RESEARCH ARTICLE

MGFuse: An Infrared and Visible Image Fusion Algorithm Based on Multiscale Decomposition Optimization and Gradient-Weighted Local Energy

HONGTAO HAO [ID], BINGJIAN ZHANG [ID], AND KAI WANG
School of Mechanical Engineering, Ningxia University, Yinchuan 750021, China
Ningxia Key Laboratory of CAE on Intelligent Equipment, Yinchuan 750021, China

Corresponding author: Hongtao Hao (haoht_03@126.com)

ABSTRACT Existing image fusion algorithms have difficulty in effectively preserving valuable target features in infrared and visible images, which easily introduces blurry edges and unremarkable notable targets during their fusion process. We propose the MGFuse algorithm as a solution to this problem, which is a novel fusion algorithm that utilizes multiscale decomposition optimization and gradient-weighted local energy. Initially, non-subsampled shearlet transform (NSST) is applied to partition both the infrared and visible images into several high-frequencies and low-frequencies components. Subsequently, the acquired low frequencies continue to be decomposed via the proposed optimization function to get base layers and texture layers, which can optimize the quality of image edges and preserve fine-grained details, respectively. In addition, we have formulated an intrinsic attribute-based energy (IAE) fusion scheme to merge the two base layers. The texture layers and high-frequencies are extracted by gradient-weighted local energy (GE) operator based on structure tensor, which is employed to construct the fusion strategy for these parts. At last, the acquired texture and base parts are linearly combined to get the integrated low-frequency layer on which the final image is acquired using inverse NSST. Numerous experimental observations demonstrate that our MGFuse algorithm achieves superior fusion capability than the reference nine advanced algorithms in both qualitative and quantitative assessment, and robustness to noisy images with different noise levels.

INDEX TERMS Image fusion, multiscale decomposition optimization, gradient-weighted local energy, structure tensor.

I. INTRODUCTION

Infrared and visible image fusion has been extensively utilized across multiple disciplines, including but not limited to remote sensing [1], clinical diagnosis [2], object detection [3], and modern military [4]. As an effective image processing technology, infrared and visible image fusion can seamlessly combine the complementary target features from multiple source images to generate an image with robustness and informativeness according to a certain fusion algorithm [5]. In recent years, there has been a surge in the number of fusion algorithms proposed to augment the effectiveness and performance of image fusion techniques. The most commonly used approaches include multiscale transform (MST) [6], [7], [8], [9], sparse representation (SR) [10], [11], and deep learning (DL) [12], [13], [14].

Currently, MST-based algorithms are the most widely studied and applied among these fusion algorithms, including wavelet transform [15], Laplacian pyramid transform [16], non-subsampled shear wave transform [17], and non-subsampled contourlet transform [18]. The typical procedure followed by these algorithms involves decomposing

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan [ID].

the source images into multiple frequency scales, and subsequently devising distinct fusion strategies for each of these parts. Generally, despite the potential for structural distortions in the fused images, these algorithms are known to produce superb visual outcomes by successfully retaining the texture details of the original images.

With the continuous development of image representation theory, SR-based algorithms continue to receive a great deal of attention and many relevant algorithms have been widely introduced over the past few years, including JSR [19], CSR [20], LRR [21], and MDLatLRR [22]. Usually, the SR-based algorithms are mainly divided into three procedures: 1) It decomposes source images into the corresponding sparse coefficients based on the learned super-complete dictionary. 2) Then well-designed fusion strategies are utilized to combine corresponding components, and 3) at last, the final image is given through image reconstruction. The key tasks of SR-based algorithms lie in construct a dictionary and sparse coefficients decomposition. However, it is hard to perform feasible and valid dictionary learning offline due to limited dictionary information representation capabilities, and dictionary learning is also computationally complex due to multiple iterations.

DL-based algorithms have been widely applied in the field of image fusion owing to their excellent feature extraction and comprehensive information representation capabilities. For example, DeepFuse [23], and DCGAN [24] algorithms are respectively based on convolutional neural networks (CNNs) [25], [26] and generative adversarial networks (GANs) [27] framework. Though these algorithms can well achieve fused effects, the existing deficiencies are also evident due to the requirement of a large training dataset and powerful hardware support. Meanwhile, non-sufficient ground truth images are not satisfactory for the training of the network model. As a result, it is difficult for DL-based algorithms to obtain the appropriate model parameters, which can result in the poor-quality fused images.

As mentioned above, these advanced fusion algorithms demonstrate outstanding fusion capabilities for image fusion tasks compared to some conventional algorithms. However, these algorithms may suffer from the following weaknesses: 1) Insufficiently thorough image decomposition would be inevitably structural or brightness distortion, 2) most existing algorithms would not develop appropriate fusion rules, resulting in a lack of valuable feature information in the fusion results, and 3) these advanced algorithms cannot satisfy the demands of both noise-free and noise interference.

To tackle the issues mentioned above and enhance the effectiveness of image fusion, we propose a novel fusion algorithm called MGFuse, which utilizes multiscale decomposition optimization and gradient-weighted local energy to thoroughly optimize and improve the quality of the fused images. This MGFuse can largely retain the detailed information and energy of source images. Firstly, the source images

are divided into several high-frequencies and low-frequencies using NSST. Subsequently, the acquired low frequencies continue to be decomposed via the proposed optimization function to get base layers and texture layers, which can optimize the quality of image edges and preserve fine-grained details, respectively. Then, an intrinsic attribute-based energy (IAE) fusion rule is designed to combine the base layer. Moreover, the high-frequencies and texture components are extracted via the gradient-weighted local energy (GE) operator based on the structure tensor. At last, the acquired texture and base parts are linearly combined to get the integrated low-frequency layer on which the final image is acquired using inverse NSST. Numerous qualitative and quantitative experiments indicate that our MGFuse outperforms nine advanced image fusion algorithms since the proposed MGFuse retains abundant meaningful visible texture information and salient infrared targets from noise-free and noise-perturbed images with different noise levels, which is beneficial to human eyes perception.

In summary, the primary contributions of our research are as follows:

(1) The multiscale decomposition optimization is designed for precise and thorough decomposition of the global and local texture details from low-frequency. Based on the decomposition of the base layer (global features and contents) and texture layer (texture and details) via the optimization function, which can obtain better fusion performance in both noise-free and noise-perturbed images.

(2) Since the base layer contain the majority of the approximate features of the source images, an intrinsic attribute-based energy (IAE) fusion scheme is designed. This fusion scheme can greatly highlight the valuable target features of the input images. Furthermore, we formulated the gradient-weighted local energy (GE) fuse strategy to fuse the texture layers and high-frequencies respectively, which can detect more target features of visual structures or objects by comparing the gradient-weighted local energy differences between the texture layers or high-frequencies, and facilitates the preservation of the global and local texture details in the texture layers.

(3) The advantage of anti-noise robustness allows our MGFuse to be flexibly used in the situation of both noise-free and noise-perturbed image fusion. Meanwhile, our MGFuse has excellent extensibility for the other three kinds of image fusion problems, which verifies the fact that it has excellent generalization ability.

The layout of the subsequent sections in this paper is as follows. In Section II, we display the brief theory of relevant works, including the definition of non-subsampled shearlet transform and structure tensor. Section III describes the multiscale decomposition optimization and fusion rules in base layers and texture layers. Then, section IV presents the research findings along with a detailed analysis of the results. Finally, brief concluding remark with prospective future tasks is shown in Section V.

II. TECHNICAL BACKGROUND

In this section, the related algorithms on which our MGFuse is based are briefly reviewed. The reasons for using these algorithms are that NSST can obtain optimized representation for target image extraction, and structure tensor can represent the structural and spatial information of an image. The content and principle of the nonsubsampled shearlet transform (NSST) and structure tensor are introduced as follows.

A. NON-SUBSAMPLED SHEARLET TRANSFORM

In comparison to earlier MST analysis algorithms for image representation, the shearlet transform (ST) [28] show big strengths over other algorithms in accurately exacting the detailed information of input images at different directions, which can acquire optimized representation for target image extraction. However, the upsampling operation in the standard ST causes its deficiency of shift-invariance and pseudo-Gibbs phenomena, which lead to artifacts effects during image fusion.

To overcome these shortcomings, we put forward a non-subsampled shearlet transform (NSST) which is based on the ST and serves as a multidimensional and multidirectional representation of wavelet transform. The NSST consists of two main parts, which are multiscale decomposition and directional localization. These processes are executed using nonsubsampled pyramid filters (NSPF) and shift-invariant shearlet filter banks (SFB), respectively. Each iteration thereafter continues until the NSPF has decomposed the upper low-frequency subband and reaches the desired decomposition layer. The SFB decomposes the high-frequency coefficients of each level of decomposition to yield multidirectional representations of these coefficients. In this paper, we use NSST as our MST analysis algorithm because of its multiscale, multidirectional, and shift-invariant properties.

