Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.063469

ARTICLE

A Mask-Guided Latent Low-Rank Representation Method for Infrared and Visible Image Fusion

Kezhen Xie[1,2], Syed Mohd Zahid Syed Zainal Ariffin[1,*] and Muhammad Izzad Ramli[1]

[1]College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Shah Alam, 40450, Malaysia
[2]Guangzhou College of Technology and Business, Guangzhou, 510850, China

*Corresponding Author: Syed Mohd Zahid Syed Zainal Ariffin. Email: zahidzainal@uitm.edu.my

Received: 15 January 2025; Accepted: 02 April 2025; Published: 09 June 2025

ABSTRACT: Infrared and visible image fusion technology integrates the thermal radiation information of infrared images with the texture details of visible images to generate more informative fused images. However, existing methods often fail to distinguish salient objects from background regions, leading to detail suppression in salient regions due to global fusion strategies. This study presents a mask-guided latent low-rank representation fusion method to address this issue. First, the GrabCut algorithm is employed to extract a saliency mask, distinguishing salient regions from background regions. Then, latent low-rank representation (LatLRR) is applied to extract deep image features, enhancing key information extraction. In the fusion stage, a weighted fusion strategy strengthens infrared thermal information and visible texture details in salient regions, while an average fusion strategy improves background smoothness and stability. Experimental results on the TNO dataset demonstrate that the proposed method achieves superior performance in SPI, MI, Qabf, PSNR, and EN metrics, effectively preserving salient target details while maintaining balanced background information. Compared to state-of-the-art fusion methods, our approach achieves more stable and visually consistent fusion results. The fusion code is available on GitHub at: https://github.com/joyzhen1/Image (accessed on 15 January 2025).

KEYWORDS: Infrared and visible image fusion; latent low-rank representation; saliency mask extraction; weighted fusion strategy

1 Introduction

Infrared and visible image fusion integrates the thermal radiation information of infrared images with the texture details of visible images to generate more informative fused images, thereby enhancing the understanding of complex scenes [1]. Infrared images can highlight thermal targets, such as pedestrians and vehicles, in low-light environments but lack rich texture details. In contrast, visible images contain structural information about objects but are highly affected by lighting conditions, which may lead to the loss of target information. Therefore, fusing these two types of images helps leverage their respective advantages and improve perception capabilities.

Existing infrared and visible image fusion methods can be categorized into traditional methods and deep learning-based methods. Traditional methods include multiscale decomposition [2,3], sparse representation [4,5], statistical feature-based methods [6], and spatial domain methods [7,8]. The main advantages of these traditional fusion approaches lie in their computational simplicity, ease of implementation, and strong interpretability of the fusion process. However, these methods have limitations in handling complex

scenes, particularly in effectively addressing the significant differences between multimodal images. As a result, they often suffer from salient target information loss and detail degradation. Compared to traditional methods, deep learning-based approaches possess powerful feature extraction capabilities, enabling them to automatically learn deep feature representations from source images, thereby enhancing fusion quality. Current deep learning-based fusion methods mainly include autoencoder-based fusion methods [9,10], convolutional neural network (CNN)-based fusion methods [11,12], and generative adversarial network (GAN)-based fusion methods [13,14], Compared to traditional fusion techniques, deep learning methods excel in extracting deep feature representations from source images. However, the performance of deep learning-based fusion methods heavily depends on the design of the loss function and the optimization of the network structure. Due to the absence of ground truth fused images, it is challenging to design a loss function that effectively balances the weights of different modalities. This imbalance may lead to overemphasis on one modality while suppressing the other, ultimately affecting local detail preservation and overall fusion performance.

Existing fusion methods often ignore the distinction between salient objects and background regions, indiscriminately fusing different areas of the source images. This leads to the texture details of salient regions being suppressed by the smoothing process of the background, resulting in detail loss. As shown in Fig. 1, this issue is evident in the fusion results of GAN-based fusion methods.

Figure 1: Example of existing GAN-based fusion method: Loss of salient target details due to background smoothing

In a fused image, salient regions (such as pedestrian targets) should retain the key texture details from the visible image, such as object edge information, to ensure clarity and distinguishability. In contrast, for background regions (such as building facades and trees), excessive texture details may not be essential, and appropriate smoothing can help reduce distractions. However, existing methods fail to effectively differentiate between these two types of regions, resulting in the suppression of salient target details due to background smoothing, which degrades fusion quality and scene interpretability. To address this issue, a mask-guided latent low-rank representation fusion method is proposed in this paper. First, GrabCut [15] is employed to extract a salient object mask, distinguishing salient regions from background regions. Then, latent low-rank decomposition is utilized to extract deep image features, enhancing the representation of key information. During the fusion stage, a weighted fusion strategy is applied to salient regions to enhance infrared thermal information and visible texture details, while an average fusion strategy is applied to background regions to improve smoothness and stability. Experimental results show that the proposed method performs well across multiple evaluation metrics.

2 Related Work

2.1 Traditional Fusion Methods

In recent years, significant progress has been made in infrared and visible image fusion methods. Researchers have conducted in-depth studies on strategies such as multiscale decomposition, probabilistic statistics, edge preservation, low-rank representation, and pixel-level fusion, introducing improvements in target enhancement, detail preservation, and background smoothness.

Li et al. (2018) proposed a fusion method based on latent low-rank representation (LatLRR) [16]. This method utilizes low-rank decomposition to extract global structural information while enhancing local details in salient regions, achieving a balance between global information and detailed features. The method outperforms traditional approaches in visual quality and objective metrics, but its high computational complexity affects real-time applications. Panda et al. (2024) proposed a multiscale feature fusion method based on Bayesian probabilistic strategy [17]. By integrating bidimensional empirical mode decomposition (BEMD) with Bayesian modelling, the method extracts salient features at different scales and utilizes statistical models for information selection, reducing redundancy and enhancing structural clarity. Compared to traditional multiscale methods, it achieves more precise feature extraction, but its high computational complexity limits real-time applicability. Panda et al. (2025) specifically proposed a novel infrared and visible image fusion method [18] integrating a modified guided edge-preserving filter with a quantum computing-based weight map generation mechanism. This method introduces a 3-qubit quantum state modelling framework to encode the uncertainty and complementary information of multimodal images. The weight maps derived from quantum probability states effectively guide the fusion of thermal and texture details, leading to enhanced clarity and reduced redundancy. Although promising in theoretical modelling and preliminary evaluation, the method's practical application remains limited due to the nascent state of quantum hardware.

