IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 18, 2020, accepted July 27, 2020, date of publication July 30, 2020, date of current version August 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013027

Two-Scale Multimodal Medical Image Fusion Based on Guided Filtering and Sparse Representation

CHUNYANG PEI[1], (Graduate Student Member, IEEE), KUANGANG FAN[1], (Member, IEEE), AND WENSHUAI WANG[2], (Student Member, IEEE)

[1]School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China
[2]School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China

Corresponding author: Kuangang Fan (kuangangfriend@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61763018, in part by the 03 Special Project and 5G Program of Science and Technology Department of Jiangxi Province under Grant 20193ABC03A058, in part by the Key Foundation of Education Committee of Jiangxi Province under Grant GJJ170493 and Grant GJJ190451, and in part by the Program of Qingjiang Excellent Young Talents of the Jiangxi University of Science and Technology.

ABSTRACT Medical image fusion techniques primarily integrate the complementary features of different medical images to acquire a single composite image with superior quality, reducing the uncertainty of lesion analysis. However, the simultaneous extraction of more salient features and less meaningless details from medical images by using multi-scale transform methods is a challenging task. This study presents a two-scale fusion framework for multimodal medical images to overcome the aforementioned limitation. In this framework, a guided filter is used to decompose source images into the base and detail layers to roughly separate the two characteristics of source images, namely, structural information and texture details. To effectively preserve most of the structural information, the base layers are fused using the combined Laplacian pyramid and sparse representation rule, in which an image patch selection-based dictionary construction scheme is introduced to exclude the meaningless patches from the source images and enhance the sparse representation capability of the pyramid-decomposed low-frequency layer. The detail layers are subsequently merged using a guided filtering-based approach, which enhances contrast level via noise filtering as much as possible. The fused base and detail layers are reconstructed to generate the fused image. We experimentally verify the superiority of the proposed method by using two basic fusion schemes and conducting comparison experiments on nine pairs of medical images from diverse modalities. The comparison of the fused results in terms of visual effect and objective assessment demonstrates that the proposed method provides better visual effect with an improved objective measurement because it effectively preserves meaningful salient features without producing abnormal details.

INDEX TERMS Medical image fusion, guided filtering, sparse representation, salient features, meaningless details.

I. INTRODUCTION

Owing to advancements in imaging technologies, medical images have been essential to clinical investigation and disease analysis. However, medical images from a single imaging modality provide limited organ/tissue information. For instance, computed tomography (CT) imaging provides clear visualization of dense bone structures, and magnetic resonance imaging (MRI) exhibits an evident advantage in capturing soft tissues and neurovascular structures with high spatial resolution. In medical imaging, the fusion of CT and MRI images enables the simultaneous visualization of bony structures and soft tissues, and thus, facilitates clinical diagnosis and treatment [1]. Single-photon emission computed tomography (SPECT) and positron emission tomography (PET) are two widely used functional imaging techniques with low resolution but can provide 3D observation of the body's metabolism information. The fusion of PET/SPECT and MRI images enables the clear observation of functional information and structural features from source images.

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan.

Therefore, by using a medical image fusion technique to obtain a single composite image with superior quality, medical professionals will no longer be required to separately analyze medical images from a single imaging device, reducing the uncertainty of lesion analysis and clinical diagnosis [2].

According to [3], the general fusion algorithm can be roughly divided into four categories: the multi-scale decomposition (MSD)-based methods, the sparse representation (SR)-based methods, the spatial domain-based methods, and the hybrid transform-based methods. In MSD-based fusion methods, multi-scale transform (MST) and spatial-filtering are two widely used fusion schemes. The MST [4]–[10] are popular techniques because of their excellent performance in extracting salient information. However, no single transform domain can completely represent all the information of the source images. Moreover, some significant features of the source images are lost during the inverse transform [11], resulting in quality degradation, such as ringing effect, contrast decreases, color distortion, or artificial edges [12]. In contrast with MST methods, the other scheme forms a fused image by using a spatial-filtering decomposition scheme. A typical approach is the guided filtering-based method [13] in which the weight average strategy captures the edge information of details layers and large intensity variations of base layers. However, similar to conventional spatial domain-based methods, spatial filtering-based methods also easily cause low contrast. SR-based methods are generally time-consuming because they rely on the training and optimization of the dictionary [14]. In general, single transform schemes with simple fusion strategies cannot always identify salient features from different decomposed coefficients. Hybrid transform-based methods simultaneously employ more than one transforms in the fusion process, aiming to combine the merits of different transforms [15], [16]. A representative hybrid transform-based example was reported in [17]. In particular, the combination of the Laplacian pyramid and sparse representation (LP-SR) was proposed to overcome the limitations of traditional MST-based and SR-based methods in medical image fusion, achieving favorable fused results. Nevertheless, this method suffers from spectral distortion and visual artifacts in medical image fusion.

In medical images, no salient features exist in flat areas, such as bone regions. However, the complementary image that joins the fusion procedure generally has abundant but meaningless details in the same region, and these meaningless details are easily distinguished as salient features [18]. Thus, most MST-based fusion methods are prone to extracting meaningless details from these regions to the fused images, leading to an inferior visual effect of the fusion results. The motivation to work with this two-scale framework is based on the fact that MST approaches have failed to identify meaningful and salient features of medical source images.

To address the aforementioned limitation and achieve perceptually good results, the current study presents a two-scale image fusion method that utilizes guided filtering and sparse representation. The main advantages of this paper can be summarized as follows:

(1) This work proposes a two-scale fusion method for multimodal medical images. This method can capture meaningful and salient information without producing abnormal details. In particular, two different schemes under the same framework are presented and performed on nine pairs of medical images from different modalities to verify the performance of the proposed framework. We theoretically analyze the significant difference between the two aforementioned schemes and conduct experimental verification to identify the best scheme in terms of both visual perception and objective assessment.

(2) The proposed method uses the guided filtering for a simple structure-texture decomposition of the source images to obtain the base and detail layers, which enhances the edge information of the fused image.

(3) Considering the extraction of salient features and less meaningless details, two different fusion rules are exploited for the base and detail layer to preserve the spatial consistency at each layer. Overall, most strong structural features are found in the base layer, and the rest of the texture details are located in the detail layer. Thus, the LP-SR rule is used to retain most of the structural information in the source images, and the guided filtering-based strategy is utilized to reduce noise sensitivity and improve the contrast level.

(4) In particular, a spatial degraded dictionary using an image patch selection-based scheme is learned by the K-singular value decomposition (K-SVD) algorithm from two source images. In this phase, the mean value of image patches determines whether an image patch is fit for selection, which is capable of excluding meaningless image patches and improving sparse representation capability for the pyramid-decomposed low-frequency layer.

The rest of this paper is organized as follows. In Section II, the recent works that are related to our methods are introduced. Section III describes the details of the proposed two-scale image fusion method using guided filter and sparse representation. Experimental results and related discussions are provided in Section IV. Finally, Section V concludes the paper and discusses future works.

II. RELATED WORK

In this section, we provide the recent researches, which have the similar work or relate to our medical image fusion framework, that is, the decomposition-based methods, the dictionary learning-based methods. Besides, several deep learning-based methods are briefly introduced. Finally, the main considerations of the proposed framework are illustrated in detail.

A. DECOMPOSITION-BASED METHODS

Unlike the MST methods, spatial filtering-based decomposition methods have become a hot topic in image fusion in the last decades. These methods tend to use edge-preserving filters to accurately extract the edge information and details at different scales, helping maintain the shift-invariance and

achieving better preservation performance of edge features; they utilize discriminative fusion strategies that consider the characteristic of each decomposed image to generate a fused image. Jian *et al.* proposed an image fusion method for via rolling guidance filter-based decomposition [19]; this method can better preserve valid structural information and details. Ma *et al.* utilized a Gaussian filter to achieve the two-scale decomposition of source images and applied different fusion rules to integrate the layers, preserving the thermal radiation and details of source images [20]. Zhu *et al.* proposed a hybrid multi-scale decomposition scheme with guided image filtering [21] and introduced three diverse fusion rules for different scales [22]. The details of source images can be assigned to different scales by setting the parameters in the guided filter.