B. STRUCTURE TENSOR

The gradient feature in an image can typically be obtained by calculating the difference between the central pixel value and the values of its surrounding pixels. For infrared and visible images, the gradient feature can precisely represent the geometrical features of the significant targets, especially for the edges and contours. The structure tensor has an intimate connection with the first-order directional derivative. The structure tensor can achieve a variety of image processing tasks, such as edge detection [29], image reconstruction [30], image registration [31], image denoising [32], and image fusion [33], [34].

For a source image $g(m, n)$, the square of variation $g(m, n)$ at the location (m, n) in direction φ over the local finite window $\Gamma(m, n)$ can be represented by Eq. (1) for any $\varepsilon \to 0^+$ [29]

$$
\begin{aligned}
(dg)^2 &= \|g(m + \varepsilon \cos \varphi, n + \varepsilon \sin \varphi) - g(m, n)\|_2^2 \\
&\approx \sum_{\Gamma(m,n)} (g_m \varepsilon \cos \varphi + g_n \varepsilon \sin \varphi)^2
\end{aligned}
\tag{1}
$$

Then, the rate of change $Q(\varphi)$ of source image $g(m, n)$ at position (m, n) denoting the local geometric features, which can be presented as

$$
\begin{aligned}
Q(\varphi) &= \sum_{\Gamma(m,n)} (g_m \cos \varphi + g_n \sin \varphi)^2 \\
&= (\cos \varphi, \sin \varphi) \begin{pmatrix} \sum\limits_{\Gamma(m,n)} (g_m)^2 & \sum\limits_{\Gamma(m,n)} g_m g_n \\ \sum\limits_{\Gamma(m,n)} g_n g_m & \sum\limits_{\Gamma(m,n)} (g_n)^2 \end{pmatrix} \\
&\quad \times (\cos \varphi, \sin \varphi)^T \\
&= (\cos \varphi, \sin \varphi)(\sum_{\Gamma(m,n)} \nabla g \nabla g^T)(\cos \varphi, \sin \varphi)^T
\end{aligned}
\tag{2}
$$

where $\nabla g = (g_m, g_n)^T$, g_m and g_n represent the horizontal and vertical gradients at the position (m, n), respectively. ∇ denotes the gradient operator to obtain the gradients in directions m and n. Mathematically, the following positive semi-definite covariance matrix P is called a structure tensor and can be expressed as

$$
P = \sum_{\Gamma(m,n)} \nabla g_i \nabla g_i^T = \begin{bmatrix} K & L \\ L & Z \end{bmatrix}
\tag{3}
$$

where $K = \sum\limits_{\Gamma(m,n)} (g_m)^2$, $L = \sum\limits_{\Gamma(m,n)} g_m g_n$ and $Z = \sum\limits_{\Gamma(m,n)} (g_n)^2$, the structure tensor-based salient detection (STD) operator is calculated as [35]

$$
STD = \sqrt{(\lambda_1 + \lambda_2)^2 + 0.5(\lambda_1 - \lambda_2)^2}
\tag{4}
$$

where λ_1 and λ_2 are the non-negative eigenvalues of the structure tensor and convey the sharp structure information about the local geometry of the image, which can be calculated by

$$
\lambda_{1,2} = \frac{1}{2}((K + Z) \pm \sqrt{(Z - K)^2 + 4L^2})
\tag{5}
$$

Generally, the value of $|\lambda_1 - \lambda_2|$ in Eq. (4) is very small, manifesting that the current pixel is flat with no significant change and is located in the isotropic region. Whereas if $|\lambda_1 - \lambda_2|$ is relatively large, it means that the current pixel is sharp with prominent structures and is located in the anisotropy region.

III. PROPOSED FUSION ALGORITHM

In this section, we elaborate proposed MGFuse algorithm in detail, including image multiscale decomposition optimization, fusion strategies of different layers, and image reconstruction, as displayed in Figure 1.

A. IMAGE MULTISCALE DECOMPOSITION OPTIMIZATION

In our proposed MGFuse algorithm, the infrared image A and visible image B are firstly employed to decompose into a series of high-frequencies (expressing the large-scale variations in intensity) and low-frequency (representing the small-scale variations in intensity) by NSST. And can be expressed as below.

$$
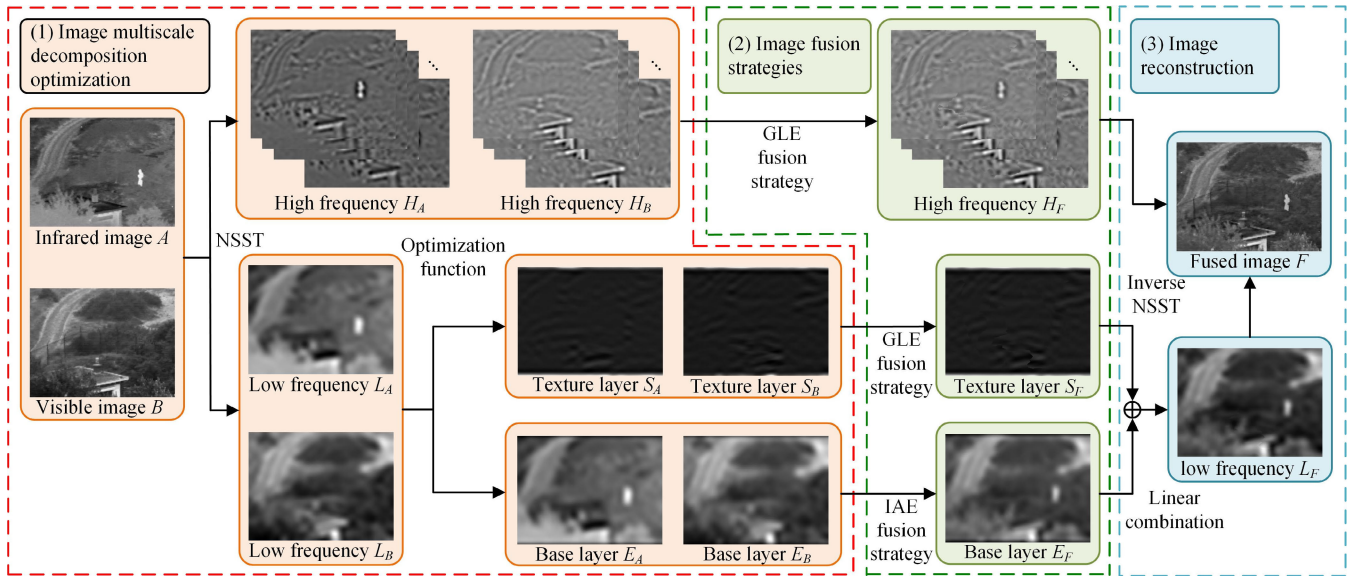\{L_I, H_I\} = NSST\_dec(I)
\tag{6}
$$

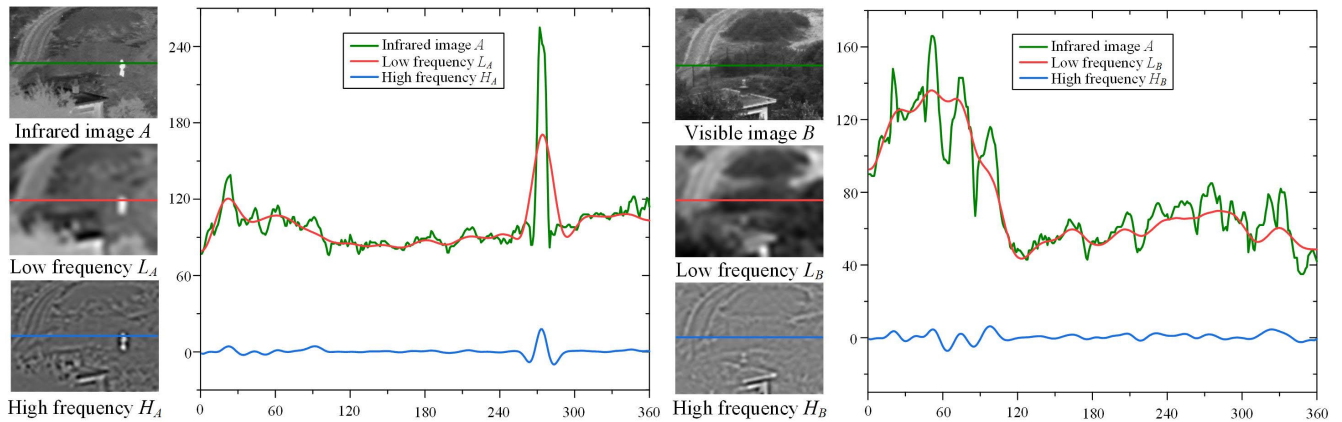**FIGURE 1.** The architecture of our proposed MGFuse.

