



# Two-Scale Multimodal Medical Image Fusion Based on Structure Preservation

Shuaiqi Liu<sup>1,2,3†</sup>, Mingwang Wang<sup>1,2</sup>, Lu Yin<sup>1,2</sup>, Xiuming Sun<sup>3\*</sup>, Yu-Dong Zhang<sup>4†</sup> and Jie Zhao<sup>1,2</sup>

<sup>1</sup> College of Electronic and Information Engineering, Hebei University, Baoding, China, <sup>2</sup> Machine Vision Technological Innovation Center of Hebei, Baoding, China, <sup>3</sup> School of Mathematics and Information Science, Zhangjiakou University, Zhangjiakou, China, <sup>4</sup> School of Computing and Mathematics, University of Leicester, Leicester, United Kingdom

Medical image fusion has an indispensable value in the medical field. Taking advantage of structure-preserving filter and deep learning, a structure preservation-based two-scale multimodal medical image fusion algorithm is proposed. First, we used a two-scale decomposition method to decompose source images into base layer components and detail layer components. Second, we adopted a fusion method based on the iterative joint bilateral filter to fuse the base layer components. Third, a convolutional neural network and local similarity of images are used to fuse the components of the detail layer. At the last, the final fused result is got by using two-scale image reconstruction. The contrast experiments display that our algorithm has better fusion results than the state-of-the-art medical image fusion algorithms.

## OPEN ACCESS

### Edited by:

Dongming Zhou,  
Yunnan University, China

### Reviewed by:

Yuanluo An,  
Beijing Jiaotong University, China  
Vipin Tyagi,  
Indian Institute of Technology  
Kharagpur, India

### \*Correspondence:

Xiuming Sun  
sunxiuming@zjku.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

**Received:** 28 October 2021

**Accepted:** 28 December 2021

**Published:** 31 January 2022

### Citation:

Liu S, Wang M, Yin L, Sun X,  
Zhang Y-D and Zhao J (2022)  
Two-Scale Multimodal Medical Image  
Fusion Based on Structure  
Preservation.  
*Front. Comput. Neurosci.* 15:803724.  
doi: 10.3389/fncom.2021.803724

**Keywords:** medical image fusion, scale decomposition, structure preservation, bilateral filter, CNN

## INTRODUCTION

Medical imaging technology has developed rapidly, and medical images of various modalities have significant applications in clinical diagnosis and disease analysis (Wang S. et al., 2020; Li et al., 2021). There are some common medical imaging technologies, such as CT, MRI, PET, and single-photon emission computed tomography (SPECT). Due to different imaging technologies, various kinds of medical images contain complementary characteristics. For the same organ or the same tissue of the human body, medical images with different imaging technologies reflect various pathological characteristics. CT can well show dense structures, such as bones, while MRI has a higher spatial resolution for soft tissues. PET and SPECT are more conducive to display functional information of different tissues of the human body. However, limited by the equipment, the quality of medical images is often unsatisfying. Two or more medical images are merged into one image with comprehensive features by fusion methods for the sake of improving the accuracy of clinical diagnosis (Li et al., 2017; Amala Rani and Lalithakumari, 2019; Huang et al., 2020). Medical image fusion aims at extracting more useful information of input images, improving the application of medical images, and helping doctors understand image contents better. Therefore, the study of medical image fusion is significant (Tirupal et al., 2020).

The existing image fusion methods include pixel-level image fusion methods, feature-level image fusion methods, and decision-level image fusion methods (Dolly and Nisa, 2019; Zhao et al., 2021). The pixel-level fusion methods fuse multiple registered images and preserve the maximum amount of information. Feature-level fusion algorithms extract different features from

input images, such as textures and details, and then merge all the features into one image. Feature-level image fusion has lower requirements for image registration so it will lose part of source images' detailed information. Decision-level fusion is performed by the high-level abstracted information, so the detailed information will be lost severely (Kim et al., 2016; Liu Z. et al., 2019). In conclusion, pixel-level fusion algorithms are the simplest methods with the lowest computation cost. Therefore, we propose a pixel-level medical image fusion algorithm.

In general, pixel-level image fusion methods contain spatial domain methods and transform domain methods (Yu and Chen, 2020). In spatial domain, the fusion operation is performed on pixels (Ashwanth and Swamy, 2020; Zhu et al., 2020). Spatial domain fusion algorithms include principal component analysis (PCA)-based image fusion method (Benjamin and Jayasree, 2018), guided filtering (GFF)-based image fusion method (Li et al., 2013), improved sum-modified-Laplacian (ISML)-based image fusion algorithm (Liu et al., 2015a), nuclear norm minimization (NNM)-based image fusion algorithm (Liu et al., 2015b), convolutional neural network (CNN)-based image fusion algorithm (Liu et al., 2017b), and so on. In transform domain image fusion algorithms, source images are first decomposed into low- and high-frequency components through multi-scale transformations. And then appropriate fusion rules are used to fuse different components. Finally, the final fused result can be obtained through inverse transformation. At present, scholars around the world have proposed many image fusion methods in the transform domain (Mao et al., 2020), such as the wavelet-based image fusion methods proposed in Sumir and Gahan (2018), Panigrahy et al. (2020b), and Rahmani et al. (2020), the image fusion algorithms using non-subsampled contourlet transform (NSCT) proposed in Das and Kundu (2012), Zhu et al. (2019), Liu J. et al. (2020), and Panigrahy et al. (2020a), and the image fusion algorithms using non-subsampled shearlet transform (NSST) (Liu et al., 2017a; Ganasala and Prasad, 2020; Panigrahy et al., 2020c). These algorithms have achieved good fusion effects. However, traditional transform domain image fusion algorithms need to perform frequency decomposition and synthesis during fusion, which often have problems, such as fused image distortion and structural information loss of the source images (Zeng et al., 2020). However, in spatial domain, image fusion often needs to block the images during fusion, which often produces serious blocking effects. To maintain the structural and detailed information, Zhu et al. (2018) proposed an image fusion method based on image cartoon-texture decomposition and sparse representation. In recent years, many structure-preserving filters have appeared (Tomasi and Manduchi, 1998; Farbman et al., 2008; Xu et al., 2011; He et al., 2019). These filters can well maintain the structural information in the images while smoothing textures, which is beneficial to restore the structures and neighborhood details of source images. Therefore, structure-preserving filters are increasingly used in image fusion. The proposed algorithm also applies a structure-preserving filter to overcome block effects and maintain structural information. Recently, deep learning-based fusion algorithms became a hotspot (Tomasi and Manduchi, 1998; Zhu et al., 2018). A multi-focus image fusion method by using CNN is proposed in

Farbman et al. (2008), which regards the generation process of the information feature focus map as a classification problem, and the fusion rule can be regarded as the classifiers used in general classification tasks. Liu S. et al. (2019) proposed a residual network-based multi-focus image fusion method. Subsequently, deep learning has also been applied in medical image fusion. Liu et al. (2015b) also proposed a CNN-based medical image fusion algorithm, which achieved satisfying fusion results. These algorithms can effectively integrate the design of fusion rules and the generation of a decision map, greatly simplifying the steps of image fusion. However, for the sake of obtaining better image fusion results, deep learning-based medical image fusion algorithms need to train a great number of samples. In the training of the network, the training data sets should have ground-true images, but it is not easy to obtain ground-true images of medical images.

Recently, deep learning-based fusion algorithms became a hotspot (Wang et al., 2021a,b). A multi-focus image fusion method by using CNN is proposed in Liu et al. (2017c), which regards the generation process of the information feature focus map as a classification problem, and the fusion rule can be regarded as the classifiers used in general classification tasks. Liu S. et al. (2019) proposed a residual network-based multi-focus image fusion method. Subsequently, deep learning has also been applied in medical image fusion. Liu et al. (2017b) proposed a CNN-based medical image fusion algorithm, which achieved satisfying fusion results. Wang K. et al. (2020) proposed a CNN-based multi-modality medical image fusion algorithm. This algorithm uses the Siamese convolutional network to generate the weight map. The source images are decomposed by contrast pyramid and then fused based on the weight map. These algorithms can effectively integrate the design of fusion rules and the generation of a decision map, greatly simplifying the steps of image fusion. However, for the sake of obtaining better image fusion results, deep learning-based medical image fusion algorithms need to train a great number of samples. In the training of the network, the training data sets should have ground-true images, but it is not easy to obtain ground-true images of medical images.

For the sake of solving the above-mentioned problems of image fusion algorithms, we proposed a two-scale medical image fusion algorithm based on structure preservation. Aiming at different imaging modes of multimodal medical images, a two-scale decomposition operation is adopted to decompose source images into base layer components and detail layer components. In this algorithm, we used different strategies to fuse different components, making the fusion process more comfortable for the human visual system. In the fusion of the base layer components, an iterative joint bilateral filter is applied to improve the decision map, and the weighted sum strategy is used to fuse the base layer components in the spatial domain. In the proposed algorithm, the CNN is adopted to fuse components of the detail layer to produce an image fusion weight map. The components of the detail layer are then fused by employing a multi-scale method and a fusion strategy based on local similarity. In the proposed algorithm, we made full use of the structure-preserving property of the iterative bilateral filter. In

addition, the application of CNN also improves the performance of medical image fusion. Compared with previous algorithms, the main contribution of our method consists of three points. (1) We use a two-scale decomposition method to decompose source images into base layer components and detail layer components, which can make full use of scale information. (2) For the sake of retaining more information, a fusion strategy based on an iterative joint bilateral filter is adopted to fuse base layer components. (3) A new fusion rule based on CNN and local similarity was applied to the detail layer to maintain more detailed information.

The rest of this article is as follows. Section “Related Works” introduces the related works. Section “The Proposed Algorithm” describes the proposed fusion method and its implementation steps. Section “Experiments and Discussion” shows the experimental results and analysis of this algorithm. Finally, in Section “Conclusion” we conclude our fusion method.