Although traditional infrared and visible image fusion methods have made significant progress, they still have certain limitations. Most methods rely on handcrafted fusion rules, making it difficult to adapt flexibly to feature variations across different scenes. Moreover, the lack of deep feature extraction capabilities may result in the loss of key information during the fusion process, affecting the detail representation and structural integrity of the final fused image.

2.2 Deep Learning-Based Fusion Methods

In recent years, deep learning has made significant progress in infrared and visible image fusion, focusing primarily on feature extraction, cross-modal information interaction, and structural preservation. Li et al. (2019) proposed DenseFuse [19], which integrates dense blocks for improved feature extraction and adopts addition and L1-norm fusion strategies. However, due to limited receptive fields in convolutional layers, its capacity to capture global structural information may be constrained. Ma et al. (2019) proposed FusionGAN [20], which leverages a generative adversarial network (GAN) to fuse infrared thermal radiation and visible gradient details, eliminating the need for handcrafted fusion rules. However, it is susceptible to the instability of GAN training, leading to mode collapse or color distortion. Zhang et al. (2020) proposed PMGI [21], which employs a dual-channel feature extraction path combined with inter-path information interaction to enhance the clarity and contrast of the fused image. However, its fixed ratio preservation strategy may lead to information loss. Li et al. (2021) proposed a novel end-to-end fusion network architecture (RFN-Nest) [22], which is based on a residual fusion network (RFN) and integrates multiscale feature extraction with nest connections to optimize fusion quality. It achieves excellent experimental performance,

but its generalization ability across different datasets remains a challenge. Yao et al. (2023) proposed HG-LPFN [23], a fusion network based on the Laplacian Pyramid and hierarchical guidance. It employs a bottom-up fusion strategy with multi-level saliency mapping to adaptively fuse low-and high-frequency details. Using cross-correlation attention and a multi-loss strategy, it enhances local details while maintaining global style consistency. However, its reliance on predefined pyramid levels and saliency maps reduces fusion performance under extreme lighting or high dynamic range (HDR) conditions. Li et al. (2024) introduced a fusion method based on Transformer and Cross Attention Mechanism (CAM) [24], which optimizes feature complementarity through two-stage training and enhances salient target details. However, it is sensitive to hyperparameter selection, which affects its robustness. Liu et al. (2025) designed a Dual-Branch Auto-Encoder [25], incorporating Invertible Neural Networks (INN) to preserve details and global information while improving structural integrity. However, the method requires a large amount of training data, making it difficult to adapt to low-data scenarios. Yao et al. (2025) proposed the Low-light Color Fusion Network (LCFN) for nighttime scenarios [26], integrating Low-Light Enhancement (LLE) and Knowledge Distillation to improve brightness and color fidelity in fused images while reducing grayscale effects. However, extreme lighting conditions may introduce overexposure or noise.

In summary, deep learning research in infrared and visible image fusion primarily focuses on feature extraction, cross-modal information interaction, and structural preservation. These methods effectively extract deep features and facilitate information exchange between different modalities, making them more adaptive and capable of learning compared to traditional approaches. However, further optimization of network structures and loss functions is still required to better balance multimodal information and enhance detail preservation and overall stability.

2.3 Existing Issues

Most fusion methods adopt a globally uniform fusion strategy, failing to effectively distinguish between salient targets and background regions. As a result, the texture details of targets are often suppressed by the smoothing process of the background. For example, GAN-based fusion methods such as FusionGAN enhance image contrast but still suffer from information loss in target region details, affecting the interpretability of the fused image. Additionally, fixed ratio preservation strategies, such as PMGI, weaken salient features across different scenes, making them less adaptable to complex environments.

To address these issues, this study utilizes the GrabCut algorithm to extract salient object masks, effectively distinguishing salient and background regions from the source images. GrabCut is a graph-cut-based image segmentation method that models foreground and background using a Gaussian Mixture Model (GMM) and optimizes an energy function via the max-flow/min-cut algorithm, enabling automatic segmentation of foreground and background [27]. Compared to traditional fixed-threshold segmentation methods, GrabCut exhibits higher adaptability, allowing it to accurately extract salient target regions in complex backgrounds, thus improving target integrity and accuracy. Next, Latent Low-Rank Representation (LatLRR) is employed for image decomposition. LatLRR constructs a low-rank subspace to represent the global structure of the image while modelling local details as sparse residuals, thereby achieving adaptive modelling of salient and background regions. LatLRR offers superior global-local feature decoupling capability, effectively preserving salient target details while optimizing background smoothness and enhancing the clarity and stability of the fused image. During the fusion process, a weighted fusion strategy is applied to salient regions, ensuring that infrared thermal information and visible texture details are enhanced, effectively preventing target feature loss. Meanwhile, an average fusion strategy is used for background regions to ensure balanced multimodal information processing, thereby enhancing background smoothness and stability.

2.4 Main Contributions

The main contributions of this study are as follows:

1. Proposed a fusion framework based on salient object extraction and latent low-rank decomposition: The GrabCut algorithm is introduced for salient object extraction, enabling the distinction between salient and background regions. Additionally, Latent Low-Rank Representation (LatLRR) is employed to decouple global and local information, enhancing detail expression in salient regions while optimizing the smoothness of background regions, thereby improving the hierarchical perception and structural consistency of the fused image.
2. Proposing an adaptive weighted fusion strategy for enhancing salient target details, combined with a structural preservation optimization method for background regions, ensuring background smoothness while improving the stability of the fused image.