In the past few years, the total variation (TV) has been introduced in image decomposition. Since then, a variety of TV decomposition models have been proposed in image fusion field due to their inherent denoising ability. These methods directly employ the original TV or other enhanced TV models to obtain the multiscale decomposition of source images and adopt complicated fusion rules to pursue good fusion performance. In [23], the total variation (TV-L1) model was applied to the two-scale decomposition of source images and a particle swarm optimization (PSO)-based adaptive weighted scheme was introduced to form a fusion image, which effectively preserves image energy. Similarly, Liu *et al.* proposed the total-variational transform into moving least squares scheme (TV-MLS) and designed different strategies [24]. Furthermore, in multi-focus image fusion, a novel TV and quad-tree based decomposition scheme was proposed to get the focus region to represent the source image completely [25].

Recently, a variety of novel image decomposition schemes have also been introduced for image fusion. Xing *et al.* utilized the Taylor expansion theory to decompose the source image into a deviation component and several energy components, which supresses the information loss caused by limited image descriptions [26]. In medical image fusion, Du *et al.* applied the intrinsic image decomposition for the two-scale decomposition of medical source images and employed three fusion rules to fuse these decomposed images [2]. It has been proven that the framework using the principle component analysis (PCA) rule achieves the best performance as it realizes the enhancement of structural information without color distortion. The PCA scheme also can be used in dimensionality reduction of hyperspectral image [27], [28]. Maqsood *et al.* introduced two gradient operators to obtain a two-scale representation of source images and then combined contrast enhancement, feature extraction algorithm and SR rule to form a fused image [29]. However, this method is prone to abnormal edges. A decomposition method based on structure tensor was introduced in [30]. The method can effectively capture gradient to distinguish individual edge information from detail information, which achieves competitive fusion performance. For decomposition-based methods, the simple fusion rule cannot always effectively identify the principal information of each decomposed image [31], causing quality degradation of the fused image. Thus, designing discriminative fusion strategies that consider the characteristic of each decomposed image is important.

B. DICTIONARY LEARNING-BASED METHODS

The sparse representation model has also been widely used in several image fusion applications [32]–[35]. A comprehensive review of this group is given in [36]. In particular, constructing an adaptively trained dictionary by using a learning method has been proven to provide adaptive sparse representation compared with other methods that use fixed dictionary models, (e.g., DCT and wavelet); it is also highly effective for improving the flexibility of dictionary. For medical images, image patches that are divided from source images possess redundant information that generates unvalued and uncertain information during sparse coding. That is, local overlapping image patches from source images appear simple and exhibit unstable structures; they cannot be used directly as a training set for dictionary learning.

To solve this problem, a novel dictionary learning approach was proposed in [37]. This approach calculates the Euclidean distance between each image patch divided from medical source images to select informative patches and uses a classification method with a clustering algorithm based on local density peaks to learn several sub-dictionaries. The dictionary trained using this scheme appears complete and informative and can effectively describe source images. Kim *et al.* proposed a multimodal image fusion method based on a joint patch clustering-based dictionary learning scheme [38]. In this scheme, all patches from different source images are clustered together with their structural similarities. The aforementioned method achieves remarkable fusion performance because the learned dictionary enables the complete description of images details. Yin *et al.* proposed a novel SR-based multi-focus image fusion method [39] that adopts the K-SVD algorithm to learn a joint dictionary and merges sparse coefficients by utilizing a maximum weighted multi-norm rule. The joint dictionary does not require any prior knowledge and can provide adaptive representation to source images. A separable dictionary learning-based method was proposed in [40]. This method combines separable dictionary optimization with a Gabor filter to solve the spatial inconsistency problem in flat regions. A novel dictionary learning scheme based on sparse and low-rank component decomposition was proposed for medical image fusion in [41]. This scheme performs well regardless of whether the source images are clear or corrupted by noise. In [42], a novel discriminative dictionary learning method is performed on the coarse-scale and fine-scale components to recover the structural information of coarse components and fine details corrupted by noises. In noisy image fusion, a novel discriminative dictionary learning method was presented to realize noise suppression and detail retention [43]. The discriminative ability of learned dictionaries is improved since the linear correlation between the sparse

coding coefficients is considered. In [44], the discrimination ability of dictionaries is enhanced by minimizing the correlation between the dictionaries of different components and a novel decomposition method using analysis-synthesis dictionary pair learning scheme was proposed to preserve image information and maintain the contrast. In brief, conducting a good dictionary plays an important role in these sparse representation-based methods.

C. DEEP LEARNING-BASED METHODS

Deep learning techniques have received extensive attention and have been successfully applied to image fusion in recent years. In [3], the authors reviewed the recent deep learning-based fusion methods. We present only several typical examples in the current study. Liu *et al.* first introduced the deep convolutional neural network (DCNN) to address the multi-focus fusion problem [45]. This method uses a DCNN model to learn a direct mapping between source images and weight maps, overcoming the difficulty of designing activity-level measurement and fusion rule. A novel medical image fusion method that combines the advantages of MST and the DCNN model was proposed in [46]. However, this method cannot sufficiently extract details from source images. Ma *et al.* introduced a novel method based on an end-to-end model of a generative adversarial network (GAN) [47]; this method was successfully applied to the fusion of infrared and visible images. Similarly, Huang *et al.* introduced a Wasserstein generative adversarial networks to color medical image fusion [48], in which a generator and two discriminators are employed to form a fused image. This method enhances the structure information and prevents the functional information from being weakened. Liu *et al.* proposed a convolutional neural network (CNN)-based multimodal medical image fusion method [49] under the Laplacian pyramid domain to enhance details and preserve image energy. Besides, most methods are not a unified network for image fusion and thus they are only applicable for specific fusion tasks. Aiming to overcome this limitation, some effective fusion networks [50], [51] are presented, which is capable of obtaining high-quality fusion images in various image fusion tasks. However, such deep learning-based methods generally take a long time to adjust the parameters of the framework [14].

D. THE MAIN CONSIDERATIONS OF THE PROPOSED FUSION FRAMEWORK

The proposed framework mainly combines three aspects: spatial filtering decomposition, dictionary learning, and two predesigned fusion strategies to extract more significant and salient information with less meaningless details from medical source images. The proposed method is primarily based on the following considerations: MST-based decomposition methods generally require a long time and can easily introduce visual artifacts; thus, our method uses a guided filter to obtain a two-scale representation, i.e., the base and detail layers. Given the edge-preserving characteristic of the guided filter, the base layer includes structural information containing most of the edge features, and the detail layer includes the remaining texture features and noises. This decomposition scheme can effectively enhance the edge information of the fused image and reduce computation complexity. LP-SR is an effective hybrid transform-based rule; it has been proven to preserve most of the image energy but is also prone to capturing meaningless details of source images. Thus, the LP-SR rule is suitable for the fusion of the base layers as the base layer contains most of the structural information and fewer details. In addition, most of the texture information with noises lies in the detail layer. A guided filtering-based strategy is conducted to simultaneously maintain spatial consistency and exclude the meaningless details by noise filtering. Intuitively, the combination of the two aforementioned models indicates that edge features can be preserved while filtering out abnormal texture details. Further discussion will be provided to explain the proposed framework in detail.

III. PROPOSED METHOD

This section summarizes the overall implementation steps. The proposed framework utilizes the guided filtering-based decomposition scheme and two fusion strategies: LP-SR and guided filtering-based weighted average. A diagram of our framework is shown in Fig. 1. The primary implementation process has the following steps: two-scale decomposition, base layer fusion, detail layer fusion, and image reconstruction. First, we obtain two-scale representations of source images by applying a guided filter and a difference operator. Then, the LP-SR and guided filtering-based weighted average strategies are applied to fuse the base and detail layers corresponding to pixel characteristics. Lastly, the fused image is obtained by reconstructing the fused base and detail layers. Furthermore, to identify the best scheme under this framework, two schemes, namely, Scheme 1 and Scheme 2, are provided. Here, we use Scheme 1 as an example to discuss the detailed implementation of our framework and then explain the significant difference between the two schemes.

