Research Article

Multimodal Medical Image Fusion Based on Multiple Latent Low-Rank Representation

Xi-Cheng Lou [ID] **and Xin Feng** [ID]

School of Mechanical Engineering, Key Laboratory of Manufacturing Equipment Mechanism Design and Control of Chongqing, Chongqing Technology and Business University, Chongqing 400067, China

Correspondence should be addressed to Xin Feng; 149495263@qq.com

A multimodal medical image fusion algorithm based on multiple latent low-rank representation is proposed to improve imaging quality by solving fuzzy details and enhancing the display of lesions. Firstly, the proposed method decomposes the source image repeatedly using latent low-rank representation to obtain several saliency parts and one low-rank part. Secondly, the VGG-19 network identifies the low-rank part's features and generates the weight maps. Then, the fused low-rank part can be obtained by making the Hadamard product of the weight maps and the source images. Thirdly, the fused saliency parts can be obtained by selecting the max value. Finally, the fused saliency parts and low-rank part are superimposed to obtain the fused image. Experimental results show that the proposed method is superior to the traditional multimodal medical image fusion algorithms in the subjective evaluation and objective indexes.

1. Introduction

Medical imaging modalities are varied, each of which provides different information about organs in the body. For example, computerized tomography (CT) has excellent resolution, enabling examination of finer details in tissue, but it is weak in showing the global organ structure and pathological changes. Magnetic resonance imaging (MRI) reveals remarkable soft tissue definition with high spatial resolution, but it is limited in detecting fractures. Furthermore, anatomical imaging techniques such as CT and MRI can not reflect the body's movement information, such as metabolism. Positron emission tomography (PET) and single-photon emission computed tomography (SPECT) can visualize metabolic processes and other physiological activities, such as blood flow and regional chemical composition absorption [1]. Nevertheless, functional imaging techniques such as PET and SPECT have a low spatial resolution. In summary, it is impossible to obtain all the details of an organ from a separate imaging modality. For improving the clinical accuracy of diagnosing based on medical images, an effective

method is multimodal medical image fusion, which combines multiple medical images from various modalities to improve quality and reduce redundancy of imaging.

Most medical image fusion algorithms are based on the multiscale transform (MST), which converts the source images into the transform domain and obtain the transformed coefficients through preset functions. Then, the processed coefficients can be converted to the fused image by inverse MST. According to the different decomposition methods of source images, MST can be divided into pyramid-based methods [2–4], wavelet-based methods [5–8], and multiscale geometric analysis- (MGA-) based methods [9–18]. Due to the limitation of preset functions in the MST-based algorithm, some essential features of the source images, such as edge and texture information, may not be well expressed and extracted, which significantly reduces the fusion performance. Moreover, the MST-based algorithm is usually sensitive to misregistration.

Yang and Li [19] first applied sparse representation (SR) [20] in image processing. SR decomposes the source images into several patches through a sliding window and
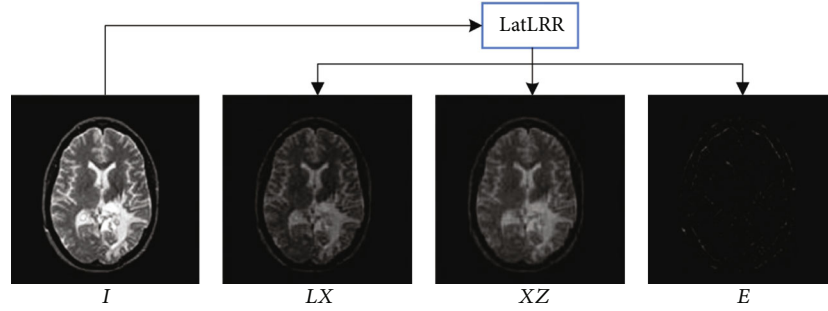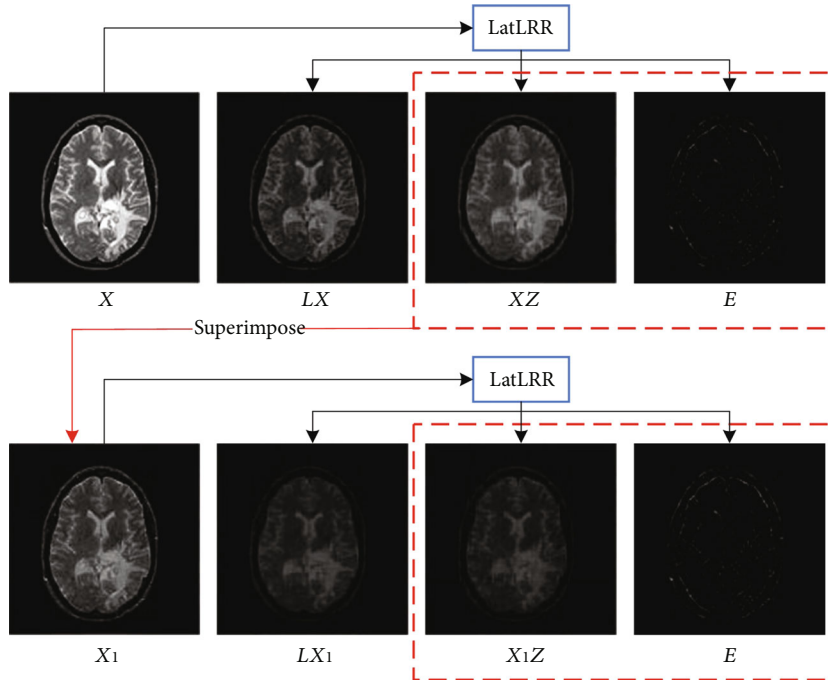
FIGURE 1: LatLRR decomposition diagram.

FIGURE 2: Two-layer LatLRR decomposition diagram.

rearranges these patches to sparse coefficient vectors, which are the linear combination of vectors in the dictionary matrix. Then, the fused image's sparse coefficient vectors can be determined through maximal l_1-norm, rearrange these vectors to patches of fused image, and put these patches return to the seat can obtain the fused image. In terms of edge feature extraction, SR has certain advantages over MST. Many improved versions of SR have appeared in recent years to increase computational efficiency or improve fusion quality. Liu and Wang [21] proposed adaptive sparse representation (ASR), with seven subdictionaries trained in advance to match the patches of images categorized by the gradient. Liu et al. [22] proposed convolutional sparse representation (CSR), which does not use the sliding window to decompose the source image but applies a globe process. Liu et al. [23] proposed convolutional sparsity-based morphological component analysis (CSMCA), which simultaneously achieves multicomponent and global SR by integrating CSR and the morphological component analysis into a unified framework. However, since the SR-based algorithm's dictionary matrix cannot fully include source image data, it fails to extract the source image's detailed texture information. Some scholars applied MST [24–26] or filter [27–29] to decompose the source images. And SR can be used to fuse the low-frequency subbands. Theoretically, such methods can preserve the edge information of the image better than using SR [30] to decompose the source images. Liu et al. [31] proposed the low-rank representation (LRR), which applies the source image as the dictionary matrix, and can solve dictionary completeness. Liu and Yan [32] proposed the latent low-rank representation (LatLRR), an improved version of LRR, which can decompose the source image to saliency part, low-rank part, and noise part. Li et al. [33] proposed MDLatLRR, which integrated SR and LatLRR by using a sliding window to sample the saliency parts of LatLRR and processed the sparse coefficient vectors just like in SR-based methods. LatLRR has an extraordinary capacity for extracting texture from the image, but the ability to extract high-frequency information is not as good as MST.
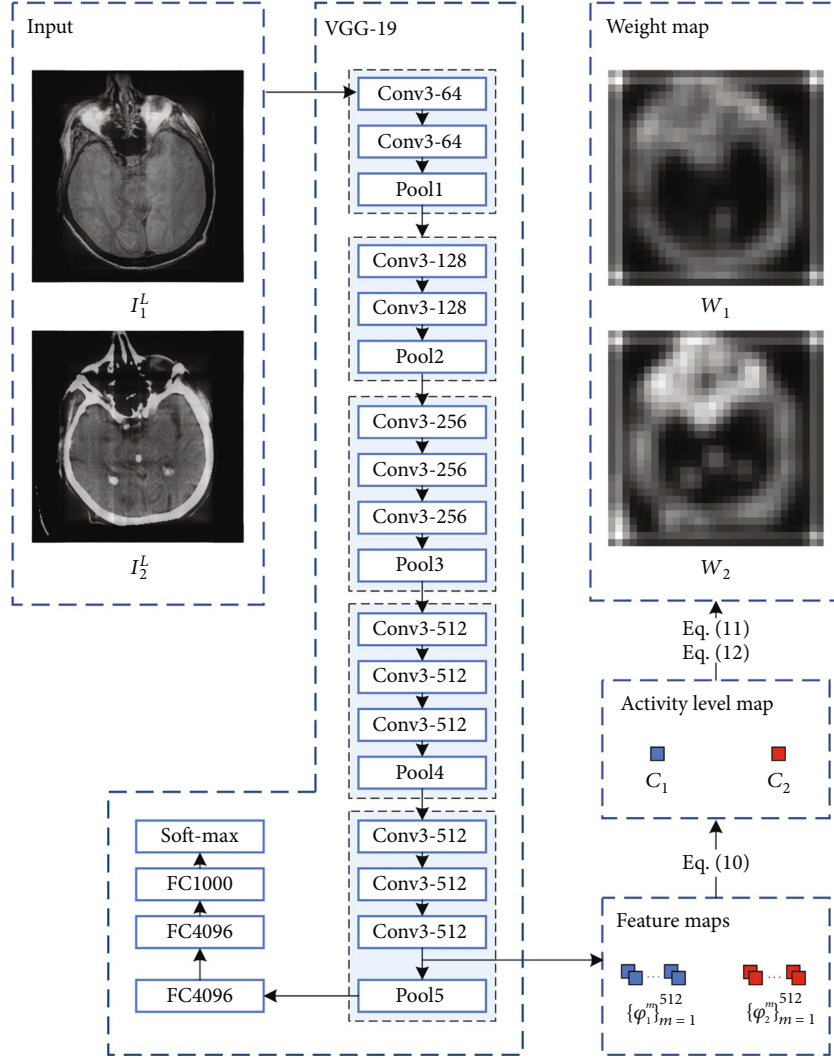
Figure 3: The procedure of low-rank part fusion.

