applied sciences

MDPI

Article

Dense-FG: A Fusion GAN Model by Using Densely Connected Blocks to Fuse Infrared and Visible Images

Xiaodi Xu [1], Yan Shen [1,*] and Shuai Han [2]

[1] College of Mathematical Sciences, Harbin Engineering University, Harbin 150001, China; xuxiaodi1500450@163.com
[2] College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; hanshuai_heu@163.com
* Correspondence: shenyan@hrbeu.edu.cn

Abstract: In various engineering fields, the fusion of infrared and visible images has important applications. However, in the current process of fusing infrared and visible images, there are problems with unclear texture details in the fused images and unbalanced displays of infrared targets and texture details, resulting in information loss. In this article, we propose an improved generative adversarial network (GAN) fusion model for fusing infrared and visible images. In the generator and discriminator network structure, we introduce densely connected blocks to connect the features between layers, improve network efficiency, enhance the network's ability to extract source image information, and construct a content loss function using four losses, including an infrared gradient, visible intensity, infrared intensity, and a visible gradient, to maintain a balance between infrared radiation information and visible texture details, enabling the fused image to achieve ideal results. The effectiveness of the fusion method is demonstrated through ablation experiments on the TNO dataset, and compared with four traditional fusion methods and three deep learning fusion methods. The experimental results show that our method achieves five out of ten optimal evaluation indicators, with a significant improvement compared to other methods.

Keywords: infrared and visible images; image fusion; generative adversarial networks; dense connection blocks; fusion evaluation

check for updates

Citation: Xu, X.; Shen, Y.; Han, S. Dense-FG: A Fusion GAN Model by Using Densely Connected Blocks to Fuse Infrared and Visible Images. *Appl. Sci.* **2023**, *13*, 4684. https://doi.org/10.3390/app13084684

Academic Editor: Habib Hamam

Received: 7 March 2023
Revised: 30 March 2023
Accepted: 4 April 2023
Published: 7 April 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Image fusion is the process of combining multiple images into a single one, which is an image containing more comprehensive information. The fusion result integrates the temporal and spatial correlations and complementarities between information of multiple images, providing a more thorough and precise description of the scene. This is advantageous for both human observation and machine automatic detection [1]. Infrared and visible image fusion is a technique that merges infrared and visible images to obtain an image with improved information content and more details.

Infrared images typically reflect thermal radiation information from objects and have strong adaptability and anti-interference capabilities, but cannot effectively reflect image texture information. Visible light images have relatively high resolution and can provide high-quality images and details, but are easily affected by external factors such as light and climate [2]. Since infrared and visible images each have different characteristics and advantages, fusing the two types of images can yield more comprehensive and accurate information, improving the ability to recognize, detect, and trace objects.

Infrared and visible image fusion technology has a variety of applications. In the military field, this technology can improve the precision and reliability of reconnaissance targets, and enhance combat capabilities in low visibility situations [3]. In the realm of aerospace, this technology holds significant potential for the navigation and monitoring of both aircraft and satellites, helping pilots and astronauts better understand the situation

of target objects, thereby improving the accuracy of navigation and monitoring. In the realm of firefighting and rescue operations, infrared imaging can identify fire origins and smoke, while visible imaging can offer clearer representations of building architecture [4]. By combining the two, rescue personnel can pinpoint fire sources and smoke more swiftly and precisely, thus enhancing rescue efficiency. Image fusion technology has also been extensively employed in the medical domain. By providing enhanced medical images with improved clarity and comprehensiveness, it ultimately leads to a higher degree of accuracy and efficiency in medical diagnoses. In addition, infrared and visible image fusion technology is extensively utilized in information security, natural resource management, and environmental monitoring [5]. In summary, the technology has broad application potential, providing more comprehensive and accurate information for many fields. Research into the technology of fusing infrared and visible images holds significant practical importance and is valuable for many applications.

Currently, there are mainly two methods for fusing infrared and visible images: traditional image transformation decomposition and synthesis methods, and deep learning methods [6]. Traditional methods have low detail presentation and significant differences in fusion results under different scenarios, thus their applicability is limited. Deep learning methods mainly use CNN, GAN, DenseNet, and other methods to construct networks for fusion [7–9], which can cover more usage scenarios, but the balance effect between infrared target brightness and visible texture details is not good. Image texture details are unclear and infrared target details are lost, requiring further improvement in fusion image quality.

This article proposes a GAN network fusion method using dense connection blocks for infrared and visible image fusion, which fully utilizes the source image information while preserving more complementary information. The texture details are more abundant and clearer, and the important features of the image pairs are retained, resulting in higher fusion quality and better visual effects.

To summarize, the main contributions of our research can be shown as follows:

(1) A generator network structure and discriminator network structure with dense connection blocks were designed so that there are paths connecting all layers of the network, enabling feature reuse and improving computational efficiency.
(2) A content loss function was constructed using four losses, an infrared gradient, visible intensity, infrared intensity, and a visible gradient, to maintain a balance between infrared radiation information and visible texture details, and to achieve an ideal fusion image.
(3) By updating the 8-direction gradient operator template and optimizing the design of the loss function, the fusion image details were made richer.
(4) Histogram comparison was used to demonstrate the fusion ability of eight fusion methods in a clear and intuitive way, providing a reference for the development and improvement of fusion image methods in the future.

The remaining sections of this article are organized as follows. Section 2 introduces some related work on image fusion from both traditional and deep learning methods. Section 3 presents our proposed method, including the fusion framework and loss function. Section 4 provides experimental settings, image evaluation metrics, fusion comparison results, and an analytical discussion, as well as ablation experiments. Finally, Section 5 concludes the paper.

2. Literature Survey

2.1. Research Status of Traditional Infrared and Visible Image Fusion

From the perspective of fusion levels, image fusion can be divided into three levels from high to low: pixel level, feature level, and decision level. Pixel-level image fusion falls under the category of low-level image fusion [10], which involves extracting and computing features from processed images from multiple sources to create a fused image. Such fusion methods can retain more detailed information but require large amounts of data processing and have high demands on equipment.

Spatial domain algorithms and transform domain algorithms are the most common pixel-level image fusion methods. Image fusion methods based on spatial domain include principal component analysis (PCA) [11], average gradient (AVG) [12], and others. The image fusion using PCA separately fuses the different components of the source images (mainly the first component). The image fusion using AVG takes the weighted average of corresponding pixels in the source images to form the fused pixel and generate a new fused image. This approach is straightforward and intuitive, but selecting relevant information can be challenging, and the fusion result is prone to instability.

Transform domain image fusion methods include pyramid transform, wavelet transform, multi-scale geometric transform, and other similar techniques. Pyramid and wavelet transform-based image fusion typically involves decomposing the source views into multi-resolution architectures and merging the high-pass residual information and low-pass average details to reconstruct the ultimate fused image. Frequently employed techniques for decomposing and reconstructing data include the discrete wavelet transform (DWT) [13], Laplacian pyramid (LP) [14], filter subtract decimate pyramid (FSDP) [15], and contrast pyramid (CP) [16], as well as their respective improved variants. The above methods can only obtain decomposition coefficients in the horizontal, vertical, and diagonal directions during the source image decomposition process, which is not conducive to preserving detailed information in the fused image. Image fusion based on multi-scale geometric transforms can overcome this limitation, and common methods include curvelet[17], contourlet [18], NSCT [19], etc. Although such algorithms can effectively fuse source image information to some extent, specific implementation requires the manual setting of decomposition layers, low-frequency fusion strategies, high-frequency fusion strategies, and other parameters, making it more complex.

Feature-level image fusion is categorized as a type of middle-level image fusion. This method collects features from the registered sources and accomplishes geometric alignment through parameter templates, statistical analysis, pattern correlation, and other methods [20]. The specific fusion algorithm and parameter selection need to be adjusted and optimized according to the specific application scenario.