FIGURE 2. Comparison results of low-frequency and high-frequency information on multiscale decomposition for source images.

where $NSST\_dec(\cdot)$ in Eq. (6) is the function for the NSST decomposition operation from the input images I. The L_I $(I = A, B)$ and H_I are the low-frequency and high-frequency of the input images I, respectively.

Usually, insufficiently thorough image decomposition makes it difficult to develop appropriate fusion rules, which may also affect the fusion quality. The low-frequency retains much energy, while the high-frequencies preserve much sharp texture details. Nevertheless, the presence of remaining global and local texture details are inevitable owing to the NSST decomposition level is not sufficient. Therefore, we should enable the low frequencies as smoothly as possible, which can enhance the fusion effect using only MST-based algorithms. To enable the capturing of image texture details more efficiently, the acquired low frequencies L_I continue to be decomposed via the proposed optimization function to get base layers E_I and texture layers S_I.

To this end, the optimization function of image multiscale decomposition is defined as follows [36]

$$L_I = \arg\min_{E_I} \|E_I - L_I\|_F^2 + \beta \cdot \|E_I \otimes Mask\|_F^2 \quad (7)$$

where the sign $\otimes$ is the convolution operation, which denotes the multiplication of corresponding elements in two matrices. The regularization parameter β is set to 0.01 in our paper, and the $Mask$ matrix is defined as

$$Mask = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (8)$$

The optimization function can be calculated efficiently in the Fourier domain. This means that the closed-form solution for minimizing this problem can be expressed using a specific

mathematical formula.

$$E_I = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(L_I)}{\beta\overline{\mathcal{F}(Mask)}\mathcal{F}(Mask) + \mathcal{F}(1)}\right) \quad (9)$$

Which $\mathcal{F}(\cdot)$, $\mathcal{F}^{-1}(\cdot)$ and $\overline{\mathcal{F}}(\cdot)$ denotes the Fast Fourier Transform, the corresponding inverse transformation, and the complex conjugate operator, respectively. After obtaining the base layer E_I, getting the texture layer S_I is easy by subtracting it from the low-frequency layer L_I.

$$S_I = L_I - E_I \quad (10)$$

B. IMAGE FUSION STRATEGY OF THE BASE LAYER

Infrared and visible images can be divided into high-frequencies, base layers, and texture layers using the multiscale decomposition optimization. As shown in Figure 2, showing the variation characteristics of pixel values at the same location among them. In particular, the low frequency (or base layer) preserves most of the global feature information of the source images, reflecting the approximate characteristics of the source images, in which the intensity of adjacent pixels changes slowly. The high frequency (or texture layer) contains global and local detailed information of the source images, reflecting edge and contour features from different directions and scales, in which the intensity of adjacent pixels changes rapidly. Consequently, to accurately obtain better information interpretation, different factors ought to be considered according to the corresponding characteristics when designing fusion rules.

The base layer contains the most energy information of the source images that has a significant impact on the fusion performance. Furthermore, it is noteworthy that the high intensity and contrast information of the objects often makes great effects on human visual perception. Therefore, the fusion strategy for the base layer is critical for fusing infrared and visible images. The average strategy is traditionally selected to combine the base layers in many fusion rules for a base layer, which may cause less noise and vignetting in the fused images. To address the above-mentioned issues, an intrinsic attribute-based energy (IAE) fusion scheme is presented in the base layer fusion. The intrinsic attribute values of the source image I are computed as

$$IA_I = \frac{\mu_I + M_I}{2} \quad (11)$$

where μ_I and M_I represent the average pixel intensity and the median pixel intensity of the whole image I respectively. Thus, $\mu_I + M_I$ can reflect the degree of pixel intensity. The initial weight map coefficients $w_I(m, n)$ at position (m, n) are as follows

$$w_I(m, n) = \exp(\alpha |E_I(m, n) - IP_I|) \quad (12)$$

where $exp(\cdot)$ represents the exponential operator, and α denotes the adjustable hyperparameter. The final weight map coefficients $W_A(m, n)$ at position (m, n) are obtained by a weighted mean

$$W_A(m, n) = \frac{w_A(m, n)}{w_A(m, n) + w_B(m, n)} \quad (13)$$

where $w_A(m, n)$ and $w_B(m, n)$ are the initial weight map coefficients of the infrared image A and visible image B at position (m, n), respectively. The fused base layer $E_F(m, n)$ coefficients can be ultimately obtained using the following formula (14).

$$E_F(m, n) = E_A(m, n) \cdot W_A(m, n) + E_B(m, n) \cdot (1 - W_A(m, n)) \quad (14)$$

where $E_A(m, n)$ and $E_B(m, n)$ are the base layer coefficients of the infrared image A and visible image B at position (m, n), respectively.

C. FUSION STRATEGY OF TEXTURE LAYER AND HIGH-FREQUENCY

Unlike the base layers, the texture layer and high-frequency usually contain rich details and structural features of the infrared and visible images, such as edges, outlines, and textures. The STD operator can obtain gradient features of the objects, which tend to can be captured by the visual attention and interest of the observers. Consequently, it is reasonable and feasible to apply STD to express the structural and spatial features of input images. However, the STD has difficulty in accurately extracting weak and small structural features, which can lead to a partial loss of detailed information. Based on the above deficiencies, we use the combination of weighted local energy (WLE) and STD operators, called gradient-weighted local energy (GE) operators, to extract the structural information of the input images as much as possible. The WLE operator can express some local energy structure features of the input images, which can effectively enhance the texture layer and high-frequency detection performance of the STD. Therefore, we employ the GE to merge the texture layer and high-frequency respectively. GE is defined as

$$GE_{Q_I}(m, n) = WLE_{Q_I}(m, n) \cdot STD_{Q_I}(m, n) \quad (15)$$

where $STD_{Q_I}(m, n)$ refers to those structural feature coefficients of the layer $Q_I(Q = S, H)$ at the position (m, n). $WLE_{Q_I}(m, n)$ represents the energy feature coefficients of the layer Q_I at position (m, n), which is calculated in a local neighborhood and can be obtained by the following formula:

$$WLE_I(m, n) = \sum_p \sum_q W(p, q)Q_I(m + p, n + q) \quad (16)$$

where $Q_I(m, n)$ is the layer coefficients of the texture layer S_I and high-frequency H_I at position (m, n), and W is a weight matrix that is defined as

$$W = \frac{1}{16}\begin{vmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{vmatrix} \quad (17)$$
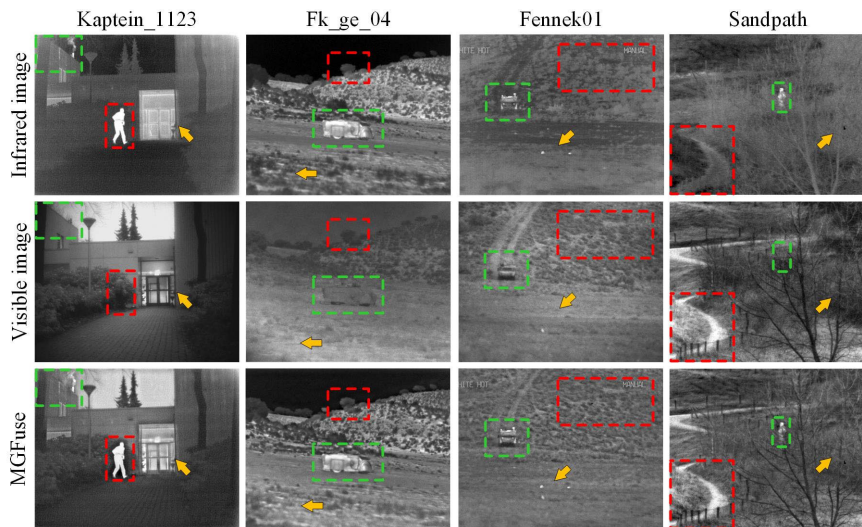
FIGURE 3. Visual fusion results of our MGFuse in 4 pairs of infrared and visible images.

By comparing the GE value of the layer Q_A and Q_B based on the WLE and STD operators at position (m, n), the fused layer $Q_F(m, n)$ coefficients are defined as follows:

$$Q_F(m, n) = \begin{cases} Q_A(m, n), & if \ \left| GE_{Q_I}(m, n) \right| > \left| GE_{Q_I}(m, n) \right| \\ Q_B(m, n), & otherwise \end{cases}$$

(18)

where $Q_A(m, n)$ and $Q_B(m, n)$ are the layer coefficients of the infrared image A and visible image B at position (m, n), respectively.