A. TWO-SCALE DECOMPOSITION

In this subsection, two-scale decomposition is performed on each of the pre-registered source images into a base layer that captures large-scale of structural information, and a detail layer containing the texture details and noises. As shown in Fig. 1, a pair of registered CT and MRI images is denoted as I_1 and I_2, respectively. Let $GF_{r,\varepsilon}(a, b)$ represents the guided filtering operation, where r and ε are the local window radius and regularization factor of guided filter, respectively; a and b are considered as the input image and guidance image, respectively. Then, the guided filtering is applied to each source image serving as input image and guidance image simultaneously to obtain the base layers B_1 and B_2 as shown

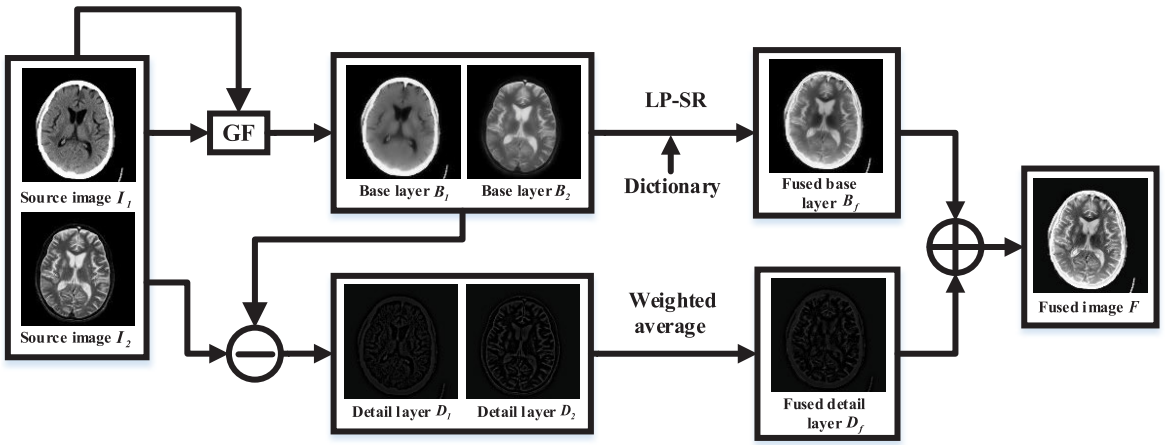


FIGURE 1. Schematic diagram of the proposed image fusion method.

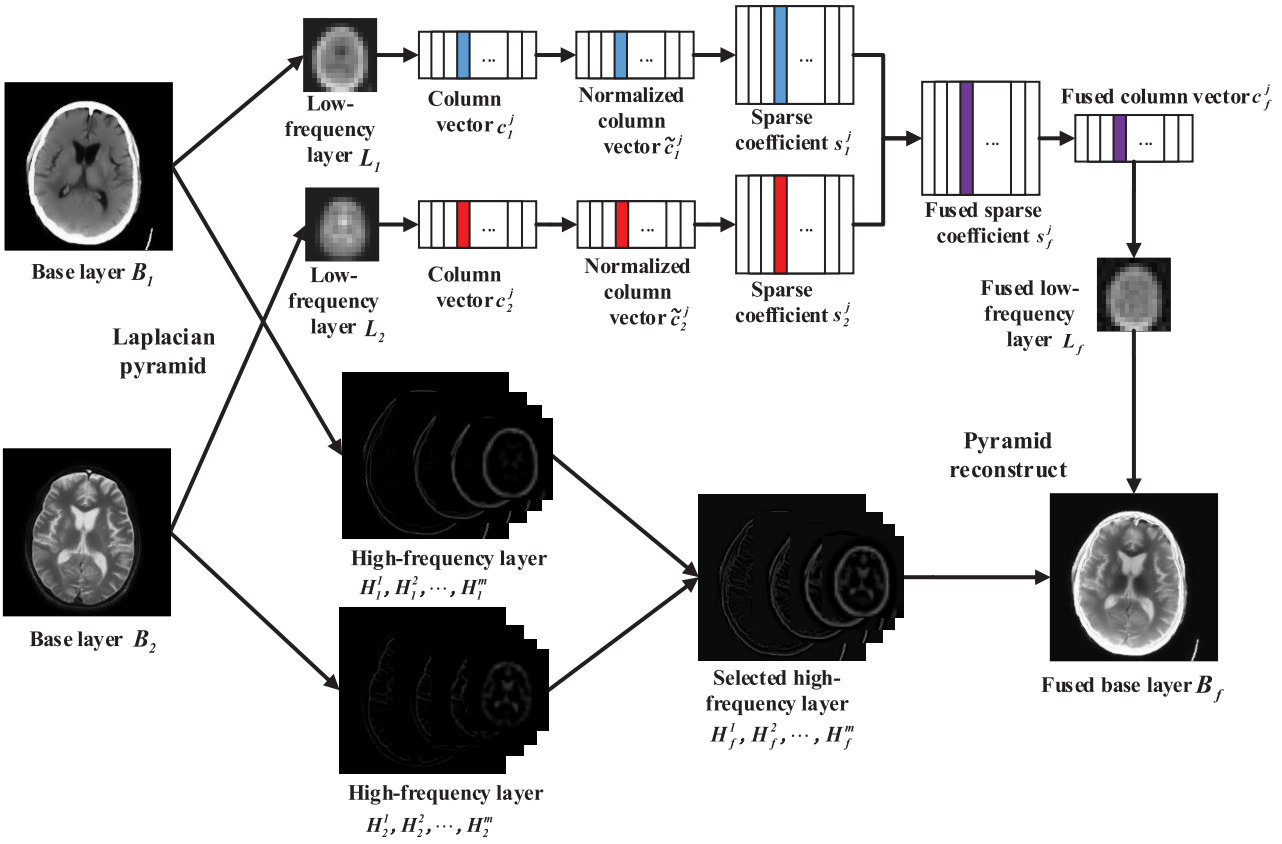


FIGURE 2. Schematic diagram of the base layer fusion.

in (1). The absolute values of the difference between the source images and the base layers are calculated to generate the detail layers D_1 and D_2, as shown in (2).

$$B_1 = GF_{r_1,\varepsilon_1}(I_1, I_1)$$
$$B_2 = GF_{r_1,\varepsilon_1}(I_2, I_2) \quad (1)$$
$$D_1 = |I_1 - B_1|$$
$$D_2 = |I_2 - B_2| \quad (2)$$

where $|\cdot|$ is the absolute value operator. Thus, large-scale information containing edge features in source images is retained in the base layers and the texture information is retained in detail layers.

B. BASE LAYER FUSION

The diagram of the base layer fusion framework is shown in Fig. 2, and the major process is presented in detail in this subsection.

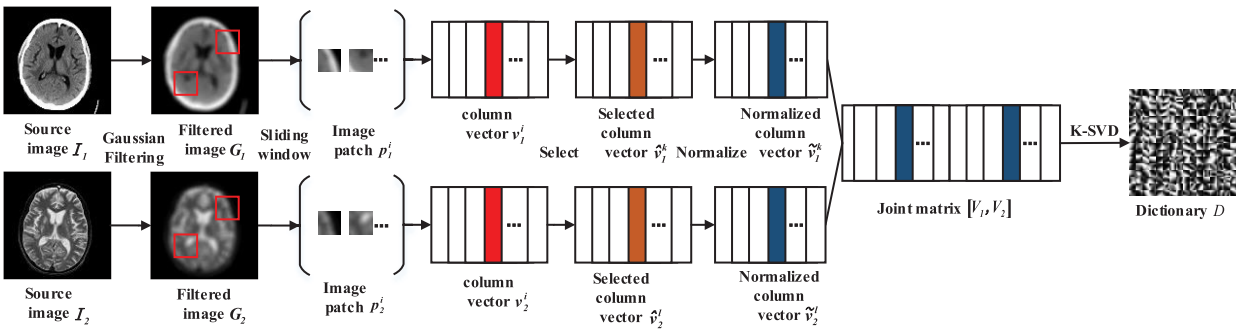


FIGURE 3. Schematic diagram of the dictionary construction method.

1) LAPLACIAN PYRAMID DECOMPOSITION

The Laplacian pyramid exhibits the advantages of effective structure, simple implementation, and high calculation efficiency. In this phase, the Laplacian pyramid is adopted for the multiscale representation of base layers. A source image can be transformed into a low-frequency layer and a series of high-frequency layers. First, we apply the Laplacian pyramid to divide B_1, B_2 into low-frequency layers L_1, L_2 and high-frequency layers $\{H_1^i, H_2^i\}(H_s^i = \{H_s^1, H_s^2, \cdots, H_s^m\}_{s=1,2})$, where m is the decomposition level of the Laplacian pyramid. Then, the high-frequency layers are selected in accordance with the pixel-wise maximum rule of $\{H_1^i, H_2^i\}$, for each level i with pixel location (x, y), which can be expressed as:

$$H_f^i(x,y) = \begin{cases} H_1^i(x,y) & if\ H_1^i(x,y) > H_2^i(x,y) \\ H_2^i(x,y) & otherwise. \end{cases} \tag{3}$$

2) DICTIONARY CONSTRUCTING

For the low-frequency layer, considering the low-frequency layer as a multi-filtered version of a source image is reasonable. As shown in Fig. 3, we construct a spatial degraded dictionary by integrating two informative column vector sets from two filtered source images. A column vector is selected when its mean value is larger than the default value. The major steps of dictionary construction are as follows.