Decision-level image fusion belongs to the highest level of image fusion. This method extracts features from multiple input images and makes decisions separately, obtaining multiple decision results. Then, these decision results are combined to obtain the final fusion result. The advantage of decision-level image fusion is that it can use the decision results of different images for fusion, making the fusion results more comprehensive and accurate. However, its disadvantages are also quite obvious. It requires feature extraction and decision-making for each set of input images, which has a high computational complexity. At the same time, it is susceptible to the influence of feature-extraction and decision-making algorithms, therefore requiring high algorithm stability [21].

In conclusion, traditional infrared and visible image fusion techniques possess their strengths and weaknesses, and suitable techniques should be selected based on the particular application scenario.

2.2. Research Status of Infrared and Visible Image Fusion Based on Deep Learning

Deep learning has exhibited outstanding results in image processing and has been extensively utilized for the integration of infrared and visible images. In 2016, Liu et al. [22] proposed a fusion method based on convolutional neural networks (CNN) for infrared and visible images. They employed a Siamese convolutional network to generate weight maps, extract pixel information from two source images, and combine the extracted information in a multi-scale way through an image pyramid.

To address the issue of most deep learning-based methods directly using deep features without feature extraction or processing, according to Li et al. [23] residual networks and zero-phase component analysis (ZCA) can be used to integrate such data. This approach utilizes residual networks for extracting profound features and normalizes these features utilizing ZCA, leading to an enhancement in the quality of the reconstructed fusion image.

To further improve the quality of the fusion image, Li et al. [24] implemented image fusion using encoding and decoding networks and introduced dense connection blocks in the encoding network to extract more source image features.

In 2019, Ma et al. [25] introduced a technique for combining infrared and visible images utilizing generative adversarial networks (GAN), which they named FusionGAN. This method continuously improves its generation and discrimination capabilities through a generator and a discriminator playing an adversarial game, obtaining fusion images containing rich source-image information.

In 2020, Ma [26] played an adversarial game with one generator and two discriminators to preserve a balance between the infrared radiation data and visible texture features and extended it to multi-resolution image fusion.

In 2021, Xu et al. [27] used a classifier to classify source images to address the limitation of manually setting fusion rules during the fusion process. The classification result was quantified as an important coefficient and presented as a saliency map. The fusion of various feature maps generates a fusion image, and even simple networks produce good results. Long et al. [28] suggested a fusion approach based on aggregated residual dense networks for infrared and visible images, which automatically assesses the information preservation of the source image, extracts hierarchical features, and achieves effective fusion. Yang et al. [29] suggested a fusion strategy that utilizes a combined texture map and an adaptive guided filter to generate multiple decision maps. The fusion image performed well in the subjective evaluation and quantitative indicators. To address the issue of non-prominent infrared targets in fusion images, Li et al. [30] incorporated an attention mechanism into the generator and discriminator based on GAN to obtain fusion images with prominent targets and rich details.

The fusion procedure of infrared and visible images based on deep learning mentioned above is mainly achieved by modifying the image fusion framework, configuring the fusion network architecture, and formulating loss functions to restrict the fusion process. Various design concepts and fusion regulations have distinct impacts on fusion outcomes. Further research can be conducted to enhance the quality of the fused image by exploring the equilibrium between the luminance of the infrared target and the texture particulars of the visible light.

3. Proposed Method

3.1. Network Architecture Design

The methodology of the fusion framework is shown in Figure 1. During training, infrared and visible images are concatenated along the channel dimension and used as the input to the generator to generate fusion images. The discriminator takes the generated fusion images and the visible images as input to discriminate whether the generated images are from the generator or real images. By engaging in an adversarial game between the generator and discriminator, visible information is continuously supplemented into the generated fusion images to optimize the generator and discriminator. During testing, the concatenated infrared and visible images are fed into the trained discriminator to generate the ultimate fusion image.

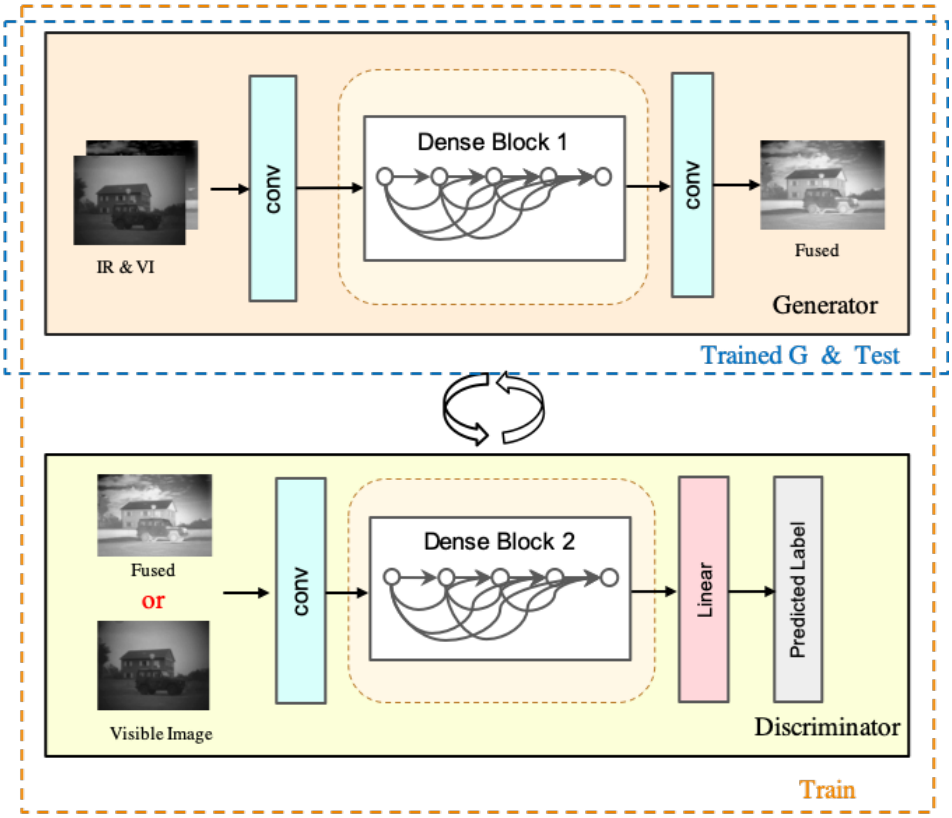

Figure 1. Fusion framework.

3.1.1. Generator Network Architecture

The generator network architecture is shown in Figure 2.

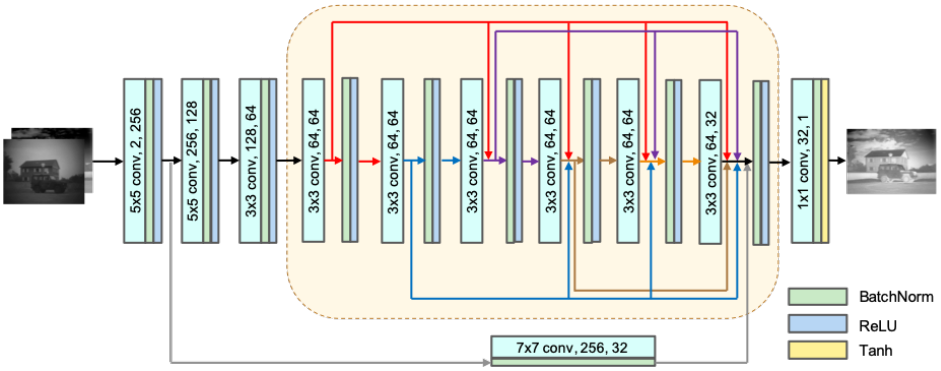

Figure 2. Generator network architecture.

