ORIGINAL PAPER

Multi-modal remote sensing image fusion method guided by local extremum maps-guided image filter

Menghui Sun[1,2] · Xiaoliang Zhu[1,2] · Yunzhen Niu[1,2] · Yang Li[1,2]

Abstract

This paper proposes a multi-modal fusion algorithm for image filtering based on the guidance of local extrema maps. The image is subjected to a smoothing process using a locally extremal maps-guided image filter, and the difference from the original image forms the detail layer, while the smoothed image serves as the base layer. To preserve image details, both the base and detail layers undergo a resolution reduction process, establishing a multi-scale decomposition hierarchy. The base layer images at each level are decomposed using wavelet transform into low-frequency and high-frequency coefficients. For the low-frequency component, a region-energy fusion rule with adaptive weight allocation based on spatial and gradient information is employed, while the high-frequency component undergoes fusion using the AGPCNN fusion rule with neuron weight assignment initialized. The detail layer is fused using a weighted averaging fusion rule. Finally, the base layer images are fused through wavelet inverse transform. For the fusion of base and detail layers, a bottom layer employs the absolute maximum principle, while the top layer utilizes an average weighted fusion rule. Through subjective and objective analysis, experimental results indicate that the proposed fusion model algorithm not only effectively preserves edge and texture details but also maintains good color fidelity and spectral distortion control. Comparative analysis with eight other advanced fusion algorithms demonstrates superior fusion performance.

Keywords Image fusion · Multi-scale decomposition · Local extrema map · Guided filtering · Region energy

1 Introduction

With the advancement of modern remote sensing technology, various Earth observation satellites continuously provide remote sensing images with different spatial, temporal, and spectral resolutions [1]. Satellite remote sensing is unable to actively monitor changes, has low spatial resolution, and cannot meet high-precision requirements. The fixed orbit and observation frequency of satellites result in a low frequency of Earth observation, making it challenging to obtain timely information for sudden events, leading to slow data collection and low operational efficiency [2]. Different sensors highlight different information in images; PAN has higher spatial resolution, while MS provides more spectral information [3]. For research purposes, the characteristic information of a single image is insufficient to meet research objectives. High spatial and spectral resolution is essential for a range of image processing applications [4]. Therefore, the fusion of MS and PAN images, also known as pansharpening, is crucial for obtaining a comprehensive image with abundant spatial and spectral information [5].

Researchers commonly classify fusion methods into three levels based on the stage in the processing workflow and the abstraction level of information, namely pixel-level, feature-level, and decision-level fusion [6]. Pixel-level fusion methods directly manipulate pixels of source images, improving image quality and referentiality through different rules. Feature-level fusion methods extract objects of interest from source images, combining features using different fusion rules to create a multi-modal fusion image with highlighted features. Decision-level fusion extracts information using interpreted/labeled data. The principle of selection is linked

---

Menghui Sun and Xiaoliang Zhu should be considered joint first author.

✉ Xiaoliang Zhu
zhuxiaoliang3721@163.com

1 School of Software, Xinjiang University, Ürümqi 830008, China

2 Xinjiang Uygur Autonomous Region Big Data and Intelligent Software Engineering Research Center, Xinjiang University, Ürümqi 830008, China

to consolidated data. The main advantage of this approach is that higher-level representations make multi-modal fusion more robust and reliable [7].

Pixel-level image fusion methods can be categorized into three types: spatial domain-based, transform domain-based, and other methods. Spatial domain methods directly operate on the pixels of source images, such as Weighted average, PCA, and IHS. [8]. An example is the multi-scale exposure fusion method with detail enhancement in the YUV color space proposed by Qiantong Wang et al. [9]. Spatial domain fusion methods are simple, efficient, and computationally fast. However, the commonly used simple overlay operations for fusion rules often significantly reduce the signal-to-noise ratio and contrast of the resulting images. Classical transform domain operations include pyramid transforms and wavelet transforms [10]. Algorithms based on multi-scale transform methods, such as Laplacian pyramids [11, 12], wavelet transforms [13, 14], and hybrid methods, neural networks [15, 16], play a crucial role in image fusion and have been proven effective decomposition algorithms [17]. Fusion methods based on Multiscale Transform (MST) are the most common traditional methods. They typically apply traditional weighted fusion rules to base layers, overlooking global contrast [18]. To address these limitations, many scholars have proposed improvements. For instance, Yu Zhang et al. introduced a multi-modal brain image fusion method guided by local extreme value maps, using two guided image filters with local minimum and maximum mappings, respectively [19]. Veshki et al. presented a multi-modal image fusion method based on coupled feature learning [20], separating the images to be fused into relevant and irrelevant components using sparse representations with the same support and Pearson correlation constraints. However, their fusion images suffer from severe color loss. Jie et al. proposed a multi-modal medical image fusion method based on multiple dictionaries and truncated Huber filtering [21], using multiple dictionaries and truncated Huber filters to separate images at different layers. However, the color fidelity of the fused images is not high, and there is significant spectral loss.