Step 1: The Gaussian low-pass filter with size 5 is used to smooth source images and generate corresponding blurred versions, denoted as G_1 and G_2.

Step 2: The sliding window technique with a step of one pixel is used to divide G_1 and G_2 into overlapping image patches with size 8×8. Then, all the image patches are converted into corresponding column vectors through lexicographic ordering. In particular, let p_1^i and p_2^i represent the *ith* image patches in G_1 and G_2, respectively. Then, the corresponding column vectors are denoted as v_1^i and v_2^i, and the informative column vectors $\{\hat{v}_1^k, \hat{v}_2^l\}$ are extracted, in which the mean value of the corresponding column vector exceeds the cutoff threshold, which can be described as:

$$\hat{v}_1^k = \begin{cases} Select & v_1^i & if\ v_1^i > 0.1 \times C \\ Ignore & v_1^i & otherwise. \end{cases} \tag{4}$$

$$\hat{v}_2^l = \begin{cases} Select & v_2^i & if\ v_2^i > 0.1 \times R \\ Ignore & v_2^i & otherwise. \end{cases} \tag{5}$$

where $\hat{v}_1^k$ and $\hat{v}_2^l$ refer to the selected column vectors with sufficient energy from G_1 and G_2, respectively; C and R are the mean value of G_1 and G_2, respectively; k and l are used for each input image G_1 and G_2 to denote the *kth* and *lth* informative column vector, respectively.

Step 3: The mean value of informative column vectors $\hat{v}_1^k$ and $\hat{v}_2^l$ is normalized to obtain $\tilde{v}_1^k$ and $\tilde{v}_2^l$, which contain only structural details

$$\begin{aligned} \tilde{v}_1^k &= \hat{v}_1^k - m_1^k \\ \tilde{v}_2^l &= \hat{v}_2^l - m_2^l \end{aligned} \tag{6}$$

where m_1^k and m_2^l are the $n \times 1$ vector, and all elements in the vector are the mean value of $\hat{v}_1^k$ and $\hat{v}_2^l$, respectively.

Step 4: Steps 2 and 3 are repeated for all the image patches. After all informative column vectors from the filtered source images are extracted and normalized, these column vectors are combined as a single matrix $V_1 = [\tilde{v}_1^k]_{k=1}^K$ and $V_2 = [\tilde{v}_2^l]_{l=1}^L$. The joint matrix V is constructed by combining the two aforementioned matrices as follows:

$$V = [V_1, V_2] \tag{7}$$

Step 5: Lastly, the dictionary D can be learned from V by applying the K-SVD algorithm [52].

3) FUSION OF LOW-FREQUENCY LAYERS

The fusion of low-frequency layers contains the following steps:

Step 1: The sliding window technique is used to divide L_1 and L_2 into corresponding patches from the upper left to the lower right with a step of one pixel.

Step 2: All the patches are transformed into column vectors $\{c_1^j, c_2^j\}_{j=1}^J$, where J is the number of column vectors, and then normalized vectors are obtained by subtracting their mean value as follows:

$$\begin{aligned} \tilde{c}_1^j &= c_1^j - m_1^j \\ \tilde{c}_2^j &= c_2^j - m_2^j \end{aligned} \tag{8}$$

Step 3: After the dictionary is obtained, we further compute the corresponding sparse coefficients $S_1 = \{s_1^j\}_{j=1}^J$ and $S_2 = \{s_2^j\}_{j=1}^J$ by orthogonal matching pursuit (OMP) approach [53], which has the following formulation

$$\begin{aligned} s_1^j &= \arg\min_s \|s\|_0 \text{ . s.t. } \left\|\tilde{c}_1^j - Ds\right\|_2^2 \leqslant \epsilon \\ s_2^j &= \arg\min_s \|s\|_0 \text{ . s.t. } \left\|\tilde{c}_2^j - Ds\right\|_2^2 \leqslant \epsilon \end{aligned} \tag{9}$$

where the parameter ϵ is used as the sparse approximation error.

Step 4: Then, the sparse coefficients are determined by applying the max-ℓ_1 rule, and the fused column vectors are calculated

$$s_f^j = \begin{cases} s_1^j & if\ \|s_1^j\|_1 > \|s_2^j\|_1 \\ s_2^j & otherwise. \end{cases} \tag{10}$$

$$c_f^j = \begin{cases} Ds_f^j + m_1^j & if\ s_f^j = s_1^j \\ Ds_f^j + m_2^j & otherwise. \end{cases} \tag{11}$$

Step 5: (8)-(11) are iterated for all patches in the low-frequency layer to produce all the fused vectors $\{c_f^j\}_{j=1}^J$. After that, each c_f^j is reshaped into an image patch with size 8×8, and all the image patches are placed to their original positions. In particular, each pixel's value in the low-frequency layer is averaged when patches are overlapped. The fused low-frequency layer L_f is obtained after the iteration. Finally, the fused low-frequency layer and several selected high-frequency layers are reconstructed by the inverse Laplacian pyramid to obtain the fused base layer B_f.

C. DETAIL LAYER FUSION

The detail layer contains the rest of the texture details and noises. For the fusion of the detail layer, the guided filtering-based weighted average strategy is used, which enhances the contrast level by noise filtering as much as possible.

Generally, since the Laplacian operator achieves edge detection by performing differentiation operations on images, leading to its sensitivity to discrete points or noises. Hence, Laplacian filtering is performed on images, and then Gaussian convolution filtering is applied to the filtered image for noise reduction, improving its robustness to noises [54]. Furthermore, the construction of the saliency map typically considers the following important factors: detection accuracy and computational complexity. However, widely used graph-based methods [55], contrast-based methods [56], and recent deep learning-based methods [57] require precise adjustment of multiple parameters to achieve the desired results. By contrast, the Gaussian function is only driven by fewer parameters (i.e., filter size and blurred degree) to estimate a more robust saliency map, which reduces the computational complexity. Considering the above-mentioned analysis, the saliency maps of two detail layers are constructed via a Laplacian filtering followed by Gaussian smoothing to highlight the small-scale edge details and denoise the detail layers. The detailed implementation is described as follows.

Step 1: The detail layers D_1, D_2 are obtained using the image decomposition method mentioned earlier. First, Laplacian filtering is applied to D_1 and D_2 to obtain the edge feature map corresponding to detail layers

$$\begin{aligned} H_1 &= D_1 * L_{3\times3} \\ H_2 &= D_2 * L_{3\times3} \end{aligned} \tag{12}$$

where $L_{3\times3}$ is a Laplacian filter with size 3×3.

Step 2: The local smoothing of edge feature map is implemented to construct the saliency map by applying a Gaussian filter

$$\begin{aligned} S_1 &= |H_1| * G_{\alpha,\sigma} \\ S_2 &= |H_2| * G_{\alpha,\sigma} \end{aligned} \tag{13}$$

where $G_{\alpha,\sigma}$ is a Gaussian filter with size $(2\alpha+1)(2\alpha+1)$ and the σ represents the blurred degree. In this work, α and σ were set as 5 and 2, respectively.

Step 3: The saliency maps are compared to generate the initial weight map

$$W_1(x,y) = \begin{cases} 1 & if\ S_1(x,y) > S_2(x,y) \\ 0 & otherwise. \end{cases} \tag{14}$$

$$W_2(x,y) = 1 - W_1(x,y) \tag{15}$$

where the initial weight map is actually a binary map that corresponds to the 'choose max' strategy in the pixel location (x, y).

Step 4: Guided filtering is applied to each initial weight map, and its corresponding detail layer is regarded as the guidance image for optimizing the initial weight map

$$\begin{aligned} W_{f1} &= GF_{r_2,\varepsilon_2}(W_1, D_1) \\ W_{f2} &= GF_{r_2,\varepsilon_2}(W_2, D_2) \end{aligned} \tag{16}$$

where W_{f1} and W_{f2} represent the final weight maps of D_1 and D_2, respectively. To restore the remaining edge information and obtain spatial consistency, small local window radius and regularization factor are chosen in this phase.