3 Methodology

3.1 GrabCut Theory

GrabCut is an image segmentation method based on GraphCut, which uses a Gaussian Mixture Model (GMM) to model the foreground and background and optimizes the energy function through the Max-Flow Min-Cut algorithm to achieve adaptive foreground extraction [28]. Its main idea is to convert the segmentation problem into an optimization problem by constructing an energy function and iteratively optimizing it based on the initial annotation provided by the user so that the foreground region gradually converges to the real target. GrabCut minimizes the energy function E(L, θ, α) to achieve this, as defined in Eq. (1).

$$\mathbf{E}(\mathbf{L}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{\mathbf{i}} -\mathbf{logP}(\mathbf{Z_i}|\boldsymbol{\alpha_i}) + \boldsymbol{\gamma} \sum_{\mathbf{i,j}} \boldsymbol{\delta}(\boldsymbol{\alpha_i}, \boldsymbol{\alpha_j}) \mathbf{exp}\left(-\boldsymbol{\beta} \|\mathbf{z_i} - \mathbf{z_j}\|^2\right) \tag{1}$$

The term $\sum_{\mathbf{i}} -\mathbf{logP}(\mathbf{Z_i}|\boldsymbol{\alpha_i})$ measures the matching degree between the pixel $\mathbf{Z_i}$ a and the foreground or background model, where $\mathbf{P}(\mathbf{Z_i}|\boldsymbol{\alpha_i})$ is estimated by the GMM, describing the probability that pixel $\mathbf{Z_i}$ belongs to the foreground or background. $\boldsymbol{\alpha_i}$ represents the class label of pixel i, and $\mathbf{Z_i}$ is the RGB value of pixel i. The term $\boldsymbol{\gamma} \sum_{\mathbf{i,j}} \boldsymbol{\delta}(\boldsymbol{\alpha_i}, \boldsymbol{\alpha_j}) \mathbf{exp}\left(-\boldsymbol{\beta} \|\mathbf{z_i} - \mathbf{z_j}\|^2\right)$ is the smoothing term, ensuring the continuity of the foreground or background boundary. $\boldsymbol{\gamma}$ is the weight parameter of the smoothing term, which controls the balance between the data term and the smoothing term. N represents the pixel neighborhood, and $\boldsymbol{\delta}(\boldsymbol{\alpha_i}, \boldsymbol{\alpha_j})$ indicates whether adjacent pixels belong to different categories. $\boldsymbol{\beta}$ controls the weight of similarity between pixels. The Max-Flow Min-Cut algorithm is used to solve this energy function, optimizing the segmentation of pixels into the foreground or background.

3.2 Latent Low-Rank Representation Theory

Low-rank representation (LRR) can capture global information about an input image, but it cannot capture local information. The method cannot perform well when the data sample is insufficient or severely corrupted. In 2014, a latent low-rank representation method was proposed [29]. Latent low-rank representation can address the issue of low-rank representation failing to capture local structural information. The latent low-rank model decomposes the input data into three components: the low-rank component, the significant component, and the sparse noise component. The rank formula of the latent low-rank model is shown in Eqs. (2) and (3).

$$\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda \|E\|_1 \tag{2}$$

s.t..X = XZ + LX + E (3)

In the equations, $\|\|_1$ denotes the l_1 norm, and $\|\|_*$ represents the nuclear norm. X is the input data matrix, typically representing an image or a feature matrix. Z is the low-rank representation matrix used to capture the global structure of the data. L is the significant component, which represents the local salient structural information, such as edges and textures. E is the sparse noise component, indicating anomalies or noise in the data. λ is the sparsity regularization parameter used to control the weight of the sparse noise component.

3.3 Proposed Fusion Method

In this work, we adopt four key steps, mask computation, low-rank decomposition, fusion weight calculation, and image fusion, to ensure the enhancement of salient target regions and balanced processing of background regions during the fusion process, thereby improving the quality and information integrity of the fused image. The framework of our fusion method is shown in Fig. 2.

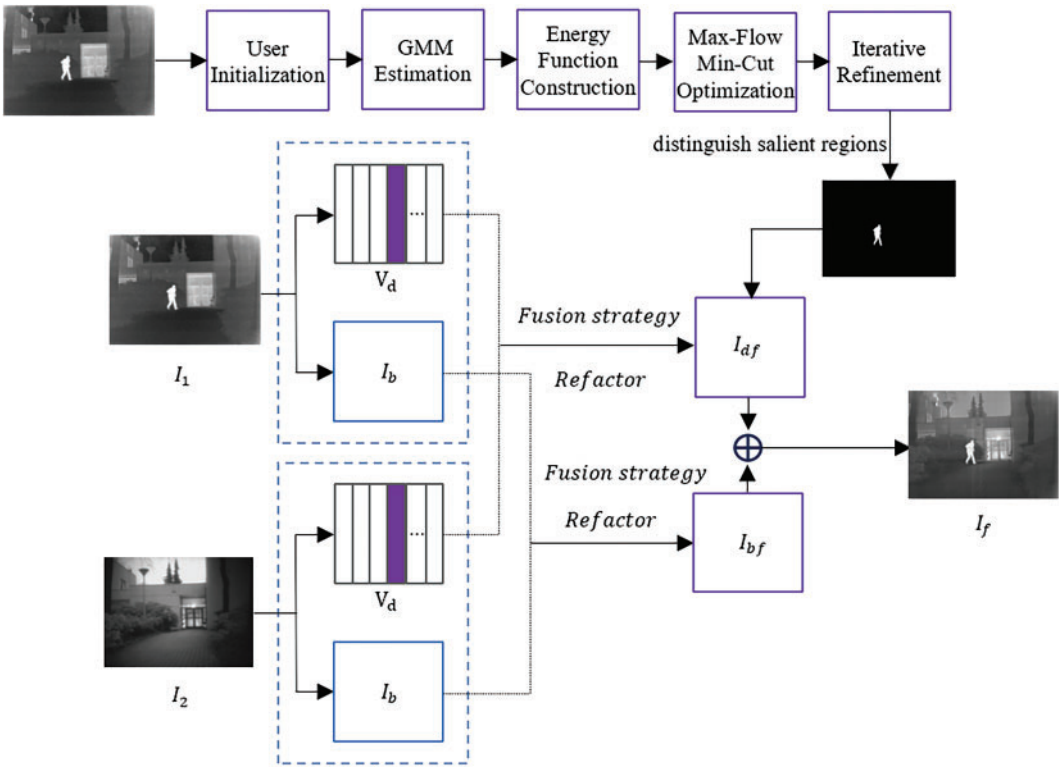

Figure 2: The proposed fusion method framework

First, this study employs the GrabCut algorithm to extract salient object masks. The user first initializes the target region by selecting a potential target area with a rectangular box, providing a foundation for the preliminary segmentation of the foreground and background. Then, GMM (Gaussian Mixture Model) is used to model the color distributions of the foreground and background, establishing their probability distributions and estimating the likelihood of each pixel belonging to a particular class. Graph Cut is applied to optimize the segmentation by minimizing the energy function based on the Max-Flow Min-Cut algorithm, generating an initial salient object mask. By continuously adjusting the GMM parameters, the mask is further

