Research Article

Multimodal Brain Image Fusion Based on Improved Rolling Guidance Filter and Wiener Filter

Hairui Lin,[1] **Yuhang Song** ⬤,[2] **Hua Wang,**[3] **Luoxin Xie,**[1] **Dongfen Li,**[1] **and Guocheng Yang** ⬤[1]

[1]*College of Computer Science and Cyber Security (Oxford Brookes College), Chengdu University of Technology, Chengdu 610059, China*
[2]*School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610059, China*
[3]*Department of Electronic Information Engineering, Rizhao Polytechnic, Rizhao 276826, China*

Correspondence should be addressed to Guocheng Yang; yangguocheng@cdut.edu.cn

Medical image fusion technology can integrate complementary information from different modality medical images to provide a more complete and accurate description of the specific diagnosed object, which is very helpful for image-guided clinical diagnosis and treatment. This paper proposes an effective brain image fusion framework based on improved rolling guidance filter (IRGF). Firstly, input images are decomposed into base layers and detail layers using the IRGF and Wiener filter. Secondly, the visual saliency maps of the input image are computed by pixel-level saliency value, and the weight maps of detail layers are constructed by max-absolute strategy and are further smoothed with Gaussian filter, the purpose of which is to make the fused image appear more naturally and more suitable for human visual perception. Lastly, base layers are fused by visual saliency map based fusion rule and the corresponding weight maps from detail layers are fused by the weighted least squares optimization scheme. Experimental results testify that our method is superior to some state-of-the-art methods in both subjective and objective assessments.

1. Introduction

Medical image fusion from multiple imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI) image, single photon emission computed tomography (SPECT) image, and positron emission tomography (PET) image, can provide a collective view of the lesion at a specific position of human body, which helps physicians to better understand the pathological conditions of the internal tissues and organs. Inherently, from any single modal image, it is difficult for the human eye to obtain complete and accurate information of the specific organ or tissue [1–3]. Multimodal medical image fusion techniques can integrate complementary information from multiple images acquired by different imaging sensors in the same target, so the expression of the fusion image in the same human body part is more accurate and reliable for diagnosis and assessment of medical problems.

Over the past two decades, many medical image fusion methods have been investigated at the pixel level [4–7]. Specially, multiscale transform- (MST-) based methods have been the hotspots of research and attracted the most widespread attention. For the MST based method, it is generally divided into three steps: decomposition, fusion, and reconstruction. To summarize, MST based method is more widely used, the process of which decomposes an image into the base layer and detail layers coefficients by applying selected transform tool. Further, a variety of fusion strategies are applied to combine different layer subimages and lastly achieves the fused results. Therefore, the selection of the MST tools and the fusion rules are two crucial factors and should be jointly considered for MST based image fusion.

In terms of image decomposition, some typical transform tools include, but are not limited to, Laplacian pyramid [8], complex wavelet transform [9], contourlet transform [10], nonsubsampled contourlet transform (NSCT) [11], shearlet transform (ST) [12], and nonsubsampled shearlet transform (NSST) [13]. With the flexible multiscale, multidirection, and shift invariant image decomposition function, the NSCT and NSST have manifested significant advantages and outstanding fusion performance compared with the other MST tools. Meanwhile, the conventional fusion rules are listed as follows: the averaging scheme, the max-absolute selection scheme, and weighted average scheme etc. In the past five years, researchers have proposed many image fusion algorithms based on NSCT and NSST. Zhu et al. [14] developed an NSCT-based medical image fusion method, in which, a phase congruency based fusion rule was introduced to merge the low-frequency component, and a local Laplacian energy rule was designed to fuse the high-frequency components. Wang et al. [15] proposed a medical image fusion method using the interscale and intrascale dependencies between multiscale image coefficients in the NSST domain. In their method, the dependencies of the NSST coefficients of the cross-scale and intersubbands are fully considered in the proposed fusion rule. In addition, pulse-coupled neural network (PCNN) model and MST based image fusion rule have been widely used in medical image fusion fields. Yin et al. [16] proposed a medical image fusion method with parameter-adaptive pulse-coupled neural network (PA-PCNN) in NSST domain. In their method, the PA-PCNN model and an energy based rule were used to fuse the high frequency and low-frequency coefficients, respectively.

Although many image fusion methods have achieved competitive performance, most approaches assume that the source images are noise-free, which may contain noise in actual tasks. For the fusion of noisy images, sparse representation- (SR-) based technique [7] is regarded as a good choice among numerous image fusion algorithms. To synchronously address the two problems of medical image fusion and denoising, recently, some joint image denoising and fusion algorithms based on SR have been proposed. Liu and Wang [17] proposed an adaptive SR method for image fusion and denoising. In this method, the appropriate dictionary is adaptively selected according to the image gradient features, and thus image fusion and denoising can be synchronously implemented. Kim et al. [18] designed a patch clustering-based dictionary learning method for multimodal image fusion. Although their method can effectively suppress noise generation, some important brightness information of source images cannot be well preserved in the fused images. In 2018, Li et al. [19] presented a joint medical image fusion, denoising, and enhancement method based on discriminative low-rank sparse dictionaries learning, to further improve the discriminative ability of the learned dictionaries. Afterwards, Li et al. [6] further developed a discriminative dictionary learning model to decompose input images into images coarse scale and fine scale components for detail-preserving noisy image fusion. In this scheme, an efficient multiple component dictionary learning model is formulated by integrating rank minimization into the low-rank component. However, this method is relatively complex and time-consuming. Li et al. [20] proposed a novel medical image fusion method based on the joint bilateral filter (JBF) and a gradient energy perception fusion rule, which decomposed a medical image into a structure layer with multilevel gradient and the energy layer containing rich intensity information by the JBF. Subsequently, Li et al. [21] proposed an image fusion method, i.e., joint image fusion and denoising via three-layer decomposition and SR, in which, the sparse reconstruct error parameter is adaptively designed according with the noise level, so as to realize the fusion and denoising for high-frequency components simultaneously.

With the rapid development of deep learning (DL) theory, meadical image fusion by DL is also one of the research focuses in recent years. Some DL based medical image fusion methods [22, 23] have emerged. Yu et al. presented a medical image fusion method based on convolutional neural networks (CNN) [24] for the first time in 2017, which achieved good results relative to the spatial domain and the transform domain. Subsequently, Liu et al. [23] also proposed a medical image fusion based on the convolutional sparse coding and morphological component analysis. However, this method has the flaw that the experiments are only tested on the gray scale medical images.

In recent years, multilevel edge-preserving filtering has been widely used in the field of image fusion. To gain better performance, researchers have proposed a variety of edge-preserving filters [25, 26] for recovering blurred image edges. However, most traditional edge-preserving filters only smooth details based on the contrast of the input image and do not take fully into account the spatial scale of input images. To address this issue, Zhang et al. [27] proposed a rolling guidance filter (RGF), which can smooth small structure information and include both scale aware and edge-preserving properties. Accordingly, many multiscale image fusion methods are introduced by applying RGF. Ma et al. [28] proposed a novel infrared and visible image fusion method that used the RGF and Gaussian filter to decompose input images into base layers and detail layers. Jian et al. [29] put forward an innovated image fusion method based on RGF and JBF filtering, in which RGF could remove small structures and recover the edges during the iterative process. Chen et al. [30] proposed a novel medical image fusion method based on RGF, and the method cannot only remain edge information but also maintain the energy of the source image.

In this paper, to resolve the synchronization problem of the fusion and denoising, we propose a medical image fusion algorithm based on improved RGF (IRGF), in which, in order to better extract the texture and edges of input images, the edge recovery process of images are repeatedly implemented using Wiener filter (WF) [31] to get the smoother images. Finally, the filtered outputs by IRGF and WF, base layers are fused by visual saliency map (VSM) based fusion rule, and the corresponding detail layers are fused by the weighted least squares (WLS) optimization scheme.

The main contributions of this paper are as follows:

(i) RGF is successfully modified to IRGF and used for brain image fusion based

(a) on multiscale edge-preserving decomposition

(ii) a novel VSM based image fusion rule is introduced for base layers, which can improve the contrast and overall details of the resultant image

(iii) a WLS optimization scheme is introduced to merge the corresponding detail layers, thus the fused image appears more natural

The remaining subsections of this paper are organized as follows. In Section 2, we give a brief review of WF and IRGF. In Section 3, we will explain in detail the overall framework and fusion strategies. Qualitative and quantitative evaluation of experimental results are given in Section 4. Lastly, in Section 5, we conclude our work.

