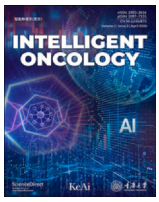
Intelligent Oncology 2 (2026) 100054

Contents lists available at ScienceDirect

Intelligent Oncology

journal homepage: www.keaipublishing.com/en/intonc/



Full Length Article

A multiscale residual dense fusion network for nuclear medical image fusion

Jun Fu, Gengyu Ge*, Yihui Tan, Miaoqiang Yang, Qian Wang, Jie Yang, Ya Wang, Aijia Ouyang

School of Information Engineering, Zunyi Normal University, Zunyi Guizhou 563006, China

ARTICLE INFO

Keywords:
Multimodal medical image
Multiscale residual dense
Fusion network

ABSTRACT

Introduction: Multimodal medical image fusion technology generates new images containing more accurate disease information by fusing different modal images. It not only improves the accuracy and efficiency of diagnosis but also provides strong support for the formulation of treatment plans. Meanwhile, it also shows great potential value in disease monitoring, personalized medicine, and clinical research. Although different multimodal medical image fusion methods have been presented, most of them are hindered by information loss, blurred edges, and low fusion efficiency.

Methods: To solve these problems, this paper proposes a multiscale residual dense fusion network (MRDFN) for multimodal medical image fusion. MRDFN integrates the strengths of both the multiscale residual network and dense network to achieve feature extraction and fusion.

Results: Experiments show that the fusion images of the proposed method are superior to the reference methods in terms of edge intensity, detail definition, and objective metrics. The comparative analysis of these fusion metrics proves that the fusion image quality of MRDFN is better than that of the reference methods. The suggested method achieves higher values in average gradient, standard deviation, spatial frequency, and visual information fidelity for fusion, with average gradient reaching 2.0 times the average of the comparison algorithms, standard deviation 1.2 times, spatial frequency 2.3 times, and visual information fidelity for fusion 1.3 times.

Conclusions: The findings in this study demonstrate that MRDFN outperforms other approaches discussed in the analysis, particularly in objective metrics and detailed information, and the average fusion time of MRDFN is lower than that of most reference methods, demonstrating effective multimodal medical image fusion.

1. Introduction

Accurate diagnosis and effective treatment of diseases, cancers in particular, are crucial for protecting human health. With the rapid advancement of medical imaging technology, various advanced imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission CT (SPECT), are widely used in clinical practice. However, each imaging technology has its unique advantages and limitations. For example, CT provides optimal visualization of bone structures, MRI offers strong soft tissue contrast, and PET and SPECT can reflect the metabolic functions of tissues. To leverage the advantages of different imaging modalities and overcome the limitations of single-modality imaging, image fusion technology[1–4] has emerged. This technology integrates information from multiple medical imaging modalities, fully exploits their respective advantages, compensates for their shortcomings, and provides doctors with richer, more comprehensive, and more accurate diagnostic information. This not only improves the reliability and accuracy of disease diagnosis but also provides stronger support for treatment planning, ultimately enhancing treatment outcomes and patients' quality of life. Therefore, an in-depth study of medical image fusion technology has significant practical importance and broad application value.

Medical image fusion aims to effectively combine multimodal image information to generate a fusion image containing more complete and precise information. Such fusion images provide a rich diagnostic basis, improve diagnostic accuracy and reliability, and support the development of precise and personalized treatment plans, thereby improving treatment outcomes and patient prognosis. For example, tumors can be more accurately localized and delineated by fusing CT and MR images,

* Corresponding author.
E-mail address: 948413217@qq.com (G. Ge).

https://doi.org/10.1016/j.intonc.2026.100054
Received 4 January 2026; Received in revised form 18 March 2026; Accepted 27 March 2026
Available online 2 April 2026
2950-2616/© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

enabling more precise surgical planning. During treatment, medical image fusion can also be used to monitor therapeutic efficacy. By comparing fusion images obtained before and after treatment, clinicians can evaluate treatment outcomes and adjust treatment plans accordingly. Furthermore, medical image fusion can be applied in telemedicine, where integrating imaging data from different locations enables doctors to provide more accurate diagnoses and treatment recommendations for remote patients.

Early medical image fusion methods realized effective information integration by decomposing source images into multiple scales or resolution levels. These methods preserve and enhance important features across different scales, thereby improving fusion quality. For example, wavelet transform[5,6] decomposes an image into multiple frequency subbands using wavelet basis functions, with each subband representing image details at a specific scale. During fusion, the most significant features at each scale are selected according to predefined rules, and the fusion image is reconstructed using the inverse wavelet transform. The Laplacian pyramid[7,8] is another classical multiscale representation method that constructs a pyramid through successive downsampling and upsampling operations. The original image is first successively downsampled to generate a series of low-resolution representations. These representations are then upsampled to the original size and subtracted from the corresponding images to obtain Laplacian pyramid representations at different levels. During the fusion stage, optimal pixel values are selected from the corresponding layers of the source images and combined. Finally, the fusion image is reconstructed through the inverse transformation process. The non-subsampled contourlet transform (NSCT)[9,10] is an improved wavelet-based method designed to overcome the shortcomings of traditional wavelet transforms in directional representation. It can more accurately capture geometric features, such as edges and textures, because NSCT supports both multiscale and multidirectional analysis. Non-subsampled shearlet transform (NSST)[11,12] is also specially designed to efficiently capture and represent anisotropic features in images. Similar to NSCT, the NSST preserves directional information; however, it can be more flexible and efficient in certain scenarios. These multiscale transform-based methods possess distinct characteristics and have improved the performance of medical image fusion to some extent.

Deep learning–based medical image fusion methods have undergone rapid development in recent years. These methods leverage the powerful feature extraction capabilities of convolutional neural networks (CNNs)[13–15] to realize effective fusion of multimodal medical images. CNNs automatically learn hierarchical feature representations from input images through stacked convolutional, pooling, and fully connected layers. In image fusion tasks, CNN models with specialized architectures can extract key features from source images and fuse them effectively based on these representations. In addition, attention mechanisms[16–18] enable models to focus on the most informative regions of input images, which is particularly important when processing medical images with complex backgrounds and significant targets. Through training, attention mechanisms can locate and fuse key information more accurately to improve the fusion effect in different modal images. Recent advances in medical image fusion include methods such as a co-occurrence filter and local extrema-based framework in the NSST domain,[19] a content-aware generative adversarial network–based approach,[20] TSJNet (a multimodality target and semantic awareness joint-driven image fusion network),[21] clustering-based multimodal fusion,[22] and directive clustering contrast-based multimodality medical image fusion for smart healthcare systems.[23] Different from the previously proposed MSRPAN (a multiscale residual pyramid attention network for medical image fusion),[24] this study introduces a multiscale residual dense fusion network, termed MRDFN, for multimodal medical image fusion. MSRPAN employed a multiscale residual pyramid attention network consisting of three components: a feature extractor, a feature fusion module, and a feature reconstruction module. It requires an additional fusion strategy to combine the high-dimensional features extracted by the network before reconstructing the fused image. In contrast, MRDFN is an end-to-end model that directly generates the fused image from the input images without requiring an explicit fusion strategy. This design simplifies the overall framework while maintaining reliability and adaptability, making it effective for multimodal medical image fusion.

However, most early fusion methods have several limitations, including complex parameter settings, low fusion efficiency, edge degradation, and information loss. Different from traditional fusion approaches, deep learning-based methods overcome these shortcomings to a certain extent and have therefore been widely studied and applied in the field of image processing. Although recent deep learning–based fusion methods have achieved notable progress, there is still significant room for improvement in terms of fusion efficiency, model complexity, edge preservation, and information retention. To address these challenges, this paper proposes an MRDFN for nuclear medical image fusion. Experimental results indicate that MRDFN outperforms existing methods in objective evaluation metrics and information richness. The contributions of this paper are described as follows:

(1) We propose a multiscale residual dense fusion network for nuclear medicine image fusion. The architecture effectively utilizes both local and global information and improves the fusion quality of multimodal complementary information.
(2) We adopt three different loss functions to enhance the quality of the fusion images. Mean squared error (MSE) loss is used to preserve basic image content, structural similarity loss improves image details, and gradient difference loss enhances edge strength.
(3) We analyze the distributions of objective evaluation metrics and pixel intensities of fusion images obtained using different methods. The experimental results demonstrate that the proposed method achieves superior fusion performance and that its pixel distributions are more uniform and reasonable compared to those of the competing methods.