D. IMAGE RECONSTRUCTION

The fused base layer E_F and texture layer S_F can be acquired by employing the respective fusion rules mentioned in Section B and Section C, as presented previously. The image reconstruction process is bifurcated into two distinct phases: firstly, linear combination, and secondly, inverse NSST. The concept of linear combination describes the addition of the fused base layer E_F and texture layer S_F, thereby facilitating the acquisition of a fused low-frequency L_F output. The linear combination formula can be expressed as follows:

$$L_F = E_F + S_F \tag{19}$$

The inverse NSST technique is employed to reconstruct the fused image F. The calculation formula is as follows Eq. (20). $NSST\_rec(\cdot)$ is the function for the NSST reconstruction operation from the source image I.

$$F = NSST\_rec(L_F, H_F) \tag{20}$$

IV. EXPERIMENTS AND ANALYSIS

We first display experimental settings, including experimental images, nine advanced algorithms, and experimental criteria. Then, we performed several experiments with the proposed algorithm to illustrate the rationality and advantages of our algorithm. Next, our MGFuse is compared with other

algorithms in qualitative and quantitative capability assessment, which demonstrate its superiority in comparison to the competing algorithms. Finally, we perform quantitative assessment experiments to extend to three kinds of image fusion using our MGFuse algorithm.

A. EXPERIMENTAL SETTINGS

We extensively evaluate the effectiveness and advancement of our MGFuse algorithm through an in-depth discussion.

In our experiments, the experimental images are from the publicly available dataset TNO [37]. The TNO comprises a collection of multispectral nighttime imagery of different military-related scenarios under various weather conditions, acquired by different multiband camera systems. There are 21 pairs of source images were elaborately picked for the tests [46]. All the source images were registered and converted to gray-scale images.

The proposed MGFuse algorithm is a novel fusion algorithm based on multiscale decomposition optimization in the NSST domain. We have configured the NSST decomposition level to be 4, and we are utilizing the "maxflat" pyramid filter and the "[2 2 3 3]" directional filter. The adjustable hyperparameter in Eq. (12) is set to 4.

For comparison, our MGFuse is compared with nine advanced and representative algorithms, including four traditional algorithms (WLS [38], LatLRR [39], GFFuse [40], Bayesian [41]), and five deep learning-based algorithms (VggML [42], ResNet [43], DenseFuse [44], GAN [45], RFNNest [46]). To ensure the fairness of the experiment, all these competing algorithms in this paper are implemented with their publicly released codes, and the corresponding configuration criterion of these are entirely conforming to their respective research.

In order to quantitatively evaluate the capability of different advanced algorithms and validate the robustness and rationality of our MGFuse algorithm, six widely recognized
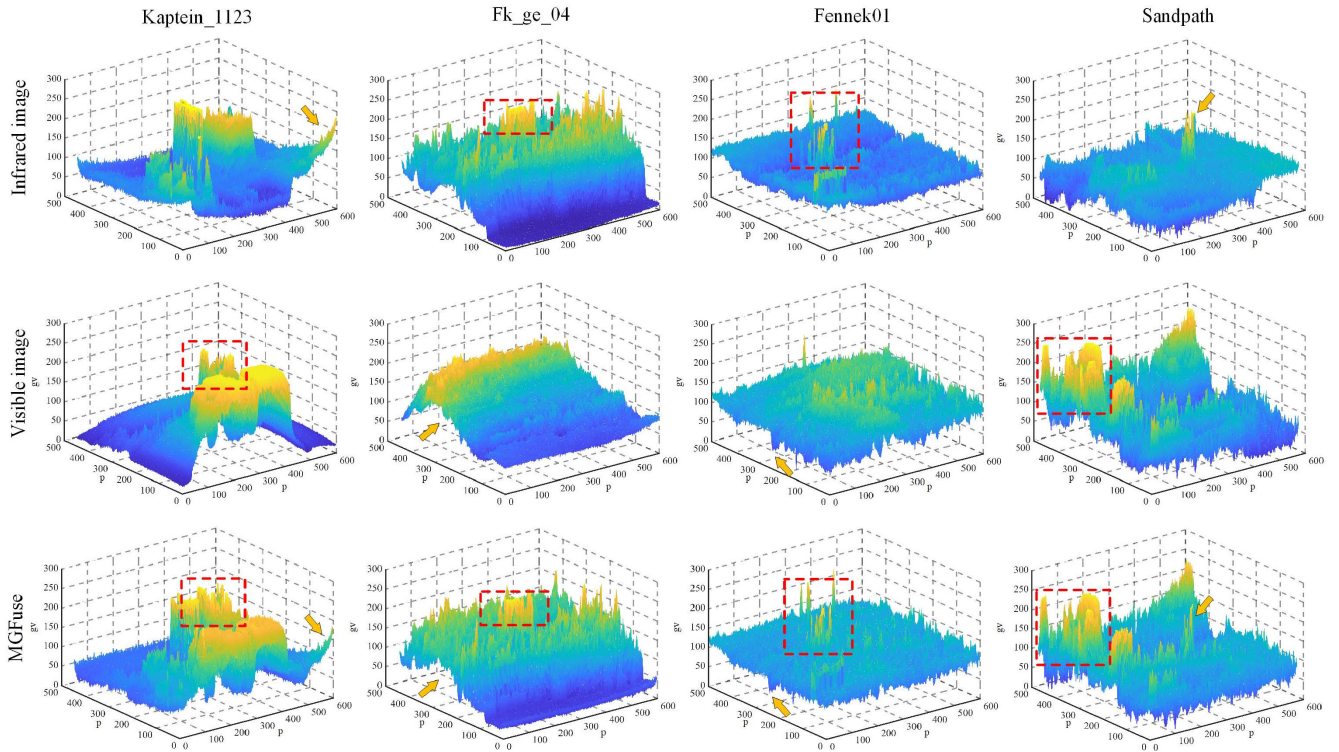
FIGURE 4. Three-dimensional gray-scale images, with "gv" referring to the pixel gray value and "p" representing the pixel position.
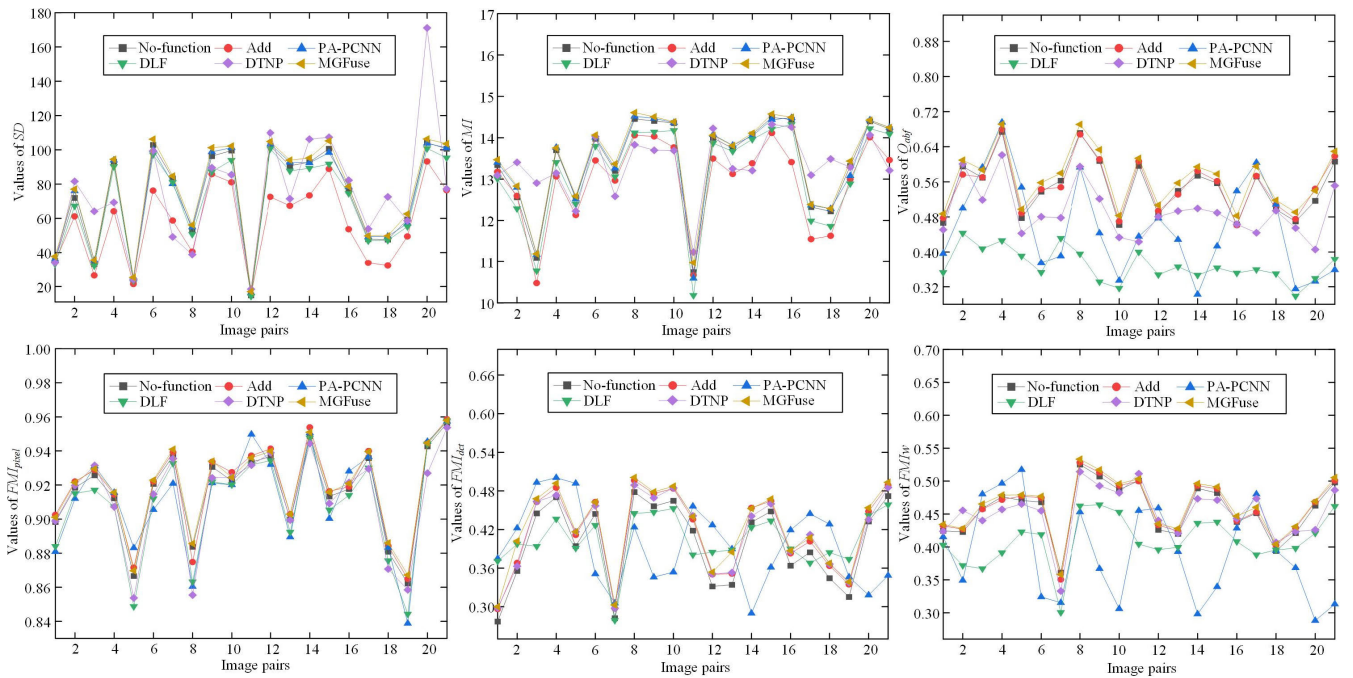
FIGURE 5. Qualitative comparison results of ablation experiments on 21 pairs of infrared and visible images.

objective fusion metrics are utilized in our experiments. They are: 1) standard deviation (SD) [47]; 2) mutual information (MI) [48]; 3) edge retentiveness (Q_{abf}) [49]; 4) pixel mutual

information (FMI_{pixel}) [50]; 5) discrete cosine mutual information (FIM_{dct}) [5]; 6) wavelet feature mutual information (FIM_w) [51]. The larger SD means source images contain

TABLE 1. The average quantitative comparison results of ablation experiments on 21 pairs of infrared and visible images.