## RELATED WORKS

### Two-Scale Image Decomposition

For a given source image  $I$ , it can be decomposed into the base layer component  $I^b$  and the detailed layer component  $I^d$ .  $I^b$  can be obtained through the following optimization problem:

$$\arg \min \|I - I_b\|_F^2 + \eta \left( \|g_x * I^b\|_F^2 + \|g_y * I^b\|_F^2 \right) \quad (1)$$

where  $g_x = [-1 \ 1]$  and  $g_y = [-1 \ 1]^T$  are horizontal and vertical gradient operators, respectively. The regularization parameter  $\eta$  is set to 5. The optimization problem is a Tikhonov regularization problem, and it can be resolved efficiently by fast Fourier transform (FFT). By subtracting the base layer

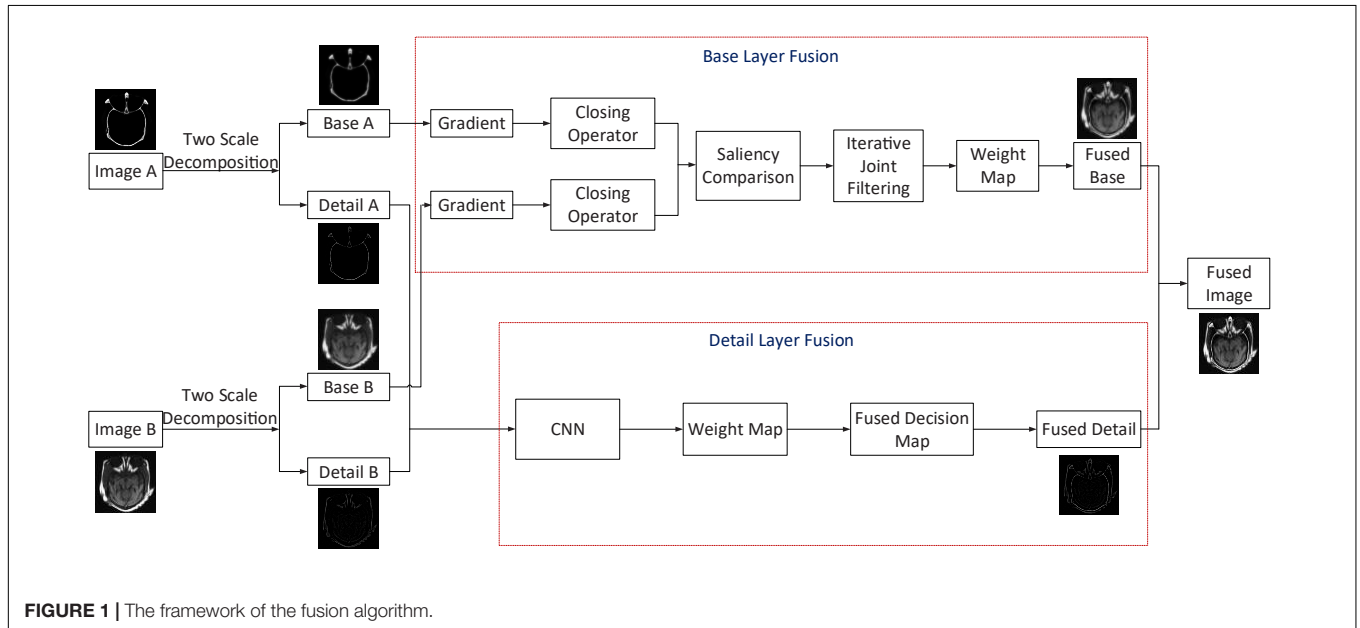


FIGURE 1 | The framework of the fusion algorithm.

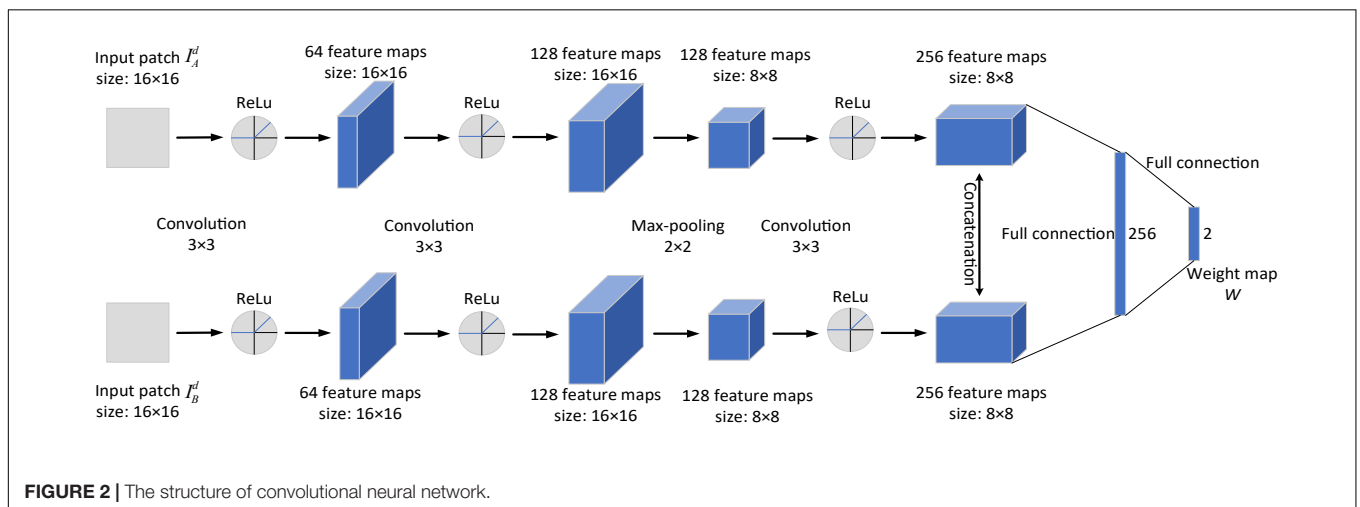


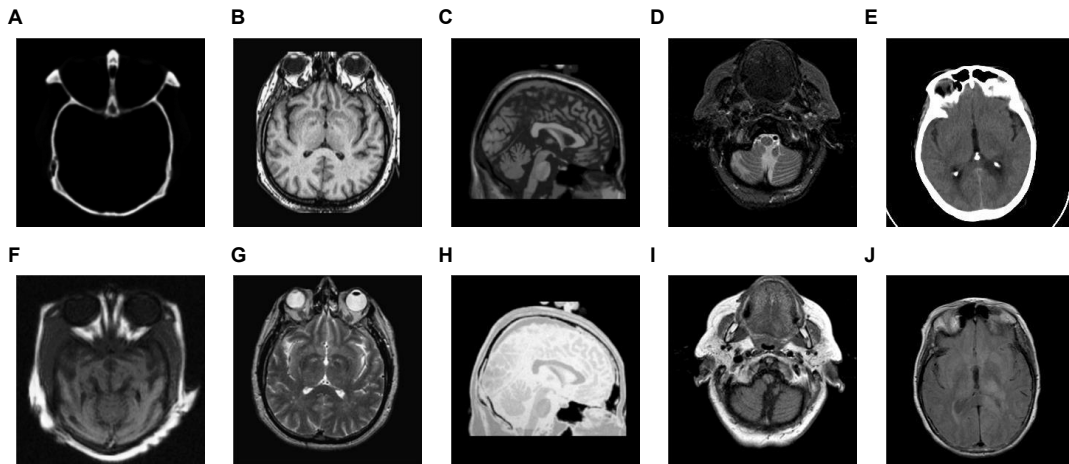
FIGURE 2 | The structure of convolutional neural network.

component from the source image, we can get the detail layer component  $I^d$ , that is

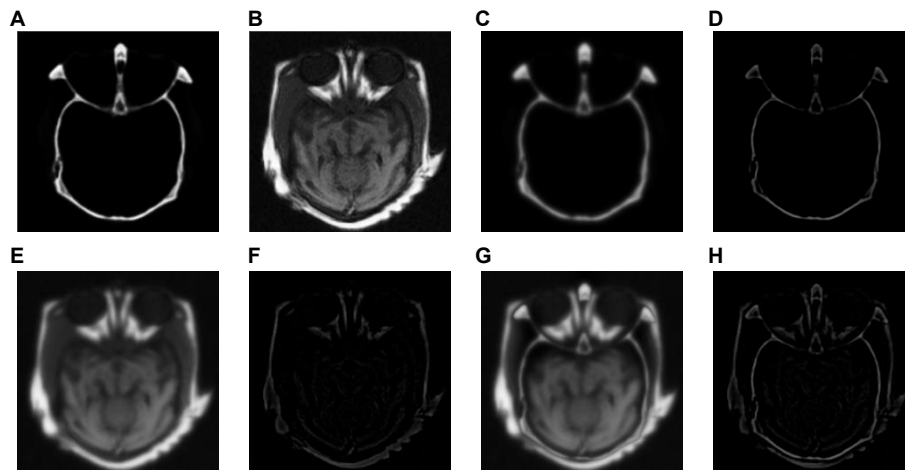
$$I^d = I - I^b \quad (2)$$

### Iterative Joint Bilateral Filter

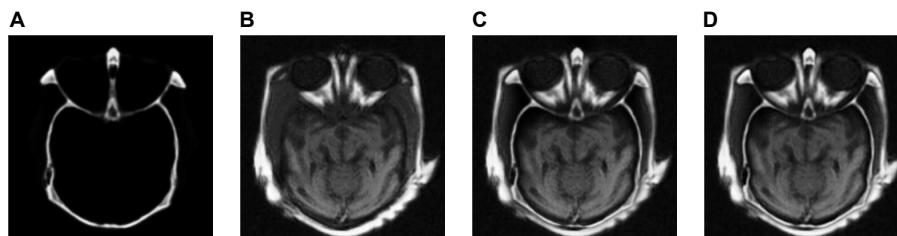
A bilateral filter is non-linear and it can achieve edge preservation during denoising (Tomasi and Manduchi, 1998). In bilateral filtering, the filtered pixel value of a point depends



**FIGURE 3 |** Source image sets for fusion testing. (A–C,E) CT images. (D) MR-T1 image. (F–H) The corresponding MRI images. (I) The corresponding MR-T2 image. (J) Proton density-weighted MR image.