2. Related works

This section introduces the principles and applications of multiscale residual networks and multiscale dense networks, which form the basis of the subsequent work.

2.1. Multiscale residual network

The multiscale residual network[25,26] adaptively extracts image features at different scales using convolution kernels of varying sizes, thereby capturing image information more comprehensively and effectively. First, it employs multiple convolution kernels of different sizes to perform convolution operations, enabling the simultaneous extraction of both large-scale and small-scale features from the image. For example, large convolution kernels capture the overall contour and texture information, while small convolution kernels focus more on fine details and edge information. Second, the residual network incorporates a skip connection mechanism that directly transfers input information to the output and fuses it with features obtained from convolution operations. This connection helps alleviate the vanishing gradient problem, making the network easier to optimize and improving the efficiency of information transmission. In addition, it enables effective fusion of features at different scales, resulting in richer and more representative image feature representations. This maximizes the utilization of image feature information and improves overall network performance.

Multiscale residual networks are widely used in image-related tasks. For example, they have been applied in image reconstruction[27,28] to better restore image details and texture information and improve image clarity and quality. They were also used in tasks such as pose estimation of moving objects, robotic navigation, and autonomous driving. By

integrating multiscale feature extraction, residual connections, and feature fusion, multiscale residual networks can process image data more effectively and improve performance and generalization ability, demonstrating broad application prospects across various fields.

2.2. Multiscale dense network

The multiscale dense network, termed MSDnet,[29,30] is a neural network architecture that combines multiscale feature extraction with dense connections, aiming to process image information at different scales. It contains multiple receptive fields, enabling it to simultaneously capture global information, context information, and fine details. For example, a larger receptive field can capture macro-level information, such as the overall structure of an image, while a smaller receptive field focuses on micro-level details, such as local textures and edges, thereby facilitating a more comprehensive understanding of image content. In addition, the multiscale dense network can effectively detect and recognize both large and small objects. For large-scale targets, the network utilizes its global receptive field for accurate localization and classification, whereas for small-scale targets, the local receptive field improves detection accuracy.

Compared with traditional network architectures, each layer in a multiscale dense network is closely connected to all preceding layers, allowing it to receive feature information from every previous layer. This dense connection strategy not only ensures maximum feature reuse and mitigates information loss but also integrates features from different levels to generate more representative feature maps. Moreover, the dense connection structure promotes efficient feature propagation and interaction throughout the network, enabling deep features to be rapidly transmitted to shallow layers and providing richer information for training. This enhanced feature interaction helps the network better capture underlying data patterns and improves its generalization capability.

Multiscale dense networks have demonstrated strong performance in various computer vision tasks. For example, in image classification tasks,[31,32] they can accurately recognize images across different scales and categories. In object detection tasks,[33,34] they effectively locate and identify objects of varying sizes. In semantic segmentation tasks,[35–37] they enable precise pixel-level classification. By integrating multiscale feature extraction with dense connectivity, multiscale dense networks demonstrate strong image processing capabilities and achieve robust performance across a wide range of applications.

3. Methods

Medical image fusion technology is indispensable in modern medicine, and its importance is self-evident. In medical diagnosis and treatment, accurate and comprehensive medical image information serves as a key basis for physicians to make correct decisions. However, due to the differing principles and characteristics of various imaging devices, the images they acquire can only provide information from a single perspective, which imposes certain limitations. Therefore, this paper proposes a MRDFN for nuclear medical image fusion. (Figure 1).

The pseudo-code of the training process is described in Algorithm 1 and the feature dimensions are presented in Table 1.

Algorithm 1. Pseudo-code of the training process.

1. Input the training set into MRDFN for training. The learning rate is set to 0.001, the training batch size is 4, the number of training epochs is 200, and the optimization algorithm is SGD.
2. The training process is completed when the number of training epochs reaches 200, and the trained MRDFN model is obtained.
3. Input the testing set into the trained model until the fusion process is completed.
4. The fusion images are obtained after the fusion process.

3.1. Fusion principle

MRDFN consists of six convolution layers and five batch normalization layers. After adding two input images, MRDFN automatically completes the feature extraction and reconstruction processes. After 200 training iterations, MRDFN is capable of completing the end-to-end fusion process of input images. With its simple structure, MRDFN can achieve the fusion of multimodal medical images without the need for a complex parameter adjustment process. The formula for MRDFN is defined as follows:

$$\begin{aligned} F_1 &= BN(N_1(M_1 + M_2)) \\ F_2 &= BN(N_2(F_1) + F_1) \\ F_3 &= BN(N_3(Con(F_2, F_1)) + F_1 + F_2) \\ F_4 &= BN(N_4(Con(F_3, F_2, F_1)) + F_1 + F_2 + F_3) \\ F_5 &= BN(N_5(Con(F_4, F_3, F_2, F_1)) + F_1 + F_2 + F_3 + F_4) \\ M_f &= N_6(F_5) \end{aligned} \tag{1}$$

where F_1, F_2, F_3, F_4, and F_5 are the deep features after extraction. N_1, N_2, N_3, N_4, N_5, and N_6 are six convolution layers. BN is a batch normalization layer. Con is a concatenation operation. M_1 and M_2 are two multimodal medical images. M_f is a fusion image.

3.2. Loss functions

Network loss functions play an important role in the training process, especially in the optimization process of neural network training. The training process goal is to minimize the difference between the real data and the generated data. Different tasks and application scenarios require different loss functions. To improve the fusion quality of MRDFN, three loss functions were employed in this study. These three loss functions are introduced below.

3.2.1. MSE loss

MSE[38] is a loss function commonly used, especially in regression problems. It measures the difference between the predicted data and the real data and uses square errors to focus on the punishment of larger errors, with good mathematical properties. In the optimization process, the gradient of MSE is proportional to the prediction error, which helps the model converge to the optimal solution quickly. The smaller the MSE loss, the more information the fusion image retains.

The MSE loss is defined as follows:

$$loss_{MSE} = \frac{1}{RC} \sum_{i=0}^{R-1} \sum_{j=0}^{C-1} (M_{i,j}^{r} - M_{i,j}^{f})^2 \tag{2}$$

where R and C are the numbers of image rows and columns, respectively. M^r and M^f denote the reference image and the fusion image, respectively.

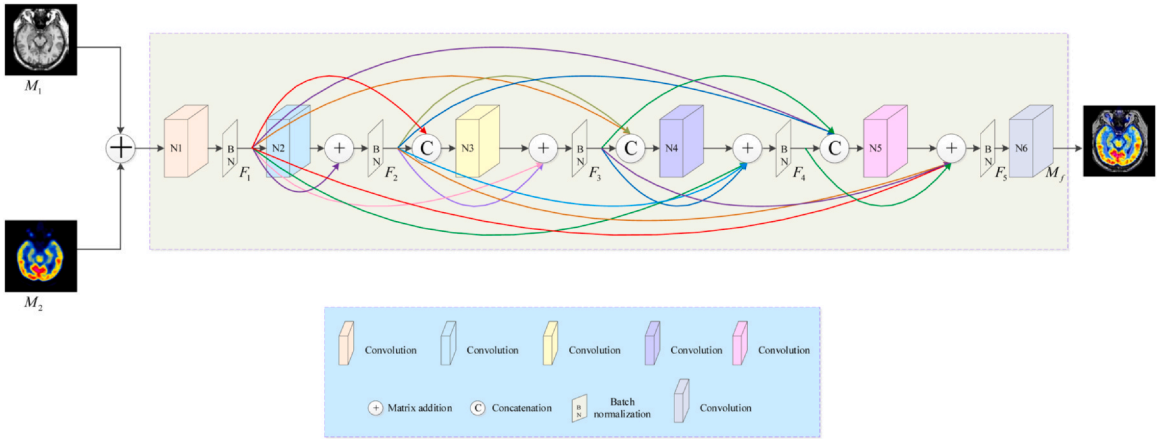


Figure 1. MRDFN schematic.

3.2.2. Structural similarity metrics loss

Structural similarity metrics (SSIM)[38] is a metric to measure the similarity of two images, which considers the brightness, contrast, and structure of the image. Different from the traditional pixel-level loss functions, such as MSE, SSIM better aligns with human visual perception of image quality. The smaller the SSIM loss, the richer the detailed information of the fusion image. The SSIM loss function is as follows:

$$loss_{SSIM} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

where x and y are the two images to be compared. μ_x and μ_y are the mean value of x and y, representing brightness. σ_x^2 and σ_y^2 are the variance of x and y, representing contrast. σ_{xy} is the covariance of x and y, representing structural similarity. C_1 and C_2 are the constants used to avoid cases where the denominator is zero. Typically, $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$. L is the dynamic range of the pixel value and is usually set to $L = 255$.