Another type of image fusion method that is more widely used is with the help of weighted maps [34, 35], and deep learning-based methods are particularly suitable for generating weighted maps due to their superior feature recognition capabilities. The deep learning-based methods [36–42] have been widely used in image fusion with the development of artificial intelligence. These methods have a prominent ability to extract feature information from the image. Therefore, it is wise to use deep learning-based methods to deal with high-frequency information after image decomposition or generate the weight map as the basis of image region fusion. Wang et al. [36] applied a convolutional neural network (CNN) to generate weight maps and decompose the source images and weight maps by contrast pyramid and Laplacian pyramid, respectively, and make the Hadamard product of each decomposition layer. Finally, the fused image can be reconstructed through the contrast pyramid. Xu et al. [37] applied LatLRR to decompose source images and processed the low-rank parts by CNN and pyramid-based methods, superimposed the fused low-rank part, and fused saliency part to obtain the fused image. Liu et al. [38] applied

CNN to generate weight maps and use the Laplacian pyramid and Gaussian pyramid to decompose the source images and weight maps, respectively, though the Hadamard product to obtain fused layers, and the fused image can be reconstructed through the Laplacian pyramid. Li et al. [39] applied the average filter to decompose the source images, and the fused base layer can be obtained by comparing the max absolute value of images in four convolutional layers of the VGG-19 neural network. The fused image can be obtained by superimposing the base layer and detail layer. Yin et al. [40], Tan et al. [41], and Panigrahy et al. [42] applied nonsubsampled shearlet transform (NSST) to decompose the source images and selected the fused high-frequency subbands by more firing times in the parameter-adaptive pulse coupled neural network (PAPCNN), bounded measured pulse coupled neural network (BMPCNN), and weighted parameter adaptive dual-channel pulse coupled neural network (WPADCPCNN), respectively.

As mentioned above, each method has its drawbacks and advantages. In this paper, the source images are repeatedly decomposed through LatLRR to extract the saliency parts.
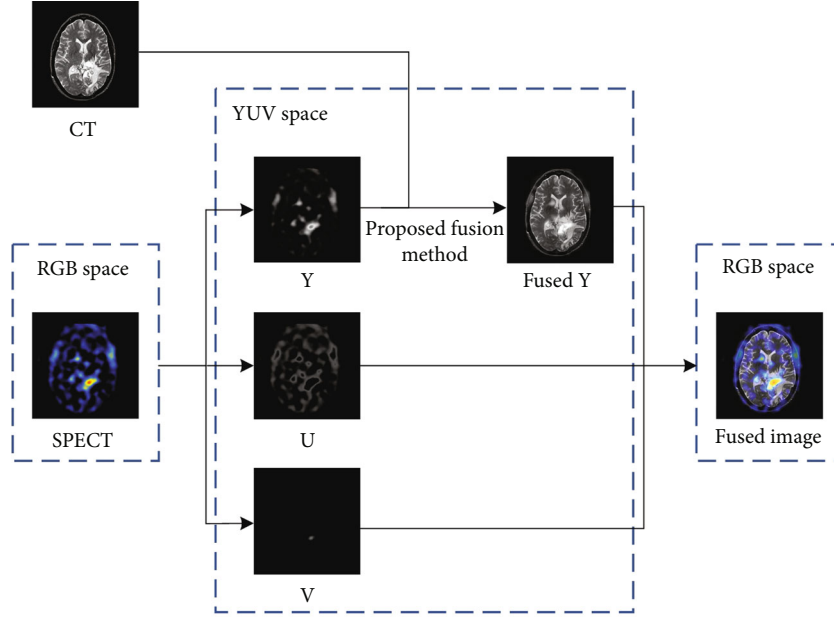
FIGURE 4: Schematic of the grayscale and color image.

The fused saliency parts can be obtained by selecting the max value. After superimposing these saliency parts, the structure and edge information of the source images will be well preserved and enhance the lesion's display. Then, the VGG-19 network is used to extract features of the low-rank part, and the weight maps can be generated to be the basis of the low-rank part's activity level. The weight maps and low-rank parts then make the Hadamard product to obtain the fused low-rank part. Finally, the fused image can be obtained by superimposing the fused saliency part and fused low-rank part. The experimental results also show that the proposed method is significantly better than the comparison method regarding image information retention. The main contributions of this paper are described as follows:

(1) The proposed method applies the image detail retention capability of LatLRR while fully extracting the high-frequency information of an image by iteratively decomposing the original image. It compensates for the deficiency of LatLRR and enhances the display of the lesion by superimposing saliency parts

(2) The feature map of the low-rank part of the original image is extracted using the VGG-19 network and then scaled up to match the size of the original image. The weight map generated in this way can well fit the low-rank part of the original image with pixel information blockwise distributed

The rest of this paper is organized as follows. Section 2 introduces the multiple LatLRR decomposition algorithm, Section 3 introduces the fusion rules, Section 4 describes the algorithmic structure of the proposed method, Section 5 provides a detailed discussion of the experimental results, and Section 6 concludes this paper.

2. Multiple Latent Low-Rank Representation

LatLRR is an improved version of LRR, whose principle is date $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$ in space $\mathbf{R}^n$ can be represented by a linear combination of vectors in an overcomplete dictionary $\mathbf{D} \in \mathbf{R}^{n \times m} (n < m)$, as

$$\mathbf{X} = \mathbf{DZ}, \tag{1}$$

where $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_M\}$ is the coefficient matrix in space $\mathbf{R}^m$; it can be determined through

$$\min_Z \|Z\|_* \quad s.t. \mathbf{X} = \mathbf{DZ}, \tag{2}$$

where $\|\cdot\|_*$ denotes the nuclear norm. The idea of this algorithm is similar to that of SR, in that it finds the coefficients of an image under certain dictionary conditions. LRR using date $\mathbf{X}$ itself as the dictionary, just like equation (3), that is the reason LRR does not have the problems of dictionary training or completeness.

$$\min_Z \|\mathbf{Z}\|_* \quad s.t. \mathbf{X} = \mathbf{XZ}. \tag{3}$$

The noise component $\mathbf{E}$ is added in equation (3); this is because the original purpose of creating the low-rank algorithm is to remove noise from the image. And Equation (4) is the formula of LRR.

$$\min_{Z,E} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{1,2} \quad s.t. \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \tag{4}$$

where $\lambda > 0$ is balance coefficient and $\|\mathbf{E}\|_{1,2}$ denotes the $l_{1,2}$-norm of $\mathbf{E}$.

However, there are two prerequisites for using $\mathbf{X}$ itself as a dictionary. One is that the data vector of $\mathbf{X}$ must be sufficiently
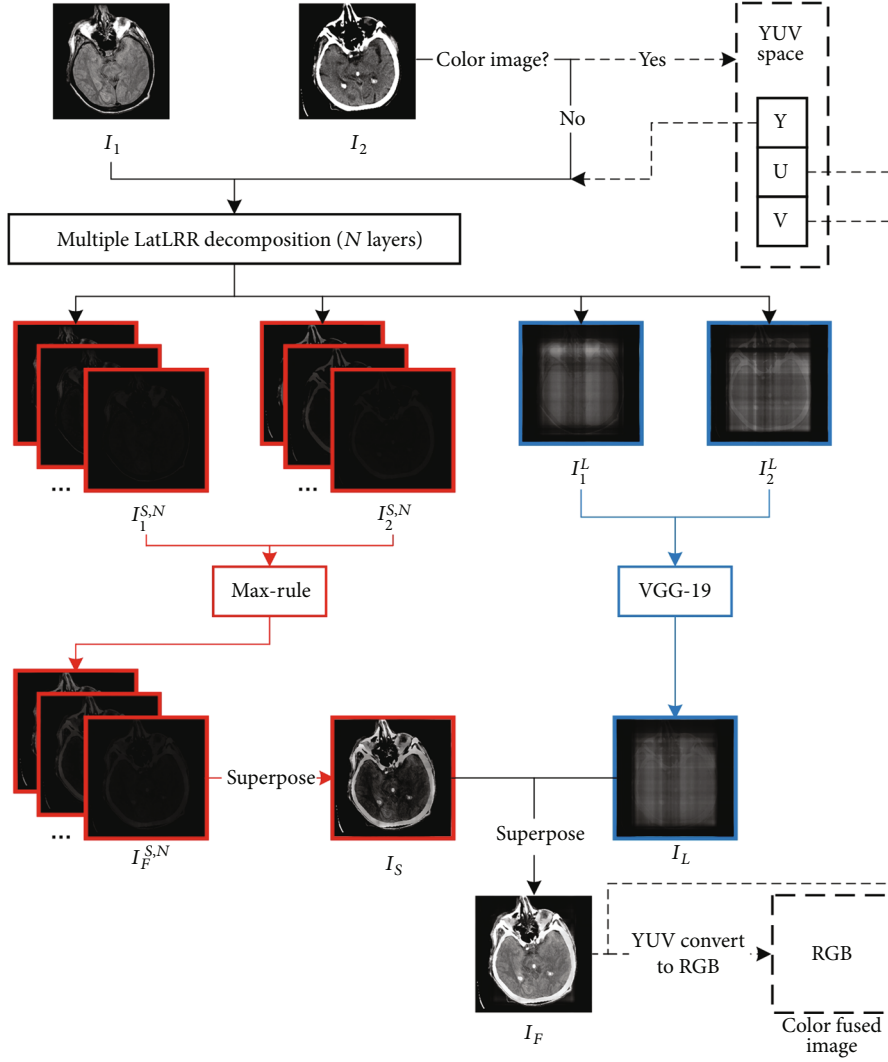
FIGURE 5: Algorithm framework.