2. Related Works

2.1. Wiener Filter. The famous mathematician Norbert Wiener proposed a new linear filter in the 1940s, namely the Wiener filter (WF) [31]. This filter is also known as least mean squared error filter or least squares error filter. Conceptually, WF is an adaptive filter with the smallest mean squared error, and the output of the filter can be adjusted according to the local variance of input image. What is more, the larger the local variance, the stronger the smoothing effect of the filter [32–34]. Generally, WF is an adaptive filter that can calculate the mean and variance of a neighborhood, and it is also a low pass filter which can be applied to different contexts to restore noise-degraded signals. Thus, the optimal WF can be considered as one of the most fundamental noise reduction approaches, which has been confirmed in different forms and adopted in various applications. The definition of WF is defined as follows:

$$Wiener(I) = \frac{Boxfilter(I)}{N} + q * \left(I - \frac{Boxfilter\left(I^2\right)}{N} \right), \quad (1)$$

where Boxfilter($\cdot$) is the filter for obtaining the difference of the matrix on the X and Y axes. N is the matrix obtained by putting an matrix of ones into Boxfilter($\cdot$). I is the input image matrix, and I^2 is the square of the image matrix. q is the weight coefficient, which is defined as follows:

$$q = \frac{m}{m + eps}, \quad (2)$$

$$m = \frac{Boxfilter\left(I^2\right)}{N} - \frac{Boxfilter(I)}{N} * \frac{Boxfilter(I)}{N}. \quad (3)$$

where eps is the hyperparameter, and the value is set to 0.05.

2.2. Improved Rolling Guidance Filter. For the last decade, edge-preserving filters such as joint bilateral filter [35], guided filter [36], and weighted least squares filter [37]. have been widely applied in image fusion. Nevertheless, most of the conventional edge preserving filters rarely consider their spatial scales while smoothing small-scale structures of images. In order to preserve edge contents while smoothing image structures according to their scales, Zhang et al. [27] proposed a scale-aware edge-preserving filter in 2014, i.e., RGF, which includes both scale-aware and edge-preserving properties for image decomposition.

In RGF, a Gaussian filter is used to remove small structures, and a guided filter is used for edge recovery. In order to better remove small structures and highlight structure and edge contents of input images, in this work, we will improve the filtering way of the traditional RGF. The implementation process is modified to remove small structures through WF instead of Gaussian filter. Similar to the typical GRF, IRGF is also composed of two main steps, namely, small structure removal and edge recovery.

Firstly, small structure removal, we use a WF to smooth the input image, resulting in a smooth image without image details. Mathematically, the filtered image G from the input image I can be expressed as

$$G = WienerFilter (I, r, \lambda, N), \quad (4)$$

where G is the smoothed image; WienerFilter ($\bullet$) is a Wiener filtering operation, and I denotes the input image. r is the radius of template in the filter. In this paper, we set it to 3; λ is the regularization parameter; we set it to 0.01; N is calculated with a box filter for the size of each local patch.

Secondly, edge recovery, this step is iteratively implemented by using a joint filter, in which the guided filter is used as a joint filter to recover the edge details. It is an iterative process, i.e., the image recovered; H^t is iteratively updated; H^1 is the image smoothed by the WF. The H^{t+1} is defined as follows:

$$H^{t+1} = GuidedFilter \left(H^t, I, \sigma_s, {\sigma_r}^2 \right). \quad (5)$$

Among them, GuidedFilter($\cdot$) is the Guided filtering operation; H^t is the smoothed image of the t-th iteration, and t is set as 4; I is the input source image, σ_s is the standard deviation served as the scale parameter, and σ_r^2 is a regularization parameter which controls the range weight.

In general, IRGF is actually realized through the above two steps, and it can be simply written as

$$U = IRGF(I, \sigma_S, \sigma_r, T), \quad (6)$$

where U represents the filtering output, and T is the iteration number which is set to 4 in this work.

In order to verify the superiority of the IRGF over RGF, we conduct the confirmatory experiment on two typical MRI and PET/SPECT image pairs, e.g. MRI-PET and MRI-SPECT fusions. For a fair comparison, we conduct the experiment using the same iterations (4 times) by applying IRGF and RGF. Figure 1 provides a comparative example of the fused results produced by the IRGF and RGF based methods. For more intuitive comparison, some close ups are enlarged in the red regions and shown in Figure 1. By

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)
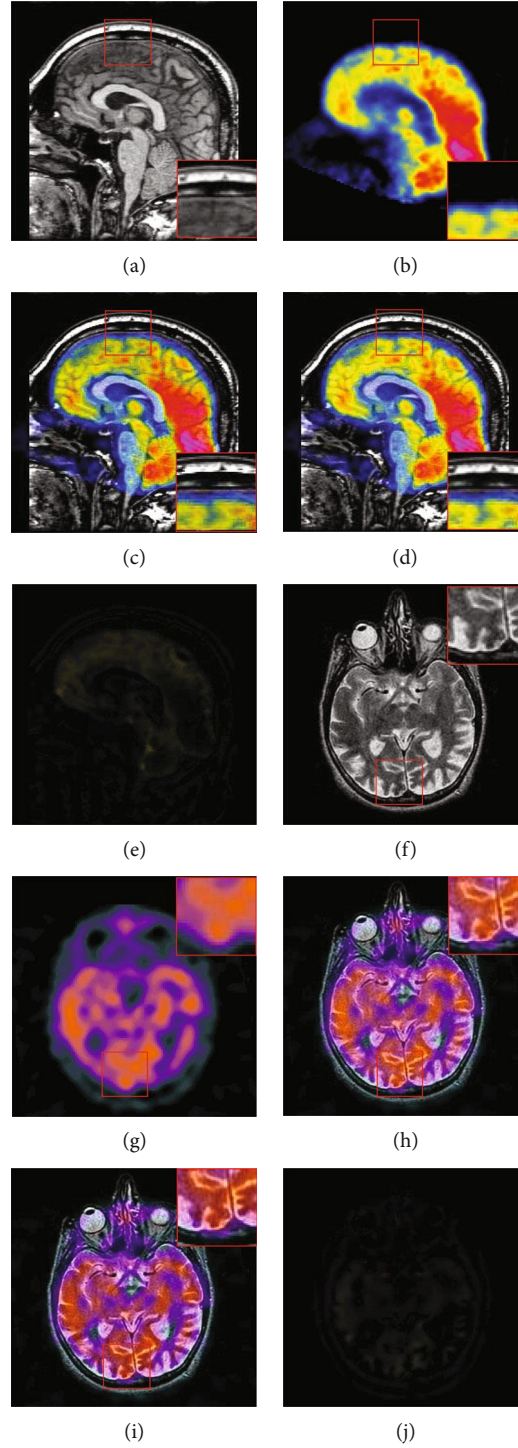
(i)   (j)

FIGURE 1: Demonstration of the fused results obtained by RGF and IRGF, (a) and (b) and (f) and (g) are the source images. (c) and (h) are the fused results produced by applying RGF. (d) and (i) are the fused results produced by applying IRGF. (e) and (j) are the difference between the fused results produced by applying IRGF and RGF method.

carefully observing the magnified version of the red regions marked of Figure 1, for the fused results produced by IRGF based method, we can see that the color information in the PET and SPECT images can better preserved and the edges are enhanced by considering the structural details of the MRI image. This experiment shows that, by applying IRGF, small-scale structures of an image are successfully removed according to their scales, while large-scale edges will be well retained. Simultaneously, using the WF to smooth the input image, all the edges of the filtered images are more blurred, which will be helpful to highlight large-scale edge details in the subsequent edge recovery.

2.3. Multiscale Decomposition Based on IRGF and WF. Accordingly, the proposed MSD based on IRGF and WF is formulated as follows:

$$u^j = IRGF\left(u^{j-1}, \sigma_S^{j-1}, \sigma_r, T\right), (j = 1, \cdots, n-1), \quad (7)$$

$$d^j = u^{j-1} - u^j, (j = 1, \cdots, n-1), \quad (8)$$

$$d^j = WienerFilter\left(u^{j-1}, r, \lambda, n\right), (j = n), \quad (9)$$

$$d^j = u^{j-1} - u^j, (j = n), \quad (10)$$

where u^j is the *j-th* level filtered image; d^j denotes the *j-th* level detail layer, and n is the number of decomposition levels. The initial image u^0 is the source image I.

Figure 2 provides the outputs obtained by repeatedly smoothing MRI image using IRGF, respectively, and the differences between adjacent images. It can be seen that, as shown in Figure 2, IRGF can effectively remove small-scale edges of input image while retain large-scale contents, and thus it can adequately extract detail features of the source images and preserve them well.