Addressing the shortcomings of the aforementioned algorithms, this paper proposes a remote sensing image multi-scale fusion model based on locally extreme maps-guided image filters. The innovation lies in the following aspects:

(1) The use of locally extreme maps-guided image filters to decompose the image into base and detail layers, combined with wavelet transforms for further extraction of details from the base layer.
(2) For low-frequency coefficients of the base layer, an improved regional energy fusion rule is applied, enhancing pixel connectivity within regions to maintain color consistency.

(3) For high-frequency coefficients of the base layer, an improved AGPCNN rule.

2 Multiscale theory

In this section, we will introduce the general multiscale models.

Scale-space image processing is a fundamental technique in computer vision for object recognition and low-level feature extraction [22]. The term "scale-space" was introduced by Witkin when proposing a method for one-dimensional signal processing through convolution with a Gaussian kernel [23]. Scale-space can be considered as an alternative to traditional statistical smoothing methods [24].

Li et al. proposed an Image Fusion Algorithm based on Laplacian Pyramid and Principal Component Analysis Transforms [12]. The Laplacian pyramid model reduces the resolution of the image through downsampling, forming a multi-scale transformation space to better extract image features. The Principal Component Analysis method adopts a dimensionality reduction approach to rank and reduce the complex data information, seeking coordinate systems that reflect patterns to the maximum extent. Both the Laplacian pyramid and Principal Component Analysis methods used in their work are multi-scale spatial transformation techniques. The multi-scale spatial transformation model employed in this paper also involves downsampling to enhance the extraction of image features.

The fundamental idea behind multiscale analysis is to map points in high-dimensional space coordinates to low-dimensional space while preserving similarity between the two as much as possible.

The general steps of multiscale analysis are as follows:

(1) Evaluate the nature of the data and, considering the data acquisition method, choose an appropriate analysis approach.
(2) Determine the appropriate dimension based on the criteria for evaluation.
(3) Assess the effectiveness and reliability of the results obtained from multiscale analysis.
(4) Name the coordinate axes and classify objects based on the spatial map of the data.

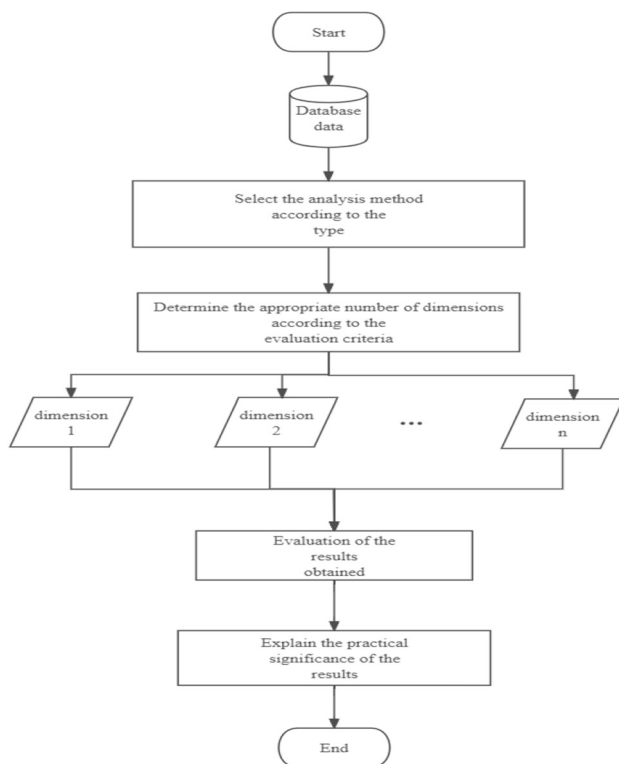The general steps of multiscale analysis are illustrated in Fig. 1.

Fig. 1 The general flow of multi-scale analysis theory

Fig. 2 Frame diagram of image fusion

Fig. 3 Comparison of filter results

3 Proposed fusion model

This section introduces the fusion model proposed in this paper, termed the local extrema maps-guided fusion model (LEMG). The fusion process of LEMG is outlined as follows:

(1) Smooth the source image using an image filter guided by local extrema maps. Generate a detail layer by subtracting the smoothed image from the source image, where the smoothed image serves as the base layer.

(2) Perform multi-scale decomposition on both the base and detail layers of two source images to extract additional details.

(3) The base layer incorporates the global information of the source image to minimize spectral loss, ensuring the fused image retains the color fidelity of the source image, decompose the base layers of multiple scales using wavelet transform into low-frequency and high-frequency components. Apply an improved regional energy fusion rule to merge the low-frequency components of each layer and an enhanced AGPCNN rule to merge the high-frequency parts. For the detail layers, use a weighted average fusion rule to create multiple levels of merged detail layers.