**FIGURE 4 |** The output images obtained after each step of Group (a). (A,B) The source images. (C) The base layer component of panel (A). (D) The detail layer component of panel (A). (E) The base layer component of panel (B). (F) The detail layer component of panel (B). (G) The fused base layer component. (H) The fused detail layer component.



**FIGURE 5 |** Ablation study of CNN. (A,B) The source images.

on the neighborhood pixels. Furthermore, the weights of the neighborhood pixels are determined by the distance and similarity of the two pixels.

The bilateral filter can be represented as

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega} \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2} - \frac{\|I_p - I_q\|^2}{2\sigma_r^2}\right) I_q \quad (3)$$

where  $I$  is the input image.  $p$  and  $q$  are the coordinates of the pixels.  $I_p$  and  $I_q$  denote the pixel values of the corresponding positions.  $\Omega$  is a sliding window centered on  $p$ .  $J_p$  is the output.  $\sigma_s$  and  $\sigma_r$  control the weight of spatial domain and range domain, respectively.  $k_p = \sum_{q \in \Omega} \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2} - \frac{\|I_p - I_q\|^2}{2\sigma_r^2}\right)$  is a normalization.

The weight of the bilateral filter is unstable in practical applications, so some artificial textures will appear near the edges. For the sake of improving the stability of weight, a joint bilateral filter (Zhan et al., 2019) is introduced. It can deal with the problem of the stability of weight by introducing a guide image as the basis for calculating the range weight.

In this article, the iterative joint bilateral filter is used to refine edges. In filtering operation,  $J^1$  represents the initial input guide image and  $J^{t+1}$  denotes the output of the  $t$ -th iteration, that is

$$J_p^{t+1} = \frac{1}{k_p} \sum_{q \in \Omega} \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2} - \frac{\|J_p^t - J_q^t\|^2}{2\sigma_r^2}\right) I_q \quad (4)$$

where  $k_p$  is used to normalize,

$$k_p = \sum_{q \in \Omega} \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2} - \frac{\|J_p^t - J_q^t\|^2}{2\sigma_r^2}\right) \quad (5)$$

where  $\sigma_s$  is set to 8 and  $\sigma_r$  is set to 0.2.

## Convolutional Neural Network

Convolutional neural network is widely used in behavior cognition, pose estimation, object recognition, neural network conversion, and natural language processing. It is one of the most commonly used models in deep learning. It is composed of input layer, hidden layer, and output layer. The hidden layer contains convolutional layer, pooling layer, and fully connected layer. There are three structural characteristics in the CNN model, which are local receptive field, weight sharing, and pooling. The local receptive field represents the connection between a single neuron and the neurons in the corresponding region

of the previous layer. As the number of layers deepens, the corresponding receptive field of a single neuron in each layer increases in the original image, and the information expressed becomes more and more abstract. Weight sharing means that the weight of the convolution kernel of feature mapping is spatially invariant. The receptive field and weight sharing decrease the number of parameters for neural network training. Pooling is also called down-sampling, which can reduce data dimensions. The core idea of CNN is to combine these three structures to obtain the deep features of images.

The calculation formula of the convolution kernel is

$$y_j^l = f\left(\sum_{i \in M_j} y_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (6)$$

where  $M_j$  represents the feature map from the input.  $k_{ij}^l$  represents the convolution kernel connected to the  $i$ -th feature map of the previous layer.  $*$  represents convolution operation.  $b_j^l$  represents the bias.  $f(\cdot)$  represents the activation function.

In CNN-based image fusion algorithms, the generation of a decision map can usually be considered as a classification problem. Specifically, the decision map can be acquired through activity level measurement. Measuring the activity level of a source image is closely related to feature extraction. The higher the activity level means that the image block is clearer. Image fusion is similar to classification tasks. Therefore, the fusion rule in this process is equal to the classifiers in image classification. According to these ideas, a CNN-based multi-focus image fusion algorithm is proposed in Liu et al. (2017c). The advantage of this algorithm is that it solves the difficulties of manually designing activity level measurement and the fusion rule. Therefore, the CNN-based image fusion algorithm solves the problem of fusion by designing the CNN model. By training a large number of image data to obtain the CNN model, it is more effective than manually designing the fusion rule. In Liu et al. (2017b), combining with the CNN model, Liu et al. used a multi-scale method and a local similarity-based fusion rule to obtain a high-quality fusion result. We applied this method to fuse detail components in our fusion algorithm.

## THE PROPOSED ALGORITHM

Aiming at obtaining a satisfying medical image fusion result, we proposed a two-scale medical image fusion algorithm based on structure preservation. The framework of this fusion method is shown in **Figure 1**. First, through the method in Section “Two-Scale Image Decomposition,” we decompose source images into

**TABLE 1** | Objective evaluation results of ablation study.

Method	MI	Q <sup>AB/F</sup>	SSIM	VIFF	UIQI	Q	Q <sub>w</sub>	Q <sub>E</sub>	R <sub>Q</sub> <sup>AB/F</sup>
Without CNN	3.9219	0.7546	0.4912	0.4600	0.4646	0.8395	0.8071	0.7854	0.3786
CNN	<b>4.3328</b>	<b>0.7869</b>	<b>0.4990</b>	<b>0.4684</b>	<b>0.4770</b>	<b>0.8636</b>	<b>0.8300</b>	<b>0.8185</b>	<b>0.3839</b>

Abbreviations: MI, mutual information; QAB/F, metric; SSIM, structural similarity; VIFF, visual information fidelity; UIQI, Universal Image Quality Index; Q, Q<sub>w</sub>, Q<sub>E</sub>, Piella index; R<sub>Q</sub><sup>AB/F</sup>, metric.

The bold values are the best metric values.

base and detail layers. After that, these scale components are fused by appropriate fusion methods. Finally, the fused image is obtained through scale reconstruction.

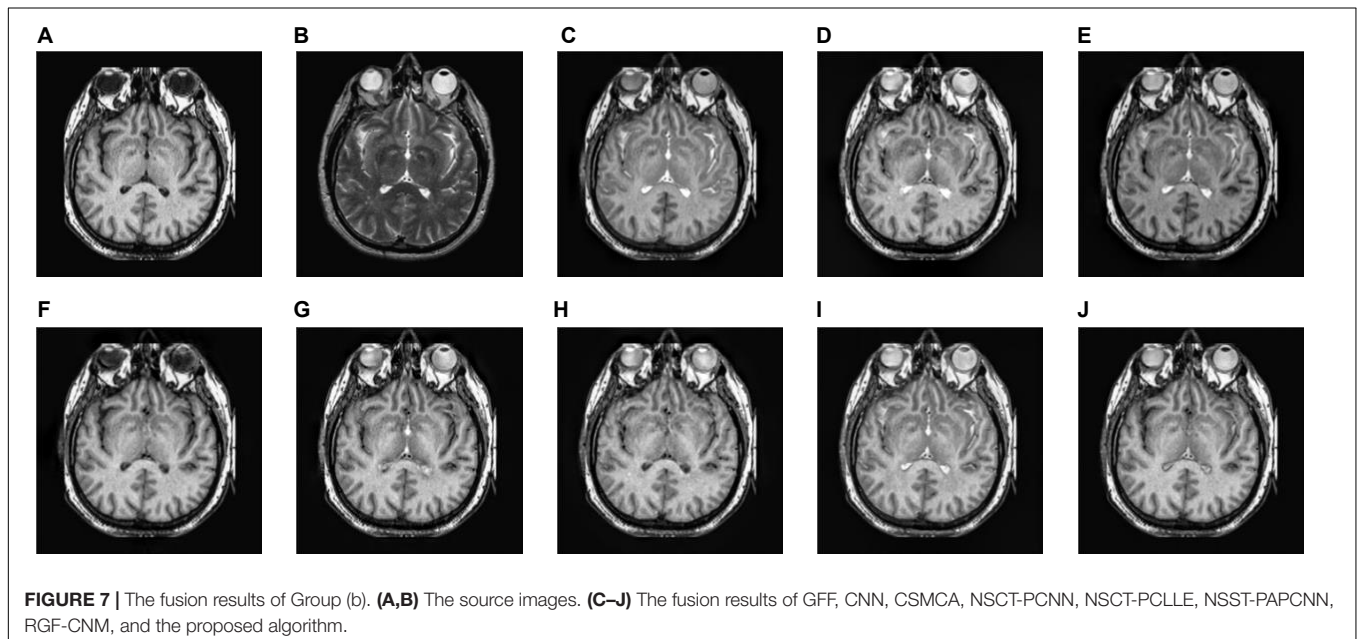
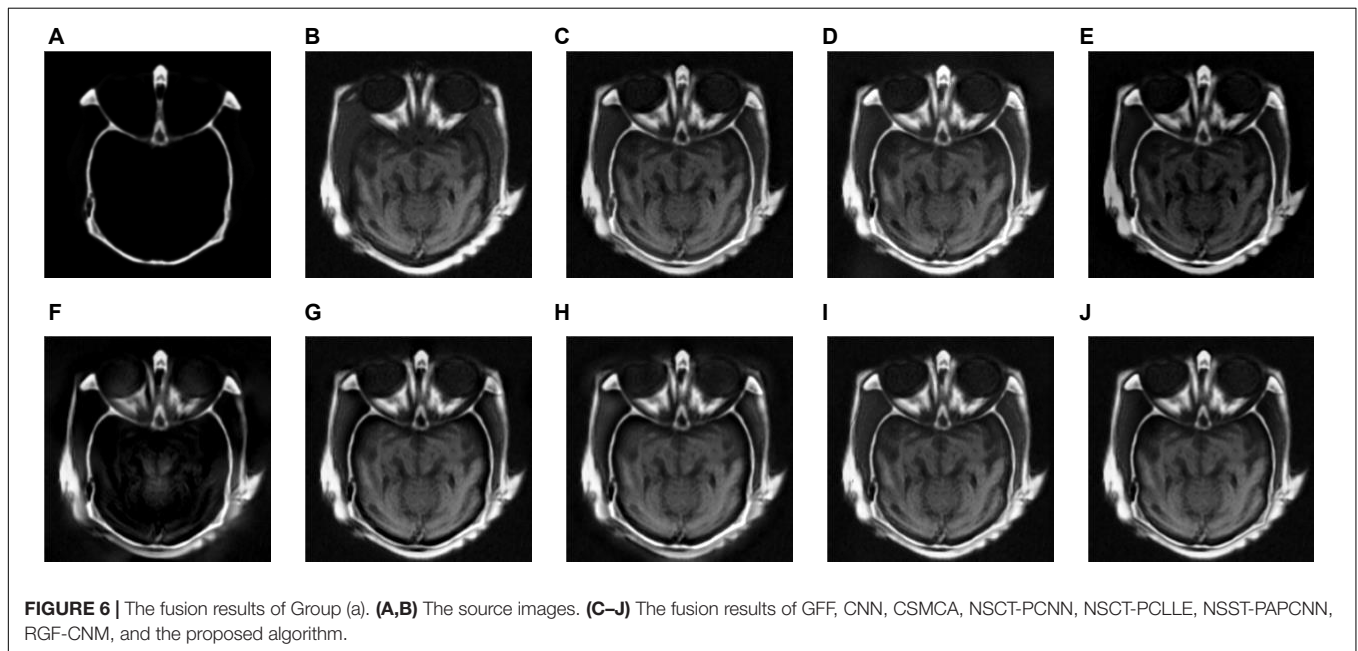
For ease of description, we suppose that the two source images are  $I_A$  and  $I_B$ , and in this article, we have given the fusion steps of two images only. The process of multiple images fusion is similar to the fusion of two images. Traditional multi-scale decomposition methods need more than two scales to get a satisfactory fusion image. In the proposed algorithm, we decomposed the source images by using the two-scale decomposition method in Liu Y. et al. (2016). Equations 1, 2 are employed to decompose source images  $I_A$  and  $I_B$  into base layer