Algorithm	No-function	Add	PA-PCNN	DLF	DTNP	MGFuse
SD	72.8311	57.4404	**73.7985**[3]	69.5128	**74.4863**[2]	**75.5790**[1]
MI	**13.4095**[1]	12.9319	**13.4361**[2]	13.1635	13.3455	**13.5033**[1]
Q_{abf}	**0.5477**[3]	**0.5520**[2]	0.4564	0.3691	0.4925	**0.5678**[1]
FMI_{pixel}	**0.9164**[3]	**0.9190**[2]	0.9125	0.9095	0.9119	**0.9194**[1]
FMI_{dct}	0.3974	0.4150	0.3948	0.4032	0.4133	**0.4190**[1]
FMI_w	**0.4563**[3]	**0.4598**[2]	0.3925	0.4095	0.4521	**0.4637**[1]

more diverse and distinctive features, indicating that the fusion algorithm has achieved an optimized fusion effect. MI utilizes both image histograms and features to compute the information preservation between the fused image and the source images. Q_{abf} evaluates edge information in terms of edge strength and orientation preservation. FMI_{pixel}, FMI_{dct}, and FMI_w can be calculated for mutual information (FMI) for the pixel features, discrete cosine, and wavelet features, respectively. As the value increases, the information correlation between the input images and the fused image also increases, indicating that there is less target features loss during the fusion process.

B. EXPERIMENT WITH OUR ALGORITHM

We first implement our MGFuse algorithm to fuse those infrared and visible images. We choose 4 pairs of infrared and visible images and perform experiments on our MGFuse algorithm according to the corresponding implementation details of Section A. To better facilitate observation and subsequent analysis, we mark it with a yellow arrow, and red and green boxes to display it.

Visual fusion results of our MGFuse in 4 pairs of source images are exhibited in Figure 3. By observing the marked regions within red and green boxes in the 'Kaptein_1123', we clearly note that the MGFuse preserves the luminance and background details to a great extent. By observing 'Fk_ge_04', it can be concluded that the input images have been fully fused even under low lighting conditions. In the fused image, it is easy to identify the car (green box) and the grasses (red box) around the hill. From the 'Fk_ge_04' and 'Sandpath', we can notice that the proposed MGFuse fully fuses those visible backgrounds and infrared objects together. The three-dimensional gray-scale images for 4 pairs of infrared and visible image fusion (corresponding to Figure 3) are displayed in Figure 4. We also can distinctly notice that those prominent infrared targets and visual background are perceptually preserved. For example, the marked within red boxed regions in the 'Kaptein_1123' is fully preserved in the fused image.

From the above-mentioned analysis of our MGFuse algorithm, we can conclude that our MGFuse attains superior fusion productivity with respect to preserving the visible background and infrared objects. Intuitively, the above test findings validate the potency and reasonableness of the proposed MGFuse.

Subsequently, we adopted additional ablation tests to confirm the effectiveness of our MGFuse through verification and discussion. "NO-function" indicates our MGFuse algorithm without the optimization function. Furthermore, we have chosen four classical handcrafted fusion strategies ("Add" [39], "PA-PCNN" [52], "DLF" [42], and "DTNP" [53]) to conduct the ablation experiments. In this experiment, "Add" and "DTNP" means the base layer fusion strategy replacing our proposed IAF fusion strategy. "PA-PCNN" and "DLF" means the texture layer fusion strategy replacing our proposed LGF fusion strategy.

To objectively evaluate the ablation experiments, six evaluation metrics (SD, MI, Q_{abf}, FMI_{pixel}, FMI_{dct}, and FMI_w) are utilized in the ablation experiments and the quantitative assessment results on 21 pairs of infrared and visible images (they are picked from [46]) are portrayed in Figure 5. We can observe from Table 1 that our MGFuse algorithm ranks first in all six fusion image quality metrics. We can conclude that the optimization function and fusion strategies of the base layers and texture layers are feasible and be applied to enhance the fusion performance of our MGFuse.

C. COMPARATIVE EXPERIMENT

Figure 6 displays the qualitative assessment results of three pairs of infrared and visible images by ten algorithms. To better facilitate observation and subsequent analysis, we mark it with a yellow arrow, and red and green boxes to display it.

In the qualitative performance assessment experiments, we can observe that the WLS and LatLRR algorithms exhibit a glaringly notable energy loss in infrared and visible images (leading to low clarity and luminance in final fused images), especially for the regions within red and green boxes. The fusion result of the VggML, ResNet, GFFuse, and Bayesian algorithms is better at fusion details, but the red and green boxes area is excessively unnatural. Furthermore, in the results of the DenseFuse, GAN, and RFNNest algorithms, the detail information is not prominent enough and there are slight artifacts at the edges of those targets, which is the result of inheriting excessively low infrared target information. The proposed MGFuse algorithm performed better than the other nine advanced algorithms with regard to detail extraction and energy preservation.
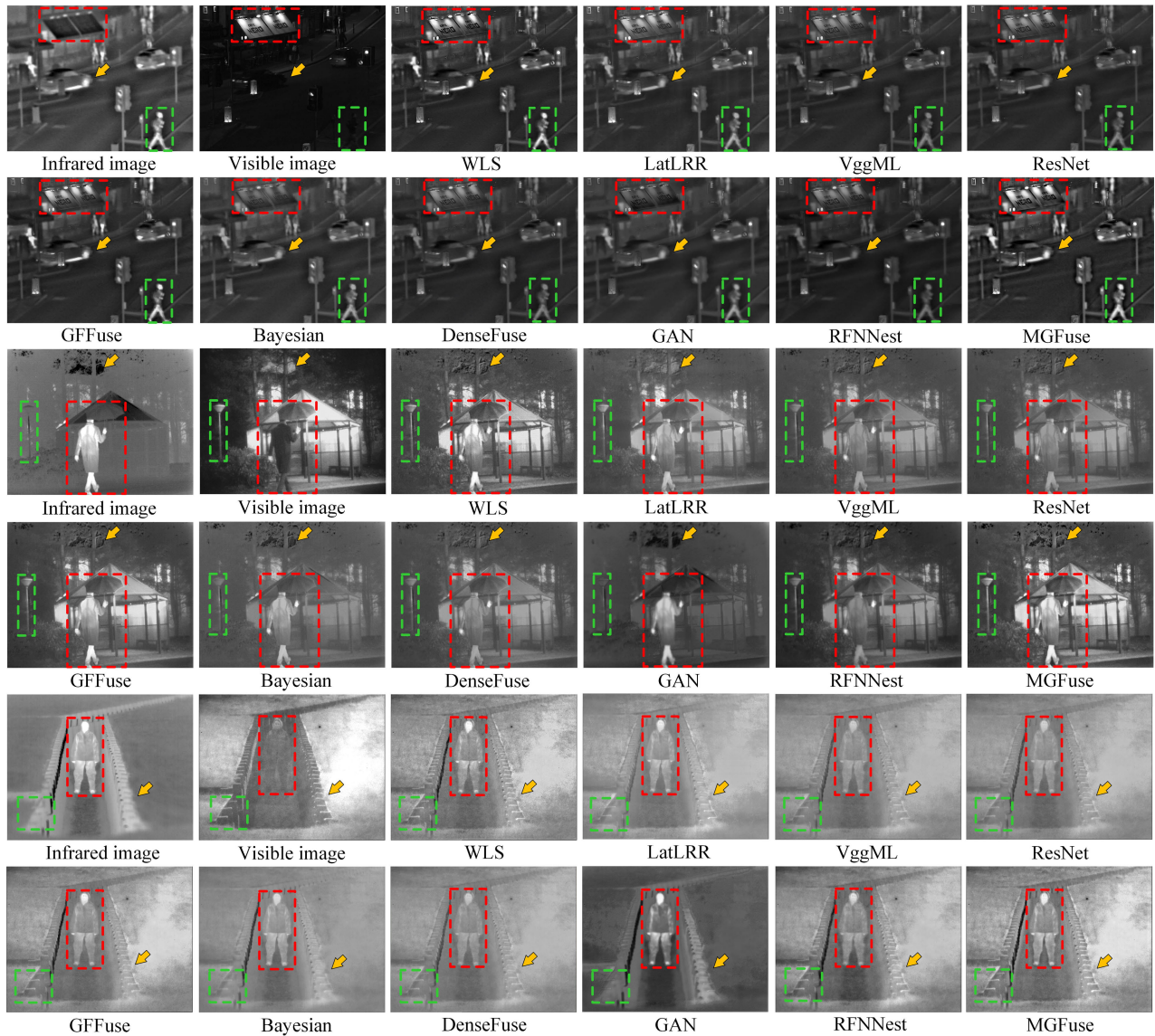
FIGURE 6. Qualitative assessment results of three pairs of infrared and visible images by ten algorithms.