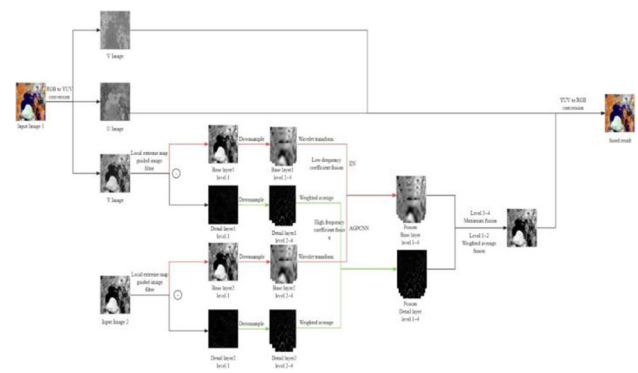(4) For each base layer, the wavelet inverse transform is applied to form the fused base layers. The fused 3rd to 4th layers of base layers and the co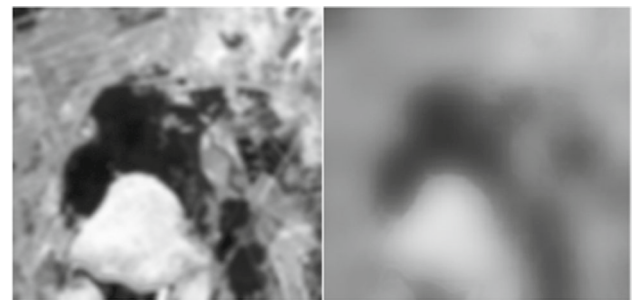rresponding detail layers are fused using the rule of taking the absolute value of the maximum to highlight the features of each image. The fused images of the 1st and 2nd layers are blended using an average weighting rule to smooth the final fused image, achieving the optimal fusion effect.

The schematic diagram of the fusion model is depicted in Fig. 2.

3.1 Local extreme maps-guided image filter

Classical filters are based on the concept proposed by the Fourier transform. Filters separate useful signals from noise, thereby improving the signal's resistance to interference and signal-to-noise ratio. At the same time, filters can filter out uninteresting signal frequencies, achieving the goal of filtering signal frequencies and improving analysis accuracy. For images that require smoothing and highlighting edges, bilateral filters are commonly used to achieve this goal. The main idea is to construct a Gaussian kernel based on color intensity distance and spatial distance. However, bilateral filters tend to cause gradient reversal and produce artifacts when processing edges, and the computation time is too long. Therefore, Kaiming He and others proposed the concept of guided filters [25]. Figure 3 illustrates the distinct filtering outcomes

of bilateral and guided filters. From left to right, the filtered images of the bilateral filter and the guide filter. The left column shows the filtering results of the bilateral filter, and the right column shows the filtering results of the guided filter.

Guided filters are essentially an optimization algorithm that constructs an energy function and then minimizes it using optimization strategies such as least squares and Markov random fields. The energy function is usually defined as follows:

$$E = U + V \qquad (1)$$

where U is the data term, V is the smoothing term.

Guided filters require two source images, namely the guided image and the input image to be processed. The function is defined as follows:

$$O_i = a_k \, G_i + b_k, \quad \forall i \in \omega_k \qquad (2)$$

where O_i and G_i are the output and guided images, a_k and b_k are unique constant linear functions in the $M \times M$ window ω_k.

The energy function is defined as follows:

$$E(a_k, b_k) = \sum_{i \in \omega_k} ((a_k \, G_i + b_k - I_i)^2 + \theta \, a_k^2) \qquad (3)$$

where I_i is the input image, θ is a constant to prevent a_k from becoming too large.

Yu Zhang et al. introduced a guided filter based on a locally extreme value map as the guided image for the input image, thereby significantly suppressing the salient features of the input image. First, the input image is filtered using the following formula:

$$I_F = G(I, I_{\min}, M) \qquad (4)$$

In the equations presented, I_F, I, $I_{\min}$, M represents the post-filtered image, the input image, the local minimum map of the input image, and the window size. The calculation formula for $I_{\min}$ is given by:

$$I_{\min} = ime(I, s) \qquad (5)$$

Here, ime denotes the morphological erosion operator, and s represents a disk-shaped structuring element.

Subsequently, the filtered image is further processed through a local maximum value mapping, with the calculation formula given by:

$$I_F = G(I_F, I_{\max}, M) \qquad (6)$$

Here, I_f, I_{max} represents the post-filtered image and the local maximum image of I_F. The calculation for $I_{\max}$ is as
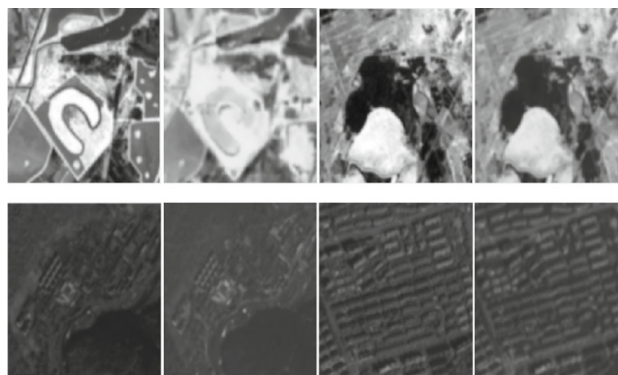

Fig. 4 The results of image filtering using local extreme maps-guided image filter

follows:

$$I_{\max} = imd(I_F, s) \qquad (7)$$

Here, imd denotes the morphological dilation operator.

The formula for the guided image filter based on local extreme maps is given by:

$$I_{\text{out}} = LG(I, s, M) \qquad (8)$$

Here, I_{out} is the filtered output image, and LG represents the guided image filter based on local extreme maps.