3.2.3. Gradient difference loss

Gradient difference (GD) loss[39] is a loss function used for image processing and computer vision tasks. The core idea of this loss function is to assess image quality by comparing the gradient spatial difference between the original image and the fusion image. GD loss can capture edge and texture information, which is crucial to maintain structural image details. The definition of gradient differential loss is as follows:

$$loss_{GD} = \frac{1}{RC}\left(\sum_{i=0}^{R-1}\sum_{j=0}^{C-1} \| \nabla_{sobel} M_{i,j}^1 - \nabla_{sobel} M_{i,j}^f \|^2 - \sum_{i=0}^{R-1}\sum_{j=0}^{C-1} \| \nabla_{sobel} M_{i,j}^2 - \nabla_{sobel} M_{i,j}^f \|^2\right) \tag{4}$$

where R and C are the numbers of rows and columns of the image. ∇_{sobel} is the Sobel gradient operator. M^1 and M^2 are the two reference images, respectively. M^f is a fusion image.

3.2.4. Total loss

In this study, MSE loss, SSIM loss, and GD loss were used to optimize the network. Therefore, the total loss function of the network is defined as follows:

$$loss_{total} = loss_{MSE} + loss_{SSIM} + k \cdot loss_{GD} \tag{5}$$

where $loss_{MSE}$ is the MSE loss, $loss_{SSIM}$ is the SSIM loss, and $loss_{GD}$ is the GD loss. Set $k = 0.35$.

4. Results

Because the principles of each fusion algorithm are different, the resulting fusion images also vary significantly. To compare the different fusion algorithms, nine image fusion methods were used in this paper. These fusion methods are GFF,[40] MLES,[41] LLFD,[42] PSF,[43] CSMCA,[44] ATBIF,[45] JBFLGE,[46] ALMFnet,[47] and LRFNet.[48] Three kinds of multimodal medical images are compared: MRT1-PET, MRT2-PET, and MRT2-SPECT image pairs. To more objectively compare the image fusion quality of different fusion methods, this study employed six different evaluation metrics, namely average gradient (*AG*),[49] entropy (EN),[49] standard deviation (SD),[50] spatial frequency (SF),[50] spatial and spectral entropies (SSEQ),[51] and visual information fidelity (VIFF).[52] and represent the image gradient and details. EN, SD, and VIFF reflect the quality of image content. SSEQ represents the fidelity of image color. The larger AG and SF are, the richer the details of the image and the higher the edge intensity. The larger EN, SD, and VIFF are, the higher the quality of the image content. The smaller SSEQ is, the higher the fidelity of image color.

Table 1
Parameters of MRDFN.

Convolution block	Kernel size	Stride	Padding	Input channel	Output channel
N1	1 × 1	1	0	1	64
N2	5 × 5	1	2	64	64
N3	7 × 7	1	3	128	64
N4	3 × 3	1	1	192	64
N5	1 × 1	1	0	256	64
N6	1 × 1	1	0	64	1

The following sections introduce these evaluation metrics, qualitative comparisons of fusion results, quantitative comparisons of evaluation metrics, and corresponding discussions. The datasets are all from the whole brain Atlas (http://www.med.harvard.edu/AANLIB/home.html). The training set includes 400 images, and the test set includes 180 images, totaling 580 images. In addition, the learning rate is 0.001, the number of learning epochs is 200, and stochastic gradient descent is utilized as an optimization algorithm. The training batch size is set to 4.

4.1. Fusion metrics

AG is a metric used to measure the image clarity and the richness of texture details. It reflects the rate of change in image gray values and is used to describe the overall gradient characteristics of an image.

$$AG = \frac{1}{(M-1)(N-1)} \sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \sqrt{[(\frac{\partial f(m,n)}{\partial m})^2 + (\frac{\partial f(m,n)}{\partial n})^2]/2} \tag{6}$$

where M and N are the image length and height, respectively. f is a fusion image.

EN is a measure of image information richness. It represents the uncertainty or information content of random variables. The more uniform and random the gray value distribution, the greater the EN and the richer the information contained in the image.

$$EN = -\sum_{i=0}^{L} h_{If}(i) \log_2 h_{If}(i) \tag{7}$$

where h_{If} represents the pixel value distribution histogram of I_f. L is the number of gray levels in the image, taken $L = 255$.

SD is a statistic used to measure the discreteness of image pixel values. It reflects the deviation degree of pixel value from the mean value in the image. The larger the SD, the more scattered the pixel values, and the image contains more details and textures.

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_{i,j} - \mu)^2} \tag{8}$$

where M and N represent the image length and width, respectively I is a fusion image and μ represents its pixel mean.

SF is a metric to describe the intensity of gray changes in an image. It reflects the richness of detail and texture in the image. The larger the SF, the more intense the change in image gray value, and thus the richer the details and textures.

$$SF = \sqrt{\frac{1}{MN} \sum_{x=1}^{M} \sum_{y=2}^{N} [I_f(x,y) - I_f(x,y-1)]^2 + \frac{1}{MN} \sum_{y=1}^{N} \sum_{x=2}^{M} [I_f(x,y) - I_f(x-1,y)]^2} \tag{9}$$

where M and N represent the image length and width, respectively I_f is a fusion image.

SSEQ uses support vector machine (SVM) to train the image distortion and quality prediction engine to obtain the quality metrics of image spatial spectral entropy. The smaller the SSEQ, the smaller the loss of spatial spectral information and the color distortion.

$$\begin{aligned} SSEQ &= SVM(E_s, E_f) \\ &= SVM(-\sum_x p(x) \log_2 p(x), -\sum_i \sum_j P(i,j) \log_2 P(i,j)) \end{aligned} \tag{10}$$

where E_s is spatial entropy and E_f is spectral entropy. p is the pixel histogram of the fusion image $P(i,j) = \frac{C(i,j)^2}{\sum_i \sum_j C(i,j)^2}$. $1 \le i \le 8$ and $1 \le j \le 8$. C is a block discrete cosine transform efficient matrix computed on 8×8 blocks.

Visual information fidelity (VIFF) is a new image fusion metric based on the fidelity of visual information. This metric emphasizes measuring image quality from the perspective of the human visual system. The larger the VIFF, the higher the fidelity of visual information and the smaller the distortion.

$$VIFF(I_1, \cdots, I_n, I_F) = \sum_k p_k \cdot VIFF_k(I_1, \cdots, I_n, I_F) \tag{11}$$

where the fusion visual information with distortion (FVID) and the fusion visual information without distortion (FVIND) are denoted in the local spatial and spectral domains.

$$VIFF_k(I_1, \cdots, I_n, I_F) = \frac{\sum_b FVID_{k,b}(I_1, \cdots, I_n, I_F)}{\sum_b FVIND_{k,b}(I_1, \cdots, I_n, I_F)} \tag{12}$$

4.2. Fusion image comparisons

In Figure 2, panels A and B represent the MRT1 and PET images, respectively. Panels D and F are similar; their brightness is low, and their color and cerebral cortex are not clear. Panels C, G, and I also exhibit similar characteristics; their color is dark, and the white matter is not clear. Panel E shows marked color distortion. The brightness of panel H is excessively high, resulting in a lack of detailed information. Panels J and K retain only color information, while failing to preserve the cerebral cortex structures present in the MR images. Compared with the other methods, panel L via the proposed MRDFN is not only more visually informative in color but also preserves clearer structural details of both white matter and gray matter.

In Figure 3, panels A and B show the MRT2 and PET images, respectively. Panels C, D, F, G, and K are similar; their colors are dim, and the details of white matter and gray matter are not clear. Panels E, H, I, and J are similar. Their color information is rich, and the details are clear. Compared with the other methods, the fusion panel L via the proposed MRDFN has richer color and detailed information, and the lesion area is clearer.

In Figure 4, panels A and B show the MRT2 and SPECT images, respectively. Panels D and F are similar; their brightness is low, and the details of the lesions are not clear. Panels C, G, H, I, and J are similar. They are rich in color information, but the details of the lesions are not clear. Marked color distortion was observed in panel E. There are a few details in panel K. Compared with the other methods, panel L via the proposed MRDFN has richer color and detailed information, and the lesion area is clearer.

