(L_1, I_1) + r(L_2, I_2) \quad (37)$$

where I_1 , I_2 and F denote the two source images and the fused image, respectively. The formula of $r(\cdot)$ function is defined as follows.

$$r(L_k, I_k) = \frac{\sum_i \sum_j (L_k(i, j) - \bar{L}_k)(I_k(i, j) - \bar{I}_k)}{\sqrt{\left(\sum_i \sum_j (L_k(i, j) - \bar{L}_k)^2 \right) \left(\sum_i \sum_j (I_k(i, j) - \bar{I}_k)^2 \right)}} \quad (38)$$

where \bar{L}_k and \bar{I}_k represent the average pixel values of L_k and I_k , $k \in \{1, 2\}$.

4.3 Qualitative Evaluation

The eight different fused images obtained by the three existing neural network methods and Laplacian-GAN are shown in [Fig. 9](#).



Fig. 9. Infrared and visible image fusion based on four typical methods. From top to bottom: infrared images, visible images, results of CNN, DLF, FusionGAN and Laplacian-GAN method.

As shown in [Fig. 9](#), the fused image obtained by CNN have more visual artifacts and artificial noise. The method causes serious loss of image background information, so this method is not suitable for image fusion of two different types. In contrast, the fused image obtained by DLF and FusionGAN look more natural, but there are some infrared details that are not extract. Thus, compared with the above methods, the fused image of our proposed method obviously keep more infrared feature and our results increase the brightness more suitable for human perception system.

4.4 Quantitative Evaluation

Since qualitative evaluation is largely dependent on visual subjective judgment, the evaluation is not accurate. Therefore, we also used some quantitative metrics to verify the comprehensive performance of Laplacian-GAN method. These metrics include MI, AG, EI, MS_SSIM, Q_{abf}

and SCD. The average values of these metrics acquired by four deep learning methods for the eight fused images are shown in Fig. 10.