Prominent bright and dark features are removed from the input image to obtain a smooth image. Subsequently, significant features are obtained by subtracting the smoothed image from the input image, according to the following formulas:

$$I_b = \max(I - I_{\text{out}}, 0) \qquad (9)$$

$$I_d = \min(I - I_{\text{out}}, 0) \qquad (10)$$

In the following, I_b, I_d denotes the bright feature map and the dark feature map, with max, min representing the maximum and minimum values, respectively.

The results of image filtering using local extreme maps-guided image filter on the experimental data of this paper are presented in Fig. 4. The images are the first to fourth filtered images from left to right, with the left column showing the filtering results of Image 1 and the right column showing the filtering results of Image 2.

3.2 Fusion rule for low-frequency coefficients

The energy algorithm involves calculating the energy of an image or pixel region. In the context of an image, the grayscale value corresponds to the energy, with higher grayscale values indicating higher energy. Generally, energy

calculation involves summing the squared grayscale values of each pixel within an image or region.

Conventional region energy algorithms typically select a window region and employ a fixed weight multiplied by the region's pixels to compute the energy of the central pixel. Iterating through the image matrix yields an energy matrix of equal size to the image matrix. However, traditional region energy algorithms employ a fixed weight matrix that cannot be adjusted based on varying region features. Consequently, these algorithms fail to adapt to diverse image characteristics, limiting the enhancement of fusion effects. This paper introduces an improved region energy algorithm that elevates the fusion effect of the region energy fusion strategy by constructing a weight matrix adaptable to each region's features.

Initially, gradient information for each region is computed to construct the weight matrix, as given by the following formula:

$$AG(F) = \sqrt{(I(i+1, j) - I(i, j))^2 + (I(i, j+1) - I(i, j))^2} \quad (11)$$

Subsequently, the spatial weight matrix for the region is created using the following formula:

$$W_{s'} = e^{\left(-\frac{x^2 + y^2}{2\varepsilon^2}\right)} \quad (12)$$

$$W_s = \frac{W_{s'}(i, j)}{\sum\limits_{(i, j)\in\omega} W_{s'}(i, j)} \quad (13)$$

The variables are defined as follows: ε embodies the standard deviation of the Gaussian kernel associated with the spatial weight matrix, x represents a $M \times M$ matrix configuration where each row encompasses M values between elements of $[-1, 1]$ within the window size, y signifies a $M \times M$ matrix setting where each column comprises M values within the window size between elements of $[-1, 1]$, $W_{s'}$ stands for the preliminary spatial weight matrix, and W_s denotes the normalized spatial weight matrix, serving as the conclusive spatial weight matrix.

Finally, the formula for the adaptive weight matrix is as follows:

$$F_1 = \text{conv}\left((AG_1)^2, W_{s1}\right) \quad (14)$$

$$F_2 = \text{conv}\left((AG_2)^2, W_{s2}\right) \quad (15)$$

$$W = \frac{F1}{(F_1 + F_2)} \quad (16)$$

Here, F_1, F_2 denotes the outcome of convolution between the spatial weight matrix and the gradient matrices of the
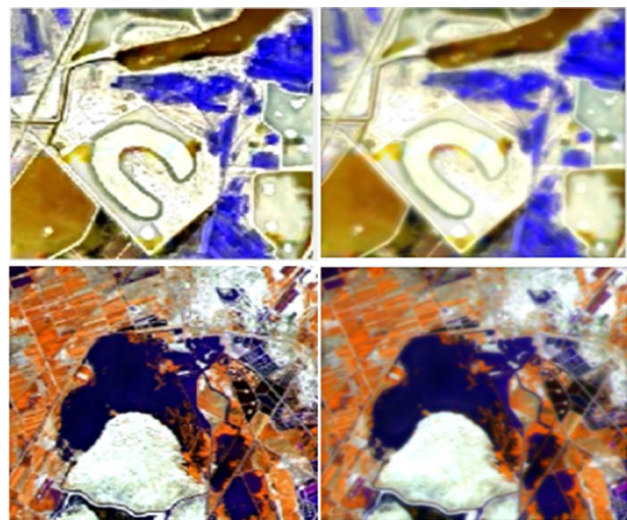

Fig. 5 Compare images with traditional fusion rules

two input images, with conv symbolizing the convolutional operation.

Figure 5 compares the results of low-frequency fusion and high-Frequency Coefficients in this paper with traditional region-based energy fusion rules. The left column is the result image of the traditional fusion rule, and the right column is the result image of the fusion rule proposed in this paper.

3.3 Fusion rules for high-frequency coefficients

The pulse coupled neural network (PCNN) is an artificial neural network inspired by physiological stimuli, providing the ability to establish a highly adaptable physiological filter. PCNN models various aspects of the primate visual cortex, including pulse duration, inter-pulse duration, and neural interconnections. This network not only meets the filtering requirements of visual models but also generates essential connections and pulses for simulating state modulation and temporal synchronization. However, due to the considerable number of parameters in the PCNN model, the AGPCNN model is introduced. This model features a simplified structure with fewer parameters and employs Gaussian filters for distributing weights among neurons.