The generator network architecture consists of regular convolution layers (Conv) and dense connection layers (Dense). The concatenated infrared and visible images are fed into the generator network. The first three layers are regular convolution layers, where the input layer and the second layer have a kernel size of 5×5, and the third layer has a kernel size of 3×3. Dense 1 to Dense 6 are introduced dense connection blocks, where the input of Dense 3 is the sum of the outputs from Dense 1 and Dense 2; the input of Dense 4 is the sum of the outputs from Dense 1, Dense 2, and Dense 3; the input of Dense 5 is the sum of the outputs from Dense 1, Dense 2, Dense 3, and Dense 4. Moreover, a standard convolutional layer with a kernel size of 7×7 is employed to link the output of the input layer to the outputs of Dense 1 to Dense 5, which then acts as the input for Dense 6. Ultimately, the merged outcome is achieved using another standard convolutional layer with a kernel size of 1×1. Batch normalization and activation functions are incorporated

into every layer of the generator network architecture, with the output layer utilizing tanh as the activation function, while the remaining layers utilize LeakyReLU.

The parameters related to the generator network architecture are presented in Table 1, with K denoting the size of the convolution kernel, S the stride, and P the padding of the convolution kernel; N1 and N2 denote the quantity of input and output channels, respectively.

Table 1. Generator network architecture parameters.

Name	Layer	K	S	Input	N1	Output	N2	Padding	Function
Input layer	Conv1	5*5	1	IR+VI	2	I1	256	VALID	LReLU
Conv layer	Conv2	5*5	1	I1	256	I2	128	VALID	LReLU
	Conv3	3*3	1	I2	128	I3	64	VALID	LReLU
Dense layer	Dense1	3*3	1	I3	64	I4	64	SAME	LReLU
	Dense2	3*3	1	I4	64	I5	64	SAME	LReLU
	Dense3	3*3	1	I4+I5	64	I6	64	SAME	LReLU
	Dense4	3*3	1	I4+I5+I6	64	I7	64	SAME	LReLU
	Dense5	3*3	1	I4+I5+I6+I7	64	I8	64	SAME	LReLU
Conv layer	Conv1_1	7*7	1	I1	256	I1_1	64	VALID	LReLU
Dense layer	Dense6	3*3	1	I1_1+I4+I5+ I6+I7+I8	64	I9	32	VALID	LReLU
Output layer	Conv4	1*1	1	I9	32	F	1	VALID	Tanh

3.1.2. Discriminator Network Architecture

The discriminator network architecture is shown in Figure 3.

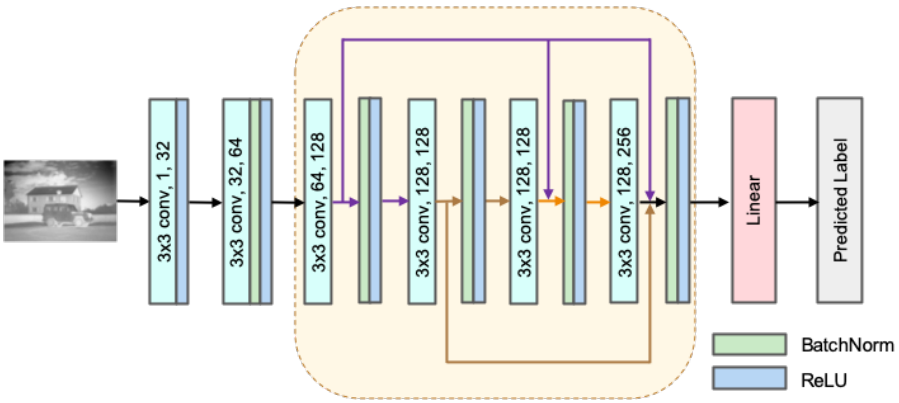

Figure 3. Discriminator network architecture.

The discriminator network architecture also consists of regular convolutional layers (Conv) and dense layers (Dense). The discriminator network receives visible images or fused images as input; the input layer, as well as the second layer, are both conventional convolutional layers, and Dense 1 to Dense 4 are dense blocks. Among them, the input of Dense 3 is the sum of the output results of Dense 1 and Dense 2, and the input of Dense 4 is the sum of the output results of Dense 1, Dense 2, and Dense 3. Finally, the output result of Dense 4 is used as the input to the linear layer to return the discrimination result. Batch normalization and activation functions are incorporated into each layer of the discriminator network architecture, with LeakyReLU being used as the activation function for all layers except for the output layer, which employs Matmul. The pertinent parameters for the discriminator network architecture can be found in Table 2.

Table 2. Discriminator network architecture parameters.

Name	Layer	K	S	Input	N1	Output	N2	Padding	Function
Input layer	Conv1	3*3	2	VI/F	1	I1	32	VALID	LReLU
Conv layer	Conv2	3*3	2	I1	32	I2	64	VALID	LReLU
Dense layer	Dense1	3*3	2	I3	64	I4	128	VALID	LReLU
	Dense2	3*3	1	I4	128	I5	128	SAME	LReLU
	Dense3	3*3	1	I4+I5	128	I6	128	SAME	LReLU
	Dense4	3*3	2	I4+I5+I6	128	I7	256	VALID	LReLU
Output layer	Line1	/	/	I7	256	Lable	1	/	Matmul

3.2. Loss Function Design

The generator loss function comprises two components, adversarial loss and content; the expression is as follows:

$$L_G = L_{\text{adv}} + \lambda L_{\text{con}}. \tag{1}$$

The coefficient for the proportion of content loss is represented by λ , and the expression for adversarial loss is:

$$L_{\text{adv}} = \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}\left(I_f^n\right) - c\right)^2, \tag{2}$$

where $D_{\theta_D}\left(I_f\right)$ represents the discriminator's discrimination result on the fused image, N denotes the number of fused images, and c is a soft label.

The content loss expression is as follows:

$$L_{con} = \frac{1}{HW}(\beta_1\left\|I_f - I_r\right\|_F^2 + \beta_2\left\|\nabla I_f - \nabla I_v\right\|_F^2 + \beta_3\left\|I_f - I_v\right\|_F^2 + \beta_4\left\|\nabla I_f - \nabla I_r\right\|_F^2). \tag{3}$$

In the above formula, the fused image is denoted by I_f , while I_r and I_v represent the visible and infrared images, respectively. H and W stand for the height and width of the image, and β_1, β_2, β_3, and β_4 represent the proportional coefficients of various losses. ∇ denotes the gradient operation of the image matrix.

The gradient loss constrained fusion image contains rich texture details; β_2 and β_4 represent the weight coefficients of the visible gradient and infrared gradient loss items, respectively. The intensity loss constrained fusion image maintains an image distribution similar to the source image; β_1 and β_3 represent the infrared intensity and visible intensity loss items weight factor, respectively.

For the fusion of infrared and visible images, it is expected that the fusion result retains the visible gradient information and infrared intensity information, while the infrared gradient information and visible intensity information are secondary, satisfying the following rules:

$$\beta_1 > \beta_3, \beta_2 > \beta_4. \tag{4}$$

In the process of solving the gradient, a Laplacian filter template with diagonal terms is used, which is defined as follows:

$$\text{filter} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{5}$$

Linear spatial filtering of size $m \times n$ (the moving step is 1, and the matrix is filled with 0 from top to bottom, left to right) is performed on the image; then:

$$\begin{cases} g(1,1) = f(1,2) + f(2,1) + f(2,2) - 8f(1,1) \\ \cdots \\ g(1,n) = f(1,n-1) + f(2,n) + f(2,n-1) - 8f(1,n) \\ \cdots \\ g(m,1) = f(m-1,1) + f(m,2) + f(m-1,2) - 8f(m,1) \\ \cdots \\ g(m,n) = f(m-1,n) + f(m,n-1) + f(m-1,n-1) - 8f(m,n) \end{cases}. \tag{6}$$

The expression of the gradient matrix of the fused image is as follows:

$$\nabla I_f = \begin{bmatrix} g_f(1,1) & g_f(1,2) & \cdots & g_f(1,n) \\ g_f(2,1) & g_f(2,2) & \cdots & g_f(2,n) \\ \vdots & \vdots & & \vdots \\ g_f(m,1) & g_f(m,2) & \cdots & g_f(m,n) \end{bmatrix}. \tag{7}$$

The discriminator loss function expression is as follows:

$$L_D = \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(I_v^n) - b\right)^2 + \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}\left(I_f^n\right) - a\right)^2, \tag{8}$$

where $D_{\theta_D}(I_v)$ stands for the discrimination result given by the discriminator on visible light images and $D_{\theta_D}\left(I_f\right)$ denotes the discrimination result of the discriminator on fused images; b and a are soft labels. To ensure that the discriminator loss achieves a minimum value, the soft label b is set to (0.7, 1.2), and a is set to (0, 0.3).