Step 5: The following pixel-wise weighted average strategy is adopted to obtain the fused detail layer as follows:

$$D_f = W_{f1}D_1 + W_{f2}D_2 \tag{17}$$

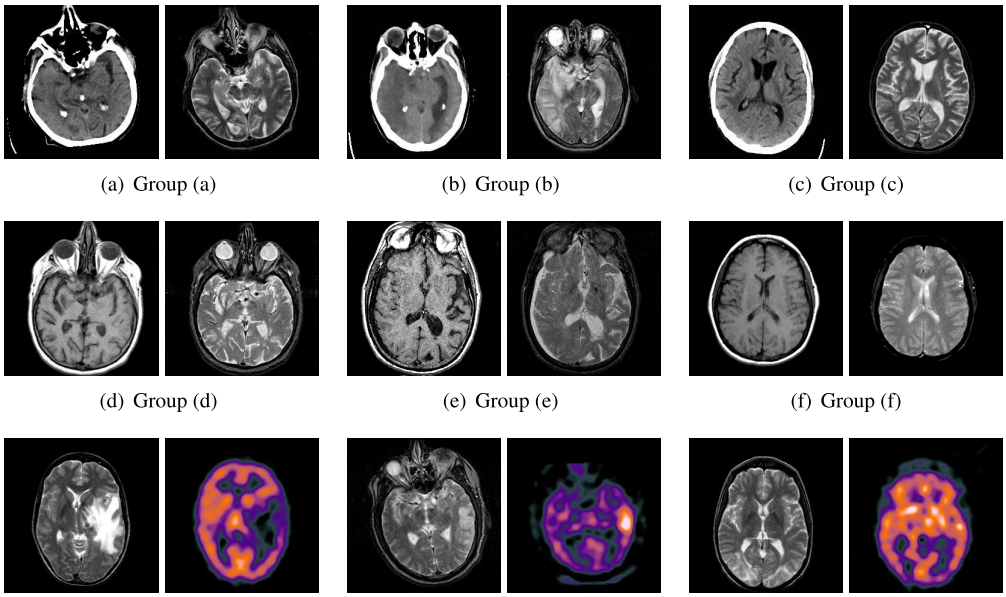
D. IMAGE RECONSTRUCTION

Finally, we obtain the final fused image by combining the fused base layer B_f and the fused detail layer D_f

$$F = B_f + D_f \tag{18}$$

E. THE SIGNIFICANT DIFFERENCE BETWEEN TWO SCHEMES

To effectively extract the salient features with less abnormal details, we propose two fusion schemes under the same framework. The difference between the two schemes is primarily focused on Scheme 2 having a smaller pyramid



(a) Group (a) (b) Group (b) (c) Group (c)

(d) Group (d) (e) Group (e) (f) Group (f)

(g) Group (g) (h) Group (h) (i) Group (i)

FIGURE 4. Source images pairs of three types of multimodal medical image. (a)-(c) three pairs of CT and MRI images; left: CT image; right: MRI image. (d)-(f) three pairs of MRI-T1 and MRI-T2 image; left: MR-T1 image; right: MR-T2 image. (g)-(i) three pairs of SPECT image and MRI image; left: MRI image; right: SPECT image.

decomposition level and a different sparse coding phase. We consider the following case for the first difference: The base layer obtained by two-scale decomposition contains large-scale structural information. However, a higher pyramid decomposition level means that more salient information will be transferred to the fused image, leading to the production of abnormal artifacts in the flat regions, particularly for those regions with high pixels. The second difference is the sparse coding phase. Notably, Scheme 1 applies the normalization procedure to the informative column vectors and dictionary atoms (see the Step 3 in dictionary constructing and Step 2 in Fusion of low-frequency layers), theoretically making the framework focus more on the extraction of structural details and disregard the global features. Thus, two schemes are adopted in comparison experiments to determine the most suitable scheme under this framework. Compared with Scheme 1, Scheme 2 has a lower pyramid decomposition level and eliminates the normalization procedure in the dictionary construction and sparse coding phases. Further experiments will provide a detailed explanation between these two schemes.

IV. EXPERIMENTS

A. SOURCE IMAGES

Experiments are performed on nine pairs of registered medical images with size 256×256 from three different modalities to verify the superiority of the proposed medical image fusion method. Three groups of registered multimodal medical images, including three pairs of CT and MRI images, three pairs of T1-weighted MR (MR-T1) and T2-weighted MR (MR-T2) images, and three pairs of SPECT and MRI images, were used in our experiments. All the source images were selected from the website of Harvard Medical School [58] and shown in Fig. 4. Relevant experimental results and discussions are presented in detail.

B. COMPARED METHODS

Eight representative image fusion algorithms including GF [13], JPCD [38], CSR [59], LP-SR [17], SR-JD [39], CSMCA [60], PAPCNN [18] and LP-CNN [49] were selected for the comparison experiments. Among these compared algorithms. GF is a spatial filtering-based method with optimization by the guided filtering. JPCD and SR-JD are two novel SR-based methods using different dictionary learning schemes. CSR utilizes a novel SR model for multimodal image fusion. LP-SR uses the Laplacian pyramid combined with the SR rule. LP-CNN is a medical fusion method using a CNN model in the Laplacian domain. CSMCA and PAPCNN were just recently proposed in the past year and have achieved good performance in medical image fusion. For parameters of the aforementioned fusion algorithms, the recommended values reported in their respective publications were adopted.

C. OBJECTIVE EVALUATION METRICS

To objectively evaluate of the performance of different methods, four widely recognized objective indexes were applied in our experiments. These indexes are as follows: the feature mutual information metric Q_{FMI} [61], the gradient-based quality metric Q_G [62], the visual information fidelity metric *VIFF* [63], and the standard deviation function *SD*. For these four indexes, a high value indicates good performance.

The metric Q_{FMI} measures global feature information between the source images and the fused image. The considered feature information is represented by image features such as gradients or edges. In [64], Q_{FMI} was proven to surpass other information theory information-based indexes, such as Q_{MI} and Q_{NCIE}, which employ the image histogram to compute mutual information. This is mainly because the image histogram only provides statistical features of specific grayscale pixels of input images and neglects other significant information that reflects structural features. The feature mutual information Q_{FMI} with two source images a, b and a fused image f, is defined as:

$$Q_{FMI} = \frac{MI_{\widetilde{a},\widetilde{f}}}{H_{\widetilde{f}} + H_{\widetilde{a}}} + \frac{MI_{\widetilde{b},\widetilde{f}}}{H_{\widetilde{f}} + H_{\widetilde{b}}} \tag{19}$$

$$MI_{\widetilde{a},\widetilde{f}} = \sum_{i,j} P_{\widetilde{a},\widetilde{f}}(i,j) log \frac{P_{\widetilde{a},\widetilde{f}}(i,j)}{P_{\widetilde{a}}(i) P_{\widetilde{f}}(j)} \tag{20}$$

$$MI_{\widetilde{b},\widetilde{f}} = \sum_{i,j} P_{\widetilde{b},\widetilde{f}}(i,j) log \frac{P_{\widetilde{b},\widetilde{f}}(i,j)}{P_{\widetilde{b}}(i) P_{\widetilde{f}}(j)} \tag{21}$$

where $\widetilde{a}$, $\widetilde{b}$, $\widetilde{f}$ are the feature maps of a, b, f, respectively. $P_{\widetilde{a},\widetilde{f}}(i,j)$ and $P_{\widetilde{b},\widetilde{f}}(i,j)$ are the joint probability density functions between images a, b and f at pixel (i,j), respectively. $P_{\widetilde{a}}(i)$ and $P_{\widetilde{b}}(i)$ are the edge probability density functions of a and b, respectively.