components  $I_A^b, I_B^b$  and detail layer components  $I_A^d, I_B^d$ . The following are descriptions of the fusion rules of base and detail layer components.

### Base Layer Fusion

First, a local structure preservation filter is used to smooth the base layer components  $I_A^b$  and  $I_B^b$ . This filter is defined as

$$S_p = \mu_k + \frac{\sigma_k^2}{\sigma_k^2 + \sigma_d^2} (I_p - \mu_k) \quad (7)$$



where  $p$  represents the pixel position.  $\mu_k$  and  $\sigma_k^2$  are the mean and variance of the image pixel values in a sliding window  $\Omega_k$  with radius  $r$ , respectively. The size of the window  $\Omega_k$  is  $(2r+1) \times (2r+1)$ .  $\sigma_d^2$  is a constant. The mean value  $\mu_k$  represents the average smoothing intensity of the filter, and its variance  $\sigma_k^2$  reflects and measures the changes in local sharpness of the image. In Eq. 7, if  $\sigma_k^2 \gg \sigma_d^2$ , there is  $S_p = I_p$  for pixel  $p$ ; if  $\sigma_k^2 \ll \sigma_d^2$ , then  $S_p = \mu_k$ . Therefore, we set  $\sigma_d^2$  as the minimum variance of the source image to achieve smoothing of tiny pixels. In this study, after many experiments, we found that our algorithm works best when  $r$  is set to 3 and  $\sigma_d^2$  is set to 0.01.

In this article, a two-stage gradient is adopted to describe the salient area of the image. The gradient image  $G_p$  can be

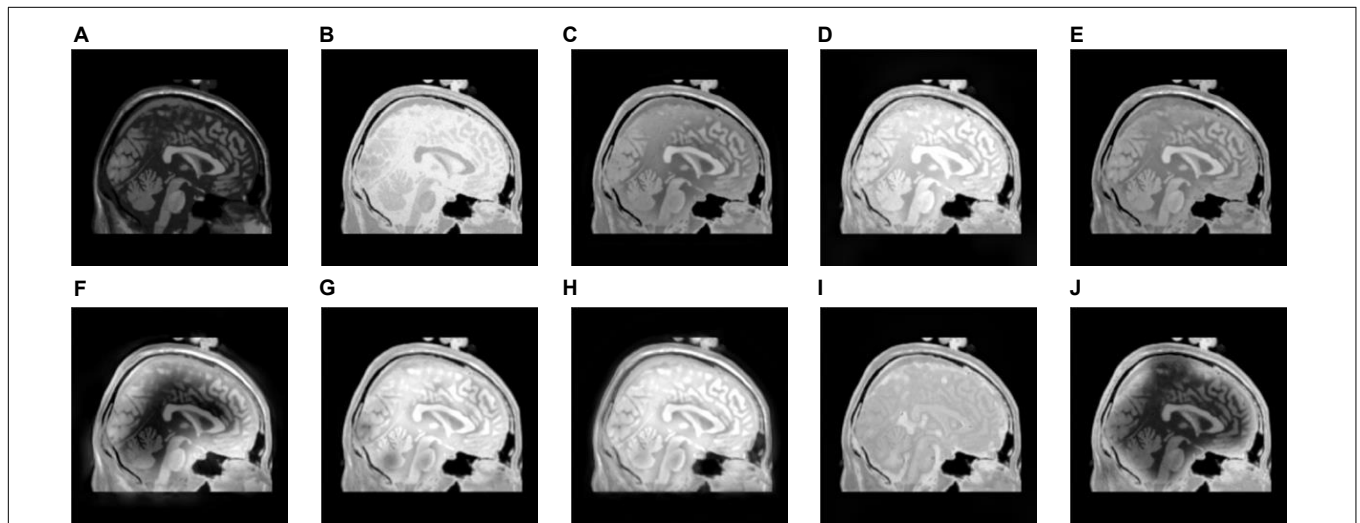
defined as

$$G_p = (S_p \otimes L)^2 \quad (8)$$

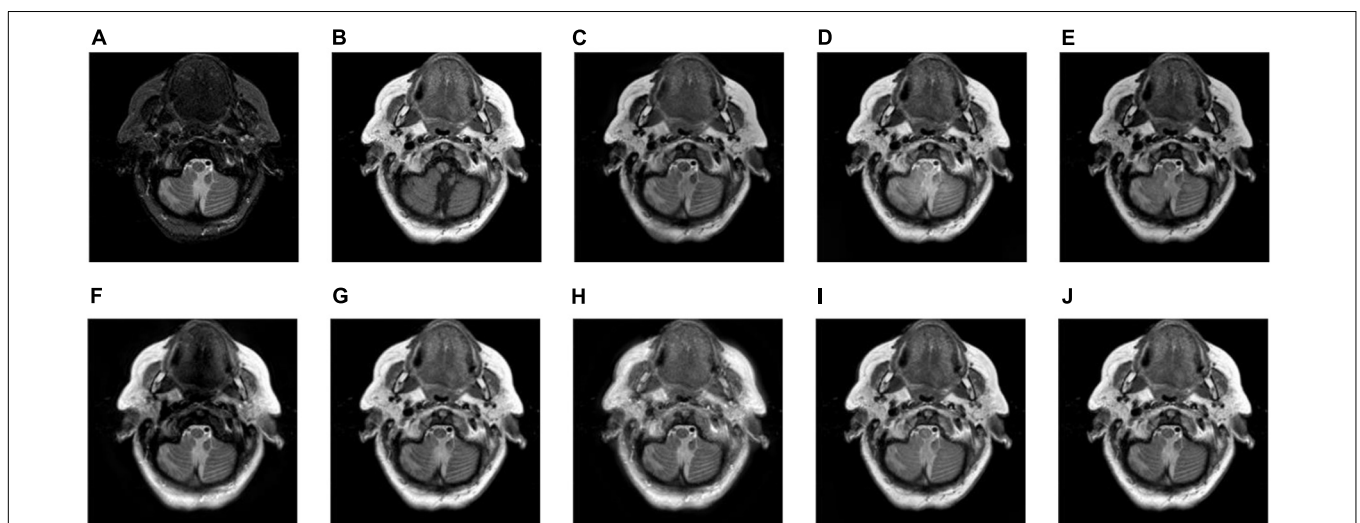
where  $\otimes$  represents convolution.  $L$  is the gradient detection template, In this article, for getting more detailed information, we introduced a gradient detection template in Zhan et al. (2019), that is

$$L = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \quad (9)$$

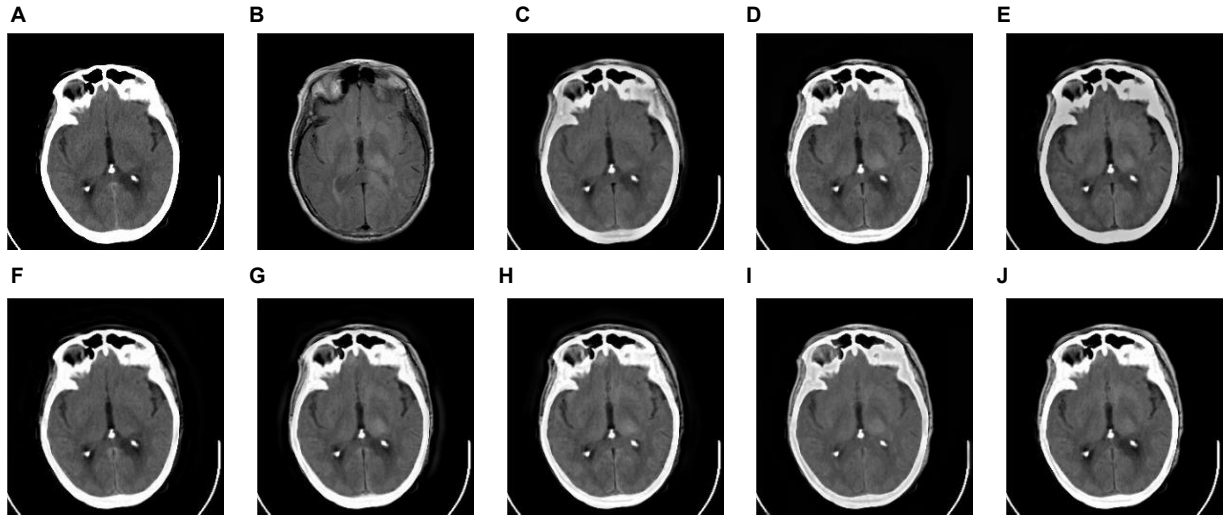
Since holes may be generated in homogeneous regions when performing gradient operation on the image, we adopted morphological closure operation to fill the image holes in this



**FIGURE 8 |** The fusion results of Group (c). (A,B) The source images. (C–J) The fusion results of GFF, CNN, CSMCA, NSCT-PCNN, NSCT-PCLLE, NSST-PAPCNN, RGF-CNM, and the proposed algorithm.