4. Experiments and Results

4.1. Experimental Design and Evaluation Metrics

The experiment in this article used some of the data from the TNO dataset [31], selecting 40 pairs of infrared and visible images as the training set, and 20 pairs of infrared and visible images as the test set.

In the generator loss function, the weight coefficient of the content loss term is set to $\lambda = 100$; λ is an experience value. The weight coefficient of the infrared and visible intensity gradient loss term is set to $\beta_1 = 1.2$, $\beta_2 = 5$, $\beta_3 = 1$, $\beta_4 = 3$. The experimental environment is based on the TensorFlow framework, CUDA version 11.6, and the graphics card used is the GeForce RTX 3090 Ti.

Objective evaluation can provide quantitative evaluation metrics, while subjective evaluation can better reflect human visual perception and preferences. Therefore, when evaluating the image fusion effect, objective evaluation and subjective evaluation can be combined to comprehensively consider the evaluation results and obtain more accurate and comprehensive evaluations. To objectively evaluate the quality of fused images, for this article we selected spatial frequency (SF), entropy (EN), sum of the correlation of differences (SCD), average gradient (AG) [32], standard deviation (SD), structural similarity index measurement (SSIM) [33], peak signal-to-noise ratio (PSNR) [34], edge preservation degree ($Q_F^{A/B}$) [35], visual information fidelity (VIF) [36,37], and mutual information (MI) [38] as evaluation indicators.

Let $A(x,y)$ and $B(x,y)$ represent two source images, and $F(x,y)$ represent the blended image. (x,y) represents the pixel coordinates, and $M \times N$ represents the size of the source images and the blended image.

(1) Spatial frequency (SF), which reflects the change rate of image grayscale, is defined as

$$SF = \sqrt{RF^2 + CF^2}. \tag{9}$$

Among them, RF and CF represent the row frequency and column frequency of the fused image, respectively, as shown in formulas (10) and (11).

$$RF = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - F(i,j-1))^2}, \tag{10}$$

$$CF = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - F(i-1,j))^2}. \tag{11}$$

Generally speaking, the larger the spatial frequency value of the image, the richer the texture detail information it contains.

(2) Entropy (EN), which is used to measure the amount of information contained in an image, is defined as follows:

$$EN = -\sum_{i=0}^{255} p_i \log_2 p_i, \tag{12}$$

where p_i represents the normalized histogram of the gray level i in the fused image. The larger the EN value, the richer the information of the fused image and the better the quality.

(3) The sum of the correlation of differences (SCD), which uses the correlation between the difference images calculated from the source image and the fused image, is defined as follows:

$$SCD = r_{D_A,A} + r_{D_B,B}, \tag{13}$$

where D_A and D_B represent the difference between the fused image and the source images B and A, respectively. r is used to calculate the correlation coefficient of D_A and D_B with the two source images. The larger the sum of differences, the more complementary information in the fused image, and the better the image fusion effect.

(4) Average gradient (AG); this indicator reflects the clarity of the image to a certain extent, and is defined as:

$$AG = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{1}{2}\left(G_x^2(i,j) + G_y^2(i,j)\right)}, \tag{14}$$

$$\begin{cases} G_x(i,j) = G(i,j) - G(i+1,j) \\ G_y(i,j) = G(i,j) - G(i,j+1) \end{cases}. \tag{15}$$

The larger the average gradient value of the image, the higher the image definition, and $G(i,j)$ represents the gradient value at the pixel point (i,j).

(5) The standard deviation (SD) is an indicator that reflects the degree of dispersion of image gray values relative to the mean gray value. Its definition is as follows:

$$SD = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - \mu)^2}. \tag{16}$$

The larger the standard deviation, the more dispersed the gray level distribution of the image, and the greater the contrast of the fused image. μ represents the average gray value of the image.

(6) Structural similarity (SSIM); this index comprehensively evaluates the degree of similarity between two images by comparing brightness, contrast, and structure information, and is defined as:

$$SSIM(A,F) = [l(A,F)]^{\alpha} \cdot [c(A,F)]^{\beta} \cdot [s(A,F)]^{\gamma}, \tag{17}$$

where $l(A,F)$, $c(A,F)$, and $s(A,F)$ represent brightness distortion, contrast distortion, and structure distortion functions, respectively, and the importance of the three functions in the SSIM factor is defined as follows:

$$l(A,F) = \frac{2\mu_A\mu_F + C_1}{\mu_A^2 + \mu_F^2 + C_1}, \tag{18}$$

$$c(A,F) = \frac{2\sigma_A\sigma_F + C_2}{\sigma_A^2 + \sigma_F^2 + C_2}, \tag{19}$$

$$s(A,F) = \frac{\sigma_{AF} + C_3}{\sigma_A\sigma_F + C_3}, \tag{20}$$

where μ_A and μ_F represent the pixel mean of the source image A and the fusion image F, respectively; σ_A and σ_F represent the standard deviations of the source image A and the fusion image F, respectively; σ_{AF} represent the covariance between the source image A and the fusion image F; and $C_i(i = 1,2,3)$ is to avoid the parameter introduced by the denominator if the above formula is zero.

The similarity between infrared image, visible image, and fusion image is expressed as:

$$SSIM(A,B,F) = \lambda_A SSIM(A,F) + \lambda_B SSIM(B,F), \tag{21}$$

where λ_A and λ_B are the proportional coefficients of the structural similarity between the infrared image and the fused image, and the structural similarity between the visible image and the fused image, respectively. The larger the index value, the higher the similarity between the fused image and the source image.

(7) Peak signal-to-noise ratio (PSNR); this metric measures the ratio between the effective information of the image and the noise, and is defined as:

$$PSNR = 10\lg\frac{2^n - 1}{MSE}. \tag{22}$$

Among them, n represents the number of image bits; MSE measures the difference between the source image and the fused image based on pixel error. PSNR can reflect whether the image is distorted to a certain extent. The larger the index value, the better the quality of the fusion image.

(8) $Q_F^{A/B}$; this metric is used to measure the amount of edge information transmitted from the source image to the fusion image, and is defined as:

$$Q_F^{A/B} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)\right)}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(w^A(i,j) + w^B(i,j)\right)}, \tag{23}$$

where $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$ represent the similarity measures between the infrared image, the visible image, and the fusion image; $w^A(i,j)$ and $w^B(i,j)$ are the weight coefficients of the source images A and B.

$$Q^{AF}(i,j) = Q_g^{AF}(i,j)Q_\alpha^{AF}(i,j). \tag{24}$$

$Q_g^{AF}(i,j)$ and $Q_a^{AF}(i,j)$ represent the edge strength and orientation value at position (i,j), and are defined as follows:

$$\begin{cases} Q_g^{AF}(i,j) = \Gamma_g\left[1 + e^{k_g\left(G^{AF}(i,j) - \sigma_g\right)}\right]^{-1} \\ Q_\alpha^{AF}(i,j) = \Gamma_\alpha\left[1 + e^{k_\alpha\left(A^{AF}(i,j) - \sigma_\alpha\right)}\right]^{-1} \end{cases}, \tag{25}$$

where Γ_g, k_g, σ_g, Γ_α, k_α, σ_α are all constants. The Sobel operator is used to calculate the edge strength and orientation value between the source images A, B and the fused image F, and are defined as follows:

$$\begin{cases} g_A(i,j) = \sqrt{S_A^x(i,j)^2 + S_A^y(i,j)^2} \\ \alpha_A(i,j) = \tan^{-1}\left[\frac{S_A^y(i,j)}{S_A^x(i,j)}\right] \end{cases}, \tag{26}$$

where $G^{AF}(i,j)$ and $A^{AF}(i,j)$ represent the correlation intensity information and correlation angle information, respectively, between the fusion image F and the source image A and are defined as follows:

$$\left(G^{AF}(i,j), A^{AF}(i,j)\right) = \left[\left(\frac{g_F(i,j)}{g_A(i,j)}\right)^M, 1 - \frac{|\alpha_A(i,j) - \alpha_F(i,j)|}{\pi/2}\right], \tag{27}$$

$$M = \begin{cases} 1 & if \quad g_A(i,j) > g_F(i,j) \\ -1 & otherwise \end{cases}. \tag{28}$$

The larger the $Q_F^{A/B}$ value, the more the edge information of the source image is preserved in the fused image.