3. Proposed Method

The schematic diagram of the proposed fusion framework is shown in Figure 3. Firstly, taking for example the fusion of MRI and SPECT images, the source MRI image and the gray image obtained from SPECT by transformation are decomposed into base layer images I_b^{MRI} and I_b^{SPECT} and detail layer images I_d^{MRI} and I_d^{SPECT} using the IRGF an WF. Then, the base layers are merged based on VSM [38–40], namely, VSM_{MRI} and VSM_{SPECT}, in terms of weighted average scheme aiming to obtain good contrast and more natural look in the final fused image. Next, the fusion of detail layers are conducted through the weighted least squares optimization scheme [28], in which the detail layers at different scales are defined by max-absolute strategy and subsequently smoothed by Gaussian filter.

After obtaining the fused subimages of the base layers and detail layers, the fused image I^F will be reconstructed by combing the fused base layer I_B^F and the fused detail layers I_D^F. Lastly, the resultant image is reconstructed through the inverse transformation.

3.1. Fusion of Base Layers. In this section, we adopt a novel VSM-based weighted average scheme to fuse the base layer images. This technique has been widely used in many computer vision and computer graphics applications. For an image, the *VSM* of which can display perceptually salient visual structures or objects highlighting from their neighbors. As a result, the *VSMs* are usually used to describe the activity level measurement of source images for fusion tasks. In view of the fact that the *VSM* can reflect salient image features, we introduce the VSM to fuse the base layers, the calculation of which refers to the literatures [28, 38, 39]. Mathematically, it is defined as follows:

$$VSM(I = i) = Sal(i + 1), i \in [0, 255], \quad (11)$$

$$Sal(j + 1) = \sum_{i=0}^{255} Sal(j + 1) + hist(i) * |j - i|, j \in [0, 255], \quad (12)$$

where I denotes the input image; $Sal(\cdot)$ is the salient pixel value of j; $hist(i)$ is the *i*-th pixel value of the histogram computed by input image.

After the VSM_A and VSM_B are obtained, the base layer images and corresponding *VSMs* are performed by applying weighted average strategy, which is described as follows:

$$I_B^F = \left(\frac{1}{2} + \frac{VSM_A - VSM_B}{2}\right) * I_b^A + \left(\frac{1}{2} + \frac{VSM_B - VSM_A}{2}\right) * I_b^B, \quad (13)$$

where I_b^A and I_b^B represent the base layers decomposed by input MRI/CT and SPECT/PET images, while VSM_A and VSM_B denote the *VSMs* of input MRI/CT and SPECT/PET images, respectively.

3.2. Fusion of Detail Layers. Basically, for the resultant image of MRI-SPECT/MRI-PET, the texture details mainly comes from MRI image while brightness information principally inherits SPECT/PET image. Since the characteristics that MRI/CT provides information with high-resolution soft tissue and SPECT/PET contains functional information with high intensity, it easily leads to loss of visually details and reduces the brightness of the fused image by using only the absolute maximum scheme. Therefore, it is necessary to make an improvement on the fusion rule which is usually used in the traditional method. With the above considerations, the WLS optimization scheme [41] is further introduced to fuse the detail layers, which is suitable for human visual perception. Firstly, the expression of the absolute maximum selection is as follows:

$$I_{abs}^{\max} = \begin{cases} 1, & \left|d_A^j < d_B^j\right| \\ 0, & otherwise \end{cases} (j = 1, 2, \cdots, n), \quad (14)$$

where d_A^j and d_b^j are the *j-th* level detail layers decomposed by input MRI/CT and SPECT/PET images, respectively.

To make the fused image look more naturally, we use a Gaussian filter to smooth the image obtained by Equation (14). The Gaussian filtering is defined as follows:

$$G = Gaussion(I_{abs}^{\max}, w), \quad (15)$$

where $Gaussion(\cdot)$ is the Gaussian filtering; w is the size of filter, which we set to 3 in this paper.
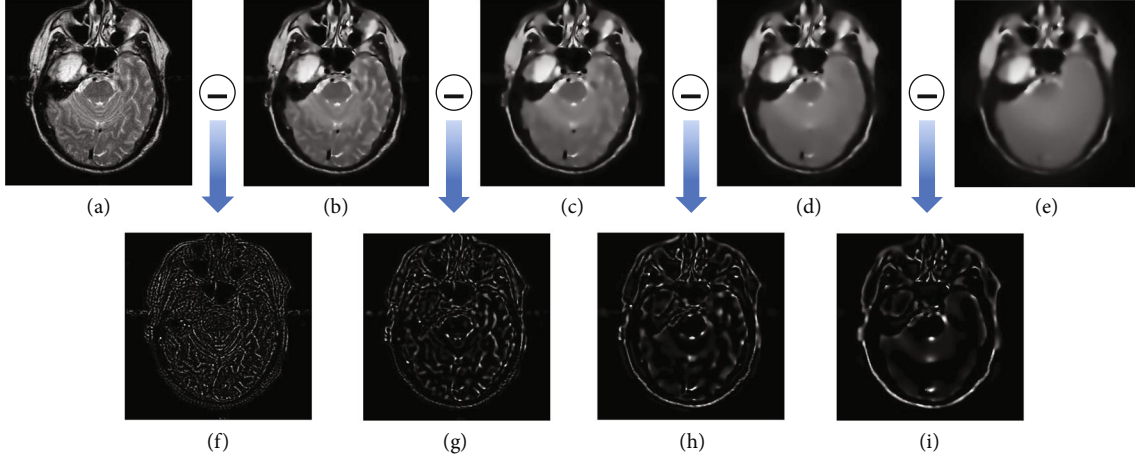
FIGURE 2: (a) MRI image. (b), (c), and (c) and (e) are the output results obtained by 4 iterations of IRGF, respectively. (f), (g), (h), and (i) are the difference between (a) and (b), (b) and (c), (c) and (d), and (d) and (e). Notice that the data of (f), (g), (h), and (i) are multiplied by 5.
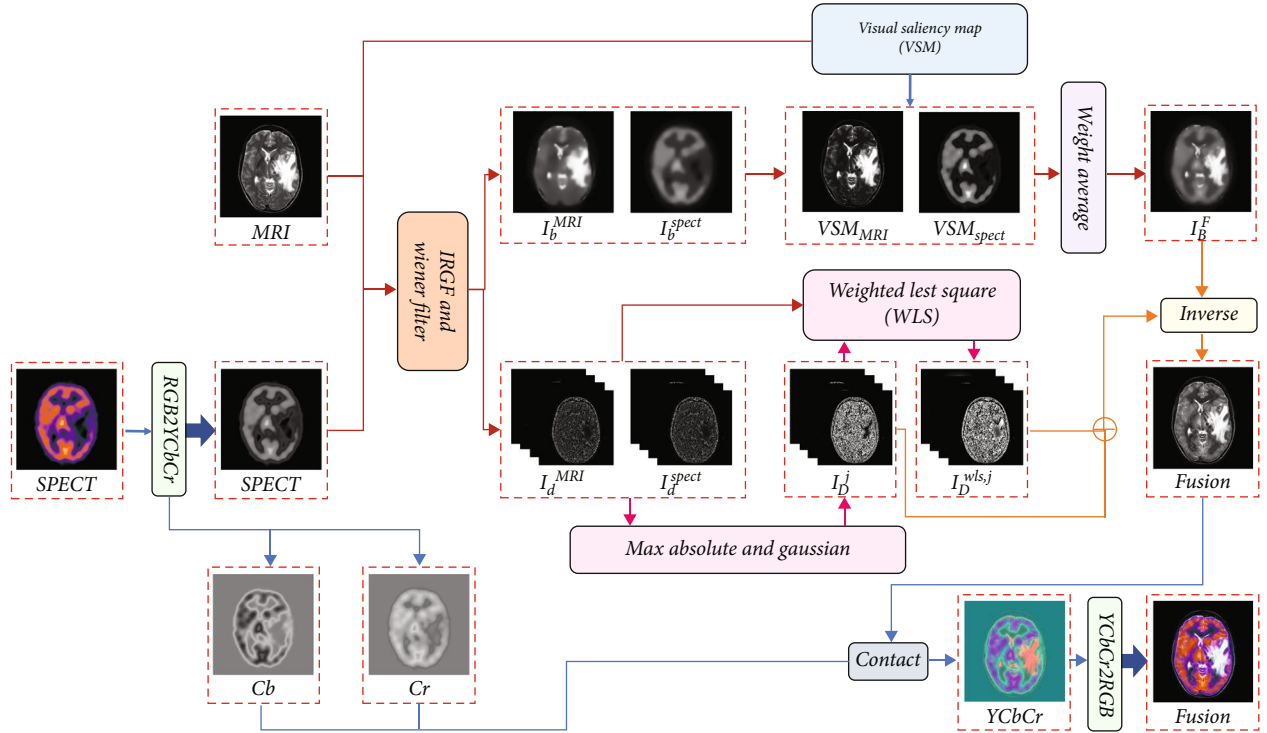

FIGURE 3: The schematic diagram of the proposed method.