complete. Second, the noise of $\mathbf{X}$ must be controlled in a small range. In many practical conditions, such requirements are challenging to achieve. For this reason, the method of adding hidden items in the dictionary is proposed in [32]

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* \quad s.t. \mathbf{X}_O = [\mathbf{X}_O, \mathbf{X}_H]\mathbf{Z}, \tag{5}$$

where $\mathbf{X}_O$ denotes the known image data and $\mathbf{X}_H$ denotes the unknown hidden data. Since the dictionary contains hidden data, this improved algorithm is called the latent low-rank representation. The influence of noise is taken into account; then, we rewrite equation (5) into

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_1 \quad s.t. \mathbf{X}_O = [\mathbf{X}_O, \mathbf{X}_H]\mathbf{Z} + \mathbf{E}, \tag{6}$$

where $\lambda > 0$ is the balance coefficient and $\|\mathbf{E}\|_1$ denotes the l_1-norm of $\mathbf{E}$. Simplify equation (6) by computing the skinny singular value decomposition (SVD) of $[\mathbf{X}_O, \mathbf{X}_H]$ [32], and

equation (7) can be obtained.

$$\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1 \quad s.t. \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{L}\mathbf{X} + \mathbf{E}, \tag{7}$$

where $\mathbf{L}$ denotes the saliency coefficient and $\mathbf{Z}$ denotes the low-rank coefficient. Equation (7) can be solved by the augmented Lagrange multiplier (ALM) [43]; the low-rank part and saliency part of the image could be represented as $\mathbf{X}\mathbf{Z}$ and $\mathbf{L}\mathbf{X}$ accordingly. The LatLRR decomposition results are shown in Figure 1. LatLRR decomposes the source image $\mathbf{X}$ to the saliency part $\mathbf{L}\mathbf{X}$, the low-rank part $\mathbf{X}\mathbf{Z}$, and the noise part $\mathbf{E}$. $\mathbf{L}\mathbf{X}$ contains the local structure information and saliency features or can be thought of as high-frequency information. $\mathbf{X}\mathbf{Z}$ contains more global structure information and brightness information or can be thought of as low-frequency information. $\mathbf{E}$ denotes the 'superfluous' part that LatLRR separates.

It is worth noting that in images with high spatial resolution, such as CT and MRI, detailed textures may contain essential diagnostic information. Therefore, denoising such images

---

Input: $\mathbf{I}_1$ and $\mathbf{I}_2$.
Output: fused image $\mathbf{I}_F$
/∗ Part 1: N layer multiple LatLRR decomposition. ∗/
1 **for** each $\mathbf{M} = [\mathbf{I}_1, \mathbf{I}_2]$**do**
2   **for** each $i = [1 : N]$**do**
3   Perform LatLRR decomposition on $\mathbf{S}$ to obtain $\{\mathbf{I}_1^{S,i}, \mathbf{I}_1^L\}$ and $\{\mathbf{I}_2^{S,i}, \mathbf{I}_2^L\}$.
4   **end**
5 **end**
   /∗ Part 2: Fusion of saliency part. ∗/
6 **for** each $i = [1 : N]$**do**
7   perform max-rule on $\{\mathbf{I}_1^{S,i}, \mathbf{I}_2^{S,i}\}$ to obtain $\mathbf{I}_S^i$ as euqation (8).
8 **end**
9 Superimpose $\{\mathbf{I}_S^i\}_{i=1}^N$ to obtain the final fused saliency part $\mathbf{I}_S$ as equation (9).
   /∗ Part 3: Fusion of low-rank part.∗/
10 **for** each $k = [1, 2]$**do**
11   Input $\mathbf{I}_k^L$ to VGG-19 network to obtain $\{\phi_k^m\}_{m=1}^{512}$ extract from the 5th layer of the network;
12   Solve the l_1-norm of $\{\phi_k^m\}_{m=1}^{512}$ to obtain activity level map $\mathbf{C}_k$ as equation (10);
13   Calculate the initial weight map $\widehat{\mathbf{W}}_k$ by equation (11);
14   Enlarge $\widehat{\mathbf{W}}_k$ to get the final weight map $\mathbf{W}_k$ as equation (12).
15 **end**
16 Calculate the final fused low-rank part $\mathbf{I}_L$ as equation (13).
   /∗Part 4: Reconstruction.∗/
17 Superimpose the fused saliency part $\mathbf{I}_S$ and the fused low-rank part $\mathbf{I}_L$ to obtain the fused image $\mathbf{I}_F$ as equation (14).

ALGORITHM 1: Proposed image fusion algorithm.

may filter out some critical information. In this case, a reasonable approach is to superimpose the noise part and low-rank part. Moreover, the saliency part **LX** contains most edge and structure information of the image so that the lesions may mainly reflect in the saliency part. If the low-rank part of the image is decomposed repeatedly, the saliency part will be further extracted. As shown in Figure 2, a two-layer LatLRR decomposition structure, after completing the first layer of LatLRR decomposition, the new object of LatLRR decomposition could be obtained through $\mathbf{X}_1 = \mathbf{XZ} + \mathbf{E}$, which can be further decomposed into saliency part $\mathbf{LX}_1$, low-rank part $\mathbf{X}_1\mathbf{Z}$, and noise part $\mathbf{E}$. The source images are denoted as $\mathbf{I}_1$ and $\mathbf{I}_2$. If the number of LatLRR decomposition layers is N, there will be N saliency parts $\{\mathbf{I}_1^{S,i}\}_{i=1}^N$ or $\{\mathbf{I}_2^{S,i}\}_{i=1}^N$ and one low-rank part $\mathbf{I}_1^L$ or $\mathbf{I}_2^L$ for each source image. The display of the edge and structure information will be strengthened in the new image to highlight the lesions by superimposing saliency parts. However, suppose the number N of LatLRR decomposition layers is blindly increased, which will reduce the efficiency of calculation. More importantly, the final fused image will display some artificial information unacceptable for medical images. In this paper, the optimal number of LatLRR decomposition layers will be determined through the experiment in Section 5.1.

3. Fusion Regulation

The saliency parts of the image include most high-frequency information. For multimodal medical images, the critical diagnostic information reflected by a single image is not the same. Therefore, the max-rule is applied to fuse the saliency parts of the image can preserve a single image's

diagnostic information as much as possible. On the other hand, Simonyan and Zisserman [44] first applied the VGG network to extract features at different layers from images and obtain a splendid result. With the development of deep learning, the operation efficiency and precision of the VGG network have been significantly improved. As the number of LatLRR decomposition layers increases, the source image's low-rank part will contain less information. If the recognition results of VGG-19 are extracted and processed, the weight map with regional emphasis can be generated. Then, the fused low-rank part can be obtained by multiplying the weight map with the source image's low-rank part. Besides, because PET and SPECT images are in color, they need to be converted into YUV color space before fusing them with the grayscale image as MRI.

3.1. Fusion of Saliency Parts. Each LatLRR decomposition of the two source images will produce a saliency part of each. By adopting max-rule for all saliency parts of N layers of LatLRR decomposition, N fused significant parts can be obtained, as

$$\{\mathbf{I}_S^i\}_{i=1}^N(x, y) = \max\left[\{\mathbf{I}_1^{S,i}\}_{i=1}^N(x, y), \{\mathbf{I}_2^{S,i}\}_{i=1}^N(x, y)\right], \quad (8)$$