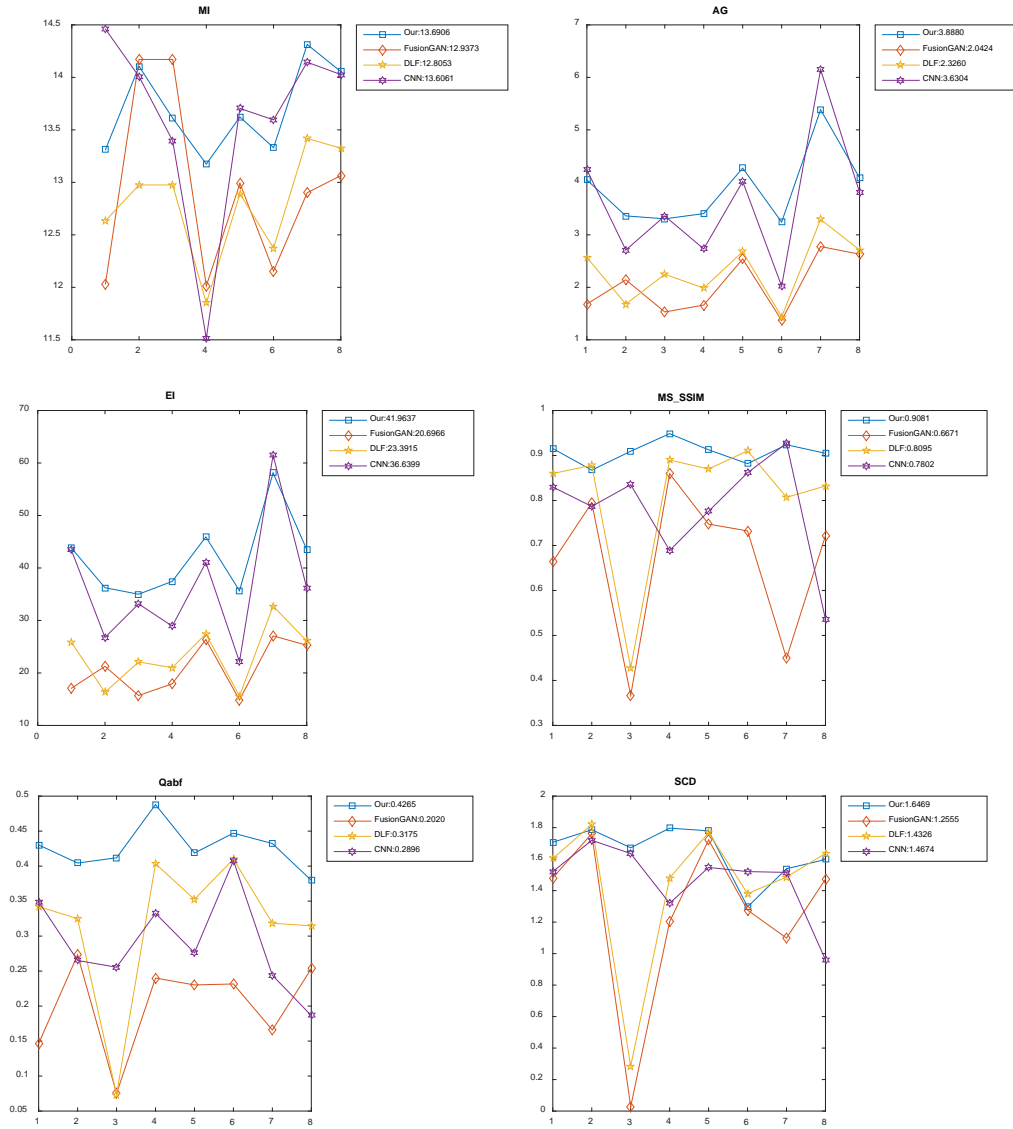


Fig. 10. Quantitative comparisons of four methods on six metrics

Among the six metrics, MI, AG, EI and Q_{abf} are used to measure the information richness in the fused image. However, there is a certain difference between the three metrics. MI is used to measure the dependence between two domain variables. In other words, the larger its value is, the more information of fused image retains about the source image. AG, EI and Q_{abf} represent the texture transformation, edge information intensity and edge information transfer in the image respectively. As shown in Fig. 10, the Laplacian-GAN has achieved the superior performance in MI, AG, EI and Q_{abf} . Compared with the CNN and DLF fusion methods, they improved by 0.62%, 7.10%, 14.46% and 34.33%, respectively. Therefore, it can indicate that

the fused images obtained by our method contain richer information. By capturing the blur across multiple scales, MS_SSIM could influence the visual perception of the human visual system by considering the influence of sampling rate, observation distance and other factors on the quality evaluation of the fused image. In terms of MS_SSIM metric, our method improves by 12.18% on the basis of the higher DLF method. SCD is the calculation of the sum of all relevant information between the fused image and the source image. The result shows that our method is superior to other methods in SCD. In summary, our method contains more detail information and the extracted infrared features are obvious, which is easier to observe and apply to the field of target recognition.

5. Conclusion

In this paper, we introduce a novel fusion method based on Laplacian pyramid and generative adversarial network. The proposed Laplacian-GAN can avoid designing complicated fusion rule manually. Compared with other three methods, the performance of fused images obtained by Laplacian-GAN method is greatly improved. The quantitative comparisons of three fusion methods demonstrate that Laplacian-GAN method preserve abundant detail information and also produce better visual effects. In future, we believe that our Laplacian-GAN can not only apply infrared and visible image fusion, but also solve other fusion problem.