(9) Visual information fidelity (VIF), which is similar to the human visual system, is used to measure the information fidelity of fusion images.

$$VIFF(A,B,F) = \sum_k p_k \cdot VIFF_k(A,B,F). \tag{29}$$

The larger the index value, the greater the fidelity of the fused visual information and the higher the quality of the fused image.

(10) Mutual information (MI), which measures the amount of source image information extracted from fusion images, is defined as:

$$MI = MI_{AF} + MI_{BF}, \tag{30}$$

where MI_{AF} and MI_{BF} represent the information amount of infrared image and visible image extracted from fusion image, respectively, as shown in formula (31).

$$\begin{cases} MI_{AF} = \sum_{a,f} p_{A,F}(a,f) \log_2 \frac{p_{A,F}(a,f)}{p_A(a)p_F(f)} \\ MI_{BF} = \sum_{b,f} p_{B,F}(b,f) \log_2 \frac{p_{B,F}(b,f)}{p_B(b)p_F(f)} \end{cases}, \tag{31}$$

where $p_{A,F}(a,f)$ and $p_{B,F}(b,f)$ represent the joint probability density distribution of source images A, B and fusion image F, respectively; $p_A(a)$ and $p_B(b)$ represent the probability density distribution of source images A and B, respectively; and $p_F(f)$ represents the probability density distribution of fusion image F. The larger the MI value of the fused image, the more information of the source image is preserved, and the better the fusion quality is.

4.2. Comparative Experiment

To validate the efficacy of the approach presented in this article, we conducted a comparative analysis with seven additional fusion methods: Average gradient (AVG), Laplacian pyramid (LP), filter subtract decimate pyramid (FSDP), discrete wavelet transform (DWT), FusionGAN, Densefuse, and RFN-Nest.

In the experiment, we applied these seven fusion methods to 10 randomly selected pairs of infrared and visible source images. Figure 4 shows the fusion results of different methods on six pairs of source images, which are "airplane_in_trees", "soldier_in_trench_1", "Sandpath", "Fennek01_005", "Kaptein_1123", and "Nato_camp" from left to right. The

sizes of the registered source image pairs are: 595 × 328, 768 × 576, 575 × 475, 749 × 551, 620 × 450, 360 × 270.

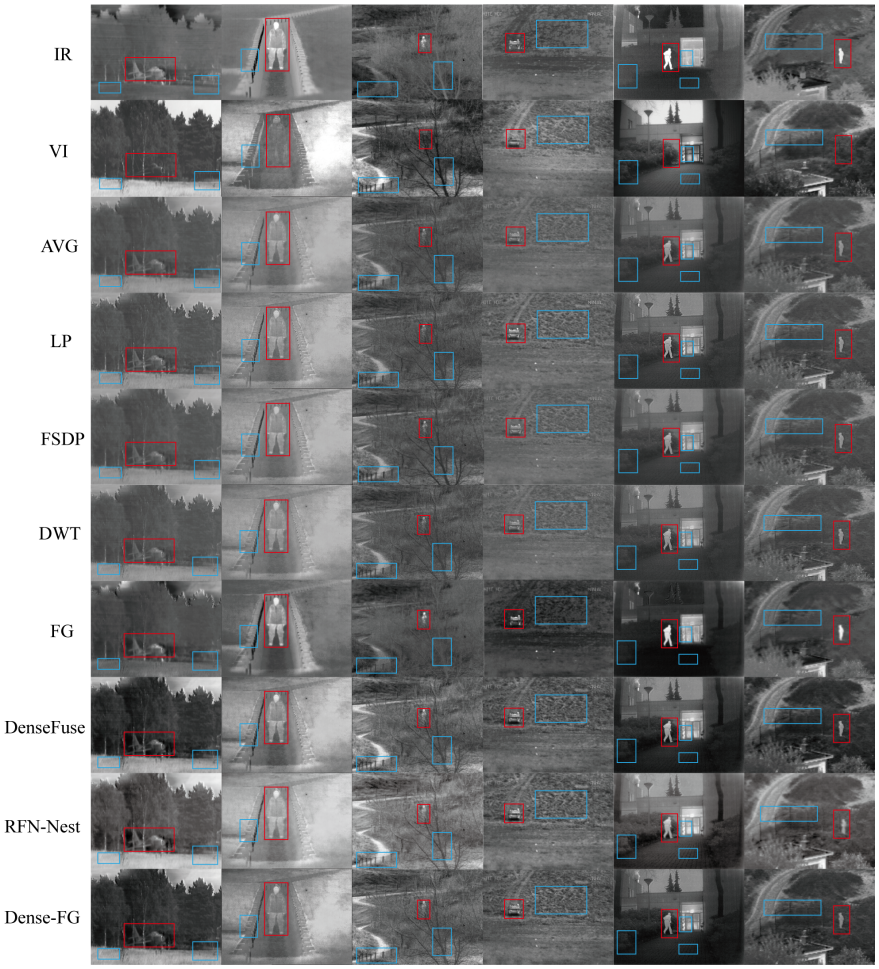

Figure 4. Comparison of fusion results of different algorithms. We label the infrared radiation information with a red box and the visible texture information with a blue box for clear comparison.

The output images of various fusion methods are shown in Figure 4. We used red boxes to mark the infrared radiation information and blue boxes to mark the visible texture information. Based on subjective analysis, traditional methods such as AVG, LP, FSDP, and DWT can effectively fuse the basic information of the source images to some extent. However, the fused images obtained have poor infrared target saliency and low richness in texture details. In addition, the overall tone of the fused images is dark and does not enable a good subjective visual perception.

Although the FusionGAN method can obtain fused images with significantly brighter infrared targets than the other methods,its feature extraction capability is insufficient, the texture details of the fused images are missing, and it cannot effectively fuse visible image information. The DenseFuse method effectively extracts prominent features from the source images through the encoding network, resulting in good target saliency of the fused images. However, due to the insufficient use of source image information during autoencoder training, the preservation of complementary information between source images is reduced, and the clarity of fused image details is reduced. The RFN-Nest method uses fusion learning through encoding–decoding and residual networks to generate images with clear infrared target contours, but lacks image detail information, resulting in blurriness of fine detail areas and high image noise. In this paper's method, dense connection blocks are introduced in both the generator and discriminator network structures to better utilize the source image information and retain more complementary

information. The texture details are more abundant and clearer, and the important features of the image pair have higher fusion quality and better visual effects.