where $\{\mathbf{I}_S^i\}_{i=1}^N(x, y)$ denote the position (x, y) of the ith layer saliency part of fused image $\{\mathbf{I}_S^i\}_{i=1}^N$, so as $\{\mathbf{I}_1^{S,i}\}_{i=1}^N(x, y)$ and $\{\mathbf{I}_2^{S,i}\}_{i=1}^N(x, y)$. The final fused saliency part can be
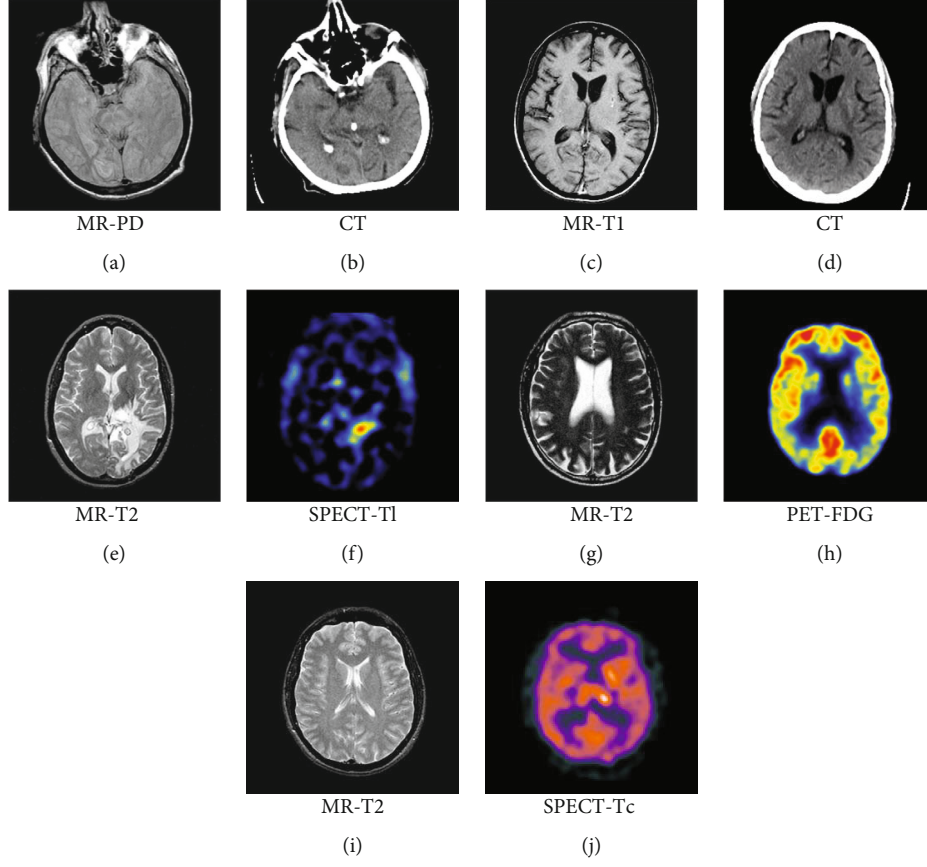
FIGURE 6: Five sets of multimodal medical images for the experiment. (a) MR-PD. (b) CT. (c) MR-T1. (d) CT. (e) MR-T2. (f) SPECT-Tl. (g) MR-T2. (h) PET-FDG. (i) MR-T2. (j) SPECT-Tc.

calculated by

$$\mathbf{I}_S = \sum_{i=1}^{N} \mathbf{I}_S^i. \tag{9}$$

3.2. Fusion of Low-Rank Parts. VGG-19 is a convolutional neural network that is 19 layers deep, including 16 convolutional layers and 3 fully connected layers. Its structure is shown in Figure 3. Each convolutional layer is denoted as conv'size of the filter'-'number of such filters,' and max-pooling layers have 2×2 filter with the stride of 2. The information of the low-rank part of the image after multiple LatLRR decompositions is relatively fuzzy and presents a regional-like distribution. According to this feature, the feature maps extracted from the fifth convolution layer of VGG-19 can match the low-rank part of the image's information distribution state after amplification.

For low-rank parts $\mathbf{I}_1^L$ and $\mathbf{I}_2^L$, $\{\phi_1^m\}_{m=1}^{512}$ and $\{\phi_2^m\}_{m=1}^{512}$ denote the feature maps extracted from the fifth convolutional layer of VGG-19. As shown in Figure 3, the 5th convolutional layer is conv3-512, so there are 512 feature maps of each low-rank part. Moreover, because of max-pooling layers, these feature maps are only $(0.5)^5$ the size of the source image. According to [39], let $\{\phi_k^m\}_{m=1}^{512}(x, y)$ denote the (x, y) position of the kth low-rank part's feature

maps, where $k \in \{1, 2\}$. The l_1-norm of $\{\phi_k^m\}_{m=1}^{512}(x, y)$ could be the activity level measure of the low-rank part. So, the activity level map $\mathbf{C}_k$ can be calculated by

$$\mathbf{C}_k(x, y) = \left\| \{\phi_k^m\}_{m=1}^{512}(x, y) \right\|_1. \tag{10}$$

Then, the initial weight map $\widehat{\mathbf{W}}_k$ can be obtained by

$$\widehat{\mathbf{W}}_k(x, y) = \frac{\mathbf{C}_k(x, y)}{\mathbf{C}_1(x, y) + \mathbf{C}_2(x, y)}. \tag{11}$$

As feature maps are only $(0.5)^5$ the size of the source image, so the initial weight map $\widehat{\mathbf{W}}_k$, which is generated by feature maps, is $(0.5)^5$ the size of the source image too. For matching the size of the source image, $\widehat{\mathbf{W}}_k$ need the upsampling procedure as

$$\mathbf{W}_k(x, y) = \widehat{\mathbf{W}}_k(x + p, y + q) \quad p, q \in \{1, 2, \cdots, 15\}. \tag{12}$$

The fused low-rank part $\mathbf{I}_L$ can be calculated by

$$\mathbf{I}_L = \mathbf{W}_1 \circ \mathbf{I}_1^L + \mathbf{W}_2 \circ \mathbf{I}_2^L, \tag{13}$$

where $\circ$ denotes the Hadamard product.
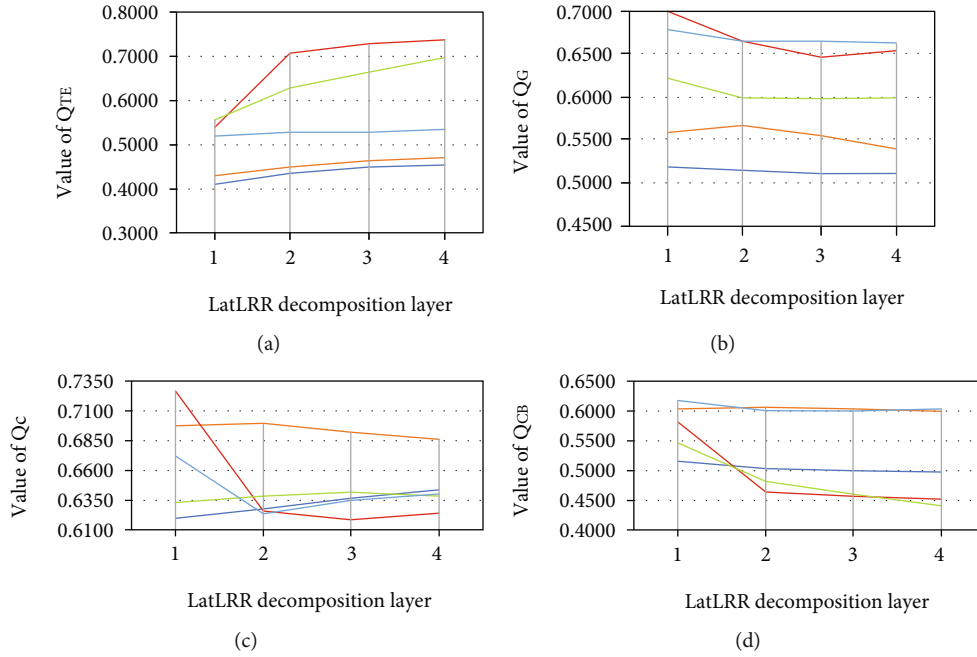
(a)

(b)

(c)

(d)

FIGURE 7: Parametric experimental results. (a) Q_{TE} of five sets of images. (b) Q_G of five sets of images. (c) Q_C of five sets of images. (d) Q_{CB} of five sets of images.
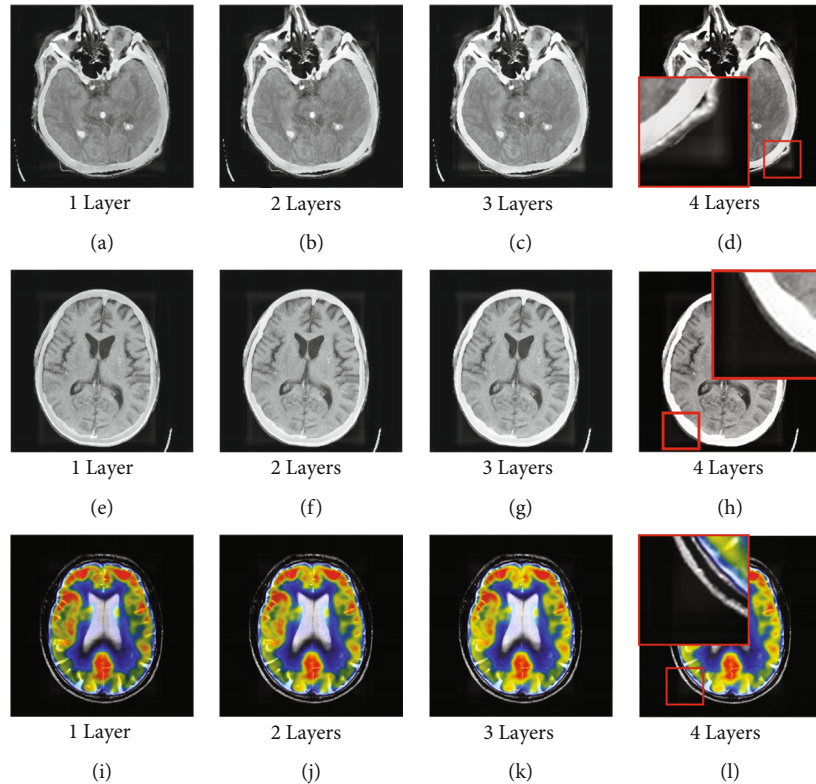
FIGURE 8: The fusion result of each LatLRR decomposition level. (a) 1 layer. (b) 2 layers. (c) 3 layers. (d) 4 layers. (e) 1 layer. (f) 2 layers. (g) 3 layers. (h) 4 layers. (i) 1 layer. (j) 2 layers. (k) 3 layers. (l) 4 layers.