References

- [1] Goshtasby A A, Nikolov S, "Image fusion: Advances in the state of the art," *Information Fusion*, vol.8, no.2, pp.114-118, 2007. [Article \(CrossRef Link\)](#)
- [2] Wu M, Li X, Liu C, et al., "Robust global motion estimation for video security based on improved k-means clustering," *Journal of Ambient Intelligence and Humanized Computing*, vol.10, no.2, pp.439-448, 2019. [Article \(CrossRef Link\)](#)
- [3] Yeh C H, Lin C H, Lin M H, et al., "Deep learning-based compressed image artifacts reduction based on multi-scale image fusion," *Information Fusion*, vol.67, pp.195-207, 2021. [Article \(CrossRef Link\)](#)
- [4] Dian R, Li S, Sun B, et al., "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Information Fusion*, vol.69, pp.40-51, 2020. [Article \(CrossRef Link\)](#)
- [5] Li H, Zhang L, Jiang M, et al., "Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network," *Pattern Recognition Letters*, vol.141, pp.45-53, 2021. [Article \(CrossRef Link\)](#)
- [6] Duan C, Wang Z, Xing C, et al., "Infrared and visible image fusion using multi-scale edge-preserving decomposition and multiple saliency features," *Optik*, vol.228, no.1, pp.165775, 2021. [Article \(CrossRef Link\)](#)
- [7] Rena K, Zhang D, Wan M, et al., "An Infrared and Visible Image Fusion Method Based on Improved DenseNet and mRMR-ZCA," *Infrared Physics & Technology*, vol.115, no.4, pp.103707, 2021. [Article \(CrossRef Link\)](#)
- [8] Liu Y, Chen X, Ward R, et al., "Image Fusion With Convolutional Sparse Representation," *IEEE Signal Processing Papers*, vol.23, no.12, pp.1882-1886, 2016. [Article \(CrossRef Link\)](#)
- [9] Ma J, Ma Y, Li C, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol.45, pp.153-178, 2019. [Article \(CrossRef Link\)](#)
- [10] Burt P J, Adelson E H, "The Laplacian Pyramid as a Compact Image Code," *Readings in Computer Vision*, vol.31, no.4, pp.671-679, 1987. [Article \(CrossRef Link\)](#)
- [11] Toet A, "A morphological pyramidal image decomposition," *Pattern Recognition Letters*, vol.9, no.4, pp.255-261, 1989. [Article \(CrossRef Link\)](#)