The AGPCNN model comprises five components: feedforward input, connectivity input, neuron internal state, binary output, and dynamic threshold. The respective computational formulas are as follows:

$$V(i, j) = I(i, j) \quad (17)$$

$$L_n(i, j) = \sum_{M=-1}^{1} \sum_{N=-1}^{1} G_\sigma(M+2, N+2) E_{n-1}(i+M, j+N) \quad (18)$$

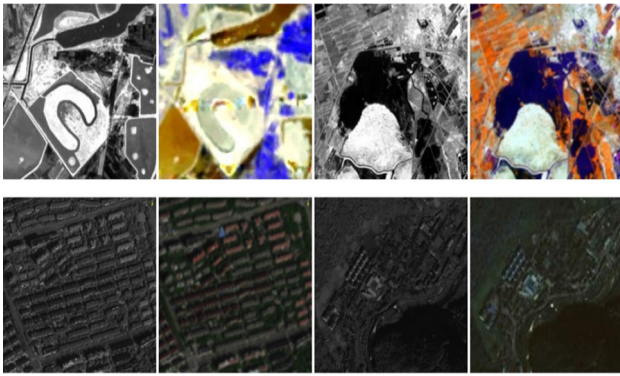$$Y_n(i, j) = V(I, j)(1 + \alpha(i, j)L_n(i.j)) \quad (19)$$

Fig. 6 Image of the experimental data

Fig. 7 Image of the result of the experimental fusion

$$E_n(i, j) = \begin{cases} 1, & \text{if } Y_n(i, j) > K_{n-1}^{(i, j)} \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

$$K_n(i, j) = d_\Theta K_{n-1}(i, j) + a_\Theta E_{n-1}(i, j) \quad (21)$$

In the provided excerpt, (i, j) signifies the position of neurons, n represents the iteration count, G_σ denotes a Gaussian filter with a standard deviation of σ and a 3×3 kernel, α stands for adaptive estimation of connection strength, d_Θ is the decay constant, and a_Θ corresponds to the normalization constant. The computational formula of α is articulated as follows:

$$\alpha(i, j) = I_s(i, j) = \sqrt{I_r(i, j)^2 + I_l(i, j)^2} \quad (22)$$

$I_s(i, j)$ represents the spatial frequency of the 3×3 local window of neurons for (i, j), $I_r(i, j)$ represents row frequency, and $I_l(i, j)$ represents column frequency. The calculation is performed as follows:

$$I_r(i, j) = \sqrt{\frac{\sum_{x=-1}^{1} \sum_{y=-1}^{1} (I(i+x, j+y) - I(i+x, j+y-1))^2}{M \times N}} \quad (23)$$

$$I_l(i, j) = \sqrt{\frac{\sum_{x=-1}^{1} \sum_{y=-1}^{1} (I(i+x, j+y) - I(i+x-1, j+y))^2}{M \times N}} \quad (24)$$

$M \times N$ is the size of the regional window.

The calculation of pulse time for the AGPCNN model after n iterations is as follows:

$$T_n(i, j) = T_{n-1}(i, j) + E_n(i, j) \quad (25)$$

The AGPCNN model employed in this study has transitioned from the prior Gaussian filter-weighted allocation for neighboring neurons to an allocation using He initialization. This adjustment, which takes into account the properties of ReLU, effectively mitigates the problem of gradient vanishing, resulting in an enhanced fusion effect.
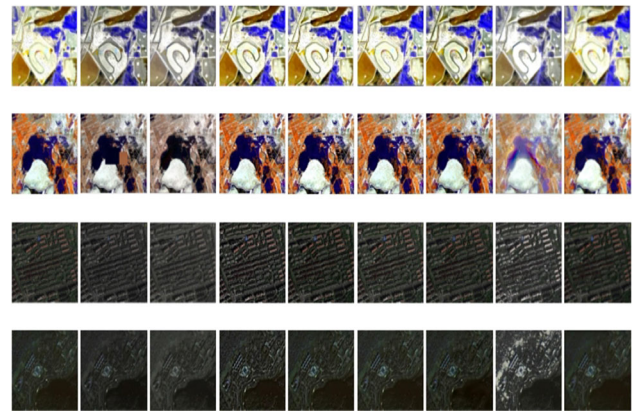
4 Experimental results and comparisons

This section presents the experimental comparative results between the algorithm model proposed in this paper and eight other existing algorithms.

4.1 Experimental setup

In order to validate the effectiveness of the proposed algorithm model, comparisons were made against eight other state-of-the-art algorithm models. The experimental data comprised four sets of remote sensing images, namely, panchromatic images and multispectral images. The experiments were conducted in the Matlab2022a environment using NVIDIA GeForce RTX 4060 Laptop GPU. Refer to Fig. 6 for the data images utilized in this study. The eight compared algorithm models are as follows:

C1: Detail-Enhanced Multi-Scale Exposure Fusion in YUV Color Space (DEMEF) [9].
C2: Medical image fusion by adaptive Gaussian PCNN and improved Roberts operator (AGPRO) [13].
C3: NSCT-DCT based Fourier Analysis for Fusion of Multimodal Images (NDFA) [14].
C4: Medical image fusion based on extended difference-of-Gaussians and edge-preserving (EDGEP) [15].
C5: Fusion of Multi-modal Images using Parametrically Optimized PCNN and DCT based Fourier Analysis (PODFA) [16].
C6: Local extreme map guided multi-modal brain image fusion (LEGFF) [19].
C7: Multi-modal image fusion via coupled feature learning (CCFL) [20].
C8: Multi-modal medical image fusion via multi-dictionary and truncated Huber filtering (MDHU) [21].