3.3. YUV Color Space. For color images such as SPECT and PET, Yin et al. [40] proposed a YUV space to solve color and grayscale images' fusion problems. The color image is first converted to YUV space and decomposed into one luminance component, 'Y' and two chrominance components, 'U' and 'V.' Then, the 'Y' component of the color image

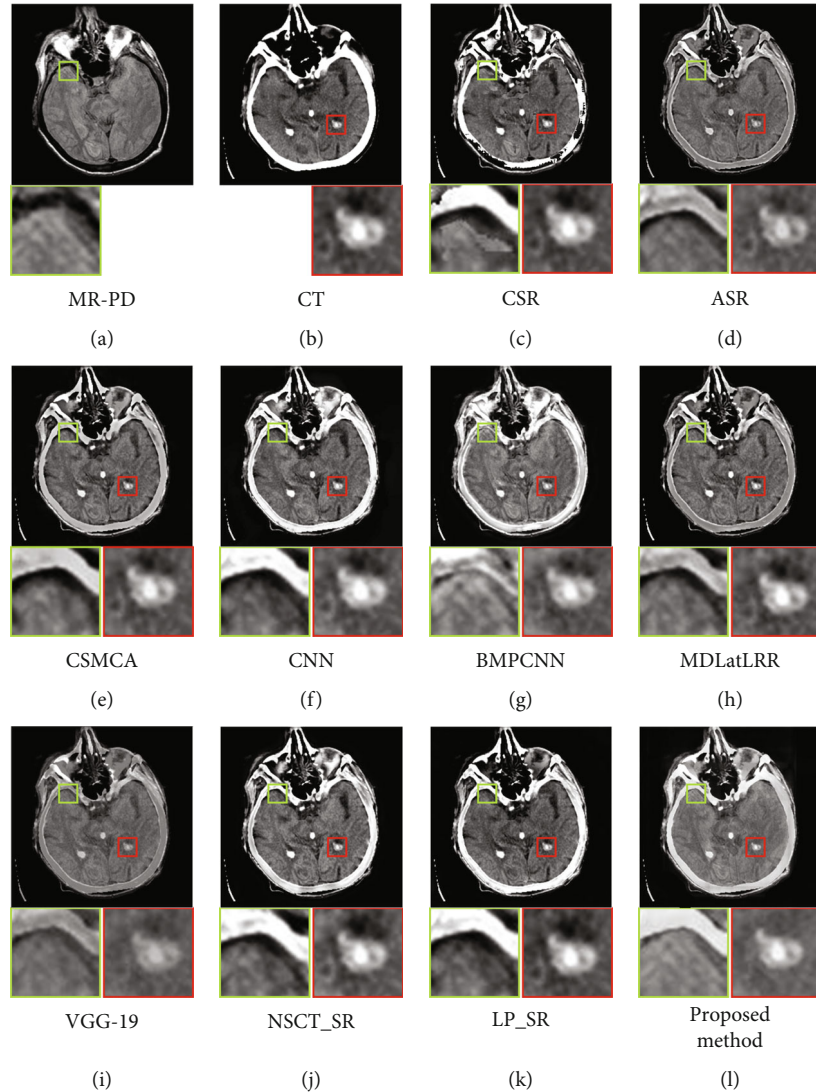FIGURE 9: Comparison of the fusion results in the first set images. (a) MR-PD. (b) CT. (c) CSR. (d) ASR. (e) CSMCA. (f) CNN. (g) BMPCNN. (h) MDLatLRR. (i) VGG-19. (j) NSCT_SR. (k) LP_SR. (l) Proposed method.

can be fused with the grayscale image by the proposed method. The final fused image can be obtained through transforming the fused component 'Y' and other two chrominance components 'U' and 'V' from YUV space to the RGB space, as shown in Figure 4.

3.4. Reconstruction. Superimpose the fused low-rank part $\mathbf{I}_L$ and fused saliency part $\mathbf{I}_S$ to reconstruct fused image $\mathbf{I}_F$ as

$$\mathbf{I}_F = \mathbf{I}_S + \mathbf{I}_L. \tag{14}$$

4. Structure of Algorithm

Source images $\mathbf{I}_1$ and $\mathbf{I}_2$ have been registered, the algorithm framework in this paper as showed in Figure 5.

The main steps of the proposed method are summarized as Algorithm 1.

5. Experiment

Five sets of registered multimodal medical images collected from The Whole Brain Atlas [45] are used to verify the effectiveness of the proposed method, as shown in Figure 6. Figures 6(a) and 6(b) are the first set of images from a 55-year-old patient with multiple embolic infarctions; Figures 6(c) and 6(d) are the second set of images from a 31-year-old man with cerebral toxoplasmosis; Figures 6(e) and 6(f) are the third set of images from a 51-year-old patient with anaplastic astrocytoma; Figures 6(g) and 6(h) are the fourth set of images from a 70-year-old patient with mild Alzheimer's disease; Figures 6(i) and 6(j) are the fifth set of images from a 36-year-old patient with infectious disease due to HIV positive.

All the experiments in this paper are conducted on a PC equipped Intel(R) Xeon(R) CPU E3-1231 v3 (3.40 GHz) and 16 GB RAM. The software environment is MATLAB R2019b installed on Win 10 64-bit operating system.
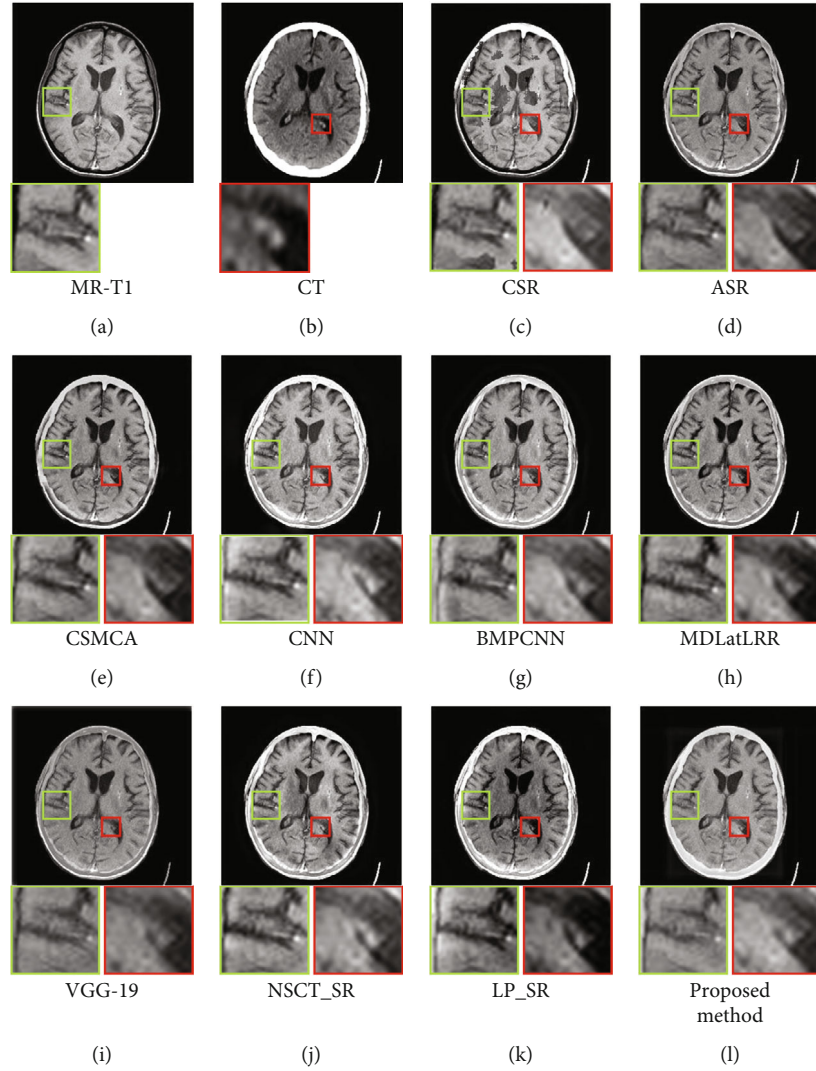
Figure 10: Comparison of the fusion results in the first set images. (a) MR-T1. (b) CT. (c) CSR. (d) ASR. (e) CSMCA. (f) CNN. (g) BMPCNN. (h) MDLatLRR. (i) VGG-19. (j) NSCT_SR. (k) LP_SR. (l) Proposed method.

5.1. Parametric Experiment. In order to determine the decomposition layers of LatLRR in this paper, five sets of images in Figure 6 were fused by the proposed method. The results are objectively evaluated by four fusion image evaluation indexes: fusion metric-based on Tsallis entropy (Q_{TE}) [46], gradient-based fusion performance (Q_G) [47], image structural similarity metric (Q_C) [48], and human perception inspired fusion metric (Q_{CB}) [49]. Q_{TE} is a divergence measure of the degree of dependence between two discrete random variables, and it calculates information from the source images is transferred to the fused image. Therefore, the larger the Q_{TE} value, the better the fusion effect. Q_G uses the Sobel edge operator to calculate the intensity and direction information of the edges in the source image and the fused image. The larger the Q_G value is, the richer the edge information of the fused image is. Q_C is used to measure the preservation degree of structure of the fused image, so it calculates how much of the salient information in each source image has been transferred into the fused

image. The larger the Q_C is, the better the structure of the source images is preserved. The calculation process of Q_{CB} is complex and consists of five steps: contrast sensitivity filtering, local contrast computation, contrast preservation calculation, saliency map generation, and global quality map computation. Q_{CB} takes the mean value of the global quality map. The larger the Q_{CB} value is, the richer the contrast information of the fused image is.