optimized to more accurately match the salient target region, ultimately obtaining a precise binary mask. The mask plays a crucial role in guiding the processing of different regions during fusion, ensuring that salient targets are prominently preserved in the fused image, while background regions undergo smoothing to reduce unnecessary interference. Then, for infrared and visible images, Latent Low-Rank Representation (LatLRR) is applied for decomposition. LatLRR models the global structure of the image through a low-rank subspace while extracting local detail information, decomposing the source image into two parts: the low-rank component retains global structural information, such as background and smooth regions, ensuring the stability and consistency of the image; the salient component extracts local high-frequency features, such as edges and texture details of targets, enhancing the detail representation of key targets during the fusion process. This decomposition method enables more refined processing of different regions, allowing subsequent fusion strategies to adaptively adjust, thereby improving the clarity and stability of the fused image. To ensure that salient target regions are enhanced during the fusion process, this approach effectively separates global and local features, optimizing the fusion quality.

Fusion weights are a critical part of the fusion process, as they determine the representation of salient target regions and background regions in the fused image. To address different regions, adaptive fusion rules are designed. Based on the generated salient object mask, the image is divided into salient regions and background regions, and a different fusion strategy is applied to each salient region for weighted fusion. In the salient object region, a mask-based weighted fusion strategy is adopted, as shown in Eq. (4).

$$F_{saliency} = M * (\alpha * I_{lrr,IR} + (1 - \alpha) * I_{saliency,VIS} \tag{4}$$

$F_{background}$ represents the fusion result of the background region. $(1 - M)$ is used to denote the mask for the background region, and $I_{lrr,IR}$ and $I_{lrr,VIS}$ are the low-rank parts of the infrared and visible images, respectively.

A mean fusion strategy was employed for the low-rank components of the background region. This method ensures smooth processing of the image background, improving overall brightness and clarity while preventing unnecessary artefacts or noise. Subsequently, after individual processing of salient and background regions, the fused low-rank and salient components are merged to produce the final fused image, as described in Eq. (5).

$$F = F_{saliency} + F_{background} \tag{5}$$

This method achieves adaptive optimization of salient target enhancement and background smoothness through four key steps: low-rank decomposition, mask computation, fusion weight calculation, and image fusion.

4 Experimental Results

4.1 Experimental Data and Parameter Configuration

In this experiment, we choose to use the TNO dataset [30], which includes a variety of complex scenes covering different lighting conditions, background complexities, and salient target types, allowing for a comprehensive evaluation of the fusion method's adaptability in different environments. Additionally, the infrared and visible images in the TNO dataset exhibit strong complementarity, effectively highlighting the differences between infrared thermal radiation information and visible image texture details. This aligns with the optimization objectives of this study, which focus on enhancing salient targets and improving background smoothness. Therefore, this dataset serves as an effective benchmark for validating the fusion method's ability to preserve salient target information and enhance the overall visual quality of the fused image. In this

study, 30 pairs of infrared and visible images from the TNO dataset covering different scenes were selected for fusion experiments to ensure data diversity and representativeness. The region of interest (ROI) in the infrared and visible images was identified, and segmentation was optimized to generate salient object masks and masked images. These masks effectively distinguish salient targets from background regions, providing crucial support for subsequent fusion experiments.

In the latent low-rank decomposition of the source images, we set mu to a small initial value of 1e − 6 to ensure that the model updates steadily during the first few iterations, avoiding oscillations or instability caused by a too-large step size. As the number of iterations increases, mu gradually grows to accelerate the convergence process. The convergence threshold was set to a small value (1e − 6) to ensure sufficient accuracy in the model's decomposition. A smaller convergence threshold demands more precise model updates but may increase computation time. Therefore, we balanced accuracy and computation time through experimentation and chose 1e − 6. The sparse noise balance parameter was set to 0.1, which helps suppress noise while retaining sufficient detail.

4.2 Quantitative Evaluation

To evaluate the enhancement of thermal information in salient target regions and the preservation of visible texture details, this study employs several objective metrics. Saliency Preservation Index (SPI) is used to measure the contrast and texture retention of salient targets in the fused image [31]. Mutual Information (MI) reflects the degree of information sharing between the fused image and the source images [32], while Quality assessment based on blur and noise factors (Qabf) assesses the contour clarity and detail preservation of salient targets.

To assess the effectiveness of the mean fusion strategy in background regions, Peak Signal-to-Noise Ratio (PSNR) is used to evaluate the overall quality of the fused image, ensuring that the background is neither excessively enhanced nor introduces artifacts [33]. Entropy (EN) measures the information content in the background region [34], preventing excessive information loss and ensuring that the fused image is neither overly blurred nor excessively sharpened.

By integrating these metrics, a comprehensive evaluation of the proposed method's ability to enhance salient targets and maintain background smoothness is achieved.

In the comparative experiments of this study, several advanced fusion methods in the field of infrared and visible image fusion were selected, including RFN-Nest [22], PMGI [21], DenseFusion [21], and FusionGAN [22]. These methods are based on different fusion strategies, including residual networks (RFN), gradient information optimization, dense connections (Dense Block), and adversarial learning, representing the major research directions in infrared and visible image fusion. Through comparison, the performance advantages and improvements of the proposed method can be comprehensively evaluated in terms of salient target detail preservation and background information balance.

As shown in Fig. 3, the fused image examples of different fusion methods are presented. The target regions within the green boxes in the results of the proposed method appear clearer, with enhanced infrared thermal radiation while preserving visible texture details. In contrast, RFN-Nest and DenseFusion exhibit blurred target edges, FusionGAN introduces artefacts, and PMGI has limited effectiveness in target enhancement. Additionally, the background regions within the red boxes in the results of the proposed method appear more uniform, reducing unnecessary detail interference and avoiding issues such as over-enhancement in FusionGAN and inconsistent background contrast in PMGI and DenseFusion.