**FIGURE 9 |** The fusion results of Group (d). (A,B) The source images. (C–J) The fusion results of GFF, CNN, CSMCA, NSCT-PCNN, NSCT-PCLLE, NSST-PAPCNN, RGF-CNM, and the proposed algorithm.



**FIGURE 10 |** The fusion results of Group (e). (A,B) The source images. (C–J) The fusion results of GFF, CNN, CSMCA, NSCT-PCNN, NSCT-PCLLE, NSST-PAPCNN, RGF-CNM, and the proposed algorithm.

article, that is

$$\hat{M}_p = 1 - (1 - G_p \otimes E) \otimes E \quad (10)$$

where  $E$  denotes the sliding square structure element with radius  $r$ .

By using Eq. 10, we can obtain the structural gradients  $\hat{M}_p^A$  and  $\hat{M}_p^B$  of each base layer component. Then, the weight map is constructed by using the strategy of taking the larger absolute value. For the image value at the pixel  $p$  at the same position, the obtained binary weight map  $B_p$  can be calculated as follows.

$$B_p = \begin{cases} 1 & \hat{M}_p^A \geq \hat{M}_p^B \\ 0 & \hat{M}_p^A < \hat{M}_p^B \end{cases} \quad (11)$$

$B_p$  is obtained by selecting the saliency structure between  $I_A^b$  and  $I_B^b$  during image fusion. For the sake of making the fused image have more spatial continuity, we can get the basic weight map  $W_p$  by performing average filtering on  $B_p$  with a window size of  $5 \times 5$ . To make the image look more natural, the iterative joint bilateral filter is employed to transfer the edge information in source images into the weight map. The final decision map  $W_p$  is obtained through three iterations, that is

$$W_p = \text{bilateral\_filter}(I_p, W_p, \sigma_s, \sigma_r) \quad (12)$$

Finally, the weighted sum fusion method is used to acquire the fused base layer component.

$$I_F^b = I_A^b W_p + I_B^b (1 - W_p) \quad (13)$$

## Detail Layer Fusion

Combined with CNN, the Laplacian pyramid and the fusion strategy based on local similarity are adopted for the fusion of detail layer components in the proposed algorithm. The CNN structure adopted is shown in **Figure 2**.

It has two identical branches and each branch has three convolutional layers and a max-pooling layer. In each convolutional layer, an ReLu is added for non-linear mapping, which has fast convergence speed and a simple gradient operation. The size of the neuron's receptive field depends on the size of the convolution kernel. Choosing an appropriate convolution kernel size is significant for CNN. If the convolution kernel is too small, local features cannot be extracted effectively; and if the kernel size is too large, the information representation capability is reduced and the image cannot be represented well. Therefore, during net training, the kernel size and step of each convolutional layer are set to  $3 \times 3$  and 1, respectively. And the scale factor and span of the max-pooling layer are set to  $2 \times 2$  and 2, respectively. The 256 feature maps are connected to a 256-dimensional feature vector through a fully connected layer. The output is a fully concatenated two-dimensional vector of the 256-dimensional feature vector.

In this algorithm, the two detail layer components  $I_A^d$  and  $I_B^d$  are input into the two branches of the CNN to generate a fusion weight map  $W$ . Then, Laplacian pyramid decomposition is performed on each detail layer component, which is denoted by  $L\{A\}^l$  and  $L\{B\}^l$ , respectively ( $l$  denotes the decomposition scale). The weight map  $W$  is decomposed by Gaussian pyramid and expressed by  $G\{W\}^l$ .

The local energy maps of  $L\{A\}^l$  and  $L\{B\}^l$  are calculated as

$$\begin{cases} E_A^l(x, y) = \sum_m \sum_n L\{A\}^l(x+m, y+n)^2 \\ E_B^l(x, y) = \sum_m \sum_n L\{B\}^l(x+m, y+n)^2 \end{cases} \quad (14)$$

Then, we calculate the local similarity measurement of  $L\{A\}^l$  and  $L\{B\}^l$ , that is

$$M^l(x, y) = \frac{2 \sum_m \sum_n L\{A\}^l(x+m, y+n) L\{B\}^l(x+m, y+n)}{E_A^l(x, y) + E_B^l(x, y)} \quad (15)$$



From Eq. 15, we can know that the range of  $M^l(x, y)$  is  $[-1, 1]$ . The closer the value is to 1, the higher the local similarity of  $L\{A\}^l$  and  $L\{B\}^l$  is. The threshold  $t$  is set to fuse pixels with different similarities using different fusion rules. If  $M^l(x, y) \geq t$ ,  $L\{A\}^l$  and  $L\{B\}^l$  are fused through the weighted average fusion method, that is

$$L\{F\}^l(x, y) = G\{W\}^l(x, y) \cdot L\{A\}^l(x, y) + (1 - G\{W\}^l(x, y)) \cdot L\{B\}^l(x, y) \quad (16)$$

If  $M^l(x, y) < t$ , the fusion will be carried out by maximizing the local energy, that is

$$L\{F\}^l(x, y) = \begin{cases} L\{A\}^l(x, y), & \text{if } E_A^l(x, y) \geq E_B^l(x, y) \\ L\{B\}^l(x, y), & \text{if } E_A^l(x, y) < E_B^l(x, y) \end{cases} \quad (17)$$

Ultimately, the fused detail layer component  $I_F^d$  is reconstructed from the Laplacian pyramid  $L\{F\}^l$ . Then the fused image  $I_F$  is reconstructed by fused base layer component  $I_F^b$  and detail layer component  $I_F^d$ , that is

$$I_F = I_F^b + I_F^d \quad (18)$$

## EXPERIMENTS AND DISCUSSION

In the study, five groups of medical images are chosen for testing, to evaluate the performance of the proposed fusion algorithm. These test images are displayed in **Figure 3** and are all from <http://www.med.harvard.edu/aanlib/home.html>. These test image sets contain five groups, among which groups (a), (b), and (c) include three CT images of bones and three MRI images of corresponding soft tissues. Group (d) includes an MR-T1 image and a corresponding MR-T2 image. Group (e) includes a CT image and a proton density-weighted MR image. All experiments are carried out under the computer configuration of 2.9 GHz CPU and 8 GB RAM, MATLAB version of R2018a, and the operating system of Win10 64-bit.

To verify the effectiveness of the proposed algorithm (abbreviated as DWTRP-CNN), we first conducted fusion on the test images and compared the proposed algorithm with commonly used fusion algorithms. These algorithms include (1) the guided filtering based image fusion method proposed in Li et al. (2013) (abbreviated as GFF); (2) the CNN-based medical image fusion method proposed in Liu et al. (2017b) (abbreviated as CNN); (3) an image fusion algorithm by using convolutional sparse morphological component analysis proposed in Liu Y. et al. (2019) (abbreviated as CSMCA); (4) a medical image fusion method based on pulse coupled neural network and improved spatial frequency in NSCT domain proposed in Das and Kundu (2012) (abbreviated as NSCT-PCNN); (5) a multimodal medical image fusion algorithm based on phase consistency and local Laplacian energy in NSCT domain proposed in Zhu et al. (2019) (abbreviated as NSCT-PCLLE); (6) a parameter adaptive pulse coupled neural network-based medical image fusion algorithm in NSST domain proposed in Yin et al. (2019) (abbreviated as NSST-PAPCNN); and (7) the rolling guidance

filtering multimodal medical image fusion by combining CNN and nuclear norm minimization proposed in Liu S. et al. (2020) (abbreviated as RGF-CNM). To make it fair, when using the above-mentioned fusion algorithms for testing, the parameter settings of these algorithms have the same parameters in the articles published by the authors.

We also choose six indicators for objective evaluation of the proposed algorithm, namely, Mutual Information (MI) (Qu et al., 2002), QAB/F metric (Xydeas and Petrovic, 2000), Structural Similarity (SSIM) (Wang et al., 2004), Visual Information Fidelity (VIFF) (Han et al., 2013), Universal Image Quality Index (UIQI) (Wang and Bovik, 2002), Piella index (Q, QW, QE) (Piella and Heijmans, 2003), and  $R_Q^{AB/F}$  metric (Sengupta et al., 2020). MI reflects the useful information that remained in a fused image. The QAB/F index is an objective measure based on edge information and is adopted to assess whether the fused image retains more edge information. SSIM represents the structural similarity between the fused image and the source images. VIFF measures the fidelity of the visual information of the fused image relative to the source images. UIQI measures image distortion through correlation loss, brightness distortion, and contrast distortion. The Piella index comprehensively reflects the similarity of intensity, contrast, and structure between the fused image and the source images.  $R_Q^{AB/F}$  metric reflects the edge and orientation strengths. For all of these indexes, the higher value means that the performance of the fusion method is better.

**Figure 4** shows the output images obtained after each step of our fusion method applied in Group (a). **Figures 4A,B** are the source images. **Figures 4C,D** are the base layer component and the detail layer component of **Figure 4A**, respectively. **Figures 4E,F** are the base layer component and the detail layer component of **Figure 4B**, respectively. **Figures 4G,H** are the fused base layer component and the fused detail layer component, respectively. It can be seen from **Figure 4** that the proposed algorithm can extract the basic information and detailed information of source images and fuse them effectively.