The test decomposition layers of LatLRR are set from 1 to 4. Parameter experimental results of indexes of four sets of images are shown in Figures 7(a)–7(d). It can be observed that as the number of LatLRR decomposition layers increases, not all indexes show a uniform trend. The value of Q_{TE} increased with the increase of decomposition layers, while Q_{CB} are optimal in the case of one-level decomposition. As for Q_G and Q_C, the changing trend is related to the image set. It is reasonable because the more decomposition layers, the image background information contained in the low-rank part will be fuzzier and more contained in the
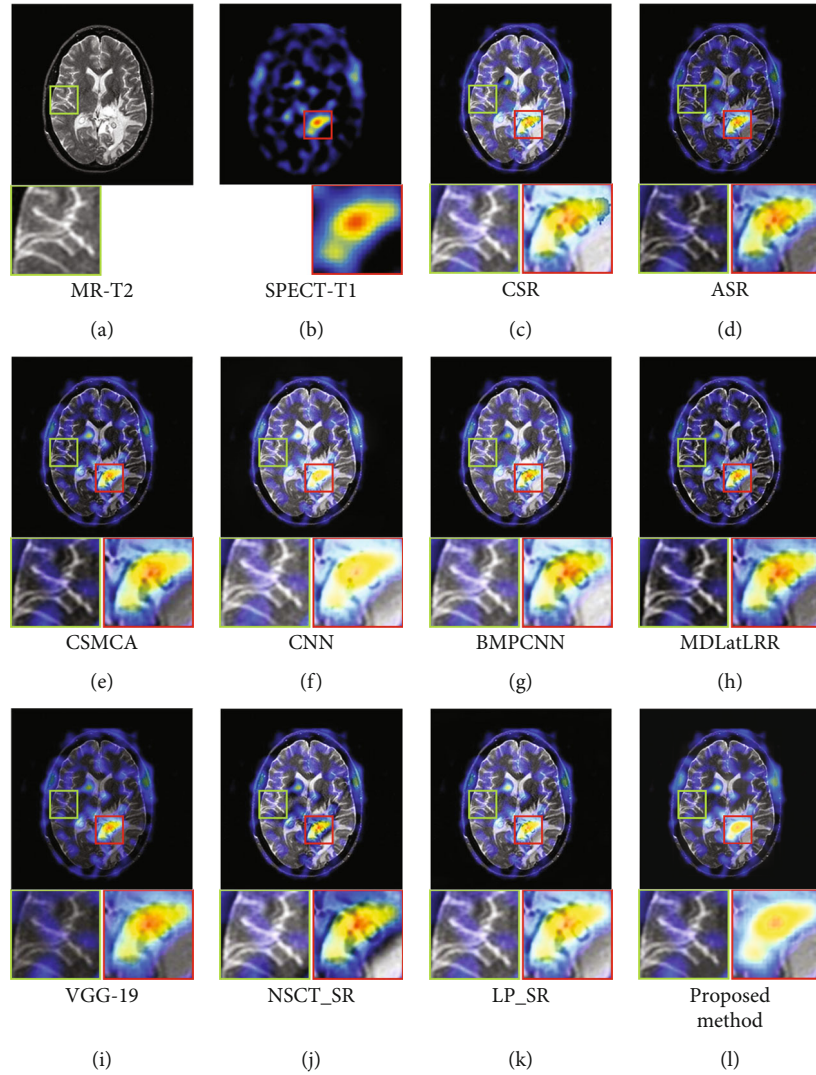
FIGURE 11: Comparison of the fusion results in the second set images. (a) MR-T2. (b) SPECT-T1. (c) CSR. (d) ASR. (e) CSMCA. (f) CNN. (g) BMPCNN. (h) MDLatLRR. (i) VGG-19. (j) NSCT_SR. (k) LP_SR. (l) Proposed method.

saliency part. The max-rule selects the saliency part of the fused image, so the background information of the source images may also be strengthened in the fused image, enhancing the appearance of the lesion. If the greater the amount of information, the larger the Q_{TE} value will be. However, strengthening image background information will weaken the boundary observation and reduce the image contrast, undoubtedly leading to a lower Q_{CB}. Medical image fusion aims to show the information of lesions in the fused image, so the Q_{TE} is more impotent than the other three indexes. It can be seen from Figure 7(a) that Q_{TE} of two decomposition layers is significantly improved than that of one decomposition layer. Still, the more decomposition layers could not considerably improve the Q_{TE}.

Besides, as Figure 8 shows, it can be seen that with the increase of the number of LatLRR decomposition layers, the artifact around the object will be aggravated in several sets of images. In order to improve the image fusion quality, the lesion in the fused image should be highlighted as much

as possible, and artifacts should be strictly controlled. Therefore, the decomposition layer of LatLRR in this paper is set as two.

5.2. Contrast Experiment. Nine typical multimodal medical image fusion methods are selected to compare with the proposed method: four sorts of SR-based methods, CSR [22], ASR [21], CSMCA [23], and MDLatLRR [33]; three sorts of deep learning-based algorithm, CNN [36], VGG-19 [39], and BMPCNN [41]; and two sorts of MST integrated SR methods are nonsubsampled contourlet transform (NSCT) combined with SR (NSCT_SR) [24] and Laplacian pyramid (LP) combined with SR (LP_SR) [24]. The MST decomposition level is set to 4.

Figure 9 shows the fusion results of the first set images, Figures 9(a) and 9(b) are the source images. It can be seen from Figure 9(l) of the proposed method, both the brain tissue texture in the green box and the lesion edge in the red box are the clearest from other fusion methods. Besides,
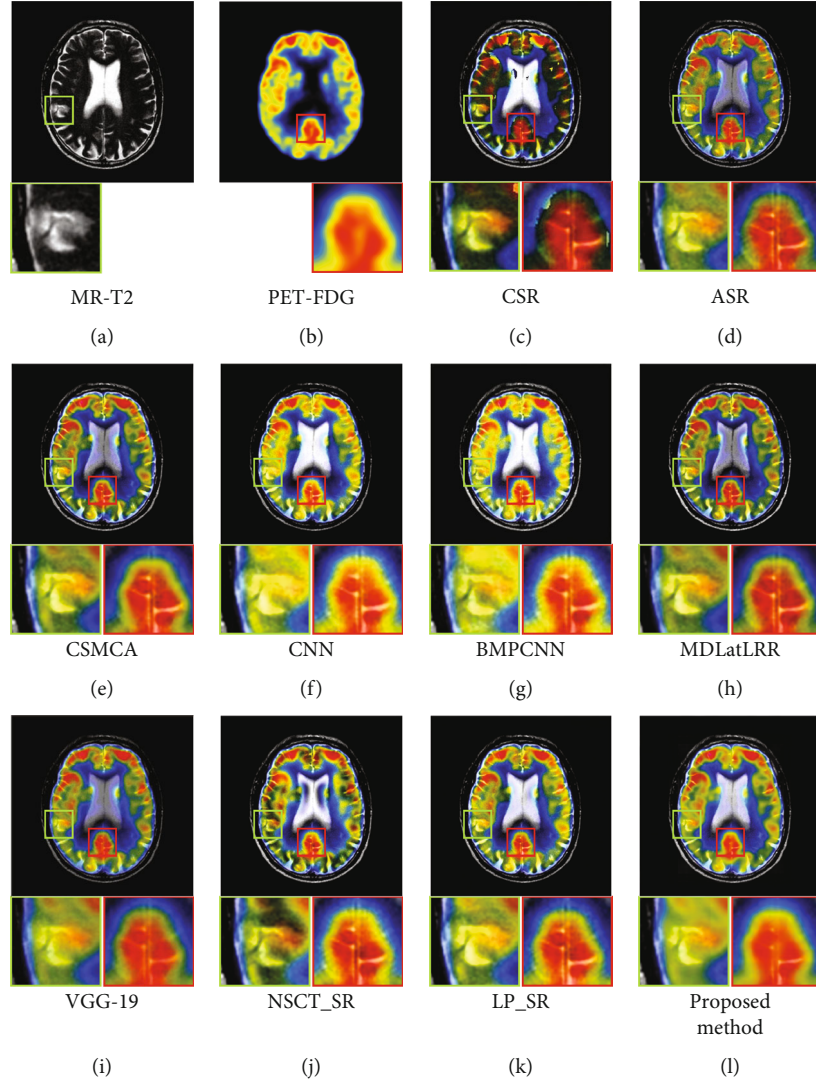
FIGURE 12: Comparison of the fusion results in the third set images. (a) MR-T2. (b) PET-FDG. (c) CSR. (d) ASR. (e) CSMCA. (f) CNN. (g) BMPCNN. (h) MDLatLRR. (i) VGG-19. (j) NSCT_SR. (k) LP_SR. (l) Proposed method.

the pixel consistency of bone in the proposed method is the best. Figure 10 shows the fusion results of the second set images; Figures 10(a) and 10(b) are the source images. As can be seen in Figure 10(l), the pixel consistency of the skeletal structure of the fused images is the best, and the widening of the brain tissue sulcus in the green box, as well as the calcified lesions and edema in the red box, are also clearly visible. Figure 11 shows the fusion results of the second set images; Figures 11(a) and 11(b) are the source images. In Figure 11(l) of the proposed method, the boundary of the metabolic abnormality in the red box is the clearest, and the chromatic aberration is most consistent with Figure 11(b). The texture information in the green box is also clearly visible. Figure 12 shows the fusion results of the third set images; Figures 12(a) and 12(b) are the source images. In Figure 12(l) of the proposed method, the boundary of the metabolic abnormality in the red box is the most distinct, and the texture in the green box also holds the best. Besides, in this set of images, some fusion methods appear

serious color distortion, as can be seen in Figures 12(c), 12(d), and 12(j), the color rendering of the proposed method is closest to the source image. Figure 13 shows the fusion results of the fourth set images, Figures 13(a) and 13(b) are the source images. Compared with Figure 13(b), the graphic structure of metabolic abnormalities (red box) in other fusion methods has been deformed to a certain extent. The structure is kept intact in the proposed method, and the chromatic aberration is most consistent with the source image.