Table 1 The first group of Evaluation index data

The first group metrics	MSE↓	CCI↑	PSNR↑	SSIM↑	DD↓	SM↑	CC↑	MI↑
DEMEF	51.3761	0.0409	13.9156	0.5736	38.4708	0.86768	0.81991	1.6315
AGPRO	49.6259	0.0669	14.2166	0.46725	35.6209	0.88247	0.79067	1.3507
NDFA	43.8365	0.0520	15.0845	0.57803	33.7217	0.93793	0.85214	1.5659
EDGEP	34.7534	0.1017	17.3109	0.54619	24.3515	0.96276	0.90808	1.7262
PODFA	53.7939	0.0409	13.2584	0.54242	40.8463	0.88887	0.80484	1.3905
LEGFF	33.2378	0.0938	17.6982	0.56337	23.02	0.95147	0.89353	1.7381
CCFL	33.6189	0.0930	17.5991	0.56578	23.6918	0.95225	0.89793	1.738
MDHU	22.2881	0.1575	21.1693	0.63088	14.6466	0.97967	0.95263	2.1133
LEMG(the proposed method)	<u>19.5162</u>	<u>0.1666</u>	<u>22.3229</u>	<u>0.74068</u>	<u>12.8174</u>	<u>0.97978</u>	<u>0.95991</u>	<u>2.2751</u>

Table 2 The second group of Evaluation index data

The second group metrics	MSE↓	CCI↑	PSNR↑	SSIM↑	DD↓	SM↑	CC↑	MI↑
DEMEF	25.9293	0.0484	19.855	0.8259	18.7852	0.91429	0.98639	2.8956
AGPRO	15.3384	0.1753	25.1206	0.82088	11.5966	0.94865	0.98047	2.5901
NDFA	55.2898	0.0453	13.2779	0.54027	42.7866	0.75666	0.84469	1.9238
EDGEP	22.1378	0.1474	21.2281	0.76842	15.1329	0.93721	0.97029	2.3923
PODFA	30.5006	0.0479	18.4653	0.85999	22.7474	0.88396	0.98442	2.7479
LEGFF	23.6062	0.1244	20.6703	0.68032	16.7896	0.92551	0.95671	2.1005
CCFL	21.3022	0.1344	21.5623	0.74848	15.7861	0.94043	0.97188	2.3807
MDHU	12.828	0.2265	25.9676	0.8423	8.7977	0.95601	0.98415	2.8051
LEMG(the proposed method)	<u>10.6061</u>	<u>0.2561</u>	<u>27.6197</u>	<u>0.94577</u>	<u>7.2288</u>	<u>0.97822</u>	<u>0.99162</u>	<u>3.4479</u>

Table 3 The third group of Evaluation index data

The third group metrics	MSE↓	CCI↑	PSNR↑	SSIM↑	DD↓	SM↑	CC↑	MI↑
DEMEF	14.9588	0.1290	8.211	0.6461	10.8781	0.2092	0.7922	0.9276
AGPRO	14.2079	0.1738	8.3601	0.6022	10.375	0.1918	0.7855	0.8594
NDFA	33.1718	0.0827	5.3989	0.5684	25.0895	0.0130	0.8413	1.2765
EDGEP	18.8736	0.1393	7.5379	0.5807	13.6641	0.1094	0.8008	0.9087
PODFA	19.3204	0.0980	5.9989	0.7122	16.6934	0.3156	0.8532	1.146
LEGFF	26.5485	0.0776	6.55	0.4188	19.3618	0.0481	0.7449	0.8153
CCFL	15.689	0.1236	8.073	0.5864	11.6336	0.1508	0.8198	0.9731
MDHU	15.3242	0.1604	8.1411	0.5965	11.3818	0 1985	0.75347	0.7854
LEMG(the proposed method)	<u>6.3362</u>	<u>0.3109</u>	<u>32.0943</u>	<u>0.8890</u>	<u>4.7435</u>	<u>0.5715</u>	<u>0.9296</u>	<u>1.556</u>

4.2 Experimental results

Figure 7 illustrates the fusion images obtained from the experiments. From top to bottom are the fusion image results of the first to fourth experimental groups. From left to right, they are C1 ~ C8 and the proposed algorithm are, respectively. Tables 1, 2, 3 and 4 provide the parameter data for evaluation. The data with the highest rankings in the evaluation parameters are marked with underlines.