To verify the effectiveness of CNN for detail layer fusion, we conducted an ablation study on Group (a). **Figure 5** shows the fusion results of the ablation study. **Figures 5A,B** show the source images. **Figure 5C** is the fusion result without CNN, that is, the detail layers are fused by the method which is the same as the method for fusing base layers. **Figure 5D** is the fusion image of the proposed algorithm.

**Table 1** shows the objective evaluation results of the ablation study. It can be seen from **Table 1** that the fusion algorithm with CNN has better objective evaluation results. The ablation study proves the effectiveness of CNN for detail layer fusion.

We tested all the methods in the test image sets in **Figure 3** and the fused images are displayed in **Figures 6–10**. **Figure 6** shows the fusion results of Group (a). **Figures 6A,B** show the source images. And **Figures 6C–J** are the fusion results of GFF, CNN, CSMCA, NSCT-PCNN, NSCT-PCLLE, NSST-PAPCNN, RGF-CNM, and the proposed algorithm, respectively.

**Figure 6** shows that the fused image of our algorithm has a higher contrast, shows more structural information and textural information, and introduces less useless information among these algorithms. And the edges of **Figure 6C** are

not clear enough, and GFF lost some structural information. **Figure 6D** retains the spatial information of source images, but some edges are lost. **Figure 6E** has low contrast, and the fusion effect is not ideal. **Figure 6F** shows that the

CSMCA does not get a good fused image, and cannot be able to retain useful information effectively. Although **Figure 6G** retains more useful information of source images, but a small number of artifacts are produced. There are unwanted artifacts

**TABLE 2** | Objective evaluation results of Group (a).

	MI	Q <sup>AB/F</sup>	SSIM	VIFF	UIQI	Q	Q <sub>W</sub>	Q <sub>E</sub>	R <sub>Q</sub> <sup>AB/F</sup>
GFF	3.4313	0.7789	0.4865	0.4865	0.4493	0.8147	0.7889	0.7843	0.3824
CNN	3.0757	0.7683	0.4764	<b>0.5453</b>	0.4451	0.7945	0.7805	0.7819	0.3810
CSMCA	2.5864	0.7280	0.4388	0.4752	0.3414	0.5721	0.6024	0.6171	0.3715
NSCT-PCNN	1.2824	0.5764	0.3990	0.3313	0.2678	0.4257	0.4381	0.4443	0.3441
NSCT-PCLLE	3.2906	0.7325	0.4861	0.4625	0.4512	0.8077	0.7878	0.7799	0.3734
NSST-PAPCNN	2.4665	0.6859	0.4637	0.4559	0.3950	0.6996	0.7030	0.5812	0.3683
RGF-CNM	4.1236	0.7808	0.4986	0.5105	0.4760	0.8475	0.8231	0.7947	<b>0.3844</b>
DWTRP-CNN	<b>4.3328</b>	<b>0.7869</b>	<b>0.4990</b>	0.4684	<b>0.4770</b>	<b>0.8636</b>	<b>0.8300</b>	<b>0.8185</b>	0.3839

The bold values are the best metric values.

**TABLE 3** | Objective evaluation results of Group (b).

	MI	Q <sup>AB/F</sup>	SSIM	VIFF	UIQI	Q	Q <sub>W</sub>	Q <sub>E</sub>	R <sub>Q</sub> <sup>AB/F</sup>
GFF	4.1027	0.6057	0.6671	0.5213	0.4734	0.5022	0.5077	0.5009	0.3544
CNN	3.8871	0.5855	0.6433	0.5775	0.4145	0.4639	0.5096	0.4884	0.3502
CSMCA	3.6764	0.6062	0.6441	0.5273	0.3774	0.3818	0.4771	0.4634	0.3525
NSCT-PCNN	4.4815	0.5787	0.6532	0.5104	0.4552	0.4019	0.4830	0.4586	0.3506
NSCT-PCLLE	4.1448	0.5618	0.6655	0.5708	0.4688	0.5022	0.5134	0.4859	0.3466
NSST-PAPCNN	3.8906	0.5228	<b>0.6814</b>	<b>0.5931</b>	0.4576	0.5028	<b>0.5374</b>	0.4858	0.3420
RGF-CNM	4.4565	0.6130	0.6800	0.5804	0.4870	0.4536	0.5119	0.4822	0.3570
DWTRP-CNN	<b>4.9866</b>	<b>0.8185</b>	<b>0.6814</b>	0.5584	<b>0.5634</b>	<b>0.5551</b>	0.5317	<b>0.5113</b>	<b>0.3575</b>

The bold values are the best metric values.

**TABLE 4** | Objective evaluation results of Group (c).

	MI	Q <sup>AB/F</sup>	SSIM	VIFF	UIQI	Q	Q <sub>W</sub>	Q <sub>E</sub>	R <sub>Q</sub> <sup>AB/F</sup>
GFF	2.9724	0.6567	0.6598	0.6146	0.4472	0.2462	0.3401	0.4045	0.3560
CNN	3.5573	0.6454	0.5050	<b>0.7817</b>	0.2930	0.2098	0.2897	0.3700	0.3528
CSMCA	3.0675	0.6509	0.6752	0.6266	0.5557	0.1904	0.3211	0.3912	0.3513
NSCT-PCNN	2.8244	0.5922	0.5382	0.6118	0.3777	0.1840	0.3243	0.3713	0.3396
NSCT-PCLLE	3.3006	0.6111	0.6392	0.7402	0.4883	0.2332	0.2917	0.3623	0.3430
NSST-PAPCNN	3.1406	0.5525	0.5564	0.6229	0.3141	0.1664	0.2522	0.3114	0.3371
RGF-CNM	3.4611	0.5414	0.6412	0.6863	0.3720	0.2323	0.3082	0.2920	0.3401
DWTRP-CNN	<b>3.8684</b>	<b>0.8569</b>	<b>0.6854</b>	0.6645	<b>0.6088</b>	<b>0.3040</b>	<b>0.3986</b>	<b>0.4272</b>	<b>0.3567</b>

The bold values are the best metric values.

**TABLE 5** | Objective evaluation results of Group (d).

	MI	Q <sup>AB/F</sup>	SSIM	VIFF	UIQI	Q	Q <sub>W</sub>	Q <sub>E</sub>	R <sub>Q</sub> <sup>AB/F</sup>
GFF	3.2966	0.5892	0.6982	0.6020	0.5022	0.4847	0.5482	0.4916	0.3398
CNN	3.6331	0.5915	0.7038	0.7405	0.5203	0.3218	0.5248	0.4476	0.3383
CSMCA	3.3221	0.6079	0.7131	0.6405	0.5570	0.483	0.5159	0.4493	0.3381
NSCT-PCNN	2.9174	0.4276	0.6266	0.6606	0.3989	0.3738	0.5048	0.4461	0.3150
NSCT-PCLLE	3.8238	0.5714	0.7303	0.7731	0.5807	0.5331	0.5754	0.5006	0.3354
NSST-PAPCNN	3.2064	0.4081	0.6527	0.6169	0.4294	0.4031	0.5250	0.4640	0.3193
RGF-CNM	4.1837	0.6338	0.7318	0.7436	0.5504	0.3717	0.5494	0.4725	0.3464
DWTRP-CNN	<b>4.5808</b>	<b>0.8423</b>	<b>0.7320</b>	<b>0.7801</b>	<b>0.6038</b>	<b>0.5570</b>	<b>0.5837</b>	<b>0.5266</b>	<b>0.3473</b>

The bold values are the best metric values.

**TABLE 6** | Objective evaluation results of Group (e).

	MI	$Q^{AB/F}$	SSIM	VIFF	UIQI	Q	$Q_w$	$Q_E$	$R_Q^{AB/F}$
GFF	3.4014	0.6171	0.7709	0.4688	0.6538	0.2954	0.4988	0.4935	0.3538
CNN	3.3405	0.6232	0.6745	0.5078	0.3448	0.2977	0.5127	0.5081	0.3576
CSMCA	3.1097	0.5933	0.7568	0.4745	0.5622	0.2731	0.4830	0.4840	0.3402
NSCT-PCNN	3.7688	0.5683	0.7257	0.4157	0.4307	0.2818	0.4623	0.4854	0.3398
NSCT-PCLLE	3.3127	0.5841	0.7153	0.4892	0.5427	0.3005	0.5002	0.4884	0.3474
NSST-PAPCNN	3.2213	0.5902	0.7180	0.5031	0.5496	0.3116	0.5207	0.5058	0.3546
RGF-CNM	3.5551	0.6131	0.7786	0.5206	0.5140	0.3247	<b>0.5352</b>	0.4896	<b>0.3597</b>
DWTRP-CNN	<b>3.7814</b>	<b>0.6326</b>	<b>0.7864</b>	<b>0.5316</b>	<b>0.6764</b>	<b>0.3352</b>	0.5286	<b>0.5319</b>	0.3556

The bold values are the best metric values.