All the fusion images are evaluated by four indexes introduced in Section 5.1. As shown in Tables 1–5, in all sets of images, the proposed method leads in Q_{TE}, especially in the first set of images, the lead is more than 30 percent. It means that the proposed method is superior to other methods in the information conversion of source images. Moreover, for the first set and the fourth set of images, the proposed method also leads in the Q_C. And it indicates that the proposed method can retain the structure of the source
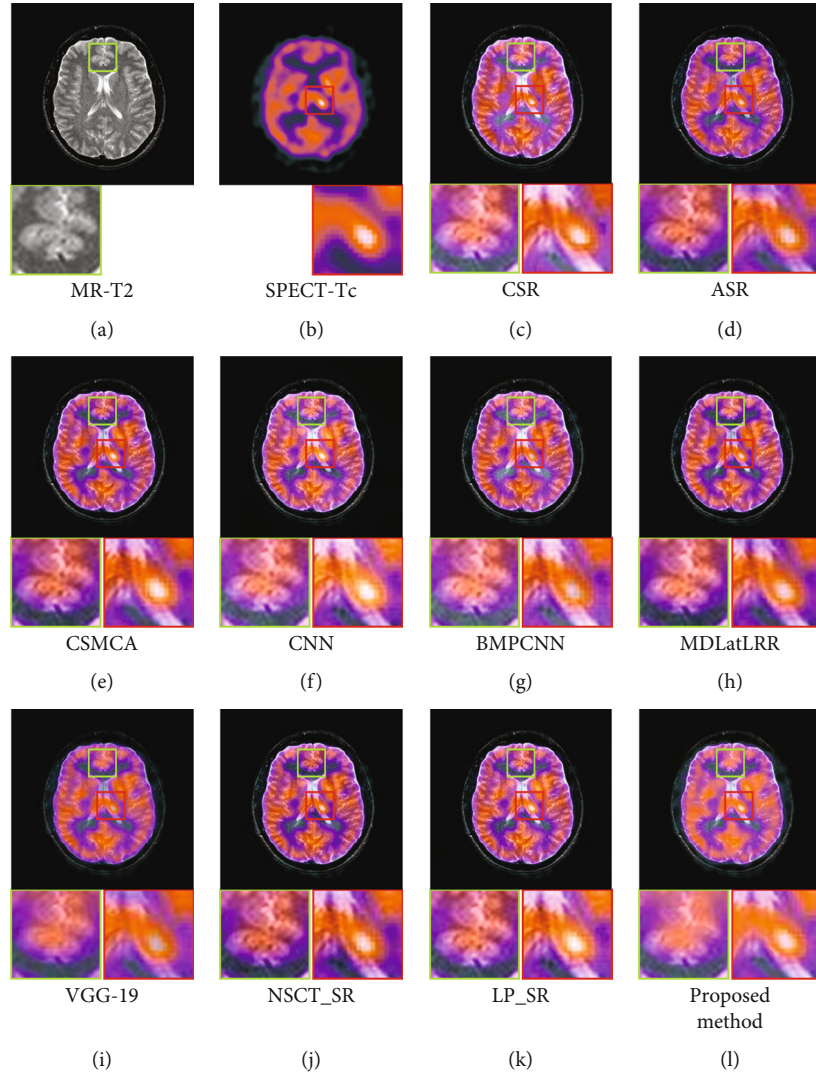
FIGURE 13: Comparison of the fusion results in the fourth set images. (a) MR-T2. (b) SPECT-Tc. (c) CSR. (d) ASR. (e) CSMCA. (f) CNN. (g) BMPCNN. (h) MDLatLRR. (i) VGG-19. (j) NSCT_SR. (k) LP_SR. (l) Proposed method.

TABLE 1: Objective evaluation of fusion methods in the first set of images.

Fusion method	Q_{TE}	Q_G	Q_C	Q_{CB}	t/s
Proposed method	*0.4359*	0.5145	*0.6436*	0.5036	158.7860
CSR	0.3096	0.6608	0.5621	*0.7137*	135.2650
ASR	0.2976	*0.6700*	0.6121	0.6622	210.1220
CSMCA	0.2945	0.5830	0.5587	0.6103	372.3040
CNN	0.3247	0.6267	0.5560	0.5916	32.7860
BMPCNN	0.3058	0.5554	0.5828	0.5943	67.0890
MDLatLRR	0.2882	0.6527	0.6077	0.6784	69.0270
VGG-19	0.2835	0.5104	0.5569	0.6273	7.4480
NSCT_SR	0.2953	0.6243	0.5895	0.6353	7.4890
LP_SR	0.2948	0.6560	0.5734	0.6588	*0.3380*

TABLE 2: Objective evaluation of fusion methods in the second set of images.

Fusion method	Q_{TE}	Q_G	Q_C	Q_{CB}	t/s
Proposed method	*0.4495*	0.5647	0.6995	0.6061	50.6440
CSR	0.3911	0.6363	0.7000	*0.7454*	33.1950
ASR	0.3595	*0.7122*	*0.7532*	0.7328	64.9900
CSMCA	0.3491	0.6315	0.7090	0.6974	78.8380
CNN	0.4131	0.4757	0.6533	0.5848	13.8630
BMPCNN	0.3726	0.5354	0.6918	0.6210	15.7120
MDLatLRR	0.3657	0.6516	0.7108	0.7357	21.1340
VGG-19	0.4211	0.4030	0.6457	0.3627	5.1570
NSCT_SR	0.3725	0.5514	0.6886	0.6389	2.2760
LP_SR	0.3582	0.6025	0.6741	0.6521	*0.2210*

images better than other methods. On the other hand, the proposed method has no advantage or even a considerable gap over the best-performing method in terms of the other

two indicators. As mentioned in Section 5.1, Q_G and Q_{CB} are mainly used to measure the degree to which the fused image retains the edge information of the source images.

Table 3: Objective evaluation of fusion methods in the third set of images.

Fusion method	Q_{TE}	Q_G	Q_C	Q_{CB}	t/s
Proposed method	*0.7297*	0.6602	0.6249	0.4645	233.0850
CSR	0.5025	*0.8111*	0.7920	*0.7010*	130.2410
ASR	0.4632	0.7901	0.7755	0.6611	188.6130
CSMCA	0.4576	0.7319	0.6790	0.5979	390.8500
CNN	0.7183	0.7822	0.7530	0.4966	38.6490
BMPCNN	0.3832	0.7900	*0.8470*	0.6505	75.5030
MDLatLRR	0.4303	0.7755	0.7298	0.6416	73.5810
VGG-19	0.4828	0.7026	0.6891	0.6179	7.8900
NSCT_SR	0.4659	0.8054	0.8200	0.6757	9.2580
LP_SR	0.5044	0.7994	0.8120	0.5991	*0.5030*

Table 4: Objective evaluation of fusion methods in the fourth set of images.

Fusion method	Q_{TE}	Q_G	Q_C	Q_{CB}	t/s
Proposed method	*0.6267*	0.5968	*0.6416*	0.4832	163.4530
CSR	0.5041	*0.8026*	0.6380	*0.7370*	149.6120
ASR	0.4431	0.7291	0.6231	0.6573	192.5610
CSMCA	0.4383	0.6821	0.6355	0.6141	344.2190
CNN	0.4197	0.6749	0.5585	0.6446	33.4460
BMPCNN	0.3437	0.5930	0.6328	0.5226	71.7970
MDLatLRR	0.4270	0.6752	0.6056	0.6567	61.6160
VGG-19	0.4730	0.6605	0.5722	0.6353	7.6620
NSCT_SR	0.4529	0.6827	0.6386	0.6001	7.5950
LP_SR	0.4457	0.6734	0.5938	0.5764	*0.3240*

Table 5: Objective evaluation of fusion methods in the fifth set of images.

Fusion method	Q_{TE}	Q_G	Q_C	Q_{CB}	t/s
Proposed method	*0.5260*	0.6594	0.6234	0.6005	46.2110
CSR	0.5175	*0.8783*	*0.8793*	*0.8374*	22.4600
ASR	0.4639	0.7902	0.7986	0.6573	53.9060
CSMCA	0.4808	0.7908	0.8237	0.7346	76.9750
CNN	0.5186	0.7668	0.8170	0.4144	11.8690
BMPCNN	0.3632	0.7289	0.8196	0.7099	14.8240
MDLatLRR	0.4591	0.7905	0.8180	0.7464	17.9870
VGG-19	0.4947	0.6974	0.6851	0.6797	5.1760
NSCT_SR	0.4962	0.7658	0.7943	0.9358	2.2660
LP_SR	0.5083	0.7843	0.8435	0.6496	*0.2280*

Q_G, in particular, is directly affected by the gradient of the fused image. As shown in Figure 12(c), the color of the CSR fusion image is seriously distorted, and the white area in the middle of the image has apparent artifacts, which is unacceptable for brain images. However, the color distortion and artifacts bring to the fused image that the 'border' does not initially exist and will undoubtedly increase the image structure's gradient value and complexity. That is why Q_G and Q_{CB}

of CSR show undeniable advantages in the contrast experiment of third set images. Q_{TE} measures the degree to which the fusion image retains the information of the source images, and the distortion of image structure or color will lower this index. Image artifacts and structural distortions may mislead medical professionals, so it is not advisable to blindly pursue high image structure indexes. The proposed method has obvious advantages in the Q_{TE} index, and the degree of color distortion and artifact in the fused images is minimal. It is crucial for medical images. Because artifacts in brain images can sometimes look very similar to lesions, false lesions in images can directly affect the judgment of medical professionals. In addition, SPECT and PET show metabolic abnormalities through chromatic aberration. If significant color distortions appear in the image, it will cause deviation from the real. Based on the above analysis, the proposed method is effective.