4.3. Comparison of fusion metrics

The quantitative comparisons of fusion metrics for MRT1-PET, MRT2-PET, and MRT2-SPECT images are presented in Table 2, respectively. The proposed MRDFN achieves the best performance in three metrics, namely,: AG, SF, and SSEQ. The proposed MRDFN attains optimal results in five metrics, namely, AG, EN, SD, SF, and VIFF. The proposed MRDFN achieves the best performance in all six metrics, namely, AG, EN, SD, SF, SSEQ, and VIFF.

From the quantitative analysis, the proposed method demonstrates superior performance across the majority of evaluation metrics

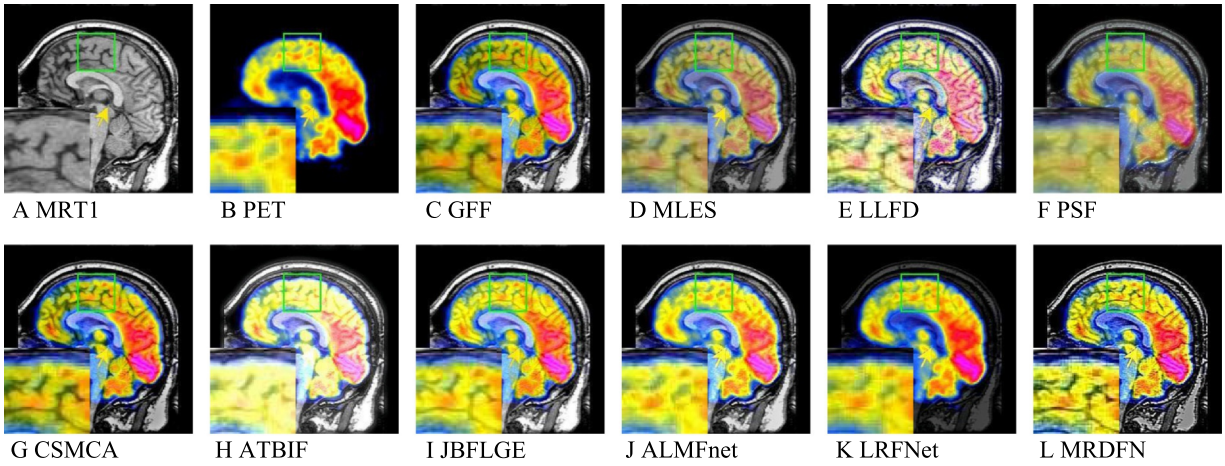


Figure 2. Comparison of MRT1-PET fusion images. Abbreviations: 1. MRT1: magnetic resonance T1 2. PET: positron emission tomography 3. GFF: image fusion with guided filtering 4. MLES: multi-scale local extrema scheme 5. LLFD: local laplacian filtering domain 6. PSF: parallel saliency features 7. CSMCA: convolutional sparsity-based morphological component analysis 8. ATBIF: adaptive two-scale biomedical image fusion 9. JBFLGE: joint bilateral filter and local gradient energy 10. ALMFnet: auto-searching light-weighted multi-source fusion network 11. LRFNet: lightweight residual fusion network.

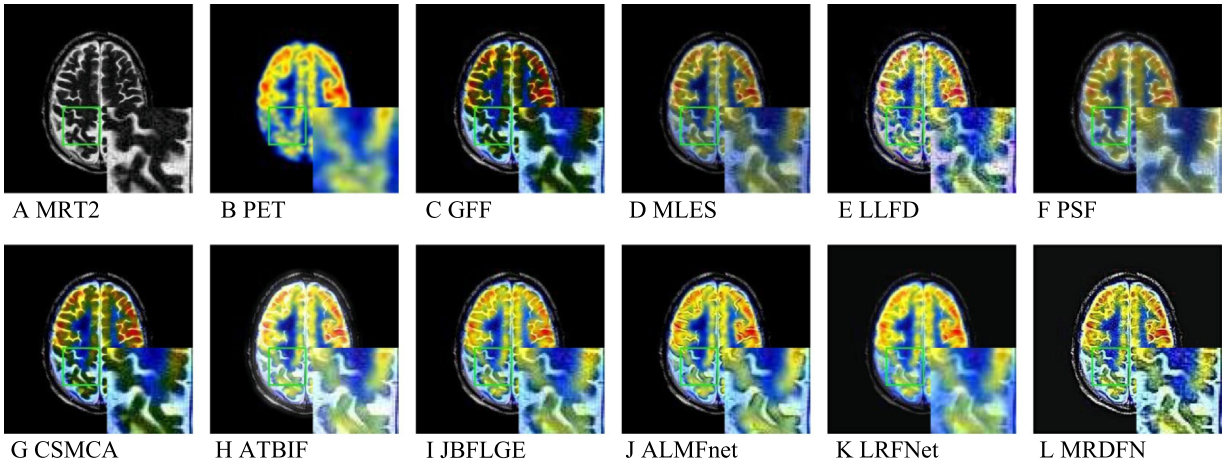


Figure 3. Comparison of MRT2-PET fusion images. Abbreviations: 1. MRT1: magnetic resonance T1 2. PET: positron emission tomography 3. GFF: image fusion with guided filtering 4. MLES: multi-scale local extrema scheme 5. LLFD: local laplacian filtering domain 6. PSF: parallel saliency features 7. CSMCA: convolutional sparsity-based morphological component analysis 8. ATBIF: adaptive two-scale biomedical image fusion 9. JBFLGE: joint bilateral filter and local gradient energy 10. ALMFnet: auto-searching light-weighted multi-source fusion network 11. LRFNet: lightweight residual fusion network.

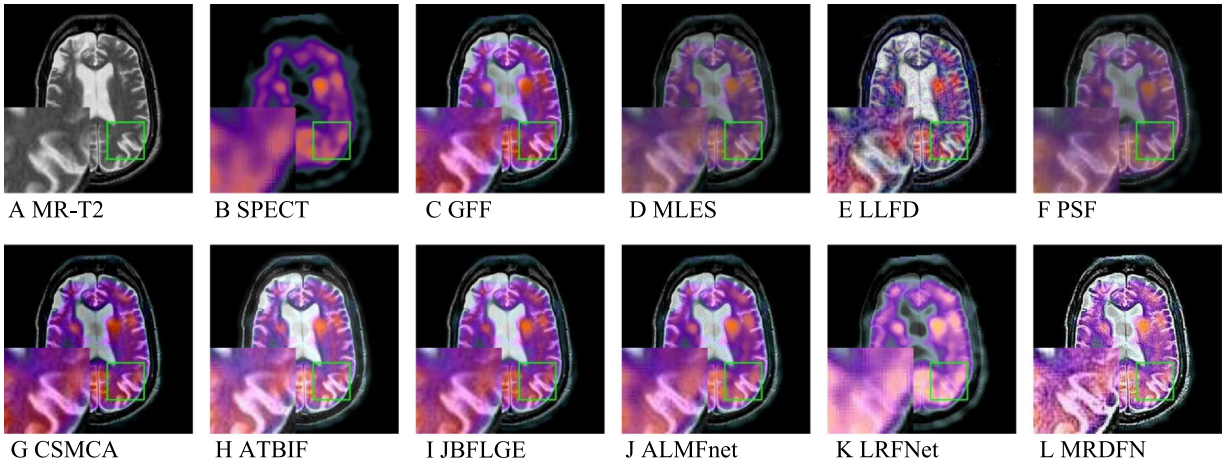


Figure 4. Comparison of MRT2-SPECT fusion images. Abbreviations: 1. MRT1: magnetic resonance T1 2. PET: positron emission tomography 3. GFF: image fusion with guided filtering 4. MLES: multi-scale local extrema scheme 5. LLFD: local laplacian filtering domain 6. PSF: parallel saliency features 7. CSMCA: convolutional sparsity-based morphological component analysis 8. ATBIF: adaptive two-scale biomedical image fusion 9. JBFLGE: joint bilateral filter and local gradient energy 10. ALMFnet: auto-searching light-weighted multi-source fusion network 11. LRFNet: lightweight residual fusion network 12. SPECT: single-photon emission computed tomography.

Table 2
Metric comparison of MRT1-PET, MRT2-PET, and MRT2-SPECT image fusion.