Then, we figure out the detail layer images, the mathematical expression of which is described as follows:

$$I_D^j = (1 - G)I_d^{1,j} + G \times I_d^{2,j}. \tag{16}$$

Next, we use WLS to optimize the detail layer images obtained by Equation (16), and the mathematical expression is as follows:

$$I_D^{wls,j} = WLS\left(I_h^j, I_d^{1,j}, I_d^{2,j}, \lambda\right), \tag{17}$$

where $WLS(\cdot)$ is the calculation of weighted least squares value; I_d is a preliminary high-frequency image; λ is a hyperparameter, and its value is set to 0.01.

Finally, we will obtain the final detail layer image, which is expressed as follows:

$$I_D^{F,j} = I_D^j + I_D^{wls,j}, j = 1, 2, \cdots, n. \tag{18}$$

Finally, the fused image I^F can be reconstructed by combing the fused base layer I_B^F and the fused detail layers

(a) MRI   (b) CT

(c) CBF   (d) CNN

(e) ConvSR   (f) DenseFuse

(g) VggML   (h) EMFusion

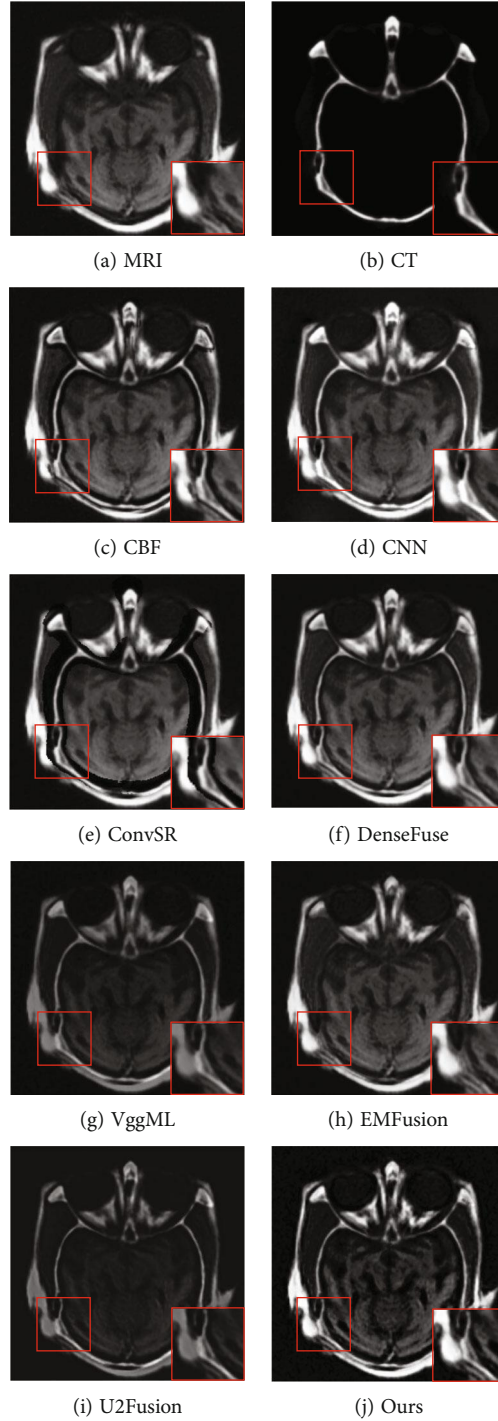(i) U2Fusion   (j) Ours

FIGURE 4: Visual comparison of our method with seven state-of-the-art methods on the first pair of MRI and CT images.

I_D^F as follows:

$$I^F = I_B^F + I_D^{F,j}, j = 1, 2, \cdots, n. \qquad (19)$$

4. Experimental Results and Analysis

The comparison experiments are conducted qualitatively and quantitatively on the publicly available dataset [42], which is a database website of the Harvard Medical School containing rich medical source images. Moreover, the runtime of different methods on three medical fusion tasks are provided and compared in the last experimental analysis.

4.1. Data Sets and Experimental Environment. To verify the superiority of the proposed method, six pairs of different types of brain images are used in our experiments, including multimodal images (CT-MRI, MRI-SPECT and MRI-PET). Furthermore, 30 pairs of MRI-SPECT images and 30 sets of MRI-PET images are selected to perform the experiment

(a) MRI

(b) CT

(c) CBF

(d) CNN

(e) ConvSR

(f) DenseFuse

(g) VggML

(h) EMFusion
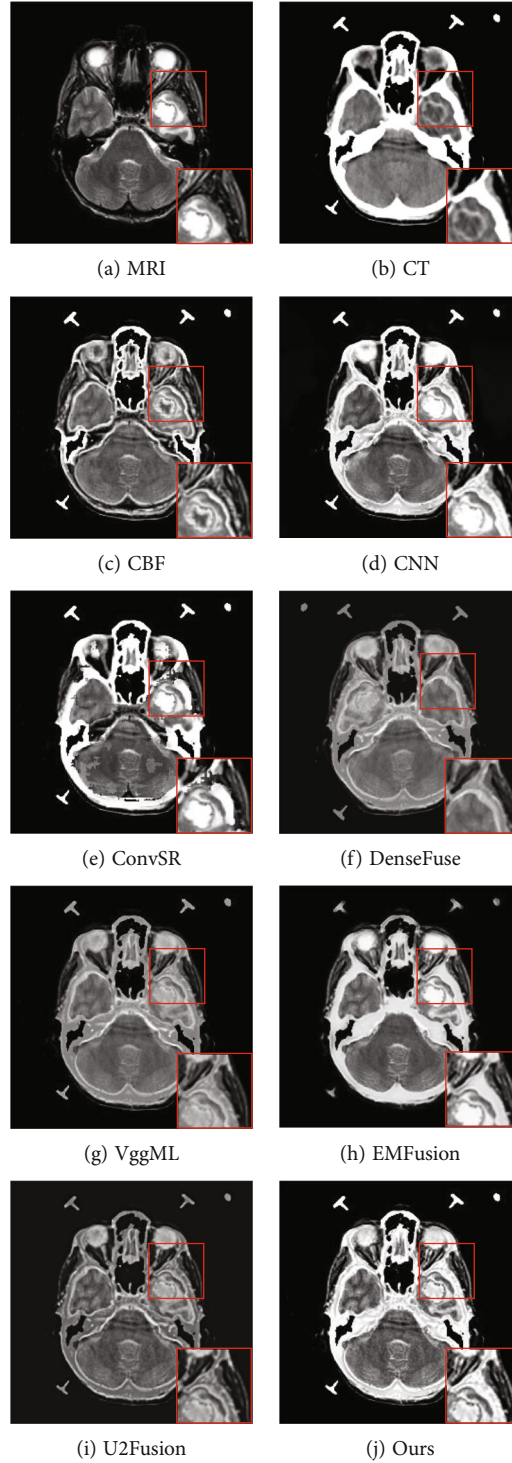
(i) U2Fusion

(j) Ours

FIGURE 5: Visual comparison of our method with seven state-of-the-art methods on the second pair of MRI and CT images.

for quantitative comparison. All the test image pairs have the spatial resolution of 256×256 pixels, and they can be downloaded from http://www.med.harvard.edu/aanlib/home.html. This experiment is implemented on both AMD Radeon (TM) Vega 8 Graphics and 1.60 GHz Intel Core i5-8250 CPU using MATLAB.