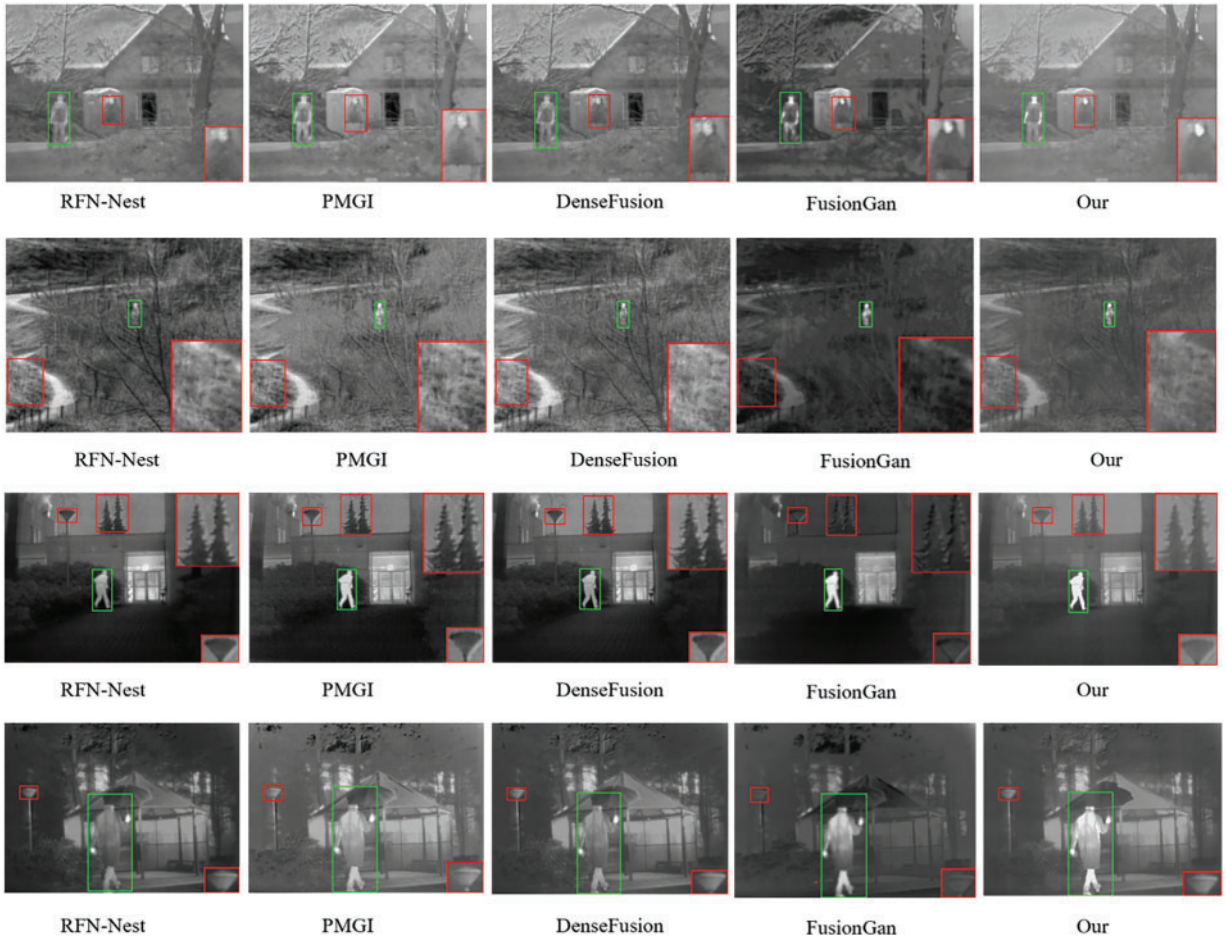

Figure 3: Our method is compared with four advanced fusion methods, including RFN-Nest. The red box indicates small regions with rich textures, while the green box highlights the salient regions

Table 1 and Fig. 4 present the comprehensive evaluation results of the fusion experiments conducted on 30 pairs of infrared and visible images selected from the TNO dataset. Table 1 summarizes the average values of five objective evaluation metrics, including Saliency Preservation Index (SPI), Mutual Information (MI), Edge Preservation Index (Qabf), Peak Signal-to-Noise Ratio (PSNR), and Entropy (EN), to quantify the performance differences among different fusion methods in terms of salient target enhancement, background information preservation, edge clarity, and overall image quality. Fig. 3 visualizes the distribution of these metrics across the 30 test image pairs using line charts, providing an intuitive representation of the stability and consistency of different fusion methods across various scenarios.

Table 1: Comparison of average evaluation metrics across different image fusion methods

Methods	RFN-Nest	PMGI	DenseFusion	FusionGan	Our
SPI	9.3525	10.4834	9.3939	10.0915	10.9715
MI	2.1095	2.3376	2.3519	2.2652	2.3410
Qabf	0.3442	0.4145	0.4442	0.2328	0.4459
PSNR	62.7877	62.4581	63.0906	61.1864	63.0922
EN	6.9611	7.0112	6.8230	6.4953	7.0175