6. Conclusion

In this paper, a multimodal medical image fusion method based on multiple latent low-rank representation is proposed. Experimental results show that the proposed method has advantages in preserving edge features and texture details and leads to objective evaluation indexes compared with other fusion algorithms. The proposed method can enhance the observer's ability to identify the lesions and contribute to practical applications such as diagnosis, treatment planning, and surgical navigation. On the other hand, the proposed method also has some drawbacks. The highlight of lesions is based on multilayer LatLRR decomposition, but with the increase of the number of decomposition layers, artifacts will become more evident. Besides, the more layers decomposed by LatLRR, the less information of the low-rank part of the source image, and the greater the error of VGG-19 network recognition. The next stage should focus on eliminating artifacts as much as possible in the case of multilayer LatLRR decomposition and improve the fusion quality of low-rank parts of source images.

Data Availability

The data supporting the study are obtained from K. A. Johnson and J. A. Becker, *The Whole Brain Atlas*, 2021. URL: http://www.med.harvard.edu/aanlib/, (accessed 12 May 2021).

Conflicts of 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

References

[1] D. L. Bailey, M. N. Maisey, D. W. Townsend, and P. E. Valk, *Positron Emission Tomography*, vol. 2, Springer, 2005.

[2] Z. Wang, Z. Cui, and Y. Zhu, "Multi-modal medical image fusion by Laplacian pyramid and adaptive sparse representation," *Computers in Biology and Medicine*, vol. 123, article 103823, 2020.

[3] J. Du, W. Li, B. Xiao, and Q. Nawaz, "Union Laplacian pyramid with multiple features for medical image fusion," *Neurocomputing*, vol. 194, pp. 326–339, 2016.

[4] F. Liu, L. Chen, L. Lu, A. Ahmad, G. Jeon, and X. Yang, "Medical image fusion method by using Laplacian pyramid and convolutional sparse representation," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 17, article e5632, 2020.

[5] J. Bhardwaj, A. Nayak, and D. Gambhir, "Multimodal medical image fusion based on discrete wavelet transform and genetic algorithm," in *International Conference on Innovative Computing and Communications*, pp. 1047–1057, Singapore, 2021.

[6] L. Zhan, Y. Zhuang, and L. Huang, "Infrared and visible images fusion method based on discrete wavelet transform," *Journal of Computers*, vol. 28, pp. 57–71, 2017.

[7] S. K. Panguluri and L. Mohan, "Discrete wavelet transform based image fusion using unsharp masking," *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 64, no. 2, pp. 211–220, 2020.

[8] A. Seal, D. Bhattacharjee, M. Nasipuri, D. Rodriguez-Esparragon, E. Menasalvas, and C. Gonzalo-Martin, "PET-CT image fusion using random forest and à-trous wavelet transform," *International journal for numerical methods in biomedical engineering*, vol. 34, no. 3, article e2933, 2018.

[9] K. Seethalakshmi and S. Valli, "A fuzzy approach to recognize face using contourlet transform," *International Journal of Fuzzy Systems*, vol. 21, no. 7, pp. 2204–2211, 2019.

[10] M. Kumar, N. Ranjan, and B. Chourasia, "Hybrid methods of contourlet transform and particle swarm optimization for multimodal medical image fusion," in *2021 International Conference on Artical Intelligence and Smart Systems (ICAIS)*, pp. 945–951, Coimbatore, India, 2021.

[11] W. Wu, S. Guo, and Q. Cheng, "Fusing optical and synthetic aperture radar images based on shearlet transform to improve urban impervious surface extraction," *Journal of Applied Remote Sensing*, vol. 14, no. 2, article 024506, 2020.

[12] A. Khare, M. Khare, and R. Srivastava, "Shearlet transform based technique for image fusion using median fusion rule," *Multimedia Tools and Applications*, vol. 80, no. 8, article 10184, pp. 11491–11522, 2021.

[13] M. Sayadi, H. Ghassemian, R. Naimi, and M. Imani, "A new composite multimodality image fusion method based on shearlet transform and retina inspired model," in *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pp. 1–5, Qom, Iran, 2020.

[14] S. Kollem, K. R. Reddy, and D. S. Rao, "Improved partial differential equation-based total variation approach to non-subsampled contourlet transform for medical image denoising," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2663–2689, 2021.

[15] J.-Y. Li and C.-Z. Zhang, "Blind watermarking scheme based on Schur decomposition and non-subsampled contourlet transform," *Multimedia Tools and Applications*, vol. 79, no. 39-40, pp. 30007–30021, 2020.

[16] T. Chu, Y. Tan, Q. Liu, and B. Bai, "Novel fusion method for SAR and optical images based on non-subsampled shearlet transform," *International Journal of Remote Sensing*, vol. 41, no. 12, pp. 4590–4604, 2020.

[17] H. Ullah, B. Ullah, L. Wu, F. Y. Abdalla, G. Ren, and Y. Zhao, "Multi-modality medical images fusion based on local-features fuzzy sets and novel sum-modified-Laplacian in non-subsampled shearlet transform domain," *Biomedical Signal Processing and Control*, vol. 57, article 101724, 2020.

[18] S. Goyal, V. Singh, A. Rani, and N. Yadav, "FPRSGF denoised non-subsampled shearlet transform-based image fusion using sparse representation," *Signal, Image and Video Processing*, vol. 14, no. 4, pp. 719–726, 2020.

[19] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.

[20] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[21] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Processing*, vol. 9, no. 5, pp. 347–357, 2015.

[22] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convo-lutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.

[23] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Processing Letters*, vol. 26, no. 3, pp. 485–489, 2019.

[24] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information fusion*, vol. 24, pp. 147–164, 2015.

[25] L. Tan and X. Yu, "Medical image fusion based on fast finite shearlet transform and sparse representation," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 3503267, 14 pages, 2019.

[26] J. Xia, Y. Chen, A. Chen, and Y. Chen, "Medical image fusion based on sparse representation and PCNN in NSCT domain," *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 2806047, 12 pages, 2018.

[27] X. Li, F. Zhou, and H. Tan, "Joint image fusion and denoising via three-layer decomposition and sparse representation," *Knowledge-Based Systems*, vol. 224, article 107087, 2021.

[28] X. Li, F. Zhou, H. Tan, W. Zhang, and C. Zhao, "Multimodal medical image fusion based on joint bilateral filter and local gradient energy," *Information Sciences*, vol. 569, pp. 302–325, 2021.

[29] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, "A novel multi-modality image fusion method based on image decomposition and sparse representation," *Information Sciences*, vol. 432, pp. 516–529, 2018.

[30] H. Li, Y. Wang, Z. Yang, R. Wang, X. Li, and D. Tao, "Discriminative dictionary learning-based multiple component decomposition for detail-preserving noisy image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1082–1102, 2020.

[31] G. Liu, Z. Lin, and Y. Yu, *Robust Subspace Segmentation by Low-Rank Representation*, vol. 1, Icml, 2010.

[32] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *2011 International Conference on Computer Vision*, pp. 1615–1622, Barcelona, Spain, 2011.

[33] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: a novel decomposition method for infrared and visible image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4733–4746, 2020.

[34] A. Sengupta, A. Seal, C. Panigrahy, O. Krejcar, and A. Yazidi, "Edge information based image fusion metrics using fractional order differentiation and sigmoidal functions," *IEEE Access*, vol. 8, pp. 88385–88398, 2020.

[35] Z. Zhu, B. Zhu, H. H. T. Liu, and K. Qin, "A model-based approach for measurement noise estimation and compensation in feedback control systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, article 5001523, pp. 1–23, 2020.

[36] K. Wang, M. Zheng, H. Wei, G. Qi, and Y. Li, "Multi-modality medical image fusion using convolutional neural network and contrast pyramid," *Sensors*, vol. 20, no. 8, p. 2169, 2020.

[37] Z. Xu, W. Xiang, S. Zhu et al., "LatLRR-FCNs: latent low-rank representation with fully convolutional networks for medical image fusion," *Frontiers in Neuroscience*, vol. 14, p. 1387, 2021.

[38] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *2017 20th international conference on information fusion (Fusion)*, pp. 1–7, Xi'an, China, 2017.

[39] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *2018 24th international conference on pattern recognition (ICPR)*, pp. 2705–2710, Beijing, China, 2018.

[40] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 49–64, 2019.

[41] W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," *Neural Computing and Applications*, pp. 1–21, 2020.

[42] C. Panigrahy, A. Seal, and N. K. Mahato, "MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN," *IEEE Signal Processing Letters*, vol. 27, pp. 690–694, 2020.

[43] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, https://arxiv.org/abs/1009.5055/.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556/.

[45] K. A. Johnson and J. A. Becker, "The whole brain atlas," 2021, http://www.med.harvard.edu/aanlib/.

[46] N. Cvejic, C. Canagarajah, and D. Bull, "Image fusion metric based on mutual information and Tsallis entropy," *Electronics Letters*, vol. 42, no. 11, pp. 626-627, 2006.

[47] C. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308-309, 2000.

[48] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *International journal of signal processing*, vol. 2, pp. 178–182, 2005.

[49] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image and Vision Computing*, vol. 27, no. 10, pp. 1421–1432, 2009.