From a subjective perspective, the fusion images between the first and third groups, as well as the fourth group of experiments reveal that C1, C4, C5, C6, C7, and the proposed fusion model algorithm exhibit superior color restoration, edge contour retention, and detail texture preservation. However, C2, C3, and C8, while maintaining good contour texture, suffer from severe color distortion and exhibit a significant color gap from the original multispectral images, resulting in poor fusion performance. The fused image results of the second

Table 4 The fourth group of Evaluation index data

The fourth group metrics	MSE↓	CCI↑	PSNR↑	SSIM↑	DD↓	SM↑	CC↑	MI↑
DEMEF	8.0207	0.1583	10.0155	0.7944	6.3067	0.4272	0.8672	1.3103
AGPRO	10.0154	0.2110	9.3725	0.7150	7.8128	0.4608	0.8045	0.9459
NDFA	37.4319	0.0797	4.9482	0.4989	23.4985	0.0280	0.6477	1.1106
EDGEP	9.2538	0.2307	10.2338	0.7255	6.4404	0.5476	0.8330	1.0774
PODFA	13.2842	0.1171	7.2987	0.7487	10.7743	0.3594	0.7873	1.1074
LEGFF	15.0047	0.1215	8.2021	0.5564	10.7575	0.2421	0.7907	0.8693
CCFL	8.6156	0.2308	9.8084	0.7555	6.013	0.4455	0.8687	1.1514
MDHU	7.7422	0.2807	10.1178	0.7241	5.4347	0.8210	0.8500	1.0622
LEMG(the proposed method)	5.0622	0.3959	34.0441	0.9018	3.2314	0.5450	0.9312	1.7411

group of experiments are similar to those of the first group of experiments. C1, C4, C5, C6, C7, and the proposed fusion model algorithm demonstrate commendable image fusion effects, while C2, C3, and C8 show subpar fusion results.

In the objective evaluation of fusion effects, this study utilizes eight fusion assessment parameters, namely mean squared error (MSE), color consistency index (CCI), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), distortion degree (DD), similarity measure (SM), correlation coefficient (CC), and mutual information (MI). MSE measures the expected value of the squared difference between estimated and true parameter values, indicating the degree of data variation. A smaller MSE implies better accuracy in describing experimental data. CCI gages the perceptual color difference between two colors, with higher CCI values indicating smaller color differences between the fusion and source images. PSNR represents the mean squared error between the fusion and original images, with higher values signifying lower image distortion. SSIM quantifies the similarity between two images, where larger SSIM values indicate greater similarity between the fusion and source images. DD assesses image distortion, with smaller values denoting less distortion in fusion images. SM measures image similarity, with larger values indicating better fusion effects. CC reflects the statistical indicator of the closeness of the relationship between variables, with larger values indicating a stronger correlation. MI measures the distance (similarity) between two distributions, with larger values indicating higher similarity between the fusion and source images.

Tables 1, 2, 3 and 4 demonstrate that the proposed fusion model algorithm holds a dominant position in all four sets of experiments across the eight evaluation metrics. In comparisons with eight other algorithms, all the data are in the first place. The algorithm exhibits superior color retention in fusion images, coupled with lower spectral distortion, demonstrating outstanding image fusion performance and applicability to various image fusion scenarios.

5 Conclusion

The experimental results highlight the efficacy of the proposed fusion model algorithm in maintaining color accuracy and minimizing spectral losses within the domain of remote sensing image fusion. Additionally, it demonstrates a robust performance in preserving image edge contours and textures during the fusion process. Future research will focus on refining and enhancing the overall image fusion performance of this model.

Author contributions M.S. and X.Z. made significant contributions to conceptualization and methodology. M.S. , X.Z. and Y.N. were involved in the software development and validation. M.S. and X.Z. actively participated in the writing of the original draft. X.Z. , M.S. and Y.L. made substantial contributions to the writing, reviewing, and editing process. X.Z. provided research materials. X.Z. played a crucial role in supervision and funding acquisition. All authors critically reviewed the manuscript.

Data Availability Statement The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest regarding the publication of this article.

References

1. Zhang, P., Jiang, Q., Cai, L., Wang, R., Wang, P., Jin, X.: Attention-based F-UNet for Remote Sensing Image Fusion. In: IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). Haikou,

Hainan, China **2021**, 81–88 (2021). https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00038

2. Chu, F., Liu, H., Wang, Z., Cao, Z.: Super-resolution Reconstruction of Airborne Remote Sensing Images based on Multi-scale Fusion. In: 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, pp. 648–651 (2022). https://doi.org/10.1109/ICBAIE56435.2022.9985886.

3. Ma, W., et al.: A multi-scale progressive collaborative attention network for remote sensing fusion classification. IEEE Trans. Neural Netw. Learn. Syst. **34**(8), 3897–3911 (2023). https://doi.org/10.1109/TNNLS.2021.3121490

4. Tan, W., Xiang, P., Zhang, J., Zhou, H., Qin, H.: Remote sensing image fusion via boundary measured dual-channel PCNN in multi-scale morphological gradient domain. IEEE Access **8**, 42540–42549 (2020). https://doi.org/10.1109/ACCESS.2020.2977299

5. Basheer, P. I., Prasad, K. P., Gupta, A. D., Pant, B., Vijavan, V. P., Kapila, D.: Optimal fusion technique for multi-scale remote sensing images based on DWT and CNN. In: 2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2022, pp. 1–6. https://doi.org/10.1109/ICSSS54381.2022.9782239.