**TABLE 7** | The running times of all fusion algorithms.

Time(s)	Group (a)	Group (b)	Group (c)	Group (d)	Group (e)
GFF	<b>0.0550</b>	<b>0.0346</b>	<b>0.0363</b>	<b>0.0449</b>	<b>0.0661</b>
CNN	10.3217	10.6671	10.3465	10.7059	10.2416
CSMCA	85.1345	84.7032	85.8902	86.4450	85.4226
NSCT-PCNN	39.5787	39.9496	39.6413	39.6515	39.7292
NSCT-PCLLE	2.1394	1.7340	1.7377	1.7313	1.7528
NSST-PAPCNN	4.0287	4.0262	3.9829	4.5434	4.1555
RGF-CNM	34.2282	34.6446	33.4955	35.2557	33.6108
DWTRP-CNN	11.3148	11.5726	10.8541	11.5949	11.1146

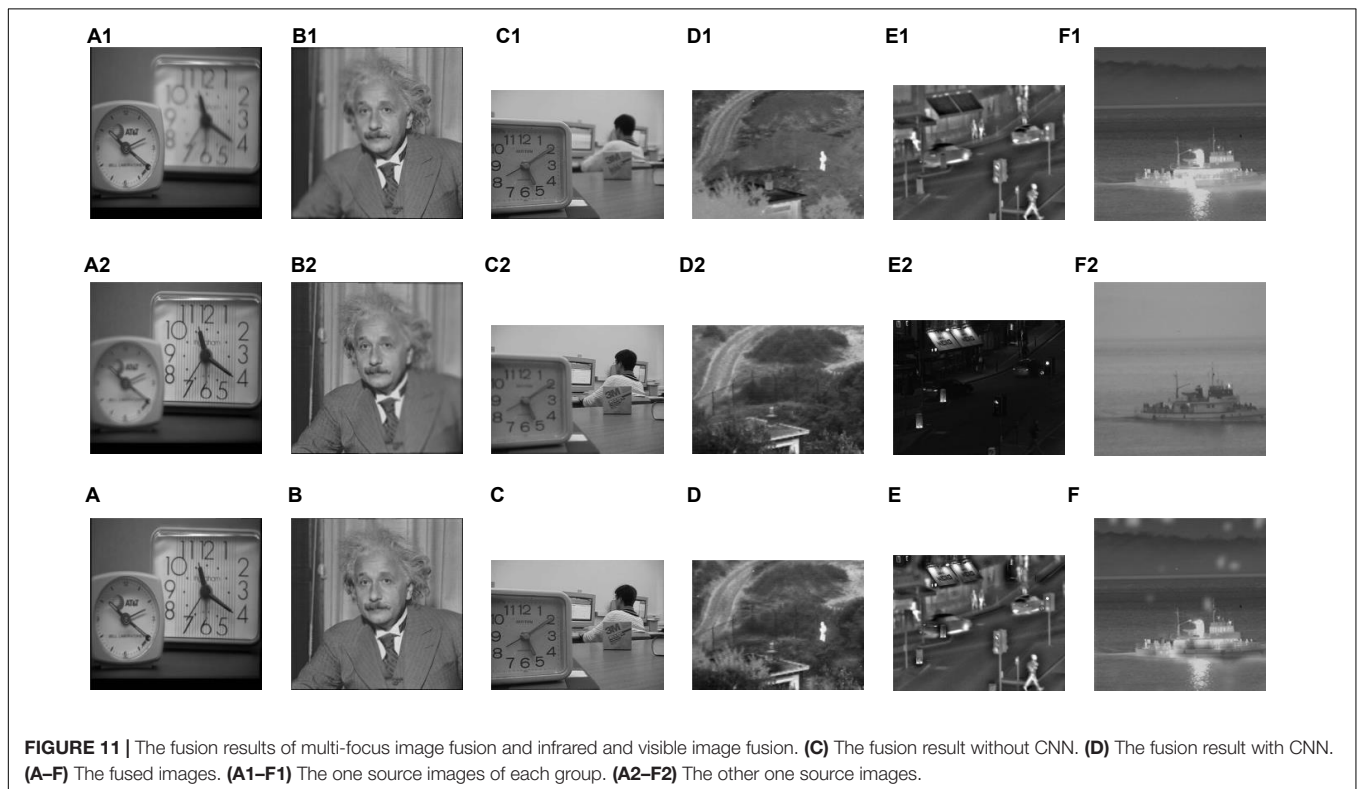
The bold values are the best metric values.

in **Figure 6H**. The contrast in **Figure 6I** is high, but the edges are blurred.

**Table 2** shows the objective evaluation results of Group (a) for all test algorithms. As seen in **Table 2**, our fusion algorithm has

the highest values on MI,  $Q^{AB/F}$ , SSIM, Piella, and UIQI. The maximum values of MI and  $Q^{AB/F}$  indicate that our algorithm retains more useful information and edges.

In **Table 2**, the highest indicators of SSIM and Piella indicate the highest structural similarity between the fused image and the source images obtained by our algorithm. The maximum value of UIQI indicates that the correlation, brightness, and contrast of the fused image and the source images obtained by our algorithm are closest. Although the VIFF of our algorithm is not the highest, from the visual effects of **Figure 6**, our fusion algorithm is the best. In all comparison algorithms, the VIFF value of CNN is the highest, which is better than our method. However, the other seven evaluation indexes are lower than our fusion algorithm, and the value of MI is far from ours, which indicates that CNN is not ideal in preserving the information of source images. Therefore, our algorithm is an effective medical image fusion algorithm.



**FIGURE 11** | The fusion results of multi-focus image fusion and infrared and visible image fusion. **(C)** The fusion result without CNN. **(D)** The fusion result with CNN. **(A–F)** The fused images. **(A1–F1)** The one source images of each group. **(A2–F2)** The other one source images.

**Figure 7** gives the fusion results of Group (b). In **Figure 7**, the contrasts of **Figures 7C,E** are low, and the image in **Figure 7F** loses more useful information. In addition, **Figures 7E,I** produce different degrees of blocking effects and artifacts. **Figures 7D,G,H** fuse the two source medical images well, but still lose a small part of detailed and textural information, and do not get the best visual effects. Compared to the contrasting algorithms, the fused image of our algorithm has the best visual effect.

**Table 3** shows the objective evaluation results of Group (b) for all test algorithms. **Table 3** shows that the proposed algorithm gives the best objective evaluation results except VIFF and QW. It indicates that the proposed algorithm retains the most useful information and edges, the structural similarity of fused image and source images is the highest, and the correlation, brightness, and contrast are the closest. In other contrast algorithms, the values of SSIM, VIFF, and QW of NSST-PAPCNN fusion algorithm are the highest, but other evaluation indexes are lower than the proposed algorithm, especially MI and QAB/F. From the visual effects and objective evaluation, the proposed algorithm achieves a satisfying fusion effect.

**Figures 8–10** show the fusion results of groups (c), (d), and (e). Similar to **Figures 6, 7**, the proposed method retains more structural and detailed information of the input medical images, has higher contrast and brightness, and introduces less useless information, such as blocking effects and artifacts from the fusion results of **Figures 8C–J, 9C–J, 10C–J**.

From the comprehensive view of the objective indicators in **Tables 4–6**, the proposed algorithm has higher objective evaluation indicators than other algorithms. In addition, **Table 7** shows the running times of these algorithms. Compared to others, the running time of our algorithm is also relatively competitive. Therefore, the proposed algorithm is an effective and robust medical image fusion algorithm.

In addition, we applied the proposed algorithm to multi-focus image fusion and infrared and visible image fusion. In **Figures 11A1–F1** are the one source images of each group. **Figures 11A2–F2** are the other one source images. **Figures 11A–F** are the fused images. As shown in **Figure 11**, the proposed fusion algorithm is also effective in multi-focus image fusion and infrared and visible image fusion.

## CONCLUSION

In this article, we proposed a two-scale multimodal medical image fusion algorithm based on structure preservation. The proposed algorithm decomposes the source images by two-scale

## REFERENCES

- Amala Rani, V., and Lalithakumari, S. (2019). Recent medical image fusion techniques: a review. *Indian J. Public Health Res. Dev.* 10, 1399–1403. doi: 10.5958/0976-5506.2019.01785.6
- Ashwanth, B., and Swamy, K. V. (2020). "Medical image fusion using transform techniques," in *Proceedings of the 2020 5th International Conference on Devices, Circuits and Systems (ICDCS)*, Coimbatore, 303–306. doi: 10.1109/ICDCS48716.2020.243604

decomposition, which fully uses the multi-scale information of the images. Our algorithm also adopts the structure preservation characteristic of the iterative joint bilateral filter and applies CNN in medical image fusion. From the visual effects and objective measures, the contrast experiments show that the proposed algorithm has better performance than the compared algorithms. However, the speed of the proposed algorithm is not ideal. In future, the computational speed of the proposed algorithm will be optimized for practical application in clinical practice.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.med.harvard.edu/aanlib/home.html>.

## ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

LY performed the computer simulations. SL, XS, and JZ analyzed the data. SL wrote the original draft. XS and Y-DZ revised and edited the manuscript. MW polished the manuscript. All authors approved the submitted version.

## FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant 62172139; the Natural Science Foundation of Hebei Province under Grant F2020201025, F2019201151, and F2018210148; the Science Research Project of Hebei Province under Grant BJ2020030; and the Open Foundation of Guangdong Key Laboratory of Digital Signal and Image Processing Technology (2020GDDSIPL-04).

## ACKNOWLEDGMENTS

We thank the High-Performance Computing Center of the Hebei University for their support.