The histograms of the output images for various fusion algorithms are shown in Figure 5. The pixel distribution information of the images can be clearly seen from the histograms. By comparison, the histograms of the fusion results of the four traditional methods, AVG, LP, FSDP, and DWT, are almost identical in shape, with discontinuous pixel distribution, resulting in brightness jumps and missing information in the images. The pixel peak values of the images generated by the four deep learning fusion methods, FG, DenseFuse, RFN-Nest, and Dense-FG, are all between those of the infrared and visible images, but their shapes are different. The fusion result of FG is closer to the infrared image, indicating that the infrared target information is more prominent in the FG fusion image, while the visible information is weaker. The histograms of the fusion results of DenseFuse, RFN-Nest, and Dense-FG are similar in shape, and they are closer to the visible image. The pixel frequency of the DenseFuse histogram is prominent and layered, indicating that the visible details in the DenseFuse fusion image are abundant, and the image details are prominent. The pixel distribution of the RFN-Nest is relatively continuous, indicating that the image is delicate and uniform, but with blurred edges. The Dense-FG histogram has a layered distribution, but the frequency values of prominent pixels are low, indicating that the image details are more prominent. At the same time, it takes the delicacy of the image into account, and the overall visual impression is good. The width of the histogram can clearly reflect the contrast of the image. Among the four deep learning fusion methods, DenseFuse has the highest width and the greatest contrast, followed by RFN-Nest, Dense-FG, and FG, in that order.

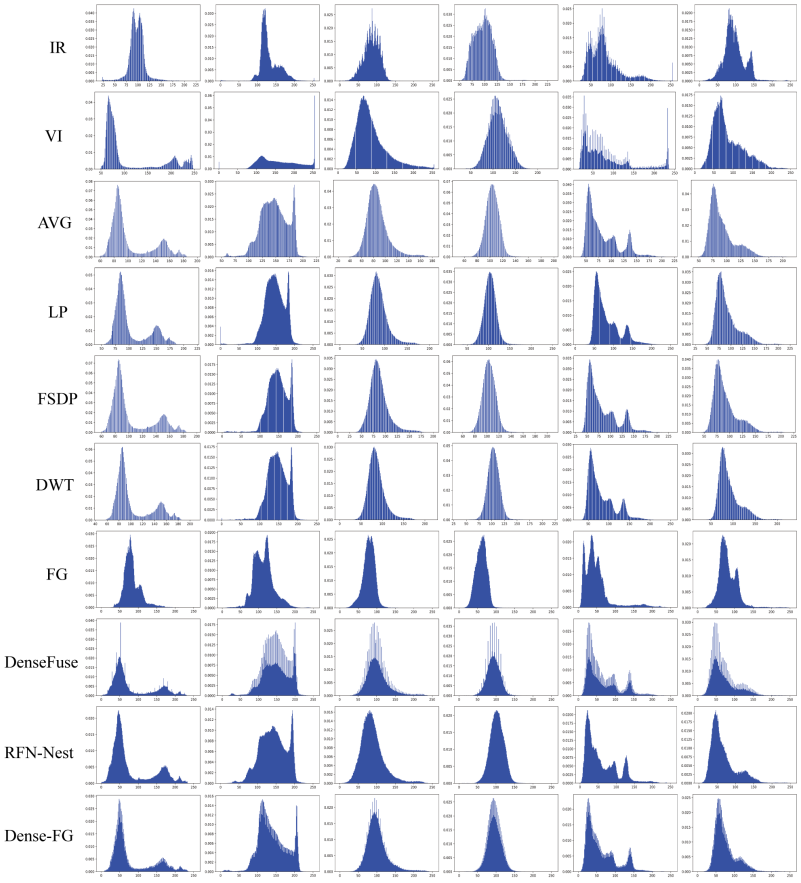

Figure 5. Comparison of histograms of fused results from different algorithms.

The average values of the fusion images generated by different algorithms under various evaluation metrics are shown in Table 3. The metrics represent the richness of

texture details, information entropy, complementary information, clarity, contrast, similarity with the source image structure, richness of edge information, image fidelity, preservation of source image information, and visual information fidelity of the fusion image. The metrics are positively correlated, with larger values indicating higher quality fusion images.

Table 3. Mean values of image metrics for different fusion algorithms .

Method	SF	EN	SCD	AG	SD	SSIM	PSNR	$Q_F^{A/B}$	VIF	MI
AVG	6.522	6.084	1.5361	3.4586	21.7940	**0.7872**	**19.1160**	0.3672	0.3590	2.0310
LP	**10.985**	6.2214	1.5774	**5.8541**	23.7810	0.7496	18.9590	**0.5113**	0.3950	1.7672
FSDP	9.247	6.1346	1.4970	4.9874	22.6290	0.7513	18.9800	**0.5045**	0.4420	1.8655
DWT	10.604	6.1636	1.5565	5.7666	22.8310	**0.7530**	**19.0090**	0.4435	0.3360	1.7352
Fusion-GAN	6.592	6.2567	1.4009	3.4929	25.6690	0.6519	16.0430	0.2893	0.2580	2.0750
Dense-Fuse	9.729	6.6278	**1.7955**	5.2541	**33.5800**	0.7150	17.4820	0.4465	0.4230	**2.1868**
RFN-Nest	7.033	**6.7754**	1.7646	3.6195	**34.6520**	0.6808	17.021	0.3951	**0.4490**	1.9810
Dense-FG	**11.276**	**6.6750**	**1.8053**	**6.1461**	32.5990	0.7227	18.2110	0.4627	**0.4970**	**2.2712**

Table 3 presents the comparative results of eight methods under 10 metrics, with the optimal values highlighted in red bold and the second optimal values highlighted in blue bold. From Table 3, it can be seen that for the SF metric, the proposed method outperforms DenseFuse by about 15.9%, FusionGAN by about 71.06%, and LP by about 2.65%, with the fusion image containing rich texture details. For the AG metric, the proposed method outperforms DenseFuse by about 16.98%, FusionGAN by about 75.96%, and RFN-Nest by about 69.81%, with high image clarity. For the SD metric, the proposed method slightly decreases compared to DenseFuse and RFN-Nest, but compared to FusionGAN and four other traditional fusion methods, the metric values are improved by about 26.99%, 49.58%, 37.08%, 44.06%, and 42.78%, with high contrast of the fusion image consistent with the histogram comparison. For the EN and SCD metrics, the proposed method achieves the optimal and second optimal values, respectively, and can extract more source image information. According to the VIF index, the method proposed in this article improved image quality by about 17.49% compared to the DenseFuse method, about 92.64% compared to the FusionGAN method, and about 10.69% compared to the RFN-Nest method. The fused image has a high visual fidelity. In terms of the MI index, the method proposed in this article improved it by about 3.86% compared to the DenseFuse method, about 9.46% compared to the FusionGAN method, and about 14.65% compared to the RFN-Nest method. The fused image retains more information from the source images.

The above analysis shows that the four traditional methods, AVG, LP, DWT, and FSDP, have strong performance in terms of structural similarity and image fidelity, but they have low image contrast and lack texture details. They can be used in special fusion scenes that require high structural similarity. The FusionGAN method tends to retain infrared image information; the DenseFuse method produces fusion images that tend to contain visible light information, with high contrast and visual fidelity, and contain more information, but the texture details at the edges are coarse, and some detailed features are submerged. RFN-Nest performs best in terms of information retention, but the details are presented in a general way, with some areas appearing blurry. Overall, our method is slightly superior to the other seven methods in terms of comprehensive performance and can meet the application requirements of more scenarios.

A comparison chart of the spatial frequency (SF), entropy(EN), sum of the correlation of differences (SCD), average gradient (AG), standard deviation (SD), structural similarity index measurement (SSIM), edge preservation degree ($Q_F^{A/B}$), peak signal-to-noise ratio

(PSNR), visual information fidelity (VIF), and mutual information (MI) index values for the fusion results of the seven algorithms mentioned above, is shown in Figure 6. The proposed method Dense-FG has the best fusion performance.