6. Li, S., Kang, X., Fang, L., Hu, J., Yin, H.: Pixel-level image fusion: A survey of the state of the art. Inform. Fusion,**33**, 100–112, ISSN 1566–2535. https://doi.org/10.1016/j.inffus.2016.05.004.

7. Kumar, K. V., Sathish, A.: A comparative study of various multimodal medical image fusion techniques—A review. In: Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII). Chennai, India, pp 1–6 (2021). https://doi.org/10.1109/ICBSII51839.2021.9445149

8. Sebastian, J., King, G.R.G.: Fusion of multimodality medical images—a review. In: Smart Technologies, Communication and Robotics (STCR). Sathyamangalam, India, pp 1–6 (2021). https://doi.org/10.1109/STCR51658.2021.9588882

9. Wang, Q., Chen, W., Wu, X., Li, Z.: Detail-enhanced multi-scale exposure fusion in YUV color space. IEEE Trans. Circuits Syst. Video Technol. **30**(8), 2418–2429 (2020). https://doi.org/10.1109/TCSVT.2019.2919310

10. Li, Y., Liu, M., Han, K.: Overview of multi-exposure image fusion. In: 2021 International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB), Yilan County, Taiwan, 2021, pp. 196–198. https://doi.org/10.1109/ICEIB53692.2021.9686453.

11. Li, H., Wang, J., Han, C.: Image mosaic and hybrid fusion algorithm based on pyramid decomposition. In: 2020 International Conference on Virtual Reality and Visualization (ICVRV), Recife, Brazil, 2020, pp. 205–208. https://doi.org/10.1109/ICVRV51359.2020.00049.

12. Li, D., Dong, X., Wang, K., Zhou, M., Chen, H., Su, J.: Image fusion algorithm based on Laplacian pyramid and principal component analysis transforms. In: 2022 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2022, pp. 31–35. https://doi.org/10.1109/ICCNEA57056.2022.00018.

13. Vajpayee, P., Panigrahy, C., Kumar, A.: Medical image fusion by adaptive Gaussian PCNN and improved Roberts operator. SIViP **17**, 3565–3573 (2023). https://doi.org/10.1007/s11760-023-02581-4

14. Jana, M., Basu, S., Das, A.: NSCT-DCT based Fourier Analysis for Fusion of Multimodal Images. In: 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Dehradun, India, 2021, pp. 1–6. https://doi.org/10.1109/UPCON52273.2021.9667618.

15. Jie, Y., Li, X., Wang, M., Zhou, F., Tan, H.: Medical image fusion based on extended difference-of-Gaussians and edge-preserving. Expert Syst. Appl.**227**, 120301. ISSN 0957–4174 (2023). https://doi.org/10.1016/j.eswa.2023.120301

16. Jana, M., Basu, S., Das, A.: Fusion of Multimodal Images using Parametrically Optimized PCNN and DCT based Fourier Analysis. In: IEEE Delhi Section Conference (DELCON). New Delhi, India **2022**, 1–7 (2022). https://doi.org/10.1109/DELCON54057.2022.9753411

17. Zhang, Y., Lee, H. J.: Infrared and visible image fusion based on multi-scale decomposition and texture preservation model. In: 2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), Shanghai, China, 2021, pp. 335–339 (2021). https://doi.org/10.1109/ICCEAI52939.2021.00067.

18. Wu, S., Zhang, K., Yuan, X., Zhao, C.: Infrared and visible image fusion by using multi-scale transformation and fractional-order gradient information. In: ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1–5 (2023). https://doi.org/10.1109/ICASSP49357.2023.10096652.

19. Yu, Z., Wenhao, X., Shunli, Z., Jianjun, S., Ran, W., Xiangzhi, B., Li, Z., Qing, Z.: Local extreme map guided multi-modal brain image fusion. Front. Neurosci. **16** (2022). https://doi.org/10.3389/fnins.2022.1055451.

20. Veshki, F. G., Ouzir, N., Vorobyov, S. A., Ollila, E.: Multimodal image fusion via coupled feature learning. Signal Process.**200**, 108637. ISSN 0165–1684 (2022). https://doi.org/10.1016/j.sigpro.2022.108637.

21. Jie, Y., Li, X., Tan, H., Zhou, F., Wang, G.: Multi-modal medical image fusion via multi-dictionary and truncated Huber filtering. Biomed. Signal Process. Control**88**(Part B), 105671. ISSN 1746–8094 (2024). https://doi.org/10.1016/j.bspc.2023.105671.

22. Lindeberg, T.: Scale-space theory in computer vision. Lecture Notes in Computer Science (1993)

23. Witkin A. P. Scale-space fltering. In: Proceedings of 8th Int. Joint Conf. Art. Intell., 1983, pp 1019–1022

24. Simonoff, J.S.: Smoothing Methods in Statistics (Springer Series in Statistics). Springer-Verlag, Berlin (1996)

25. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397–1409 (2013). https://doi.org/10.1109/TPAMI.2012.213