TABLE 2. The average quantitative values of ten algorithms in three pairs of infrared and visible images.

Algorithm	WLS	LatLRR	VggML	ResNet	GFFuse	Bayesian	DenseFuse	GAN	RFNNest	MGFuse
SD	**83.7196**[3]	63.3500	55.7576	56.9396	**83.8226**[2]	49.2470	54.8478	50.3584	67.2955	**91.5109**[1]
MI	12.2408	12.7595	12.5234	12.5788	**13.3009**[1]	12.3207	12.4985	12.2291	**12.9825**[2]	**12.6297**[3]
Q_{abf}	**0.5576**[3]	0.4427	0.3815	0.3615	**0.5727**[2]	0.3410	0.3386	0.2004	0.3045	**0.5988**[1]
FMI_{pixel}	0.9104	0.9184	0.9230	0.9223	**0.9250**[3]	0.9217	0.9181	0.9100	**0.9287**[1]	**0.9275**[2]
FMI_{dct}	0.3578	0.3515	**0.4173**[2]	**0.4161**[3]	0.3582	0.3660	**0.4155**	0.3806	0.3005	**0.4268**[1]
FMI_w	0.3832	0.3851	**0.4190**[3]	0.4187	**0.4325**[2]	0.3994	0.4167	0.3514	0.2834	**0.4538**[1]

In the quantitative performance assessment experiments, we adopted six performance indicators (described in Section A) to assess the fusion performance of our MGFuse. The average quantitative values of ten algorithms in three pairs of infrared and visible images are shown in Table 2. The bold metrics value are the top three metric scores of ten algorithms. And the superscript value for the values are their rank. Figure 7 demonstrates objective performance
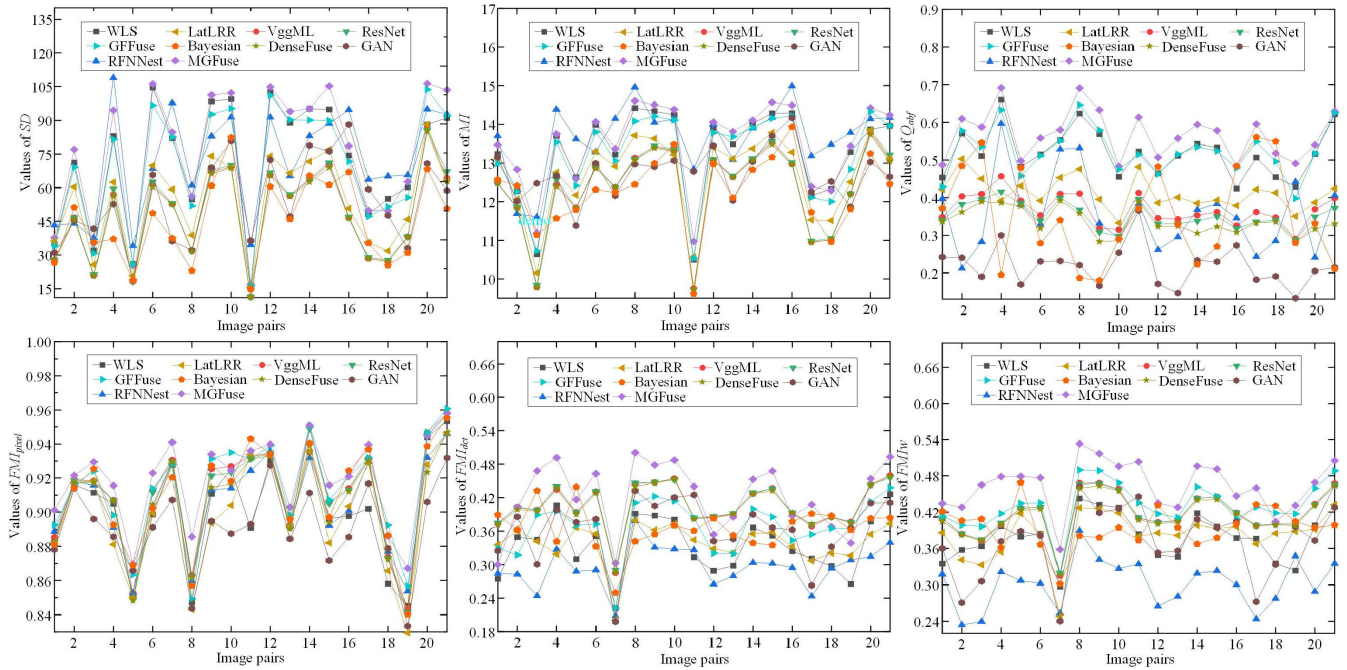
FIGURE 7. Objective performance qualitative comparison of ten fusion algorithms on 21 pairs of infrared and visible images.

TABLE 3. The average quantitative values of ten algorithms in three pairs of infrared and visible images.

Algorithm	WLS	LatLRR	VggML	ResNet	GFFuse	Bayesian	DenseFuse	GAN	RFNNest	MGFuse
SD	**71.4841**[3]	54.1365	48.1385	48.7180	69.9844	45.1077	47.6073	54.3575	**71.9013**[2]	**75.5790**[1]
MI	**13.2757**[3]	12.7149	12.3679	12.3905	13.1765	12.2713	12.3403	12.7257	**13.6827**[1]	**13.5033**[2]
Q_{abf}	**0.5198**[3]	0.4128	0.3677	0.3510	**0.5242**[2]	0.3543	0.3360	0.2183	0.3590	**0.5678**[1]
FMI_{pixel}	0.9010	0.8997	**0.9107**[3]	0.9092	**0.9137**[2]	0.9095	0.9053	0.8906	0.9048	**0.9194**[1]
FMI_{dct}	0.3349	0.3382	**0.4050**[3]	**0.4058**[2]	0.3679	0.3649	0.4039	0.3634	0.2967	**0.4217**[1]
FMI_w	0.3803	0.3826	0.4168	**0.4169**[2]	**0.4310**[2]	0.3948	0.4143	0.3708	0.3023	**0.4636**[1]

qualitative comparison of ten fusion algorithms on 21 pairs of infrared and visible images.

For the three images, our MGFuse achieves a wonderful fusion performance in the almost six evaluation metrics. Our MGFuse surpasses the nine comparison algorithms with respect to SD, Q_{abf}, FMI_{dct}, and FMI_w metrics, especially for the SD metric of the proposed algorithm with a marked improvement in 8.40% (0.0840 = (91.5109−83.8226) / 91.5109) compared to that of the GFFuse. In terms of the FMI_{pixel} metric, the MGFuse algorithm (FM_{pixel} = 0.9275) is slightly lower than the RFNNest (FMI_{pixel} = 0.9287) but slightly superior to the GFFuse (FMI_{pixel} = 0.9250).

From Table 3, the proposed MGFuse algorithm maintains the optimal values for SD, Q_{abf}, FMI_{pixel}, FMI_{dct}, and FMI_w metrics. It is noted that our MGFuse (MI = 13.5033) is slightly lower than the RFNNest (MI = 13.6827) but slightly superior to the WLS (MI = 13.2757) in terms of MI metric.

From the above-mentioned quantitative discussion, it can be concluded that our MGFuse has the optimal structure and edge information preservation for all testing images. Compared to the comparison algorithms, our MGFuse can achieve superior fusion performance, and is more in accord with the visual mechanism of human eyes.