The metric Q_G mainly calculates the amount of edge information transferred from the input images to the fused images. It considers the edge strength associated with the human visual system (HVS), such that the quality of visual information can be reflected. The gradient-based metric Q_G is defined as:

$$Q_G = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (Q^{af}(i,j) W^a(i,j) + Q^{bf}(i,j) W^b(i,j))}{\sum_{i=1}^{I} \sum_{j=1}^{J} (W^a(i,j) + W^b(i,j))} \tag{22}$$

where $Q^{af}(i,j) = Q_e^{af}(i,j) Q_o^{af}(i,j)$, $Q_e^{af}(i,j)$ and $Q_o^{af}(i,j)$ denote the edge strength and orientation preservation values at pixel (i,j). The definition of Q^{bf} is the same as that of Q^{af}. The weight coefficients $W^a(i,j)$ and $W^b(i,j)$ indicate the importance of Q^{af} and Q^{bf}, respectively.

The metric *VIFF* is a newly proposed index that measures the visual information fidelity between the fused image and each source image using the Gaussian scale mixture model (GSM), the distortion model, and the HVS model. In the calculation of *VIFF*, source images are decomposed into blocks and visual information from each block is captured by applying the three aforementioned models. Finally, the visual information from each block is integrated to obtain an overall quality measure. For more details about this metric, kindly refer to [53].

The function *SD* is generally used to evaluate the overall contrast of a fused image.

$$SD = \sqrt{\frac{1}{I \times J} \sum_{i=1}^{I} \sum_{j=1}^{J} (f(i,j) - m)^2} \tag{23}$$

where m is the mean value of a fused image.

D. ANALYSIS AND SETTING OF ALGORITHM PARAMETER

In the proposed framework, the parameter setups are determined through a series of experiments that are performed on three types of medical images. The proposed framework has three key parameters: the local window radius r_1, the regularization parameter ε_1 of the guided filter that is used in two-scale decomposition, and the decomposition level m of Laplacian pyramid. Here, we use Scheme 1 as an example to comprehensively analyze the impacts on fusion performance with three key parameters. In this subsection, three different types of medical images, i.e., Group (a), Group (d), and Group (g) are used as test images, and the four objective indexes are adopted to evaluate the impact of three groups of parameters on fusion performance. Fig. 5 shows the objective index value on the three types of source images when parameters are changed.

In the first group of experiments, the regularization parameter ranges from 10^{-5} to 10^{0}. The first row displays the impact of ε_1 on fusion performance, and an inference can be made that when ε_1 is higher than 10^{-1}, the metrics Q_{FMI}, Q_G and *SD* generally tend to decline. Notably, the metric *VIFF* appears to increase for Group (d) and (g). Also, some objective metrics exhibit a slightly increasing tendency when ε_1 increase from 10^{-5} to 10^{-1}, such as the metric Q_{FMI} of Group (a), the metric Q_G of Group (g) and the metric *SD* of Group (a). Based on the preceding observation, we set ε_1 as 10^{-1}.

In the second group of experiments, the local window radius ranges from 5 to 40. The second row shows the impact of r_1 on fusion performance, and most metrics tend to be stable as r_1 increases. It should be noted that the scores of the four indexes decrease slightly as r_1 increases in Group (g). Thus, we fixed the local window radius to 5.

The impact of m on the four indexes is exhibited in the third row. The decomposition level ranges from 1 to 5. Notably, when m increases, the metrics *VIFF* and *SD* increase substantially since more spatial information are extracted, the metric Q_{FMI} exhibits a slight tendency to decrease. Moreover, the fused image suffers from deep artifacts and color distortion when m is less than 2. When m exceeds 3, the metric Q_G tends to increase except in Group (g). Balancing that case by considering the preceding analysis is reasonable. Thus, we set m as 4.

In conclusion, the local window radius 5, the regularization parameter 0.1, and the decomposition level 4 were adopted in our experiments. For the proposed method, the remaining parameters were set as follows: the parameter r_2, ε_2 were set

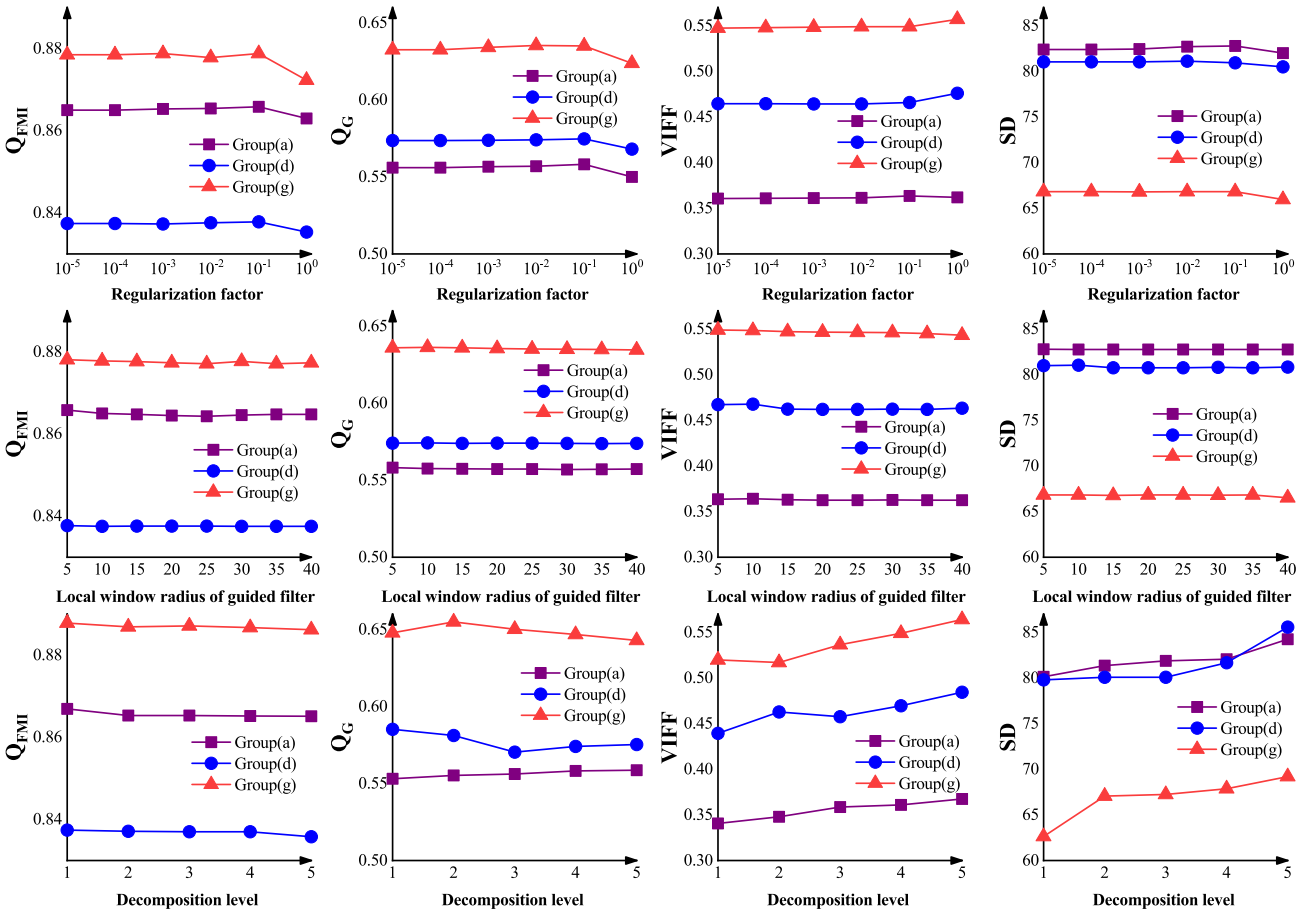


FIGURE 5. Objective performance of the proposed method with three parameters.

as 2, 1^{-6}, respectively, the dictionary size was set as 64×256 with a sparse level of 6. The sparse approximation error ϵ was set as 0.1, and the image patch of size 8×8 was adopted. For Scheme 2, the pyramid decomposition level was set as 1, and the other parameters are the same as those in Scheme 1.