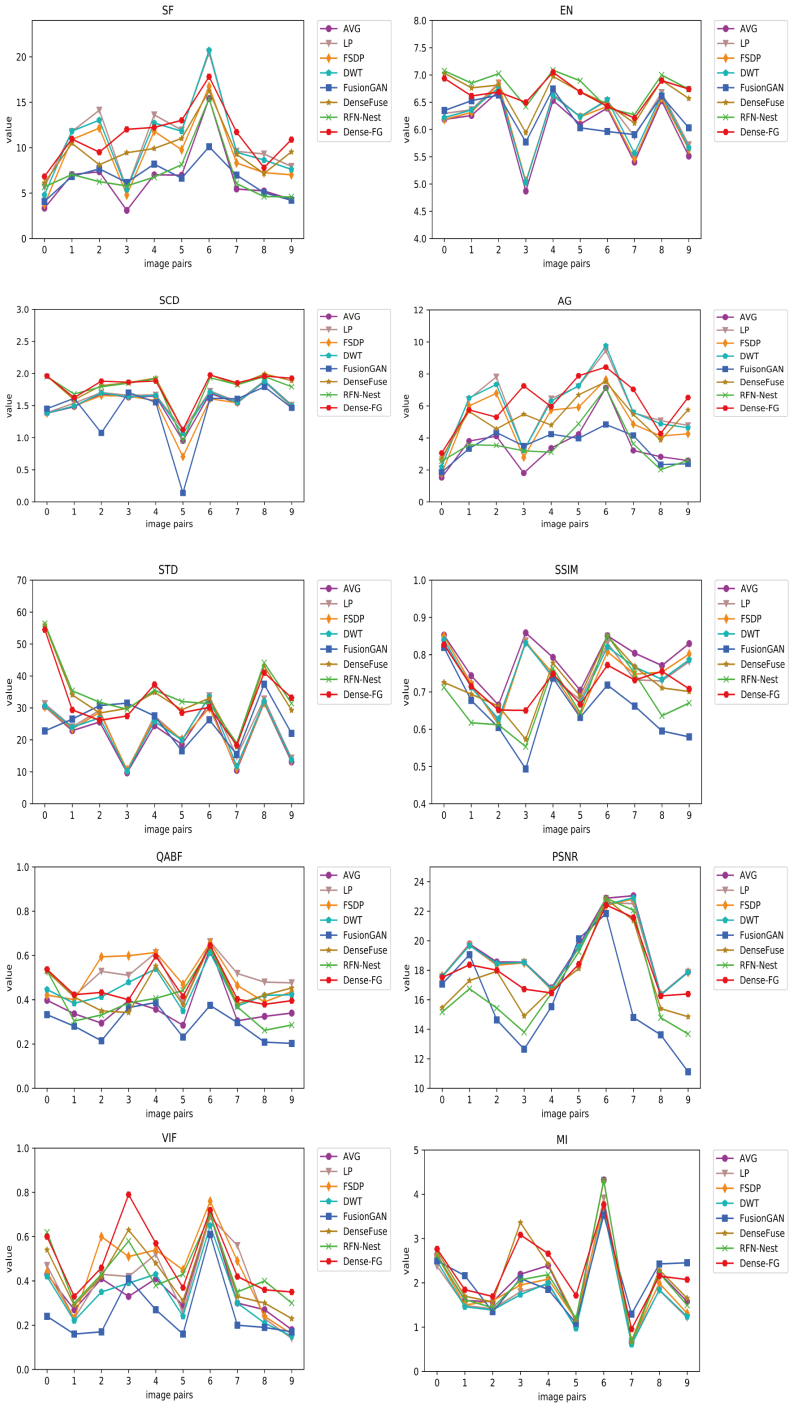

Figure 6. Objective evaluation of the fusion results.

4.3. Ablation Experiment

The image fusion method in this article includes two key designs, namely the introduction of dense connection blocks in the generator and the introduction of dense connection blocks in the discriminator. To verify the impact of these two designs on fusion performance, we conducted ablation experiments from two aspects: the generator network structure and

the discriminator network architecture. The experiments were divided into three groups and compared and analyzed with the method, which is described in this chapter.

Experiment A introduced dense connection blocks in the generator network architecture during network training, while the discriminator network architecture used conventional convolutional layers. Experiment B used conventional convolutional layers in the generator during network training, while the discriminator network architecture introduced dense connection blocks. Experiment C used conventional convolutional layers in both the generator and the discriminator network architecture during network training. Ten sets of source images were chosen at random for testing, and both subjective and objective evaluations were conducted on the fusion outcomes obtained from the three experimental groups as well as the method proposed in this article.

The findings from the experiment are presented in Figure 7. Compared to the method proposed in this paper, the fusion images generated by the other three groups of experiments have more blurred texture details and mediocre object saliency.

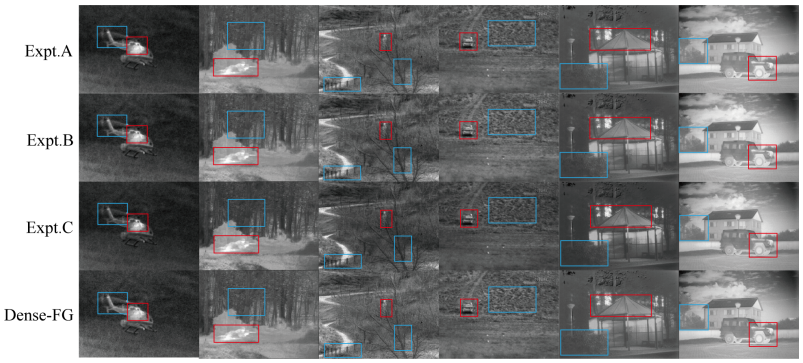

Figure 7. Partial results of ablation experiments. We label the infrared radiation information with a red box and the visible texture information with a blue box for clear comparison.

The average evaluation metrics for fused images are compared in Table 4. The results have shown that our Dense-FG approach outperforms all other methods in six key metrics, namely SF, SCD, AG, SD, $Q_F^{A/B}$, and PSNR. This indicates that introducing dense connection blocks can effectively enhance the network's ability to extract feature information from the source images and further improve the clarity of the fused images.

Table 4. Mean values of image metrics for different fusion algorithms .

Expt.	SF	EN	SCD	AG	SD	SSIM	PSNR	$Q_F^{A/B}$	VIF	MI
A	8.569	6.7146	1.5124	4.7516	30.954	**0.6897**	15.460	0.4247	0.436	1.8662
B	8.728	6.6034	1.7125	4.8224	31.597	0.6863	15.630	0.4295	**0.477**	**1.9733**
C	8.489	**6.8226**	1.6903	4.6403	31.308	0.6704	14.589	0.4106	0.461	1.8960
Dense-FG	**10.355**	6.756	**1.8142**	**5.8026**	**31.766**	0.6804	**15.959**	**0.4387**	0.447	1.8412

Bold format represents optimal value for each column.

In conclusion, the incorporation of dense connection blocks in both the generator and discriminator are the two crucial design features of our image fusion approach discussed in this chapter. These design choices have considerable impacts on enhancing the performance of our image fusion method.

5. Conclusions

This article proposes a fusion model that uses dense connection blocks to improve the performance of generative adversarial networks in infrared and visible image fusion. Dense connection blocks are introduced in both the generator and discriminator networks to effectively extract complementary information from the source images and enhance

the fusion performance. The proposed method was experimentally validated using the TNO dataset.

In the comparative experiment section, seven other methods and our proposed method, Dense-FG, were used to fuse the same image pairs, and qualitative analysis was performed by calculating the comparative histogram and ten indicators, including SF, EN, SCD, AG, SD, SSIM, $Q_F^{A/B}$, PSNR, VIF, and MI. The results of the comparative histogram and indicators show consistency in their conclusions.

Traditional methods, including AVG, LP, DWT, and FSDP, exhibit strong performance in structural similarity and image fidelity, but have low image contrast and lack texture details. These methods can be used in special fusion scenarios that require high structural similarity. The FusionGAN method tends to preserve infrared image information, while the DenseFuse method tends to preserve visible information, with high contrast and visual fidelity, but with coarse texture details, and is overwhelmed by details. RFN-Nest performs best in preserving information, but presents average detail with some blurred areas. Overall, our proposed method slightly outperforms the other seven methods and can meet the application requirements of many scenarios.