- [12] Li H, Manjunath B S , Mitra S K , “Multisensor Image Fusion Using the Wavelet Transform,” *Graphical Models and Image Processing*, vol.57, no.3, pp.235-245, 1995. [Article \(CrossRef Link\)](#)
- [13] Lewis J J , Robert J. O’Callaghan, Nikolov S G, et al., “Pixel- and region-based image fusion with complex wavelets,” *Information Fusion*, vol.8, no.2, pp.119-130, 2007. [Article \(CrossRef Link\)](#)
- [14] Zhang Q, Guo B L, “Multifocus image fusion using the nonsubsampling contourlet transform,” *Signal Processing*, vol.89, no.7, pp.1334-1346, 2009. [Article \(CrossRef Link\)](#)
- [15] Piella G, “A general framework for multiresolution image fusion: from pixels to regions,” *Information Fusion*, vol.4, no.4, pp.259-280, 2003. [Article \(CrossRef Link\)](#)
- [16] Liu Y, Chen X , Peng H , et al., “Multi-focus image fusion with a deep convolutional neural network,” *Information Fusion*, vol.36, pp.191-207, 2017. [Article \(CrossRef Link\)](#)
- [17] H. Li, X.-J. Wu, and J. Kittler, “Infrared and Visible Image Fusion using a Deep Learning Framework,” in *Proc. of International Conference on Pattern Recognition*, 2018. [Article \(CrossRef Link\)](#)
- [18] Ma J, Yu W, Liang P, et al., “FusionGAN: A generative adversarial network for infrared and visible image fusion,” *Information Fusion*, vol.48, pp.11-26, 2019. [Article \(CrossRef Link\)](#)
- [19] Li S, Kang X, Hu J, “Image fusion with guided filtering,” *IEEE Transactions on Image Processing*, vol.22, no.7, pp.2864-2875, 2013. [Article \(CrossRef Link\)](#)
- [20] Baker, K. D., and G. D. Sullivan, “Multiple bandpass filters in image processing,” *IEE Proceedings E-Computers and Digital Techniques*, vol.127, no.5, pp.173-184, 1980. [Article \(CrossRef Link\)](#)
- [21] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S. & Bengio, Y, “Generative adversarial networks,” *Communications of the ACM*, vol.63, no.11, 2020. [Article \(CrossRef Link\)](#)
- [22] Li S, Yang B, Hu J, “Performance comparison of different multi-resolution transforms for image fusion,” *Information Fusion*, vol.12, no.2, pp.74-84, 2011. [Article \(CrossRef Link\)](#)
- [23] Dogra A, Goyal B, Agrawal S, “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. [Article \(CrossRef Link\)](#)
- [24] Santurkar S, Tsipras D, Ilyas A, et al., “How Does Batch Normalization Help Optimization?,” 2019. [Article \(CrossRef Link\)](#)
- [25] Wang S H, Phillips P, Sui Y, et al., “Classification of Alzheimer’s disease based on eight-layer convolutional neural network with leaky rectified linear unit and max Pooling,” *Journal of Medical Systems*, vol.42, no.5, pp.85, 2018. [Article \(CrossRef Link\)](#)
- [26] Ma J, Chen C, Li C, et al., “Infrared and visible image fusion via gradient transfer and total variation minimization,” *Information Fusion*, vol.31, pp.100-109, 2016. [Article \(CrossRef Link\)](#)
- [27] Kingma D, Ba J, “Adam: A method for stochastic optimization,” in *Proc. of conference paper at ICLR 2015*, 2015. [Article \(CrossRef Link\)](#)
- [28] Qu G, Zhang D, Yan P, “Information measure for performance of image fusion,” *Electronics Papers*, vol.38, no.7, pp.313, 2002.
- [29] Yu S, Zhongdong W, Xiaopeng W, et al., “Tetrolet transform images fusion algorithm based on fuzzy operator,” *Journal of Frontiers of Computer Science and Technology*, vol.9, no.9, pp.1132-1138, 2015. [Article \(CrossRef Link\)](#)
- [30] Petrovic V, Coates T, “Information representation for image fusion evaluation,” in *Proc. of 2006 9th International Conference on Information Fusion*, pp.1-7, 2006. [Article \(CrossRef Link\)](#)
- [31] Wang Z, Simoncelli E P, Bovik A C, “Multiscale structural similarity for image quality assessment,” in *Proc. of The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, vol.2, pp.1398-1402, 2003. [Article \(CrossRef Link\)](#)
- [32] Xydeas C S, Petrovic V, “Objective image fusion performance measure,” *Electronics letters*, vol.36, no.4, pp.308-309, 2000.
- [33] Aslantas V, Bendes E, “A new image quality metric for image fusion: the sum of the correlations of differences,” *Aeu-international Journal of electronics and communications*, vol.69, no.12, pp.1890-1896, 2015. [Article \(CrossRef Link\)](#)



Juan Wang received her BS degree from Nanyang Normal University in 2007 and MS degree from Xihua University in 2010, and her PhD from Tianjin University in 2015. She is currently a faculty member at Hubei University of Technology. Her research interests include image processing, pattern recognition, and computer vision.



Cong Ke is a graduate student at Hubei University of Technology. His research interest is image processing. He received his BS degree from Hubei University of Technology Engineering and Technology College in 2018.



Minghu Wu received his BS degree from Communication University of China in 1998 and MS degree from Huazhong University of Science and Technology in 2003, and his PhD from Nanjing University Of Posts And Telecommunications in 2013. His research interests include artificial intelligence and image processing.



Min Liu received her BS degree from Zhengzhou University in 2004, her MS and PhD degrees from Huazhong University of Science and Technology in 2007 and 2013, respectively. Her research interests include image processing and machine learning.



Chunyan Zeng received her BS degree from Hubei University of Technology in 2007, her MS and PhD degrees from South China University of Technology in 2009 and 2013, respectively. Her research interests include artificial intelligence and image processing.