E. EXPERIMENTAL RESULTS AND ANALYSIS

1) VISUAL QUALITY

The fusion of CT and MRI images can generate a single image, in which the complementary features of the two types of images are preserved well. The first experiment is conducted on three pairs of CT and MRI images. Three sets of the fusion results of the CT and MRI images are shown in Fig. 6, and two representative regions are enlarged in each image to make better comparisons. The fusion results of the GF and SR-JD methods show serious artifacts, particularly in the bone regions [see the right regions in Fig. 6 (a3), (b7), (b3), (b7) and (c7)]. The fusion results of the JPCD and CSR methods suffer from the loss of image energy [see the enlarged region in Fig. 6 (a4), (a5), (b4), (b5), (c4) and (c5)]. The bone regions of the CT images seem to be blurred by the CSMCA method [see Fig. 6 (a8), (b8) and (c8)]. The salient features can be nearly preserved via the LP-SR and LP-CNN methods, but these methods lose a few details and introduce more meaningless details into bone regions [see Fig. 6 (a6), (a11), (b6) and (b11)]. The PAPCNN method achieves the best visual effect since it preserves nearly all the soft tissue information from MRI images. However, a few artifacts are introduced into bone regions [see the right region in Fig. 6 (a9) and (b9)]. Meanwhile, the result appears that the PAPCNN method focuses more on the extraction of details, leading to the spatial inconsistency problem [see the left region in Fig. 6 (c9)]. Compared to other methods, Scheme 1 and Scheme 2 show evident advantages in the preservation of bone regions as they filter out more abnormal details via noise filtering [see the enlarged region in Fig. 6 (a11), (a12), (b11), (b12) and (c12)]. However, Scheme 1 still weakens the soft tissue information in the MRI image [see the left region in Fig. 6 (a11), (b11) and (c11)]. Scheme 2 simultaneously extracts most of the salient information with less meaningless details that exist in the MRI image. However, a few details seem to be blurred by Scheme 2 [see the right region in Fig. 6 (a12) and (c12)].

Soft tissues, such as adipose tissue, can be easily observed in MR-T1 images. Blood vessels are clearer in the MR-T2 images than in the MR-T1 images. Three sets

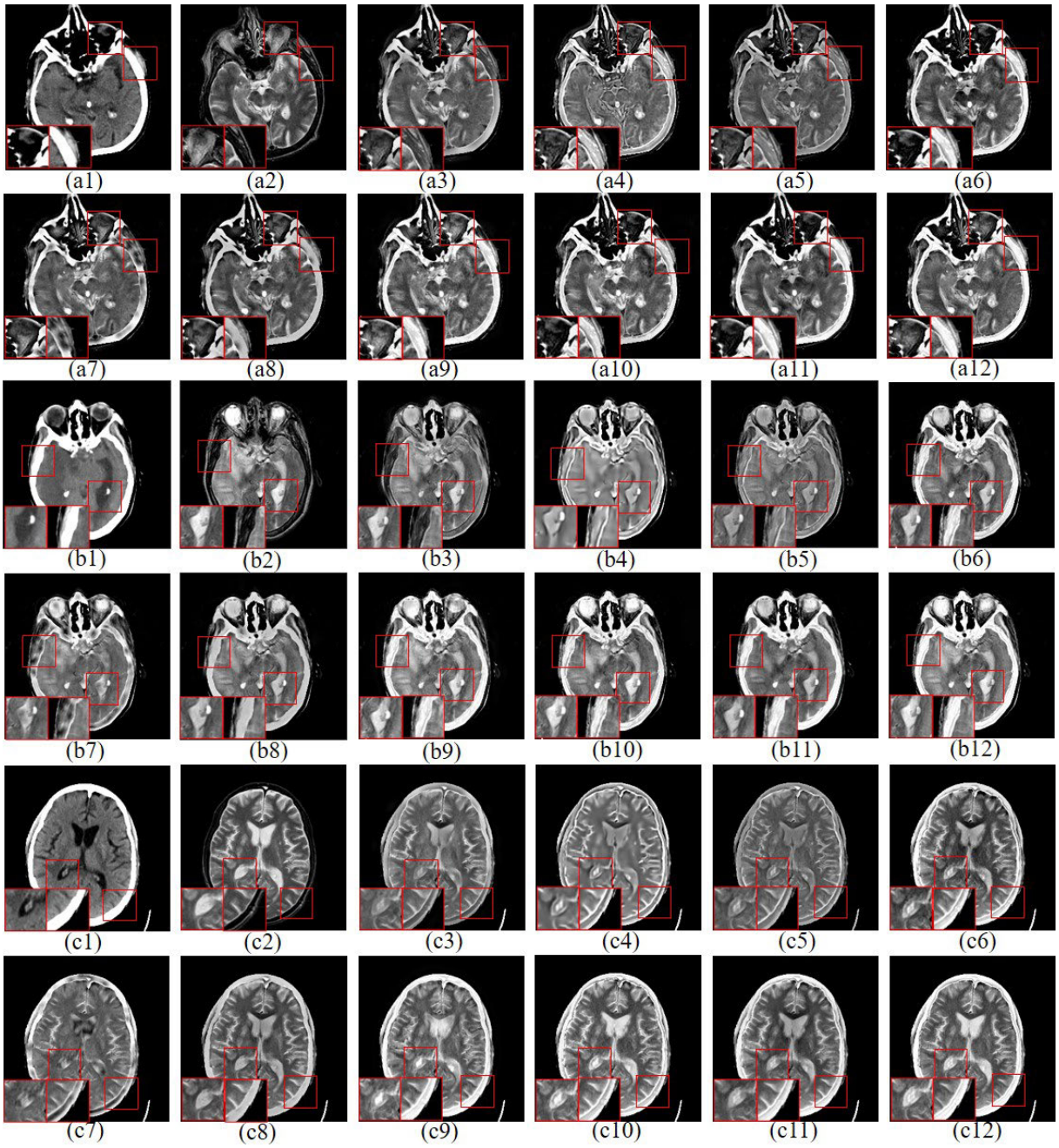


FIGURE 6. Three sets of CT and MRI image fusion results. Two close-ups are provided in each set for better comparisons. (a1) CT. (a2) MRI. (a3) GF. (a4) JPCD. (a5) CSR. (a6) LP-SR. (a7) SR-JD. (a8) CSMCA. (a9) PAPCNN. (a10) LP-CNN. (a11) Scheme1. (a12) Scheme2. (b1) CT. (b2) MRI. (b3) GF. (b4) JPCD. (b5) CSR. (b6) LP-SR. (b7) SR-JD. (b8) CSMCA. (b9) PAPCNN. (b10) LP-CNN. (b11) Scheme1. (b12) Scheme2. (c1) CT. (c2) MRI. (c3) GF. (c4) JPCD. (c5) CSR. (c6) LP-SR. (c7) SR-JD. (c8) CSMCA. (c9) PAPCNN. (c10) LP-CNN. (c11) Scheme1. (c12) Scheme2.

of fusion results of MR-T1 and MR-T2 images are shown in Fig. 7. We can easily see that the GF method extracts almost all the information of the MR-T2 image to the fused image, but it achieves low contrast [see Fig. 7 (a3)] since it loses a large amount of energy. Besides, some significant features of MR-T2 cannot be completely retained in the fused image [see the right region in Fig. 7 (b3) and (c3)]. The JPCD and CSR methods fail to preserve the salient features, resulting in a poor visual effect [see Fig. 7 (a4), (a5), (b4), (b5), (c4) and (c5)]. The LP-SR, SR-JD, CSMCA, and LP-CNN methods do not perform well in preserving the important details of the MR-T2 images to the fused image [see the right region in Fig. 7, (a6), (a8), (a10), (b6), (b7), (b8), (b10), (c6), (c7), (c8) and (c10)]. The PAPCNN method can transfer almost all the salient features from the source image to the fused image. As a result, more artifacts are introduced to the fused