4.2. Comparison Methods. In the experiments, we compared the proposed method with seven state-of-the-art fusion methods. These methods are as follows: fusion algorithm based on cross-bilateral filtering (CBF) [41], fusion algorithm based on neural networks (CNN) [43], image fusion based on sparse representation (ConvSR) [22], fusion algorithm based on dense

(a) MRI

(b) SPECT

(c) CBF

(d) CNN

(e) ConvSR

(f) DenseFuse

(g) VggML

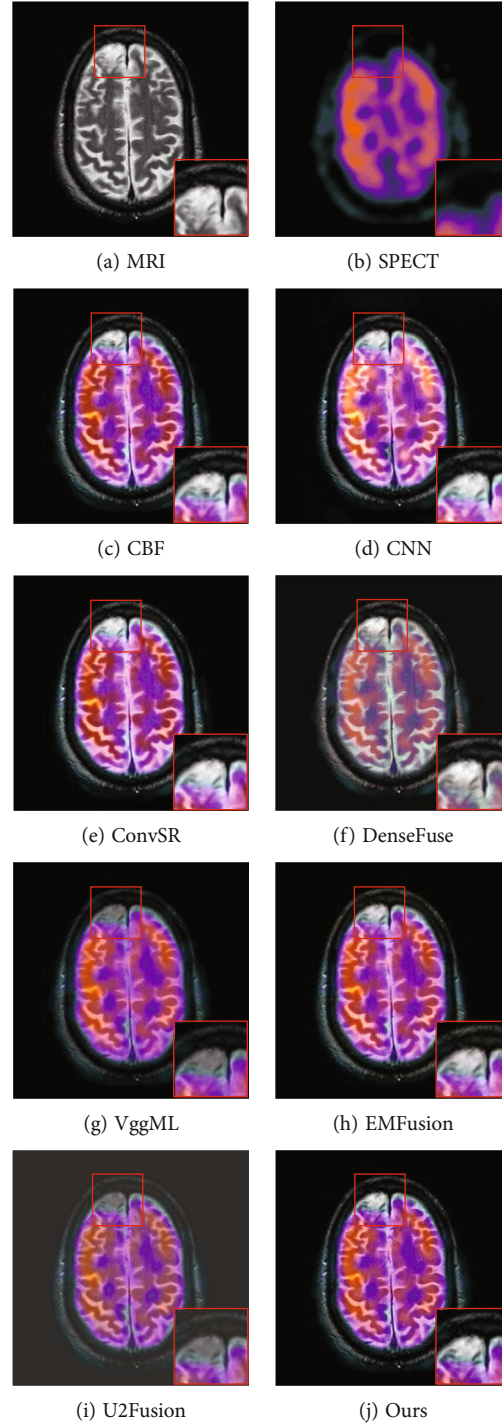(h) EMFusion

(i) U2Fusion

(j) Ours

FIGURE 6: Visual comparison of our method with seven state-of-the-art methods on the first pair of MRI and SPECT images.

network blocks and self-encoded structures (DenseFuse) [44], fusion Algorithm using a deep learning framework (VggML) [45], medical image fusion method based on unsupervised deep learning framework (EMFusion) [46], and a unified unsupervised image fusion network (U2Fusion) [47]. The source codes of these methods are all provided by their original authors, and all the parameters in the experiments are set as the default values.

4.3. Evaluation Metrics. For quantitative comparison, we select six metrics to evaluate the performance of the fused images, including spatial frequency (SF) [48], structural similarity (SSIM) [49], clarity measurement (CF) [50], extended spatial frequency (ESF) [51], average gradient (AG) [52] and visual information fidelity (VIF) [53]. Among them, SF, ESF, and CF are all image quality index that can effectively describe spatial features of an image, which indicates the

(a) MRI

(b) SPECT

(c) CBF

(d) CNN

(e) ConvSR

(f) DenseFuse

(g) VggML

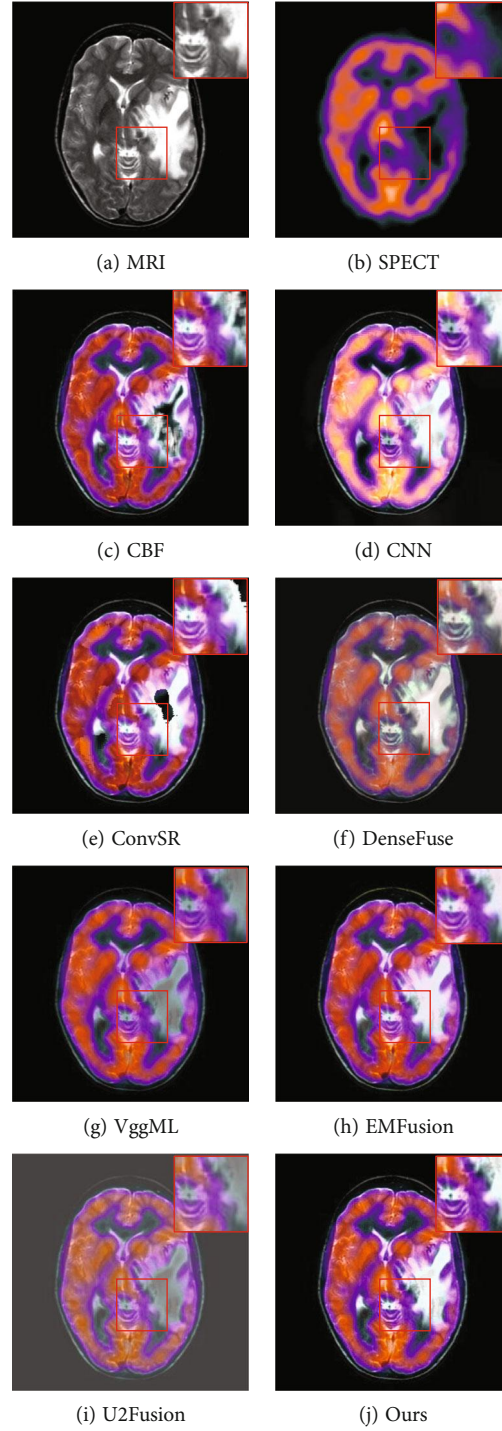(h) EMFusion

(i) U2Fusion

(j) Ours

FIGURE 7: Visual comparison of our method with seven state-of-the-art methods on the second pair of MRI and SPECT images.

overall clarity level of an image. SSIM measures the structural similarities between source images and the fused image. AG is also used to measure the clarity of the image and further analyze the texture details of the final image. VIF evaluates the percentages of visual information which is reserved between the source images and fused image. More detailed information can be provided in related references.

4.4. Parameters Analysis and Setting. It is necessary to investigate three free parameters in the proposed fusion mothod, namely, the hyperparameter λ of WF, the standard deviation σ_s, and the regularization parameter σ_r of GuidedFilter in Equation (5). In Equation (4), the hyperparameter λ with the value range from 0.01 to 1 is used to control the trade-off between the saliency and abundance measurement

(a) MRI

(b) PET

(c) CBF

(d) CNN

(e) ConvSR

(f) DenseFuse

(g) VggML

(h) EMFusion
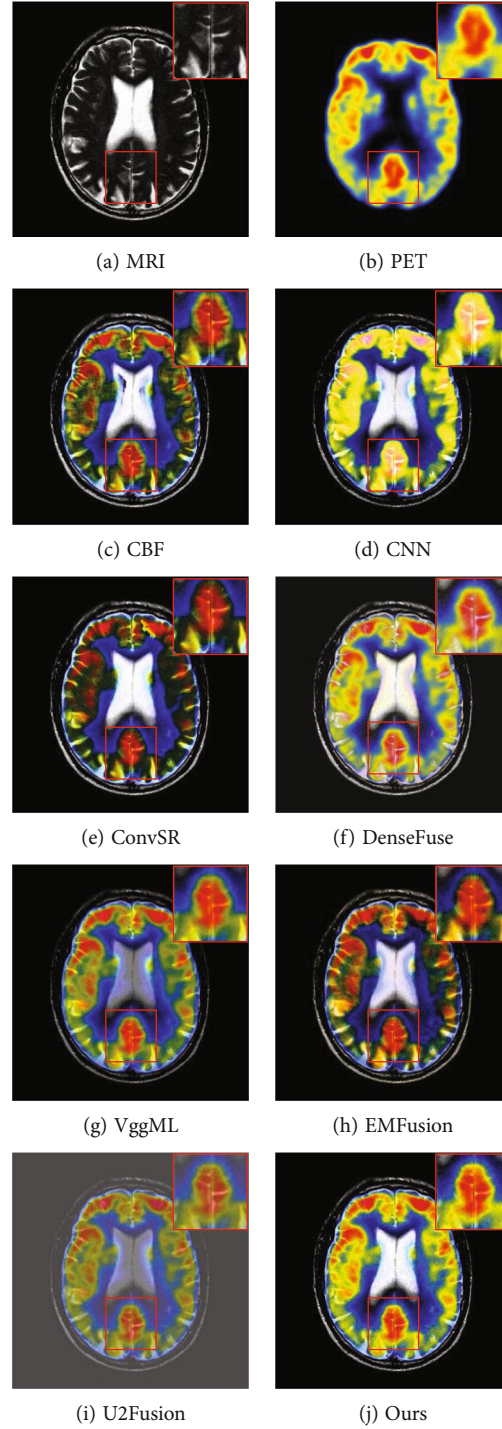
(i) U2Fusion

(j) Ours

FIGURE 8: Visual comparison of our method with seven state-of-the-art methods on the first pair of MRI and PET images.

results. When $\lambda = 0.01$, the fused results cannot only preserve the abundant details in MRI image but also retain more color information in the SPECT/PET image. When σ_s increases, the regions with high pixel intensities are not preserved more completely. As a result, we set it to 0.01. For the parameter σ_s with the its range from 1 to 10 and the parameter σ_r with the range from 0.05 to 2, the experiment demonstrates that preserve both abundant details and color

information at the same time when their values are set as 2 and 0.05. As a general rule, when we analyze one of the parameters, the other one will be fixed. Similar ways of discussing the impacts of multiple parameters have been widely adopted in some publications [36, 54, 55].