- Benjamin, J. R., and Jayasree, T. (2018). Improved medical image fusion based on cascaded PCA and shift invariant wavelet transforms. *Int. J. Comput. Assist. Radiol. Surg.* 13, 229–240. doi: 10.1007/s11548-017-1692-4
- Das, S., and Kundu, M. K. (2012). NSCT-based multimodal medical image fusion using pulse-coupled neural network and modified spatial frequency. *Med. Biol. Eng. Comput.* 50, 1105–1114. doi: 10.1007/s11517-012-0943-3
- Dolly, J. M., and Nisa, A. K. (2019). "A survey on different multimodal medical image fusion techniques and methods," in *Proceedings of the 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, Chennai, 1–5. doi: 10.1109/ICIICT.2019.8741445

- Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 1–10. doi: 10.1145/1360612.1360666
- Ganasala, P., and Prasad, A. D. (2020). “Functional and anatomical image fusion based on texture energy measures in NSST domain,” in *Proceedings of the 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, Raipur, 417–420. doi: 10.1109/ICPC2T48082.2020.9071494
- Han, Y., Cai, Y., Cao, Y., and Xu, X. (2013). A new image fusion performance metric based on visual information fidelity. *Inf. Fusion* 14, 127–135. doi: 10.1016/j.inffus.2011.08.002
- He, K., Sun, J., and Tang, X. (2019). Guided image filtering. *IEEE Trans. Patt. Anal. Mach. Intell.* 11935, 114–127.
- Huang, B., Yang, F., Yin, M., Mo, X., and Zhong, C. (2020). A review of multimodal medical image fusion techniques. *Comput. Math. Methods Med.* 2020:8279342. doi: 10.1155/2020/8279342
- Kim, M., Han, D. K., and Ko, H. (2016). Joint patch clustering-based dictionary learning for multimodal image fusion. *Inf. Fusion* 27, 198–214. doi: 10.1016/j.inffus.2015.03.003
- Li, S., Kang, X., Fang, L., Hu, J., and Yin, H. (2017). Pixel-level image fusion: a survey of the state of the art. *Inf. Fusion* 33, 100–112. doi: 10.1016/j.inffus.2016.05.004
- Li, S., Kang, X., and Hu, J. (2013). Image fusion with guided filtering. *IEEE Trans. Image Process.* 22, 2864–2875. doi: 10.1109/TIP.2013.2244222
- Li, Y., Zhao, J., Lv, Z., and Li, J. (2021). Medical image fusion method by deep learning. *Int. J. Cogn. Comput. Eng.* 2, 21–29. doi: 10.1016/j.ijcce.2020.12.004
- Liu, J., Duan, M., Chen, W. B., and Shi, H. (2020). “Adaptive weighted image fusion algorithm based on NSCT multi-scale decomposition,” in *Proceedings of the 2020 International Conference on System Science and Engineering (ICSSE)*, Kagawa, 1–5. doi: 10.1109/ICSSE50014.2020.9219295
- Liu, S., Wang, J., Lu, Y., Hu, S., Ma, X., and Wu, Y. (2019). Multi-focus image fusion based on residual network in non-subsampled shearlet domain. *IEEE Access* 7, 152043–152063. doi: 10.1109/ACCESS.2019.2947378
- Liu, S., Yin, L., Miao, S., Ma, J., Cong, S., and Hu, S. (2020). Multimodal medical image fusion using rolling guidance filter with CNN and nuclear norm minimization. *Curr. Med. Imaging* 16, 1243–1258. doi: 10.2174/1573405616999200817103920
- Liu, S., Zhao, J., and Shi, M. (2015a). Medical image fusion based on improved sum-modified-Laplacian. *Int. J. Imaging Syst. Technol.* 25, 206–212. doi: 10.1002/ima.22138
- Liu, S., Zhang, T., Li, H., Zhao, J., and Li, H. (2015b). Medical image fusion based on nuclear norm minimization. *Int. J. Imaging Syst. Technol.* 25, 310–316. doi: 10.1002/ima.22145
- Liu, Y., Chen, X., Cheng, J., and Peng, H. (2017b). “A medical image fusion method based on convolutional neural networks,” in *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, Xi’an, 1–7. doi: 10.23919/ICIF.2017.8009769
- Liu, X., Mei, W., and Du, H. (2017a). Structure tensor and nonsubsampling shearlet transform based algorithm for CT and MRI image fusion. *Neurocomputing* 235, 131–139. doi: 10.1016/j.neucom.2017.01.006
- Liu, Y., Chen, X., Peng, H., and Wang, Z. (2017c). Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* 36, 191–207. doi: 10.1016/j.inffus.2016.12.001
- Liu, Y., Chen, X., Ward, R. K., and Wang, Z. J. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Process. Lett.* 23, 1882–1886. doi: 10.1109/LSP.2016.2618776
- Liu, Y., Chen, X., Ward, R. K., and Wang, Z. J. (2019). Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Process. Lett.* 26, 485–489. doi: 10.1109/LSP.2019.2895749
- Liu, Z., Song, Y., Sheng, V. S., Xu, C., Maere, C., Xue, K., et al. (2019). MRI and PET image fusion using the nonparametric density model and the theory of variable-weight. *Comput. Methods Programs Biomed.* 175, 73–82. doi: 10.1016/j.cmpb.2019.04.010
- Mao, Q., Zhu, Y., Lv, C., Lu, Y., Yan, X., Wei, D., et al. (2020). Image fusion based on multiscale transform and sparse representation to enhance terahertz images. *Opt. Express* 28, 25293–25307. doi: 10.1364/OE.396604
- Panigrahy, C., Seal, A., Mahato, N., Krejcar, O., and Herrera-Viedma, E. (2020b). Multi-focus image fusion using fractal dimension. *Appl. Opt.* 59, 5642–5655. doi: 10.1364/AO.391234
- Panigrahy, C., Seal, A., and Kumar Mahato, N. (2020a). Fractal dimension based parameter adaptive dual channel PCNN for multi-focus image fusion. *Opt. Lasers Eng.* 133:106141. doi: 10.1016/j.optlaseng.2020.106141
- Panigrahy, C., Seal, A., and Kumar Mahato, N. (2020c). MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN. *IEEE Signal Process. Lett.* 27, 690–694. doi: 10.1109/LSP.2020.2989054
- Piella, G., and Heijmans, H. (2003). “A new quality metric for image fusion,” in *Proceedings of the 2003 International Conference on Image Processing*, Barcelona.
- Qu, G., Zhang, D., and Yan, P. (2002). Information measure for performance of image fusion. *Electron. Lett.* 38, 313–315. doi: 10.1049/el:20020212
- Rahmani, A. I., Almasi, M., Saleh, N., and Katouli, M. (2020). Image fusion of noisy images based on simultaneous empirical wavelet transform. *Traitement Signal* 37, 703–710. doi: 10.18280/ts.370502
- Sengupta, A., Seal, A., Panigrahy, C., Krejcar, O., and Yazidi, A. (2020). Edge information based image fusion metrics using fractional order differentiation and sigmoidal functions. *IEEE Access* 8, 88385–88398. doi: 10.1109/ACCESS.2020.2993607
- Sumir, R. M., and Gahan, B. G. (2018). “Image fusion using wavelet transform and GLCM based texture analysis for detection of brain tumor,” in *Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bengaluru, 1884–1888. doi: 10.1109/RTEICT42901.2018.9012136
- Tirupal, T., Mohan, B. C., and Kumar, S. S. (2020). Multimodal medical image fusion techniques - a review. *Curr. Signal Transduct. Ther.* 16, 142–163. doi: 10.2174/1574362415666200226103116
- Tomasi, C., and Manduchi, R. (1998). “Bilateral filtering for gray and color images,” in *Proceedings of the 1998 Sixth International Conference on Computer Vision*, Bombay, 839–846. doi: 10.1109/ICCV.1998.710815
- Wang, K., Zheng, M., Wei, H., Qi, G., and Li, Y. (2020). Multi-modality medical image fusion using convolutional neural network and contrast pyramid. *Sensors* 20:2169. doi: 10.3390/s20082169
- Wang, S., Wu, X., Zhang, Y., Tang, C., and Zhang, X. (2020). Diagnosis of COVID-19 by wavelet Renyi entropy and three-segment biogeography-based optimization. *Int. J. Comput. Intell. Syst.* 13, 1332–1344. doi: 10.2991/ijcis.d.200828.001
- Wang, S. H., Govindaraj, V. V., Górriz, J. M., Zhang, X., and Zhang, Y. D. (2021a). Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf. Fusion* 67, 208–229. doi: 10.1016/j.inffus.2020.10.004
- Wang, S., Nayak, D. R., Guttery, D. S., Zhang, X., and Zhang, Y. D. (2021b). COVID-19 classification by CSHNet with deep fusion using transfer learning and discriminant correlation analysis. *Inf. Fusion* 68, 131–148. doi: 10.1016/j.inffus.2020.11.005
- Wang, Z., and Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Process. Lett.* 9, 81–84. doi: 10.1109/97.995823
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861
- Xu, L., Lu, C., Xu, Y., and Jia, J. (2011). Image smoothing via L0 gradient minimization. *ACM Trans. Graph.* 30, 1–12. doi: 10.1145/2070781.2024208
- Xydeas, C. S., and Petrovic, V. (2000). Objective image fusion performance measure. *Electron. Lett.* 36, 308–309. doi: 10.1049/el:20000267
- Yin, M., Liu, X., Liu, Y., and Chen, X. (2019). Medical image fusion with parameter-adaptive pulse coupled-neural network in nonsubsampling shearlet transform domain. *IEEE Trans. Instrum. Meas.* 68, 49–64. doi: 10.1109/TIM.2018.2838778
- Yu, S., and Chen, X. (2020). Infrared and visible image fusion based on a latent low-rank representation nested with multiscale geometric transform. *IEEE Access* 8, 110214–110226. doi: 10.1109/ACCESS.2020.3001974
- Zeng, X., Luo, Z., and Xiong, X. (2020). A fast fusion method for visible and infrared images using fourier transform and difference minimization. *IEEE Access* 8, 213682–213694. doi: 10.1109/ACCESS.2020.3041759
- Zhan, K., Kong, L., Liu, B., and He, Y. (2019). Multimodal image seamless fusion. *J. Electron. Imaging* 28:023027. doi: 10.1117/1.JEI.28.2.023027
- Zhao, W., Yang, H., Wang, J., Pan, X., and Cao, Z. (2021). Region- and pixel-level multi-focus image fusion through convolutional neural networks. *Mob. Netw. Appl.* 26, 40–56. doi: 10.1007/s11036-020-01719-9
- Zhu, Z., Wei, H., Hu, G., Li, Y., and Qi, G. (2020). A novel fast single image dehazing algorithm based on artificial multiexposure image fusion. *IEEE Trans. Instrum. Meas.* 70, 1–23. doi: 10.1109/TIM.2020.3024335

- Zhu, Z., Yin, H., Chai, Y., Li, Y., and Qi, G. (2018). A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inf. Sci.* 432, 516–529. doi: 10.1016/j.ins.2017.09.010
- Zhu, Z., Zheng, M., Qi, G., Wang, D., and Xiang, Y. (2019). A phase congruency and local laplacian energy based multi-modality medical image fusion method in NSCT domain. *IEEE Access* 7, 20811–20824. doi: 10.1109/ACCESS.2019.2898111

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Wang, Yin, Sun, Zhang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.