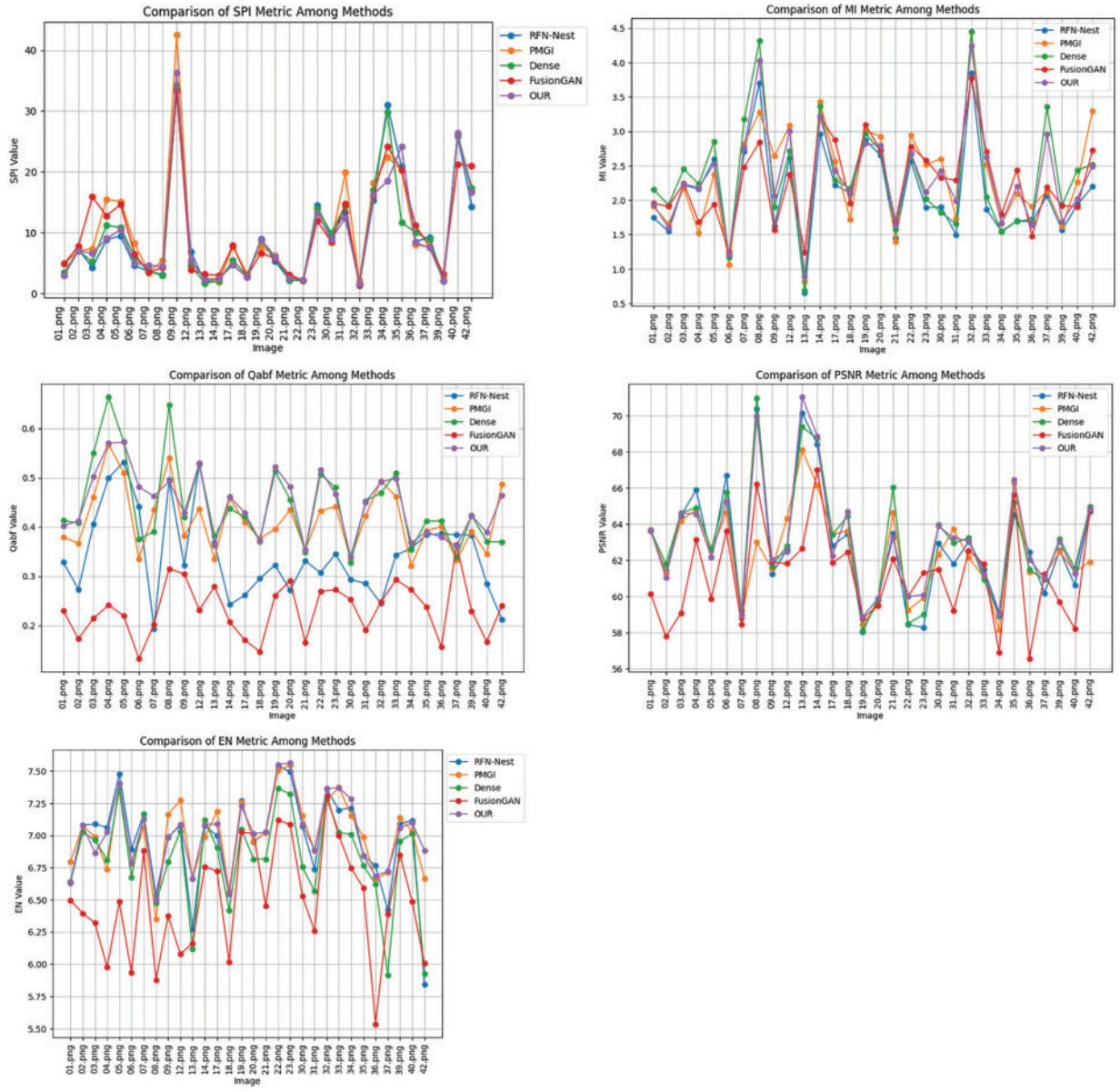

Figure 4: Quantitative evaluation of image fusion methods based on four objective metrics

From the results, the proposed method demonstrates superior performance in salient target detail enhancement, background information balance, fusion stability, and overall image quality. As an essential metric for evaluating saliency preservation, SPI shows that the proposed method achieves the highest value (10.9715), surpassing PMGI (10.4834) and FusionGAN (10.0915). This indicates that the proposed method excels in enhancing salient target information while effectively balancing gradient preservation and stability. Although PMGI also achieves a high SPI value, its heavy reliance on gradient information may lead to excessive enhancement or instability in certain regions. Observing the trend in the line charts, the proposed method exhibits lower fluctuation in SPI values, demonstrating greater stability in preserving saliency while effectively suppressing information loss. Moreover, in terms of Qabf, the proposed method achieves 0.4459, outperforming RFN-Nest (0.3442) and FusionGAN (0.2328), and slightly surpassing DenseFusion (0.4442). This indicates that the proposed method exhibits superior performance in edge detail preservation, effectively reducing target blurring and minimizing edge information loss. As a crucial metric for assessing multimodal information fusion, MI results show that DenseFusion achieves the highest MI value (2.3519), followed by PMGI (2.3376) and the proposed method (2.3410), while FusionGAN (2.2652) and RFN-Nest (2.1095) perform relatively lower. The MI performance of the proposed method is close to that of PMGI and DenseFusion, indicating that it effectively retains mutual information between infrared and visible images, ensuring the completeness of feature representation in the fused image. The line charts further demonstrate that the proposed method exhibits minimal fluctuations in MI values, suggesting robust performance in multimodal information fusion across different scenarios while avoiding information bias caused by modality imbalance. PSNR, as a key metric for evaluating the quality of fused images, shows that the proposed method achieves the highest value (63.0922), slightly outperforming DenseFusion (63.0906) and significantly exceeding RFN-Nest (62.7877) and FusionGAN (61.1864). This demonstrates that the proposed method effectively suppresses noise during fusion, improving the overall quality of the fused image. The line charts further validate that the proposed method exhibits lower fluctuation in PSNR values, highlighting its robustness and ability to maintain high signal-to-noise ratios across different scenarios, thereby ensuring stable visual quality. Entropy (EN) reflects the information content of the fused image. The experimental results indicate that the proposed method achieves the highest EN value (7.0175), followed by PMGI (7.0112), while FusionGAN has the lowest EN value (6.4953). This suggests that the proposed method effectively preserves more source image information during fusion, thereby improving information retention while maintaining stability in background regions. The line charts further confirm that the proposed method demonstrates superior EN performance, ensuring high information integrity across different scenarios while preventing information loss due to inappropriate fusion strategies.

In conclusion, the experimental results fully validate the advantages of the proposed method in salient target enhancement, background smoothness, and overall fusion image quality. Compared to existing methods, the proposed approach achieves a better balance between enhancing infrared thermal information and preserving visible texture details, exhibiting strong generalization capability across different scenarios. Despite its advantages, the proposed method still has certain limitations. The computational complexity of latent low-rank decomposition is relatively high, affecting its real-time applicability. Moreover, the performance of saliency mask extraction relies on the accuracy of the GrabCut algorithm, which may introduce errors in highly complex backgrounds, thereby affecting the fusion quality. In addition to the limitations of saliency region segmentation, the proposed method also faces certain challenges in computational efficiency.