Metrics/methods	AG	EN	SD	SF	SSEQ	VIFF
GFF	11.6822	5.9497	74.9128	33.6770	42.9567	0.4913
MLES	7.5197	6.0336	67.2775	20.7875	42.1698	0.5024
LLFD	13.5391	6.2550	85.8510	37.6024	37.9091	0.5723
PSF	6.8967	**6.3328**	67.4298	20.4555	42.5929	0.3485
CSMCA	11.9188	5.5856	73.5466	35.8804	39.1573	0.5017
ATBIF	10.5787	6.2222	**95.6699**	33.3955	44.2224	**0.5985**
JBFLGE	12.0151	5.9614	80.1008	34.3337	42.5947	0.4932
ALMFnet	11.6101	5.8555	82.9956	33.9586	40.7189	0.5363
LRFNet	6.0702	5.4022	60.9167	14.6301	38.1090	0.2606
Proposed	**16.3725**	5.4804	75.7800	**52.8648**	**34.6798**	0.4800
GFF	6.7339	4.5853	53.9212	22.8519	45.6522	0.4648
MLES	4.3141	4.5903	47.2814	14.0426	42.1422	0.4237
LLFD	8.2834	4.9945	65.1701	26.2177	36.9652	0.5589
PSF	4.4368	5.0208	51.6015	15.5409	44.6872	0.3485
CSMCA	6.5108	3.9808	50.4016	22.6421	44.0579	0.4306
ATBIF	6.4950	5.1958	73.5538	21.8720	42.6507	0.6088
JBFLGE	7.1238	4.7366	64.4705	23.2705	40.5873	0.5465
ALMFnet	6.8090	4.6096	65.7824	22.7314	41.0217	0.5445
LRFNet	4.9678	5.1573	64.1623	14.3782	**35.8056**	0.4015
Proposed	**12.6050**	**5.2157**	**75.6185**	**46.9227**	37.1530	**0.6247**
GFF	5.7359	4.7982	56.1347	17.6339	38.5479	0.5464
MLES	3.7183	4.5986	44.1006	11.5079	38.7629	0.4303
LLFD	7.9394	5.0740	59.4611	23.3741	36.8333	0.5776
PSF	3.6451	5.0097	44.5338	11.8351	41.1879	0.3274
CSMCA	4.6639	3.8279	39.5576	14.8773	40.9755	0.4040
ATBIF	5.8938	5.2697	68.2010	18.2479	40.2452	0.6528
JBFLGE	6.0279	4.8311	60.8441	18.3491	36.3931	0.5698
ALMFnet	5.7773	4.7001	59.4282	18.4536	39.0987	0.6166
LRFNet	4.6438	5.2008	63.2877	14.5038	46.6718	0.4280
Proposed	**13.3055**	**5.2790**	**79.3592**	**47.2491**	**35.6210**	**0.7476**

Abbreviations: 1. GFF: image fusion with guided filtering 2. MLES: multi-scale local extrema scheme 3. LLFD: local laplacian filtering domain 4. PSF: parallel saliency features 5. CSMCA: convolutional sparsity based morphological component analysis 6. ATBIF: adaptive two-scale biomedical image fusion 7. JBFLGE: joint bilateral filter and local gradient energy 8. ALMFnet: auto-searching light-weighted multi-source fusion network 9. LRFNet: lightweight residual fusion network 10. AG: average gradient 11. EN: entropy 12. SD: standard deviation 13. SF: spatial frequency 14. SSEQ: spatial and spectral entropies 15. VIFF: visual information fidelity.

compared with the reference methods. These results indicate that the proposed method achieves higher-quality fusion images than the competing approaches.

5. Discussion

To further distinguish the differences among various fusion methods, this study conducted a pixel histogram analysis of the fusion images, and the results are shown in Figure 5. The pixel distribution of panel A is uneven and concentrated in the lower intensity range, whereas the pixels in panel B are more evenly distributed. The pixel distributions of panels C, G, and I are similar and relatively uniform, resulting in clearer color representation and better detail preservation. Panels D and F exhibit similar characteristics; their pixel distributions are concentrated in the mid-intensity range, with fewer values in both low- and high-intensity regions. Consequently, these images have low brightness and lack clear details. The pixel distribution of panel E is concentrated in the higher intensity range, with limited representation in low-intensity regions, leading to color distortion and unclear details. Similarly, panel H shows a concentration in the higher intensity range, resulting in excessive brightness and loss of detail. Panel J also has pixel values concentrated in the higher intensity range with fewer low-intensity values, leading to weak contrast. In contrast, panel K has pixel values concentrated in the lower intensity range, with limited high-intensity distribution, causing it to retain color information but lose structural details. Compared with other methods, panel L via the proposed MRDFN demonstrates a more uniform pixel distribution, richer color representation, and enhanced detail clarity.

To validate the impact of the parameter k in the loss function of Eq. (5) on fusion images, we conducted an experiment and performed a comparative analysis. The experimental results and comparisons are shown in Table 3. In this table, the bold numbers represent optimal metrics, and the underlined numbers represent suboptimal metrics. After comparisons, we can see that the optimal metrics are the most when $k = 0.4$. However, the color distortion of the fusion images is obvious, and the brightness is the lowest when $k = 0.4$. In contrast, the fusion images have richer detail information and higher brightness than the fusion images with $k = 0.2$ and $k = 0.3$. Therefore, the parameter k is set to 0.35in the loss function of this research to better balance the details and brightness of fusion images.

The ablation experiment comparisons of different loss functions for MRT1-PET, MRT2-PET, and MRT2-SPECT are shown in Table 4, where the bold numbers in this table represent the optimal metrics. In the MRT1-PET image fusion experiment, MRDFN using MSE, SSIM, and GD loss functions achieved four optimal metrics, while MSRPAN achieved only two optimal metrics.

In the MRT2-PET image fusion experiment, MRDFN using MSE and SSIM loss functions achieved three optimal metrics. Similarly, MRDFN using MSE, SSIM, and GD loss functions also attained three optimal metrics, whereas MSRPAN achieved none. In the MRT2-SPECT image

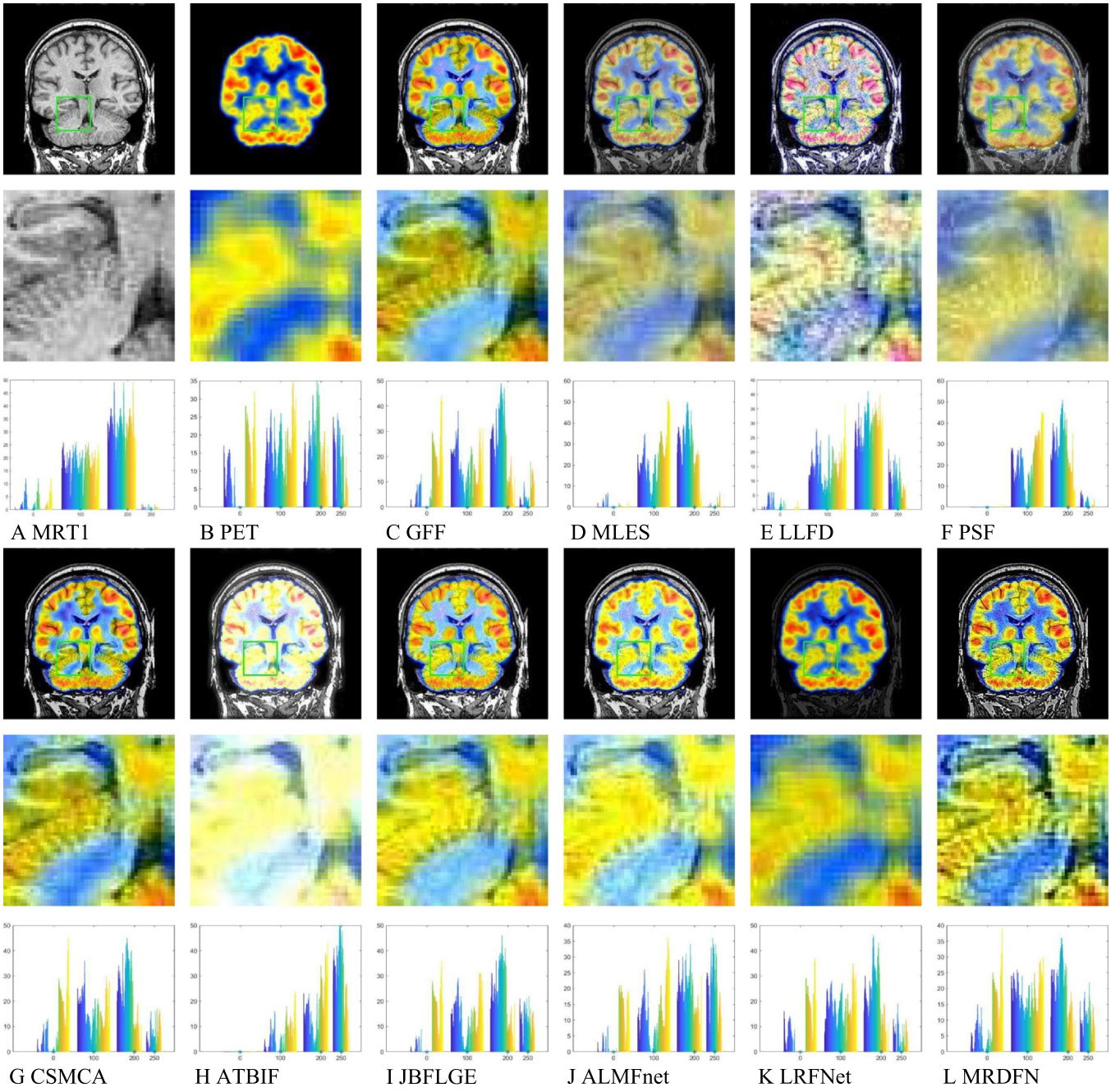
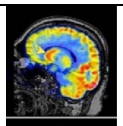
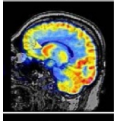
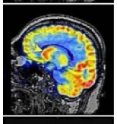
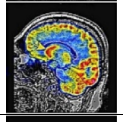
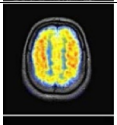
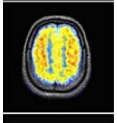
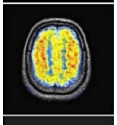
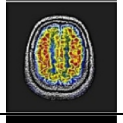
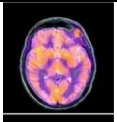
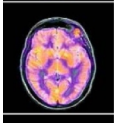
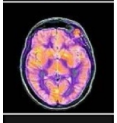