D. EXPERIMENTS OF THE NOISE ROBUSTNESS

With the aim of validating the noise robustness of the proposed MGFuse algorithm, we chose one pair of infrared and visible images with noise standard deviation $\sigma = 20$. Figure 8 demonstrates the noise-perturbed qualitative assessment results among different algorithms at standard deviation $\sigma = 20$. To better facilitate observation and subsequent analysis, we mark it with a yellow arrow, red and green boxes to display it. We can clearly see that WLS and LatLRR algorithms introduce more noise and vignetting into the final image in a noisy environment. The fusion result of VggML, ResNet, GFFuse, and Bayesian algorithms shows blurry details and structure distortion, resulting in unreal visual effects. The shortcomings of DenseFuse, GAN, and RFNNest algorithms are that the contrast is too low and lacks detailed information. In contrast, our MGFuse is more suitable for human visual observation (for the textures and edges preservation of the source image) than the reference
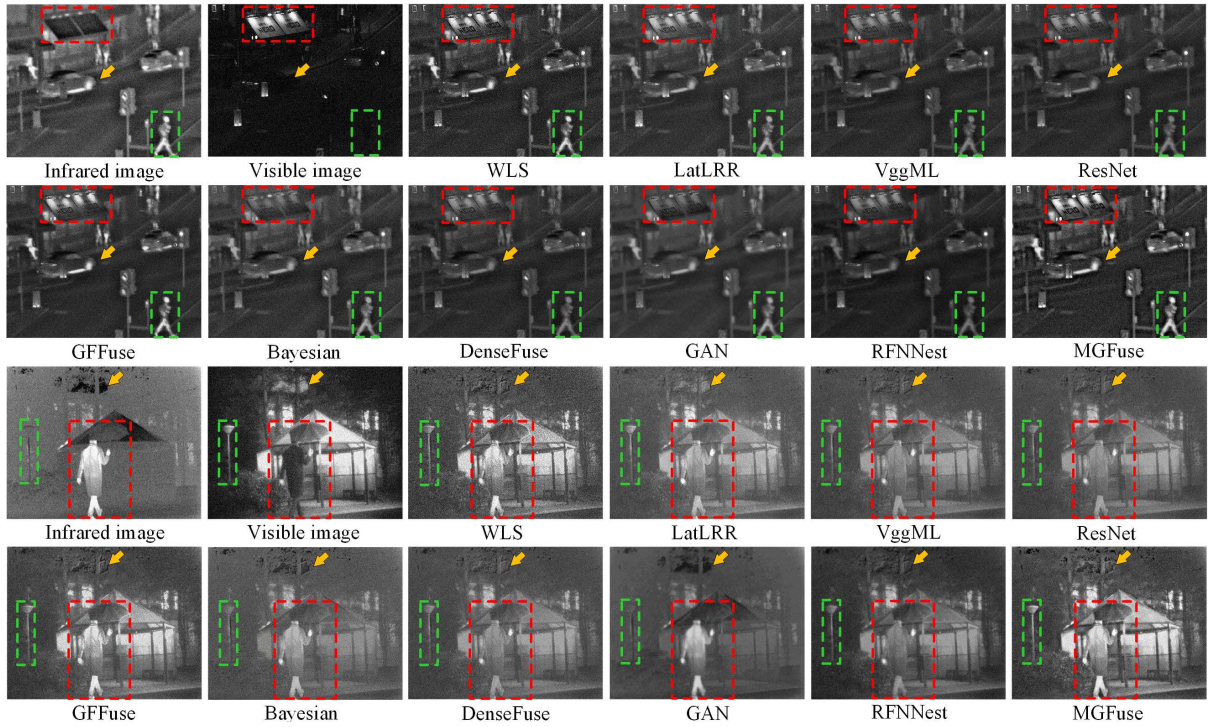
Infrared image	Visible image	WLS	LatLRR	VggML	ResNet
GFFuse	Bayesian	DenseFuse	GAN	RFNNest	MGFuse
Infrared image	Visible image	WLS	LatLRR	VggML	ResNet
GFFuse	Bayesian	DenseFuse	GAN	RFNNest	MGFuse

FIGURE 8. Qualitative comparisons of ten algorithms for noisy infrared and visible images with noise standard deviation $\sigma = 20$.

TABLE 4. The average quantitative comparison results on the noisy images ($\sigma = 20$).

Algorithm	WLS	LatLRR	VggML	ResNet	GFFuse	Bayesian	DenseFuse	GAN	RFNNest	MGFuse
SD	**93.0247[2]**	69.7584	58.4417	58.7235	**83.4142[3]**	58.8974	57.8592	47.5880	55.0139	**93.4354[1]**
MI	**14.4246[2]**	13.5634	13.3101	13.3182	**13.8777[3]**	13.4564	13.2869	12.3742	12.6445	**14.4779[1]**
Q_{abf}	**0.3206[3]**	0.3063	0.2663	0.2633	0.3154	**0.3982[1]**	0.2604	0.1020	0.1095	**0.3855[2]**
FMI_{pixel}	0.8517	0.8394	0.8498	0.8478	0.8438	**0.8563[3]**	0.8464	0.8532	**0.8667[2]**	**0.8668[1]**
FMI_{dct}	0.3338	0.3896	0.3862	0.3860	0.3635	**0.4345[3]**	0.3863	**0.4613[1]**	0.3738	**0.4384[2]**
FMI_w	0.3456	0.3845	0.3903	0.3902	0.3876	**0.4396[2]**	0.3899	**0.4522[1]**	0.3139	**0.4250[3]**

nine advancing algorithms. Table 4 demonstrates the corresponding noise-perturbed quantitative assessment results among different algorithms at standard deviation $\sigma = 20$. We can observe that our MGFuse basically obtains objective performance metrics values, which validates the feasibility and advancement of our MGFuse.

From the above qualitative and quantitative assessment, our MGFuse has the best robustness of the nine advanced algorithms with respect to fusing noisy source images. This is because the designed optimization function can well separate the global features and contents and texture details, which can greatly reduce noise generation. Through the GE fusion strategy, the fusion of texture or high-frequency layers can fuse the source images satisfactorily with clear texture details and acceptable visual effects, which can greatly avoid the effect of noise.

E. THE EXPERIMENT WITH OTHER IMAGE FUSION TYPES

A qualitative performance assessment experiment of the other three kinds of image fusion is conducted to assess the comprehensive capability and generalization ability of our MGFuse algorithm. These types of image fusion are multi-focus image fusion, multi-modal medical image fusion, and multi-exposure image fusion, respectively. Figure 9 gives the fusion examples of three kinds of image fusion.

For the multi-focus image fusion, we can notice that the child's open mouth in Figure 9(a1) and the doll's head behind the child in Figure 9(b1) can well preserve in Figure 9(c1). Similar to the above fusion results, the man's hands in Figure 9(a2) and the black and white flag in Figure 9(b2) have been copied into Figure 9(c2). For the multi-modal medical image fusion, we can notice some faint and tiny structural features (e.g. edges and shape), and those global and local energies are transferred into the fused results (see Figure 9(c3)-9(c4)). For the multi-exposure image fusion, rich detail, high clarity, and contrast existed in the fusion results (see Figure 9(c5)). As a whole, our MGFuse algorithm can effectively extract sharp edges and detailed features of target objects.
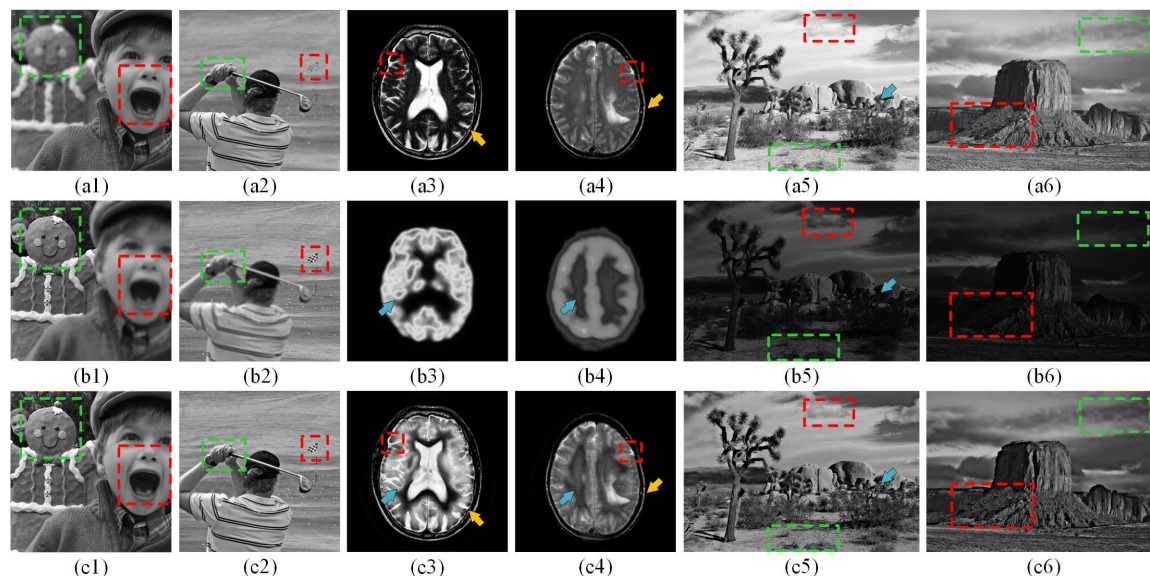
FIGURE 9. Visual results of using our MDOGFuse algorithm for three kinds of image fusion.