4.5. Visual Analysis. Figure 4 shows the MRI and CT images fused results produced by different methods. From these

(a) MRI   (b) PET

(c) CBF   (d) CNN

(e) ConvSR   (f) DenseFuse

(g) VggML   (h) EMFusion
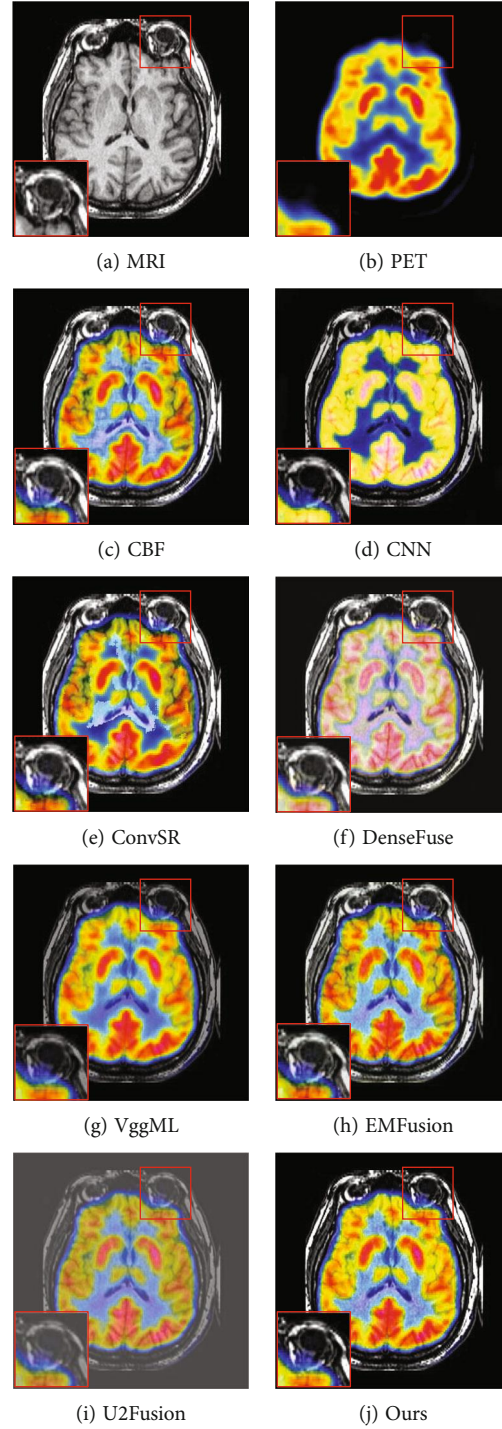
(i) U2Fusion   (j) Ours

FIGURE 9: Visual comparison of our method with seven state-of-the-art methods on the second pair of MRI and PET images.

fusion results, we can easily see that the proposed scheme outperformed the other fusion methods, and the fused image has the best visual representation of both bone structure and soft tissue information for all eight methods. The fusion results of the other seven methods are shown in Figures 4(c–i), it can be seen that fused results of the VggML and U2Fusion methods not only have low contrast but also cannot clearly show the tissue characteristics of source MRI image. The fusion results produced by the CNN and EMFusion methods suffer from the oversaturated problem of intensity (see Figures 4(d) and 4(h)). Although the fusion images generated by the CBF, ConvSR and DenseFuse methods have better visual quality, there are artifacts at the edges of these images. Furthermore, by observing the red

TABLE 1: Quantitative evaluation for 30 pairs of MRI-CT fused images (bold: optimal).

| Metrics	CBF	CNN	ConvSR	DenseFuse	VggML	EMFusion	U2Fusion	Ours
SF	34.22894	32.43961	**37.26421**	17.48996	20.15031	21.05623	18.64274	34.49011
SSIM	0.80156	**0.94015**	0.83193	0.31859	0.87559	0.88552	0.85524	0.92976
CF	9.69240	8.34933	8.84954	5.08303	5.38481	6.16628	6.10750	**9.78925**
ESF	0.03506	0.10140	0.04621	**0.50830**	0.43985	0.39173	0.46486	0.50422
AG	**8.30717**	7.05477	7.60956	4.27056	4.54141	5.32467	5.20886	8.03297
VIF	0.26443	0.45139	0.35529	0.05059	0.37849	0.40803	0.29462	**0.47872**
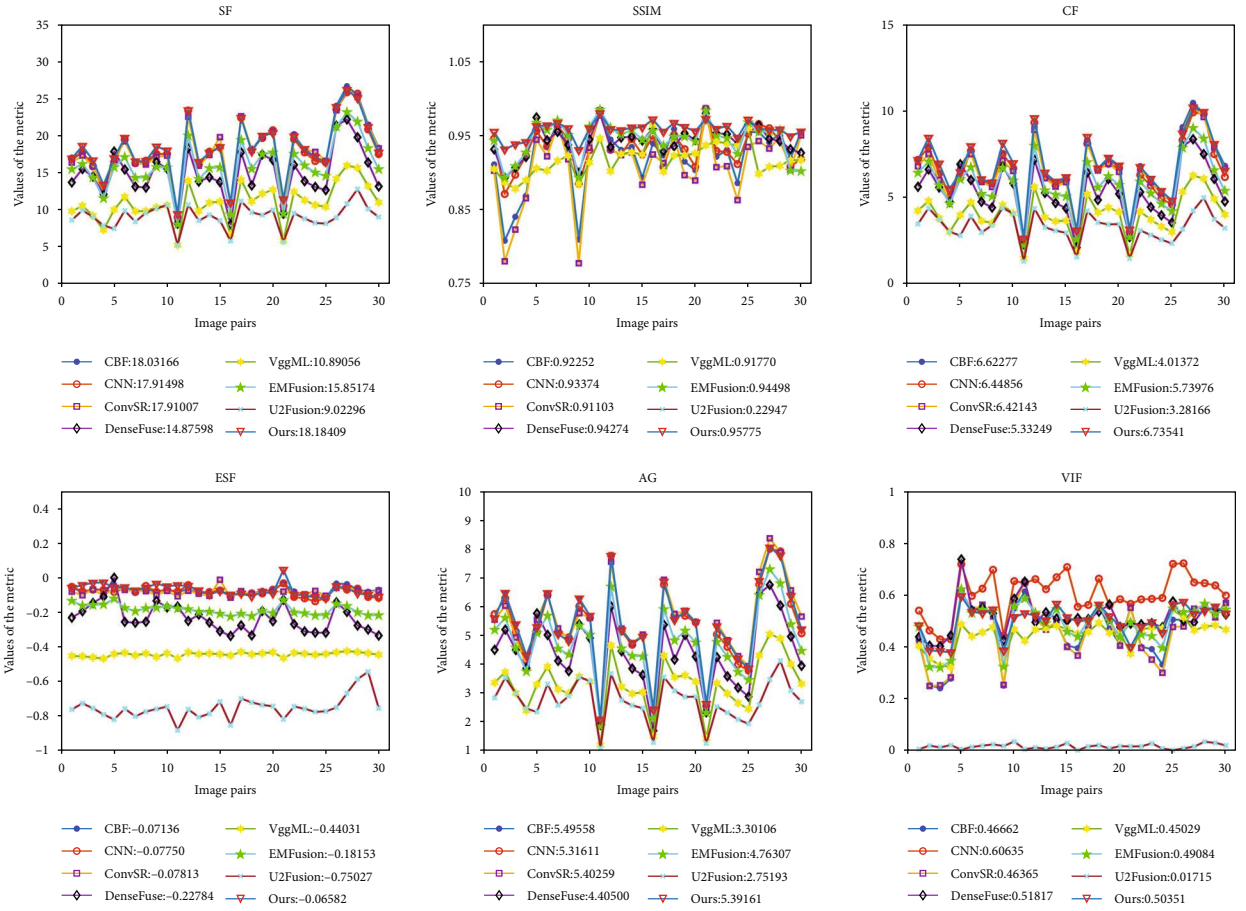
FIGURE 10: Quantitative comparisons of six metrics, i.e., SF, SSIM, CF, ESF, AG, and VIF, on 30 test image pairs of MRI-SPECT images.

box regions, the proposed method introduces fewer artifacts and offers superior image quality in terms of resolution, brightness, and contrast. Figures 5(c–i) provide qualitative comparison results of the proposed method with seven competitors on the second pair of MRI and CT image fusion. It can be clearly seen that the final results on this pair of images are similar with the descriptions in Figure 4.