4.3 Complexity Analysis

This section provides an empirical analysis of the computational complexity and runtime performance of the proposed method during the inference stage under varying image resolutions. The potential low-rank representation (LatLRR), which relies on singular value decomposition (SVD), has a computational complexity of approximately $O(n^2)$ in the optimal case. As the image resolution increases, the computational load grows exponentially. In contrast, most efficient deep learning-based methods typically maintain a complexity of $O(n^2)$, and can be accelerated through GPU-based parallel computation, thus offering a certain advantage in computational efficiency.

To evaluate the efficiency of different image fusion methods under multi-pixel inputs, this study conducts a complexity analysis based on 30 pairs of infrared and visible images from the TNO dataset. Two commonly used resolutions (320 × 240 and 640 × 480) are adopted to test the computational performance on original images. In addition, to further analyze the complexity trend of the algorithms under higher resolutions, the original images are upsampled using bilinear interpolation to 512 × 384 and 1024 × 768, simulating the variation in runtime with increasing image size. In the corresponding table, image height is used to indicate the trend of resolution change. The runtime evaluated in this section refers solely to the inference time, excluding the training process, to ensure the comparability of different methods during the deployment stage. The comparison of inference time is shown in Table 2.

Table 2: Comparison of average inference time (in seconds) for different fusion methods under varying image resolutions

Image resolution	LatLRR	RFN-Nest	PMGI	DenseFusion	FusionGAN
320 × 240 (Original)	0.500	0.400	0.500	0.350	0.600
512 × 384 (Upsampled)	1.954	1.024	1.280	0.896	1.536
640 × 480 (Original)	3.732	1.600	2.000	1.400	2.400
1024 × 768 (Upsampled)	14.585	4.096	5.120	3.584	6.144

As shown in the table, the proposed method exhibits significantly higher inference time across different resolutions compared to other deep learning-based methods. Moreover, its computational time increases exponentially with the rise in resolution, indicating a computational bottleneck when processing high-resolution images. To alleviate this issue, future work may consider optimizing the computational pipeline through GPU-based parallel acceleration or region-based fusion strategies (e.g., applying LatLRR only within salient regions), aiming to improve runtime efficiency while maintaining fusion quality. Although the proposed method has certain limitations in terms of real-time performance, it still holds high practical value in offline applications where fusion quality is critical, such as infrared image annotation and battlefield situation analysis.

5 Conclusion and Future Work

This paper proposes a mask-guided latent low-rank representation fusion method to address the issue of salient target detail suppression caused by background smoothing in existing infrared and visible image fusion methods. The GrabCut algorithm is employed to extract a saliency mask, enabling the separation of salient targets from background regions. Additionally, latent low-rank representation (LatLRR) is utilized for image decomposition, preserving global structural information while enhancing critical details. During the fusion process, a weighted fusion strategy is applied to salient regions to enhance infrared thermal

information and visible texture details, while an average fusion strategy is applied to background regions to improve smoothness and stability. Experimental results on the TNO dataset demonstrate that the proposed method performs well across multiple evaluation metrics, including SPI, MI, Qabf, PSNR, and EN, effectively enhancing salient target details while ensuring background smoothness and stability, thereby achieving superior fusion quality and robustness.

Despite its advantages, the proposed method still has certain limitations. First, the computational complexity of latent low-rank decomposition is relatively high, which may affect real-time applications in large-scale image processing. Second, the accuracy of saliency mask extraction relies on the performance of the GrabCut algorithm, which may introduce segmentation errors in complex backgrounds. Additionally, although the method effectively enhances target details and optimizes background smoothness, its fusion weight parameters are manually set, which may limit its adaptability under varying illumination conditions.

Future research will focus on optimizing computational efficiency by developing a more lightweight decomposition model to improve real-time performance. Furthermore, integrating a deep learning-based salient target detection mechanism can enhance the accuracy of target extraction and reduce dependency on GrabCut segmentation. Additionally, exploring adaptive fusion weight learning strategies will enable a more flexible and automated fusion process, improving the generalization ability of the method across different imaging conditions.

Acknowledgement: Not applicable.

Funding Statement: This study is supported by Universiti Teknologi MARA through UiTM MyRA Research Grant, 600-RMC 5/3/GPM (053/2022).

Author Contributions: Study conception and design: Kezhen Xie, Syed Mohd Zahid Syed Zainal Ariffin, Muhammad Izzad Ramli; Data collection and experimental implementation: Kezhen Xie; Draft manuscript preparation: Kezhen Xie; Manuscript revision and academic supervision: Syed Mohd Zahid Syed Zainal Ariffin, Muhammad Izzad Ramli. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article. The source code is available on GitHub at: https://github.com/joyzhen1/Image (accessed on 15 January 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Zhang H, Xu H, Tian X, Jiang J, Ma J. Image fusion meets deep learning: a survey and perspective. Inf Fusion. 2021;76:323–36.
2. Sun L, Li Y, Zheng M, Zhong Z, Zhang Y. MCnet: multiscale visible image and infrared image fusion network. Signal Process. 2023;208:108996. doi:10.1016/j.sigpro.2023.108996.
3. Li X, Chen H, Li Y, Peng YM. AFusion: multiscale attention network for infrared and visible image fusion. IEEE Trans Instrum Meas. 2022;71:1–16. doi:10.1109/tim.2022.3181898.
4. Li Q, Wu W, Lu L, Li Z, Ahmad A, Jeon G. Infrared and visible images fusion by using sparse representation and guided filter. J Intell Transp Syst. 2020;24(3):254–63. doi:10.1080/15472450.2019.1643725.
5. Ding W, Bi D, He L, Fan Z. Infrared and visible image fusion method based on sparse features. Infrared Phys Technol. 2018;92:372–80. doi:10.1016/j.infrared.2018.06.029.
6. Wang Lj, Han J, Zhang Y, Bai Lf. Image fusion via feature residual and statistical matching. IET Comput Vis. 2016;10(6):551–8. doi:10.1049/iet-cvi.2015.0280.