Figure 5. Pixel histograms of fusion images. The first row of images represents the original image and the fused images generated by different algorithms. The second row of images are local enlarged views, and the third row shows the pixel distribution histograms of the fusion images.Abbreviations: 1. MRT1: magnetic resonance T1 2. PET: positron emission tomography 3. GFF: image fusion with guided filtering 4. MLES: multi-scale local extrema scheme 5. LLFD: local laplacian filtering domain 6. PSF: parallel saliency features 7. CSMCA: convolutional sparsity-based morphological component analysis 8. ATBIF: adaptive two-scale biomedical image fusion 9. JBFLGE: joint bilateral filter and local gradient energy 10. ALMFnet: auto-searching light-weighted multi-source fusion network 11. LRFNet: lightweight residual fusion network.

fusion experiment, MRDFN using MSE and SSIM loss functions had one optimal metric, MRDFN using MSE, SSIM, and GD loss functions had five optimal metrics, and MSRPAN had none. In summary, MRDFN with MSE, SSIM, and GD loss functions performed best in terms of objective fusion metrics, significantly outperforming the reference method MSRPAN, particularly in detail information and brightness.

To verify the impact of different network structures on fusion quality, we conducted ablation experiments on MRT1-PET, MRT2-PET, and MRT2-

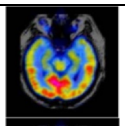
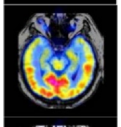
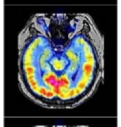
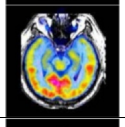
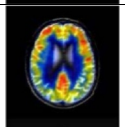
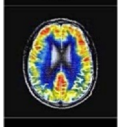
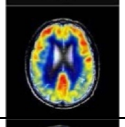
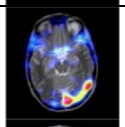
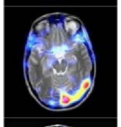
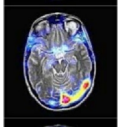
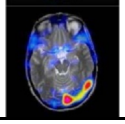
Table 3
Ablation experiment of MRT1-PET, MRT2-PET, and MRT2-SPECT image fusion based on different k values.

Imaging modality	k	Fusion image	AG	EN	SD	SF	SSEQ	VIFF
MRT1-PET	k=0.2		11.0532	5.5922	68.6332	32.9199	35.4908	0.4818
	k=0.3		13.9383	5.5039	74.8367	44.3209	34.5476	**0.5059**
	k=0.35		16.3725	5.4804	75.7800	52.8648	34.6798	0.4800
	k=0.4		**28.4866**	**6.1498**	**82.9550**	**89.1100**	**29.7760**	0.3296
MRT2-PET	k=0.2		8.6615	5.3008	71.7639	30.1275	37.2444	0.6699
	k=0.3		10.6889	5.0951	**77.0467**	39.4107	**36.8769**	**0.6717**
	k=0.35		12.6050	5.2157	75.6185	46.9227	37.1530	0.6247
	k=0.4		**21.2115**	**5.4515**	74.1414	**76.7770**	41.6643	0.4158
MRT2-SPECT	k=0.2		8.9551	5.3045	76.0501	30.6164	37.7510	**0.8061**
	k=0.3		11.2184	5.1707	**80.6973**	39.7484	36.0334	0.7996
	k=0.35		13.3055	5.2790	79.3592	47.2491	**35.6210**	0.7476
	k=0.4		**21.2115**	**5.4515**	74.1414	**76.7770**	41.6643	0.4158

Abbreviations: 1. MRT1-PET: magnetic resonance T1-positron emission tomography 2. MRT2-PET: magnetic resonance T2-positron emission tomography 3. MRT2-SPECT: magnetic resonance T2-single-photon emission computed tomography 4. AG: average gradient 5. EN: entropy 6. SD: standard deviation 7. SF: spatial frequency 8. SSEQ: spatial and spectral entropies 9. VIFF: visual information fidelity.

SPECT, and the experimental results are shown in Table 5. Without the multi-scale or residual module, MRDFN lost most of the source image information, retaining only a small amount of edge and color information and failing to achieve optimal metrics. Therefore, these modules play a crucial role in the fusion quality of MRDFN. Only with these modules can MRDFN achieve better fusion quality and objective metrics.

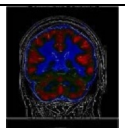
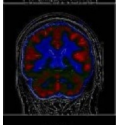
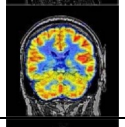
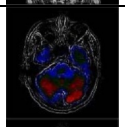
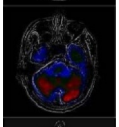
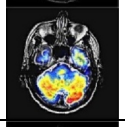
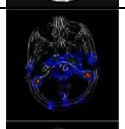
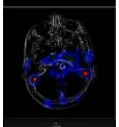
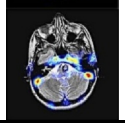
Table 4
Ablation experiment of MRT1-PET, MRT2-PET, and MRT2-SPECT image fusion based on different loss functions.

Imaging modality	Loss function	Fusion image	AG	EN	SD	SF	SSEQ	VIFF
MRT1-PET	MSE		5.5442	5.3422	50.2483	13.2275	47.7657	0.2540
	MSE + SSIM		8.0594	5.6404	65.2111	21.4684	39.6942	0.4433
	MSE + SSIM + GD		**16.3725**	5.4804	75.7800	**52.8648**	**34.6798**	**0.4800**
	MSRPAN		11.3262	**5.8636**	**77.7488**	33.5858	39.2019	0.4424
MRT2-PET	MSE		4.7245	4.6771	56.6772	14.3982	42.1736	0.4236
	MSE + SSIM		6.4109	**5.2545**	70.1718	20.7046	**34.6262**	**0.6378**
	MSE + SSIM + GD		**12.6050**	5.2157	**75.6185**	**46.9227**	37.1530	0.6247
	MSRPAN		6.3018	4.6307	62.2299	21.2552	41.0574	0.4819
MRT2-SPECT	MSE		4.5315	4.7799	60.3983	13.4426	44.4381	0.5482
	MSE + SSIM		6.3072	5.2062	74.4420	19.9787	40.8495	**0.7736**
	MSE + SSIM + GD		**13.3055**	**5.2790**	**79.3592**	**47.2491**	**35.6210**	0.7476
	MSRPAN		5.5077	4.7257	58.8697	17.2167	37.1879	0.5246

Abbreviations: 1. MRT1-PET: magnetic resonance T1-positron emission tomography 2. MRT2-PET: magnetic resonance T2-positron emission tomography 3. MRT2-SPECT: magnetic resonance T2-single-photon emission computed tomography 4. AG: average gradient 5. EN: entropy 6. SD: standard deviation 7. SF: spatial frequency 8. SSEQ: spatial and spectral entropies 9. VIFF: visual information fidelity 10. MSE: mean squared error 11. SSIM: structural similarity metrics 12. GD: gradient difference.