V. CONCLUSION

Existing image fusion algorithms have difficulty in effectively preserving valuable target features in infrared and visible images, which easily introduces blurry edges and unremarkable notable targets during their fusion process. In this paper, the source images are firstly divided into base layers, texture layers, and several high-frequencies via image multiscale decomposition optimization. The optimization function is designed to provide a precise and thorough decomposition of the base layer (global features and contents) and texture layer (texture details), which effectively retains the rich details and fully optimizes and enhances the fusion quality. Furthermore, the designed fusion strategies not only highlight the most significant targets, but also effectively retain the global and local abundant detailed information from the source images. Numerous related experiments demonstrate that our MGFuse can generate better visual effects than the other nine advanced approaches in both qualitative assessment and quantitative assessment. Finally, we further analyze the generalization capability of our MGFuse, which is applied to three kinds of image fusion problems. In the future, we will focus on those preservations for visible texture details and infrared notable targets in a harsh environment, and try to obtain more worthwhile information from low-quality images to validate the generalization capability of our MGFuse. Other direction of our future direction will spotlights those parameters optimization of MST-based algorithms to improve the operation speed.

REFERENCES

[1] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.

[2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[3] M. A. Smeelen, P. B. W. Schwering, A. Toet, and M. Loog, "Semi-hidden target recognition in gated viewer images fused with thermal IR images," *Inf. Fusion*, vol. 18, pp. 131–147, Jul. 2014.

[4] C. Chen, Y. Li, W. Liu, and J. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2760–2765.

[5] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[6] P. Chai, X. Luo, and Z. Zhang, "Image fusion using quaternion wavelet transform and multiple features," *IEEE Access*, vol. 5, pp. 6724–6734, 2017.

[7] J. Chen, X. Li, L. Luo, X. Mei, and J. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Inf. Sci.*, vol. 508, pp. 64–78, Jan. 2020.

[8] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.

[9] H. Yan, J.-X. Zhang, and X. Zhang, "Injected infrared and visible image fusion via L_1 decomposition model and guided filtering," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 162–173, 2022.

[10] J. Wang, J. Peng, X. Feng, G. He, and J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infr. Phys. Technol.*, vol. 67, pp. 477–489, Nov. 2014.

[11] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.

[12] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[13] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020.

[14] L. Jian, X. Yang, Z. Liu, G. Jeon, M. Gao, and D. Chisholm, "SEDRFuse: A symmetric encoder–decoder with residual block network for infrared and visible image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2020.

[15] A. Yehia, M. Safy, and A. S. Amein, "Fusion of high-resolution SAR and optical imageries based on a wavelet transform and IHS integrated algorithm," *Int. J. Eng. Res. Afr.*, vol. 52, pp. 62–72, Jan. 2021.

[16] T. Rabie, M. Baziyad, and I. Kamel, "Enhanced high capacity image steganography using discrete wavelet transform and the Laplacian pyramid," *Multimedia Tools Appl.*, vol. 77, no. 18, pp. 23673–23698, 2018.

[17] M. Zhao and Y. Peng, "A multi-module medical image fusion method based on non-subsampled shear wave transformation and convolutional neural network," *Sens. Imag.*, vol. 22, no. 1, pp. 1–16, 2021.

[18] X. Huang, G. Qi, H. Wei, Y. Chai, and J. Sim, "A novel infrared and visible image information fusion method based on phase congruency and image entropy," *Entropy*, vol. 21, no. 12, p. 1135, 2019.

[19] Q. Zhang, Y. Fu, H. Li, and J. Zou, "Dictionary learning method for joint sparse representation-based image fusion," *Opt. Eng.*, vol. 52, no. 5, 2013, Art. no. 057006.

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

[21] H. Li and X.-J. Wu, "Multi-focus image fusion using dictionary learning and low-rank representation," in *Proc. Int. Conf. Image Graph.* Cham, Switzerland: Springer, 2017, pp. 675–686.

[22] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: A novel decomposition method for infrared and visible image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4733–4746, 2020.

[23] P. K. Ram, S. V. Sai, and B. R. Venkatesh, "DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4714–4722.

[24] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4980–4995, 2020.

[25] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12797–12804.

[26] J. Ma, L. Tang, M. Xu, H. Zhang, and G. Xiao, "STDFusionNet: An infrared and visible image fusion network based on salient target detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[27] S. Yi, J. Li, and X. Yuan, "DFPGAN: Dual fusion path generative adversarial network for infrared and visible image fusion," *Infr. Phys. Technol.*, vol. 119, Dec. 2021, Art. no. 103947.

[28] X. Liu, Y. Zhou, and J. Wang, "Image fusion based on shearlet transform and regional features," *AEU, Int. J. Electron. Commun.*, vol. 68, no. 6, pp. 471–477, 2014.

[29] L. Jin, H. Liu, X. Xu, and E. Song, "Improved direction estimation for Di Zenzo's multichannel image gradient operator," *Pattern Recognit.*, vol. 45, no. 12, pp. 4300–4311, 2012.

[30] V. Estellers, S. Soatto, and X. Bresson, "Adaptive regularization with the structure tensor," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1777–1790, Jun. 2015.

[31] Z. Li, D. Mahapatra, J. A. W. Tielbeek, J. Stoker, L. J. van Vliet, and F. M. Vos, "Image registration based on autocorrelation of local structure," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 63–75, Jan. 2016.

[32] Z. Wu, Q. Wang, J. Jin, and Y. Shen, "Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *Signal Process.*, vol. 131, pp. 202–219, Feb. 2017.

[33] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-enhanced multi-scale exposure fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.

[34] J. Du, W. Li, and H. Tan, "Three-layer medical image fusion with tensor-based features," *Inf. Sci.*, vol. 525, pp. 93–108, Jul. 2020.

[35] Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus images," *Inf. Fusion*, vol. 20, pp. 60–72, Nov. 2014.

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

[37] A. Toet, "The TNO multiband image data collection," *Data Brief*, vol. 15, pp. 249–251, Dec. 2017, doi: 10.1016/j.dib.2017.09.038.

[38] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infr. Phys. Technol.*, vol. 82, pp. 8–17, May 2017.

[39] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, *arXiv:1804.08992*.

[40] J. Ma and Y. Zhou, "Infrared and visible image fusion via gradient-let filter," *Comput. Vis. Image Understand.*, vols. 197–198, Aug. 2020, Art. no. 103016.

[41] Z. Zhao, S. Xu, C. Zhang, J. Liu, and J. Zhang, "Bayesian fusion for infrared and visible images," *Signal Process.*, vol. 177, 2020, Dec. 107734.

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

[43] H. Li, X. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infr. Phys. Technol.*, vol. 102, Nov. 2019, Art. no. 103039.

[44] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019.

[45] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.

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

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

[48] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, p. 1, 2002.

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

[50] M. Haghighat and M. A. Razian, "Fast-FMI: Non-reference image fusion metric," in *Proc. IEEE 8th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, 2014, pp. 1–3.

[51] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[52] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2018.

[53] B. Li, H. Peng, and J. Wang, "A novel fusion method based on dynamic threshold neural P systems and nonsubsampled contourlet transform for multi-modality medical images," *Signal Process.*, vol. 178, Jan. 2021, Art. no. 107793.

HONGTAO HAO received the M.E. degree in vehicle engineering and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006 and 2015, respectively. He is currently an Associate Professor with the College of Mechanical Engineering, Ningxia University. He has authored or coauthored more than 20 journal articles. His research interests include deep learning, fault diagnosis, vehicle advanced transmission, and control.

BINGJIAN ZHANG received the B.E. degree in mechanical engineering from Xuchang University, Xuchang, China, in 2019. He is currently pursuing the master's degree with Ningxia University, Yinchuan, China. His research interests include machine vision, image fusion, and fault diagnosis.

KAI WANG received the B.E. degree in mechanical engineering from the Hefei University of Technology, Hefei, China, in 2021. He is currently pursuing the master's degree with Ningxia University, Yinchuan, China. His research interests include image processing, deep learning, and fault diagnosis.

● ● ●