7. Zhang X, Dai X, Zhang X, Jin G. Joint principal component analysis and total variation for infrared and visible image fusion. Infrared Phys Technol. 2023;128:104523. doi:10.1016/j.infrared.2022.104523.
8. Zhao L, Zhang Y, Dong L, Zheng F. Infrared and visible image fusion algorithm based on spatial domain and image features. PLoS One. 2022;17(12):e0278055. doi:10.1371/journal.pone.0278055.
9. Ren L, Pan Z, Cao J, Liao J. Infrared and visible image fusion based on variational auto-encoder and infrared feature compensation. Infrared Phys Technol. 2021;117:103839. doi:10.1016/j.infrared.2021.103839.
10. Su W, Huang Y, Li Q, Zuo F, Liu L. Infrared and visible image fusion based on adversarial feature extraction and stable image reconstruction. IEEE Trans Instrum Meas. 2022;71:1–14. doi:10.1109/tim.2022.3177717.
11. Ren X, Meng F, Hu T, Liu Z, Wang C, editors. Infrared-visible image fusion based on convolutional neural networks (CNN). In: Proceedings of the Intelligence Science and Big Data Engineering: 8th International Conference, IScIDE 2018; 2018 Aug 18–19; Lanzhou, China. Berlin/Heidelberg, Germany: Springer; 2018. p. 301–7. doi:10.1007/978-3-030-02698-1_26.
12. Wang Z, Wang J, Wu Y, Xu J, Zhang XU. NFusion: a unified multi-scale densely connected network for infrared and visible image fusion. IEEE Trans Circuits Syst Video Technol. 2021;32(6):3360–74. doi:10.1109/tcsvt.2021.3109895.
13. Li Q, Lu L, Li Z, Wu W, Liu Z, Jeon G, et al. Coupled GAN with relativistic discriminators for infrared and visible images fusion. IEEE Sens J. 2019;21(6):7458–67. doi:10.1109/jsen.2019.2921803.
14. Rao Y, Wu D, Han M, Wang T, Yang Y, Lei T, et al. AT-GAN: a generative adversarial network with attention and transition for infrared and visible image fusion. Inf Fusion. 2023;92:336–49. doi:10.1016/j.inffus.2022.12.007.
15. Lu YW, Jiang JG, Qi MB, Zhan S, Yang J. Segmentation method for medical image based on improved GrabCut. Int J Imaging Syst Technol. 2017;27(4):383–90. doi:10.1002/ima.22242.
16. Li H, Wu X. Infrared and visible image fusion using latent low-rank representation. arXiv:1804.08992. 2018.
17. Panda MK, Thangaraj V, Subudhi BN, Jakhetiya V. Bayesian's probabilistic strategy for feature fusion from visible and infrared images. Vis Comput. 2024;40(6):4221–33. doi:10.1007/s00371-023-03078-4.
18. Parida P, Panda MK, Rout DK, Panda SK. Infrared and visible image fusion using quantum computing induced edge preserving filter. Image Vis Comput. 2025;153:105344. doi:10.1016/j.imavis.2024.105344.
19. Li H, Wu XJ. DenseFuse: a fusion approach to infrared and visible images. IEEE Trans Image Process. 2018;28(5):2614–23. doi:10.1109/TIP.2018.2887342.
20. 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. doi:10.1016/j.inffus.2018.09.004.
21. Zhang H, Xu H, Xiao Y, Guo X, Ma J, editors. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity. AAAI. 2020;34(7):12797–804. doi:10.1609/aaai.v34i07.6975.
22. Li H, Wu X-J, Kittler J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images. Inf Fusion. 2021;73:72–86. doi:10.1016/j.inffus.2021.02.023.
23. Yao J, Zhao Y, Bu Y, Kong SG, Chan JC-W. Laplacian pyramid fusion network with hierarchical guidance for infrared and visible image fusion. IEEE Trans Circuits Syst Video Technol. 2023;33(9):4630–44. doi:10.1109/tcsvt.2023.3245607.
24. Li H, Wu X-J. CrossFuse: a novel cross attention mechanism based infrared and visible image fusion approach. Inf Fusion. 2024;103:102147. doi:10.1016/j.inffus.2023.102147.
25. Liu H, Mao Q, Dong M, Zhan Y. Infrared-visible image fusion using dual-branch auto-encoder with invertible high-frequency encoding. IEEE Trans Circuits Syst Video Technol. 2024;35(3):2675–88. doi:10.1109/tcsvt.2024.3493254.
26. Yao J, Zhao Y, Bu Y, Kong SG, Zhang X. Color-aware fusion of nighttime infrared and visible images. Eng Appl Artif Intell. 2025;139:109521. doi:10.1016/j.engappai.2024.109521.
27. Wang Z, Lv Y, Wu R, Zhang Y. Review of GrabCut in image processing. Mathematics. 2023;11(8):1965. doi:10.3390/math11081965.
28. Nie F, Pei S, Zheng Z, Wang R, Li X. A greedy strategy for graph cut. arXiv:2412.20035. 2024.
29. Liu G, Yan S. Latent low-rank representation. Low-Rank Sparse Model Vis Anal. 2014:23–38. doi:10.1007/978-3-319-12000-3_2.

30. Toet A. The TNO multiband image data collection. Data Brief. 2017;15:249–51. doi:10.1016/j.dib.2017.09.038.
31. Xiong J, Liu G, Tang H, Gu X, Bavirisetti DP. SeGFusion: a semantic saliency guided infrared and visible image fusion method. Infrared Phys Technol. 2024;140:105344. doi:10.1016/j.infrared.2024.105344.
32. Guo P, Xie G, Li R, Hu H. Multimodal medical image fusion with convolution sparse representation and mutual information correlation in NSST domain. Complex Intell Syst. 2023;9(1):317–28. doi:10.1007/s40747-022-00792-9.
33. Singh S, Mittal N, Singh H. Review of various image fusion algorithms and image fusion performance metric. Arch Comput Methods Eng. 2021;28(5):3645–59. doi:10.1007/s11831-020-09518-x.
34. Roberts JW, Van Aardt JA, Ahmed FB. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. J Appl Remote Sens. 2008;2(1):023522. doi:10.1117/1.2945910.