In addition, the average fusion time of each pair of medical images was computed to compare the complexity of the different fusion methods. The comparison results are shown in Table 6, where the bold numbers represent the best and the underlined numbers represent the second best. By comparison, GFF consumes the least average fusion time, and the proposed MRDFN consumes less average fusion time. Therefore, the proposed method has a certain real-time application value.

Table 5
Ablation experiment of MRT1-PET image fusion based on different networks of MRDFN.

Imaging modality	Network module	Fusion image	AG	EN	SD	SF	SSEQ	VIFF
MRT1-PET	MRDFN without the multiscale module		11.4655	4.5823	36.6364	44.1428	30.8873	0.0216
	MRDFN without the residual module		8.4097	4.9595	28.0612	31.4220	33.4158	0.0124
	MRDFN with the multiscale residual d ense module		**16.3725**	**5.4804**	**75.7800**	**52.8648**	**34.6798**	**0.4800**
MRT2-PET	MRDFN without the multiscale module		8.5375	3.8302	32.9835	37.3119	34.3103	0.0403
	MRDFN without the residual module		5.2288	3.8941	22.7227	23.2663	34.9123	0.0235
	MRDFN with the multiscale residual d ense module		**12.6050**	**5.2157**	**75.6185**	**46.9227**	**37.1530**	**0.6247**
MRT2-SPECT	MRDFN without the multiscale module		8.5801	3.8481	34.3427	37.0113	34.8550	0.0685
	MRDFN without the residual module		5.1840	3.9076	24.3992	22.9373	37.7915	0.0469
	MRDFN with the multiscale residual d ense module		**13.3055**	**5.2790**	**79.3592**	**47.2491**	**35.6210**	**0.7476**

Abbreviations: 1. MRT1-PET: magnetic resonance T1-positron emission tomography 2. MRT2-PET: magnetic resonance T2-positron emission tomography 3. MRT2-SPECT: magnetic resonance T2-single-photon emission computed tomography 4. AG: average gradient 5. EN: entropy 6. SD: standard deviation 7. SF: spatial frequency 8. SSEQ: spatial and spectral entropies 9. VIFF: visual information fidelity.

Table 6
Average fusion time comparison of different methods (unit: second).

Methods/modality	MRT1-PET	MRT2-PET	MRT2-SPECT
GFF	**0.05**	**0.05**	**0.05**
MLES	7.05	5.62	5.76
LLFD	91.89	93.72	96.64
PSF	2.69	2.73	2.58
CSMCA	62.39	60.72	57.66
ATBIF	2.41	2.35	2.35
JBFLGE	0.31	0.32	0.32
ALMFnet	0.13	0.12	0.13
LRFNet	0.11	0.11	0.11
MRDFN	0.07	0.07	0.07

Abbreviations: 1. GFF: image fusion with guided filtering 2. MLES: multi-scale local extrema scheme 3. LLFD: local laplacian filtering domain 4. PSF: parallel saliency features 5. CSMCA: convolutional sparsity based morphological component analysis 6. ATBIF: adaptive two-scale biomedical image fusion 7. JBFLGE: joint bilateral filter and local gradient energy 8. ALMFnet: auto-searching lightweighted multi-source fusion network 9. LRFNet: lightweight residual fusion network 10. MRT1-PET: magnetic resonance T1-positron emission tomography 11. MRT2-PET: magnetic resonance T2-positron emission tomography 12. MRT2-SPECT: magnetic resonance T2-single-photon emission computed tomography.

6. Conclusions

Multimodal medical fusion images not only retain the key details of source images but also enhance the image contrast and clarity. They can provide a comprehensive and accurate basis for clinical diagnosis and facilitate the formulation of effective treatment plans. However, most of the traditional fusion methods suffer from several limitations, such as information loss, edge degradation, insufficient contrast, and unclear details. To address these issues, this research proposes an MRDFN for nuclear medical image fusion. First, the two input images were combined. The combined input was then fed into the multiscale residual dense fusion network to perform feature extraction and fusion, thereby generating the fused image. A multi-scale residual dense architecture was used to construct feature extractors and reconstructors to extract deep features of multimodal medical images and generate fusion images. Experimental results demonstrate that the proposed method outperforms reference methods in terms of color fidelity, detail preservation, and contrast enhancement, achieving superior performance across most objective metrics. In future work, improving the adaptability of multimodal image fusion methods and reducing noise in fusion images will be key research directions. The main contributions of

this paper include the development of a novel network architecture, MRDFN, and its application to multimodal medical image fusion. The proposed method adopted an end-to-end framework with a simple fusion strategy and strong modality adaptability. Future research should also focus on reducing computational time to further enhance its potential for real-time applications.

CRediT authorship contribution statement

Gengyu Ge: Resources, Project administration, Methodology, Investigation, Funding acquisition. **Jun Fu:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Yihui Tan:** Writing – original draft, Resources, Methodology. **Miaoqiang Yang:** Funding acquisition, Conceptualization. **Qian Wang:** Writing – review & editing, Supervision, Resources. **Jie Yang:** Writing – original draft, Resources, Methodology. **Ya Wang:** Conceptualization, Formal analysis. **Aijia Ouyang:** Formal analysis, Investigation.

Ethical statement

All medical images used in this study were downloaded from The Whole Brain Atlas (https://www.med.harvard.edu/aanlib/home.html) to ensure that no patient information is included. The dataset consists of de-identified images, and no personal or sensitive data was accessed or used during the research. Therefore, this study does not involve any ethical concerns or require ethical approval.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the anonymous editors and reviewers for their valuable advice and help that improved this paper. This work was supported by the Research Project of Zunyi Normal University (Grant No.: ZSBS[2023]3), the “Top 100 Schools and Thousand Enterprises in Science and Technology Research and Development” Project of Guizhou Provincial Department of Education (Grant No.: Qianjiaoji [2025] 015), the Major Science and Technology Special Project of Guizhou Provincial Artificial Intelligence Laboratory (Grant No.: Qiankehe Platform RSSYS[2025] Major 004), the Decision Consultation Project of Guizhou Association for Science and Technology (Grant No.: QKX2026-ZX-012), and the Guizhou Province Graduate Education Teaching Reform Project (Grant No.: 2025YJSJGXX095).