The introduction of Dense-FG dense residual blocks enables feature connections between different layers, improves network efficiency, enhances the network's ability to extract source image information, and, compared to other fusion methods, the fused image has rich texture details, high clarity, high retention of source image feature information, and good subjective perception. The introduction of the GAN network's adversarial thinking enables a better balance between the brightness of infrared targets and visible details. Furthermore, the ablation experiment results indicate that the introduction of dense connection blocks in both the generator and discriminator plays a crucial role in improving fusion performance.

The fusion technology of infrared and visible images is in extensive demand in medical, remote sensing, and other engineering fields. In the future, we will also apply the method proposed in this paper to engineering practice and enable infrared and visible fusion technology to play an important role.

Author Contributions: X.X. wrote the draft; Y.S. gave professional guidance and edited; S.H. gave advice and edited. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by National Natural Science Foundation of China General Program 51979062.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data openly available in a public repository.

Conflicts of Interest: All authors disclosed no relevant relationships.

References

1. Wang, Z.; Ziou, D.; Armenakis, C.; Li, D.; Li, Q. A comparative analysis of image fusion methods. *IEEE Trans. Geoence Remote Sens.* **2005**, *43*, 1391–1402. [CrossRef]
2. He, C.; Liu, Q.; Li, H.; Wang, H. Multimodal medical image fusion based on IHS and PCA. *Procedia Eng.* **2010**, *7*, 280–285. [CrossRef]
3. Xu, D.; Wang, Y.; Xu, S.; Zhu, K.; Zhang, N.; Zhang, X. Infrared and Visible Image Fusion with a Generative Adversarial Network and a Residual Network. *Appl. Sci.* **2020**, *10*, 554. [CrossRef]
4. Simone, G.; Farina, A.; Morabito, F.C.; Serpico, S.B.; Bruzzone, L. Image fusion techniques for remote sensing applications. *Inf. Fusion* **2002**, *3*, 3–15. [CrossRef]
5. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [CrossRef]
6. Sun, C.; Zhang, C.; Xiong, N. Infrared and Visible Image Fusion Techniques Based on Deep Learning: A Review. *Electronics* **2020**, *9*, 2162. [CrossRef]
7. Shukla, P.; Nasrin, S.; Darabi, N.; Gomes, W.; Trivedi, A.R. MC-CIM: Compute-in-Memory With Monte-Carlo Dropouts for Bayesian Edge Intelligence. *IEEE Trans. Circuits Syst. Regul. Pap.* **2022**, *70*, 884–896. [CrossRef]

8. Bekman, T.; Abolfathi, M.; Jafarian, H.; Biswas, A.; Banaei-Kashani, F.; Das, K. Practical Black Box Model Inversion Attacks Against Neural Nets. In Proceedings of the Machine Learning and Principles and Practice of Knowledge Discovery in Databases: International Workshops of ECML PKDD 2021, Virtual Event, 13–17 September 2021; pp. 39–54. [CrossRef]
9. Nejatishahidin, N.; Fayyazsanavi, P.; Košecka, J. Object pose estimation using mid-level visual representations. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 13105–13111. [CrossRef]
10. Li, S.; Yang, B.; Hu, J. Performance comparison of different multi-resolution transforms for image fusion. *Inf. Fusion* **2011**, *12*, 74–84. [CrossRef]
11. Pu, T.; Ni, G. Contrast-based image fusion using the discrete wavelet transform. *Opt. Eng.* **2000**, *39*, 2075–2082. [CrossRef]
12. Burt, P.J.; Adelson, E.H. The Laplacian Pyramid as a Compact Image Code. *Readings Comput. Vis.* **1987**, *31*, 671–679. [CrossRef]
13. Zhang, Q.; Guo, B.l. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Process.* **2009**, *89*, 1334–1346. [CrossRef]
14. Yang, Y.; Tong, S.; Huang, S.; Lin, P. Multifocus Image Fusion Based on NSCT and Focused Area Detection. *IEEE Sensors J.* **2015**, *15*, 2824–2838. . [CrossRef]
15. Liu, G.; Yang, W. Multiscale contrast-pyramid-based image fusion scheme and its performance evaluation. *Acta Optica Sinica* **2021**, *21*, 1336–1342. .:0253-2239.2001.11.013. [CrossRef]
16. Nencini, F.; Garzelli, A.; Baronti, S.; Alparone, L. Remote sensing image fusion using the curvelet transform. *Inf. Fusion* **2007**, *8*, 143–156. [CrossRef]
17. Meher, B.; Agrawal, S.; Panda, R.; Abraham, A. A survey on region based image fusion methods. *Inf. Fusion* **2019**, *48*, 119–132. [CrossRef]
18. Stathaki, T. *Image Fusion: Algorithms and Applications*; Elsevier: Amsterdam, The Netherlands, 2011.
19. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]
20. Hogervorst, M.A.; Toet, A. Fast natural color mapping for night-time imagery. *Inf. Fusion* **2010**, *11*, 69–77. [CrossRef]
21. Li, H.; Wu, X.J. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Trans. Image Process.* **2019**, *28*, 2614–2623. [CrossRef] [PubMed]
22. Li, H.; Wu, X.j.; Durrani, T.S. Infrared and visible image fusion with ResNet and zero-phase component analysis. *Infrared Phys. Technol.* **2019**, *102*, 103039. [CrossRef]
23. Liu, Y.; Chen, X.; Cheng, J.; Peng, H.; Wang, Z. Infrared and visible image fusion with convolutional neural networks. *Int. J. Wavelets, Multiresolution Inf. Process.* **2018**, *16*, 1850018. [CrossRef]
24. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [CrossRef]
25. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 4980–4995. [CrossRef]
26. Long, Y.; Jia, H.; Zhong, Y.; Jiang, Y.; Jia, Y. RXDNFuse: A aggregated residual dense network for infrared and visible image fusion. *Inf. Fusion* **2021**, *69*, 128–141. [CrossRef]
27. Yang, Y.; Liu, J.; Huang, S.; Wan, W.; Wen, W.; Guan, J. Infrared and Visible Image Fusion via Texture Conditional Generative Adversarial Network. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 4771–4783. [CrossRef]
28. Li, J.; Huo, H.; Li, C.; Wang, R.; Feng, Q. AttentionFGAN: Infrared and Visible Image Fusion Using Attention-Based Generative Adversarial Networks. *IEEE Trans. Multimed.* **2021**, *23*, 1383–1396. . [CrossRef]
29. Xu, H.; Zhang, H.; Ma, J. Classification Saliency-Based Rule for Visible and Infrared Image Fusion. *IEEE Trans. Comput. Imaging* **2021**, *7*, 824–836. [CrossRef]
30. Jiang, Q.; Jin, X.; Hou, J.; Lee, S.J.; Yao, S. Multi-Sensor Image Fusion Based on Interval Type-2 Fuzzy Sets and Regional Features in Nonsubsampled Shearlet Transform Domain. *IEEE Sensors J.* **2018**, *18*, 2494–2505. [CrossRef]
31. Toet, A. The TNO multiband image data collection. *Data Brief* **2017**, *15*, 249–251. [CrossRef]
32. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
33. Sheikh, H.; Bovik, A. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef] [PubMed]
34. Yim, C.; Bovik, A.C. Quality Assessment of Deblocked Images. *IEEE Trans. Image Process.* **2011**, *20*, 88–98. [CrossRef] [PubMed]
35. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003. [CrossRef]
36. Wang, Z.; Bovik, A. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
37. Xydeas, C.S.; Pv, V. Objective image fusion performance measure. *Mil. Tech. Cour.* **2000**, *56*, 181–193. [CrossRef]
38. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and Panchromatic Data Fusion Assessment Without Reference. *ASPRS J. Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.