Figures 6 and 7 display the fused images of two pairs for MRI-SPECT image fusion using eight methods, respectively. It can be seen from Figure 6 that DenseFuse method produces low contrast and the color information is obviously distorted in the result image. The fusion image obtained by the U2Fusion method is low brightness and clarity.

Although the VggML method gets better fusion performance, the contrast of the fused image is still not high, which may be caused by the average fusion rule. The CNN method introduces color distortion to a large extent. What is more, a lot of texture information in MRI image is lost, and there exists an overexposure phenomenon in the result of CNN. In contrast, the CBS, ConvSR, and EMFusion and our method have better visual effects and retain rich texture information in the source MRI image. The fused results of our method clearly have better contrast. Furthermore, compared with the red box regions which are enlarged, our method shows the optimal result in terms of color information, the contrast and texture details.
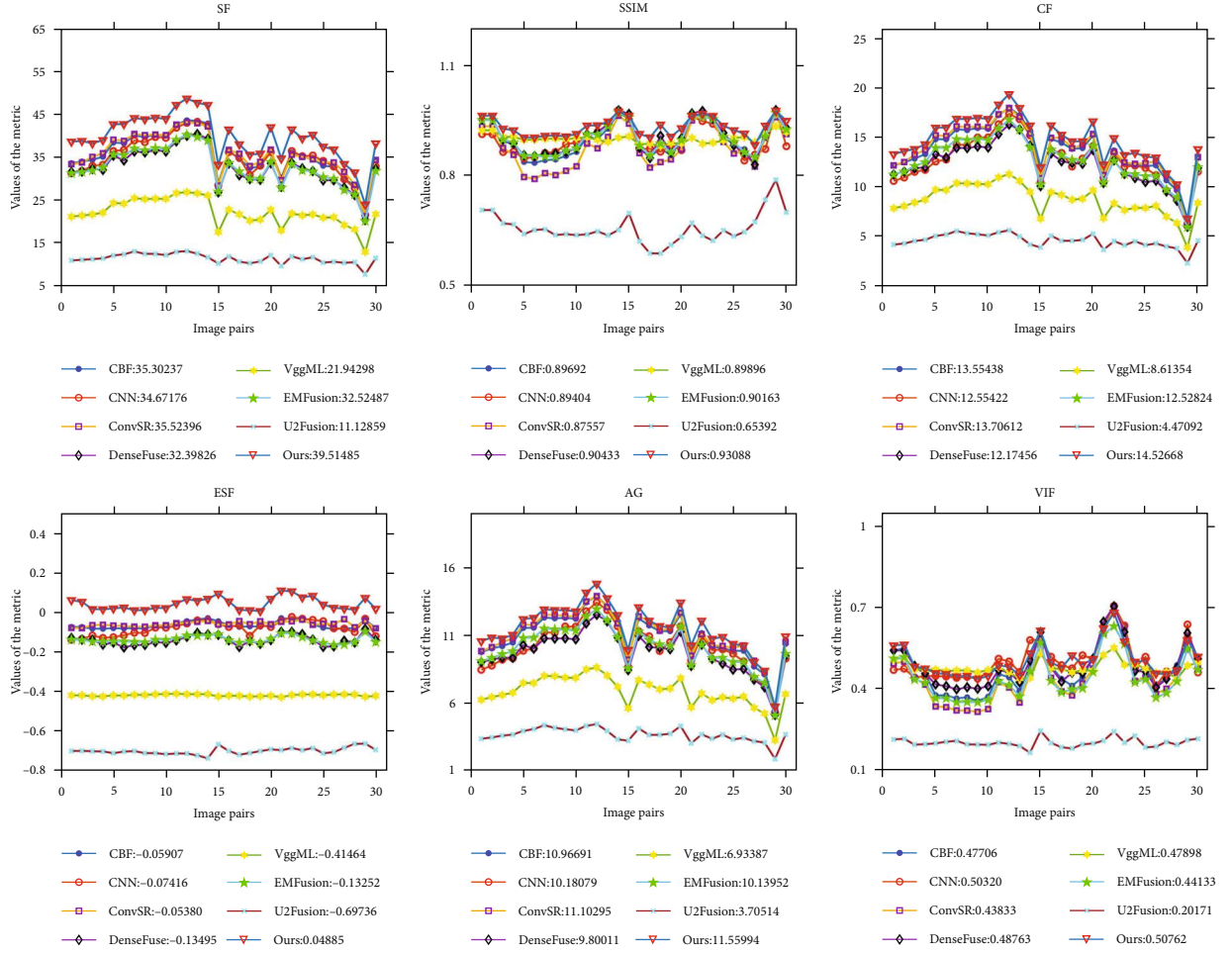
FIGURE 11: Quantitative comparisons of six metrics, i.e., SF, SSIM, CF, ESF, AG, and VIF, on 30 test image pairs of MRI-PET images.

TABLE 2: Runtime comparison of different methods on three types of medical fusion tasks.

Method/type	CBF	CNN	ConvSR	DenseFuse	VggML	EMFusion	U2Fusion	Ours
MRI-CT	2.67	16.52	36.49	0.64	1.89	0.56	2.02	0.30
MRI-SPECT	3.73	17.57	30.87	4.09	4.02	0.42	1.82	0.30
MRI-PET	3.73	17.57	30.87	4.09	4.02	0.45	1.67	0.38

For MRI-PET images, we randomly select two pairs of images to show the effectiveness of the proposed method. As shown in Figures 8 and 9, the fused images of the CBF, ConvSR, and EMFusion decrease the brightness and contrast of the source images. Besides, the color information in the PET image is also distorted in the results generated by the aforementioned methods. The result obtained by CNN is overexposed, which seems to indicate that this method is not suitable for color medical images. As observed from Figure 8(d) and Figure 9(d), the edges in the MRI image are weakened to some extent in some competitors. For the U2Fusion method, both the background and the target area of the fused image appear dim, and the contrast is too low (see Figure 8(i) and Figure 9(i)). The DenseFuse method produces serious color distortion, which leads to the loss of a lot of functional information in PET image.

Although the VggML method achieves relatively better visual effects, it still loses some details such as edges and textures. By comparison, the fusion results of our method are better than the others, in which the color information and texture details of the PET and MRI image are well enhanced and preserved. Besides, by comparing the enlarged regions marked of Figures 8–9, correspondingly. It can be clearly seen that the proposed method not only contains color information from the input PET image but also captures the texture details of the source MRI image.

4.6. Quantitative Analysis. In this work, six metrics are used to synthetically evaluate the performance of different fusion methods, which are, respectively, six metrics, i.e., SF, SSIM, CF, ESF, AG, and VIF. In theory, the bigger these metrics, the better the fusion result is. For better evaluation, three test

datasets including 30 pairs of MRI-CT images, 30 pairs of MRI-SPECT images, and 30 pairs of MRI-PET images are selected to conduct experiments. The quantitative results obtained by eight different fusion methods on six metrics are shown in Table 1, Figures 10 and 11, respectively. Among them, the plotting for data visualization of the MRI-SPECT and MRI-PET fused results are shown in Figures 10 and 11. Note that the quantitative assessment values produced by different methods on six metrics are the average runtime in two graphs.

From Table 1, we notice that the values of CBF, CNN, ConvSR, and DenseFuse methods are slightly larger than ours in AG, SSIM, SF, and ESF evaluation metrics for MRT-CT fusion problems, but the fused results of the proposed method achieves the largest value in CF and VIF evaluation metrics. Furthermore, the values of our method, respectively, ranks second or third in SF, SSIM, ESF, and AG evaluation metrics. Overall, our algorithm achieved the excellent fusion performance for the MRI-CT fusion problems. Simultaneously, as shown in the quantitative results in Figure 10, we can observe that CBF is optimal on AG and suboptimal on SF, CF, and ESF. CNN is the third best on VIF, SF, DF, and ESF. ConvSR is suboptimal on AG. DenseFuse is suboptimal on VIF and ranks third on SSIM. EMFusion is suboptimal on MS-SSIM. Neither VggML nor U2Fusion has a good score on six metrics. Our results achieve the optimal results on SF, SSIM, DF, and ESF, with AG and VIF ranked third. The fused results on six metrics show that our method can well preserve the more valuable texture details in MRI/CT images and inherit the more color and brightness information in SPECT/PET images.

Similarly, the data visualization of the measurement values of MRI-PET image fusion by seven typical fusion algorithms and our methods is shown in Figure 11. It can be seen that our method achieves the optimal result on SF, ESF, AD, and CF and the pretty good results on SSIM and VIF. It shows that our fused result has comparative similarity with source images with less lost information. For the rest seven competitors, CBF ranks third on SF, CF, ESF, and AG. CNN is suboptimal on VIF. ConvSR is suboptimal on CF, ESF, and AG. DenseFuse is suboptimal on SSIM, the third on VIF. EMFusion is the second best on SF, and SSIM ranks third. Neither VggML nor U2Fusion achieves good results on six metrics. The quantitative comparison results indicate that our results can well retain the structural information in MRI images and the functional information in PET images by keeping high similarities with source images. Thus, it outperforms the other seven methods.