References

1. Wang H, Zhang J, Wu S, et al. MMSupcon: an image fusion-based multi-modal supervised contrastive method for brain tumor diagnosis. *Artif Intell Med.* 2025;170:103253.
2. Yi X, Tang L, Zhang H, Xu H, Ma J. Diff-IF: multi-modality image fusion via diffusion model with fusion knowledge prior. *Inf Fusion.* 2024;110:102450.
3. Huo T, Wu W, Chen X, Xue M. etc. Deep learning-based multimodal data fusion in bone tumor management: Advances in clinical decision support. *Intell Oncol.* 2025;1(3):204–215.
4. Zhou M, Huang J, Yan K, et al. A general spatial-frequency learning framework for multimodal image fusion. *IEEE T Pattern Anal.* 2025;47(7):5281–5298.
5. Lieb F, Knopp T. A wavelet-based sparse row-action method for image reconstruction in magnetic particle imaging. *Med Phys.* 2021;48(7):3893–3903.
6. Ji S, Jeong B, Kamhoua C, Leslie N, Jeong D. Forecasting network events to estimate attack risk: integration of wavelet transform and vector auto regression with exogenous variables. *J Netw Comput Appl.* 2022;203:103392.
7. Jiang Q, Jin X, Cui X, Yao S, Li K, Zhou W. A lightweight multimode medical image fusion method using similarity measure between intuitionistic Fuzzy sets joint laplacian pyramid. *IEEE Tetci.* 2023;7(3):631–647.
8. Ramarao G, Bindu C, Murthy T. Convolutional laplacian gaussian pyramid approach multimodal medical image fusion. *Multimed Tools Appl.* 2024:1–23.
9. Ibrahim S, Makhlouf M, El-Tawel G. Multimodal medical image fusion algorithm based on pulse coupled neural networks and nonsubsampled contourlet transform. *Med Biol Eng & Comput.* 2023;61(1):155–177.
10. Ibrahim S, El-Tawel G, Makhlouf M. Brain image fusion using the parameter adaptive-pulse coupled neural network (PA-PCNN) and non-subsampled contourlet transform (NSCT). *Multimed Tools Appl.* 2024;83(9):27379–27409.
11. 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):317–328.
12. Sharma D, Sachdeva J, Ahuja C, Singh A. A novel hybrid multimodal medical image fusion scheme based on non-subsampled shearlet transform. *Circ Syst Signal Pr.* 2024;43(6):3627–3648.
13. Zhang R, Chung A. EfficientQ: an efficient and accurate post-training neural network quantization method for medical image segmentation. *Med Image Anal.* 2024;97:103277.
14. Qi LL, Li X, Liu Z, Zhang P, Liu L. A depth-wise separable residual neural network for PCDH8 status prediction in thyroid cancer pathological images. *Intell Oncol.* 2025;1(4):290–298.
15. Dritsas S, De Chua K, Goh Z, Simpson R. Classification, registration and segmentation of ear canal impressions using convolutional neural networks. *Med Image Anal.* 2024;94:103152.
16. Liu T, Bai Q, Torigian D, Tong Y, Udupa J. VSmTrans: A hybrid paradigm integrating self-attention and convolution for 3D medical image segmentation. *Med Image Anal.* 2024;98:103295.
17. Shi H, Shao W, Song J, Che G. etc. Graph attention network enables multipurpose prediction of imaging mass cytometry in a hepatocellular carcinoma clinical trial. *Intell Oncol.* 2025;1(4):326–340.
18. Zhang Y, Luo L, Dou Q, Heng P. Triplet attention and dual-pool contrastive learning for clinic-driven multi-label medical image classification. *Med Image Anal.* 2023;86:102772.
19. Diwakar M, Singh P, Shankar A. Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain. *Biomed Signal Proces.* 2021;68:102788.
20. Das M, Gupta D, Bakde A. *An end-to-end content-aware generative adversarial network based method for multimodal medical image fusion. Data Analytics for Intelligent Systems: Techniques and Solutions*. IOP Publishing Bristol, UK; 2024:1–7.
21. Jie Y, Xu Y, Li X, Tan H. TSJNet: A Multi-modality Target and Semantic Awareness Joint-driven Image Fusion Network. *arXiv preprint arXiv: 2402.01212.* Published online 2024.
22. Dhaundiyal R, Tripathi A, Joshi K, Diwakar M, Singh P. *Clustering based multi-modality medical image fusion. Journal of Physics: Conference Series.* 1478. IOP Publishing; 2020012024.
23. Diwakar M, Singh P, Shankar A, et al. D. Directive clustering contrast-based multi-modality medical image fusion for smart healthcare system. *Netw Model Anal Hlth.* 2022;11(1):15.
24. Fu J, Li W, Du J, Huang Y. A multiscale residual pyramid attention network for medical image fusion. *Biomed Signal Proces.* 2021;66:102488.
25. Chen H, Han Q, Li Q, Tong X. Image steganalysis with multi-scale residual network. *Multimed Tools Appl.* 2023;82(14):22009–22031.
26. Fu X, Jiang C, Li C, Li J, Zhu X, Li F. A hybrid approach for Android malware detection using improved multi-scale convolutional neural networks and residual networks. *Expert Syst Appl.* 2024;249:123675.
27. Qiu D, Cheng Y, Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey. *Comput Meth Prog Bio.* 2023;238:107590.
28. Bai Y, Li D, Duan Q, Chen X. Analysis of high-resolution reconstruction of medical images based on deep convolutional neural networks in lung cancer diagnostics. *Comput Meth Prog Bio.* 2022;217:106592.
29. Zhang J, Zhang Y, Jin Y, Xu J, Xu X. MDU-Net: multi-scale densely connected U-Net for biomedical image segmentation. *Health Inf Sci Syst.* 2023;11(1):13.
30. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med Image Anal.* 2019;51:21–45.
31. Davila A, Colan J, Hasegawa Y. Comparison of fine-tuning strategies for transfer learning in medical image classification. *Image Vis Comput.* 2024;146:105012.
32. Sadr H, Nazari M, Yousefzade-Chabok S, Emami H, Rabiei R, Ashraf A. Enhancing brain tumor classification in MRI images: a deep learning-based approach for accurate classification and diagnosis. *Image Vis Comput.* 2025:105555.
33. Luo X, Lin X, Ouyang W, Zheng S, Chen J, Liu Z. Bi-DenseNet: automatic recognition of ocular surface disease using smartphone imaging. *Biomed Signal Proces.* 2024;96:106622.
34. Huang H, You Z, Cai H, Xu J, Lin D. Fast detection method for prostate cancer cells based on an integrated ResNet50 and YoloV5 framework. *Comput Meth Prog Bio.* 2022;226:107184.
35. Jin R, Tang HY, Yang Q, Chen W. LA-ResUNet: attention-based network for longitudinal liver tumor segmentation from CT images. *Comput Med Imag Grap.* 2025;123:102536.
36. Gao H, Gu J, Liu S, Xu S, Zhao Q. MSFA-Net: multi-scale feature aggregation and attention-enhanced U-Net for microscopic hyperspectral pathology images segmentation. *Opt Laser Technol.* 2025;187:112652.
37. Shi L, Nie R, Cao J, Liu X, Li X. SCMFusion: Semantic Constrained Multi-Scale Fusion Network for infrared and visible image fusion. *Opt Laser Technol.* 2025;190:113097.
38. Zhao H, Gallo O, Frosio I, Kautz J. Loss Functions for Neural Networks for Image Processing. *arXiv preprint arXiv: 1511.08861*. Published online 2015.

39. Fu J, Li W, Peng X, et al. MDRANet: a multiscale dense residual attention network for magnetic resonance and nuclear medicine image fusion. *Biomed Signal Proces.* 2023;80:104382.
40. Li S, Kang X, Hu J. Image Fusion With Guided Filtering. *IEEE T Image Process.* 2013;22(7):2864–2875.
41. Du J, Li W, Xiao B, Nawaz Q. Medical image fusion by combining parallel features on multi-scale local extrema scheme. *Knowl-Based Syst.* 2016;113:4–12.
42. Du J, Li W, Xiao B. Anatomical-functional image fusion by information of interest in local laplacian filtering domain. *IEEE T Image Process.* 2017;26(12):5855–5866.
43. Du J, Li W, Xiao B. Fusion of anatomical and functional images using parallel saliency features. *Inf Sci.* 2018;430:567–576.
44. Liu Y, Chen X, Ward R, Wang Z. Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Proc Let.* 2019;26(3):485–489.
45. Du J, Fang M, Yu Y, Lu G. An adaptive two-scale biomedical image fusion method with statistical comparisons. *Comput Methods Programs Biomed.* 2020;196:105603.
46. Li X, Zhou F, Tan H, Zhang W, Zhao C. Multimodal medical image fusion based on joint bilateral filter and local gradient energy. *Inf Sci.* 2021;569:302–325.
47. Mu P, Wu G, Liu J, Zhang Y, Fan X, Liu R. Learning to search a lightweight generalized network for medical image fusion. *IEEE T Circ Syst Vid.* 2023;34(7):5921–5934.
48. He D, Li W, Wang G, Huang Y, Liu S. LRFNet: a real-time medical image fusion method guided by detail information. *Comput Biol Med.* 2024;173:108381.
49. Yang Y, Park DS, Huang S, Rao N. Medical image fusion via an effective wavelet-based approach. *EURASIP J Adv Sig Pr.* 2010;2010:1–13.
50. Naidu V, Raol J. Pixel-level image fusion using wavelets and principal component analysis. *Def Sci J.* 2008;58(3):338–352.
51. Liu L, Liu B, Huang H, Bovik A. No-reference image quality assessment based on spatial and spectral entropies. *Signal Process-Image.* 2014;29(8):856–863.
52. Han Y, Cai Y, Cao Y, Xu X. A new image fusion performance metric based on visual information fidelity. *Inf Fusion.* 2013;14(2):127–135.