4.7. Execution Efficiency Comparison. The runtime comparison of different methods on three test datasets is reported in Table. It should be noted that the values in Table are the average execution time of different methods. As shown in Table 2, we can intuitively observe that the execution time of ConSR is the longest of all methods, and the execution time of CNN is also relatively long, slightly shorter than ConSR. By contrast, our fusion algorithm takes the shortest time and shows the optimal computational efficiency.

5. Conclusion

In this paper, we present an improved multimodal brain image fusion method based on multilevel edge-preserving filtering. Aiming to achieve a better scale separation while preserving information of image edges and reducing halos, the proposed method utilize the IRGF and Wiener filter to decompose source images into base and detail layers. Considering the conventional average fusion rule does not take full advantage of the residual information of base layers, the VSM-based weighted average strategy is referred to fuse base layers, which facilitates to achieve better overall appearance and high contrast in the final image. For the fusion rule of the detail layers, a novel WLS optimization scheme is introduced, which can overcome the limitation of conventional absolute maximum scheme by taking fully into account the characteristics of different modal medical images. In the meantime, this optimization process can effectively control the trade-off between the saliency and abundance measurement results of source images, and thus useful visual information and detail features can be better extracted and transferred into the resultant image while suppressing the noise from input images. Furthermore, the researches and applications of multimodal medical image fusion will be extended to the disease detection and diagnosis in the field of the precision medicine.

Data Availability

(1) All the test image pairs can be downloaded from the Whole Brain Atlas dataset; they are publicly available at http://www.med.harvard.edu/AANLIB/home.html. (2) The codes of our work are, respectively, provided in the submitted program code. (i) the main program, code; (ii) the related functions called by the main program, code; (iii) the code of three comparative examples (Figure 1, Figure 2, Figure 3) using RGF and IRGF decomposition, code; (iv) quantitative comparisons of spect_evaluate, code; (v) quantitative comparisons of pet_evaluate, code; (vi) runtime of fusion on 30 pairs of MRI-SPECT images and code; (vii) runtime of fusion on 30 pairs of MRI-PET images, which is explained in detail. (3) Six metrics, i.e., SF, SSIM, CF, ESF, AG, and VIF and the results for the comparison were obtained from the source codes provided by the authors. All the parameters in the experiments are set to the default values. (4) The experiments are implemented on both AMD Radeon (TM) Vega 8 Graphics and 1.60 GHz Intel Core i5-8250 CPU using MATLAB. (5) The resultant images obtained by our method on 30 pairs of MRI-SPECT images and 30 pairs of MRI-PET images are provided as a supplementary material in "Data Availability" statement. (6) The data used to support the findings of this study are currently protected. Requests for data, (6/12 months) after publication of this article, will be considered by the first author or corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: a survey," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[2] A. P. James and B. V. Dasarathy, "Medical image fusion: a survey of the state of the art," *Information fusion*, vol. 19, pp. 4–19, 2014.

[3] F. E. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egyptian Informatics Journal*, vol. 17, no. 1, pp. 99–124, 2016.

[4] H. Li, M. Yang, and Z. Yu, "Joint image fusion and super-resolution for enhanced visualization via semi- coupled discriminative dictionary learning and advantage embedding," *Neurocomputing*, vol. 422, pp. 62–84, 2021.

[5] Y. Yang, J. Wu, S. Huang, Y. Fang, P. Lin, and Y. Que, "Multi-modal medical image fusion based on fuzzy discrimination with structural patch decomposition," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1647–1660, 2019.

[6] 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.

[7] 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.

[8] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *Readings in computer vision*, vol. 31, no. 4, pp. 532–540, 1987.

[9] R. Singh and A. Khare, "Fusion of multimodal medical images using Daubechies complex wavelet transform - a multiresolution approach," *Information fusion*, vol. 19, no. 1, pp. 49–60, 2014.

[10] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[11] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: theory, design, and applications," *IEEE transactions on image processing*, vol. 15, no. 10, pp. 3089–3101, 2006.

[12] W. Q. Lim, "The discrete shearlet transform: a new directional transform and compactly supported shearlet frames," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1166–1180, 2010.

[13] G. Easley, D. Labate, and W. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.

[14] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019.

[15] L. Wang, B. Li, and L. Tian, "Multi-modal medical image fusion using the inter-scale and intra-scale dependencies between image shift-invariant shearlet coefficients," *Information fusion*, vol. 19, pp. 20–28, 2014.

[16] 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.

[17] 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.

[18] M. Kim, D. K. Han, and H. Ko, "Joint patch clustering-based dictionary learning for multimodal image fusion," *Information Fusion*, vol. 27, pp. 198–214, 2016.

[19] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low- rank sparse dictionaries learning," *Pattern Recognition*, vol. 79, pp. 130–146, 2018.

[20] 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.

[21] 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.

[22] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional 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] L. Yu, C. Xun, J. Cheng, and P. Hu, "A medical image fusion method based on convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)*, Xi'an, China, 2017.

[25] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image processing*, vol. 22, no. 7, pp. 2864–2875, 2013.

[26] J. Jeon, H. Lee, H. Kang, and S. Lee, "Scale-aware structure-preserving texture filtering," *Computer graphics forum*, vol. 35, no. 7, pp. 77–86, 2016.

[27] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *13th European Conference on Computer Vision, ECCV 2014*, pp. 815–830, Zurich, Switzerland, 2014.

[28] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017.

[29] L. Jian, X. Yang, Z. Zhou, K. Zhou, and K. Liu, "Multi-scale image fusion through rolling guidance filter," *Future Generation Computer Systems*, vol. 83, pp. 310–325, 2018.

[30] J. Chen, L. Zhang, L. Lu, Q. Li, M. Hu, and X. Yang, "A novel medical image fusion method based on rolling guidance filtering," *INTERNET OF THINGS*, vol. 14, no. 100172, 2021.

[31] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE transactions on pattern analysis and machine intelligence*, vol. 7, no. 2, pp. 165–177, 1985.

[32] W. Li, Y. Xie, H. Zhou, Y. Han, and K. Zhan, "Structure-aware image fusion," *Optik*, vol. 172, pp. 1–11, 2018.

[33] M. Izzetoglu, A. Devaraj, S. Bunce, and B. Onaral, "Motion artifact cancellation in NIR spectroscopy using Wiener filtering," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 934–938, 2005.

[34] M. Kazubek, "Wavelet domain image denoising by thresholding and Wiener filtering," *IEEE Signal Processing Letters*, vol. 10, no. 11, pp. 324–326, 2003.

[35] R. S. M. A. G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM transactions on graphics (TOG)*, vol. 23, no. 3, pp. 664–672, 2004.

[36] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[37] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5638–5653, 2014.

[38] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[39] R. Achantay, S. Hemamiz, F. Estraday, and S. Susstrunky, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, Miami, FL, United states, 2009.

[40] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *14th Annual ACM International Conference on Multimedia, MM 2006*, pp. 815–824, Santa Barbara, CA, United states, 2006.

[41] B. K. Shreyamsha Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1193–1204, 2015.

[42] "Whole brain atlas," https://www.med.harvard.edu/aanlib/.

[43] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 16, no. 3, p. 1850018, 2018.

[44] H. Li and X. Wu, "DenseFuse: a fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2019.

[45] H. Li, X. 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.

[46] H. Xu and J. Ma, "EMFusion: an unsupervised enhanced medical image fusion network," *Information Fusion*, vol. 76, pp. 177–186, 2021.

[47] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: a unified unsupervised image fusion network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 502–518, 2022.

[48] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[50] H. N. Wang, W. Zhong, J. Wang, and D. Xia, "Research of measurement for digital image definition," *Journal of Image and Graphics*, vol. 9, no. 7, pp. 828–831, 2004.

[51] Y. Zheng, E. A. Essock, B. C. Hansen, and A. M. Haun, "A new metric based on extended spatial frequency and its application to DWT based fusion algorithms," *Information Fusion*, vol. 8, no. 2, pp. 177–192, 2007.

[52] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Optics Communication*, vol. 341, pp. 199–209, 2015.

[53] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[54] K. He, S. Jian, and X. Tang, "Guided image filtering," in *European Conference on Computer Vision*, Heraklion, Crete, Greece, 2010.

[55] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Transactions on Image processing*, vol. 24, no. 1, pp. 120–129, 2015.