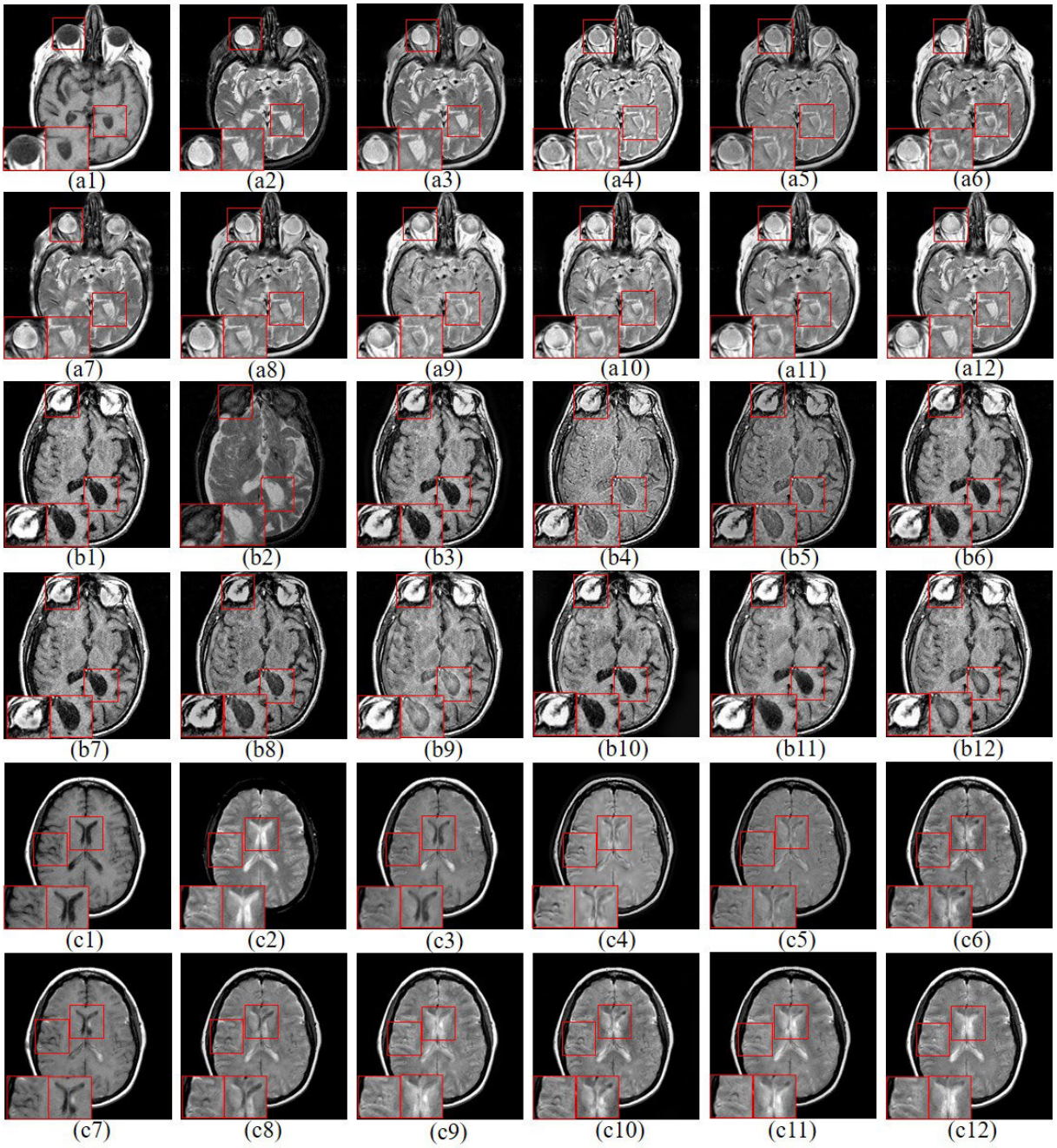


FIGURE 7. Three sets of MR-T1 and MR-T2 image fusion results. Two close-ups are provided in each set for better comparisons. (a1) CT. (a2) MRI. (a3) GF. (a4) JPCD. (a5) CSR. (a6) LP-SR. (a7) SR-JD. (a8) CSMCA. (a9) PAPCNN. (a10) LP-CNN. (a11) Scheme1. (a12) Scheme2. (b1) CT. (b2) MRI. (b3) GF. (b4) JPCD. (b5) CSR. (b6) LP-SR. (b7) SR-JD. (b8) CSMCA. (b9) PAPCNN. (b10) LP-CNN. (b11) Scheme1. (b12) Scheme2. (c1) CT. (c2) MRI. (c3) GF. (c4) JPCD. (c5) CSR. (c6) LP-SR. (c7) SR-JD. (c8) CSMCA. (c9) PAPCNN. (c10) LP-CNN. (c11) Scheme1. (c12) Scheme2.

image as it cannot distinguish the meaningful and significant features [see the enlarged region in Fig. 7 (a9) and (c9)]. Scheme 1 also does not successfully extract the significant features in the MR-T2 image [see the right region in Fig. 7 (a11) and (b11)] and fails in completely preserving textures [see the right region in Fig. 7 (a11) and the left region in (c11)]. Compared with Scheme 1, Scheme 2 achieves better performance in the extraction of significant features from the MR-T2 image [see the right region in Fig. 7 (a12) and (c12)] and some undesired details are excluded [see the bone region in Fig. 7 (a12)]. But, Scheme 2 does not exhibit an advantage in completely extracting textures [see Fig. 7 (c12)].

SPECT is a new imaging technology that can construct a three-dimensional image of a tracer concentration

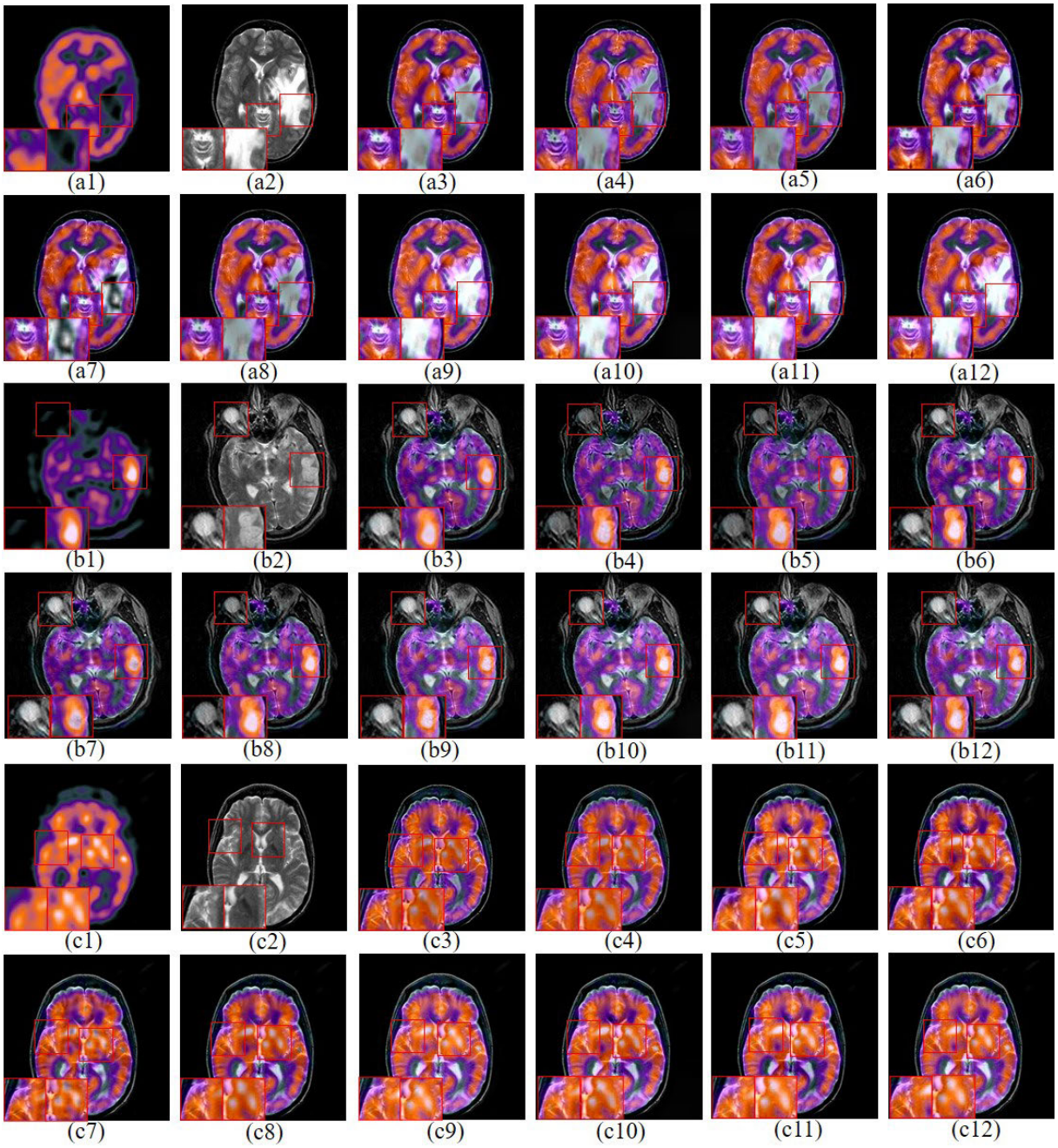


FIGURE 8. Three sets of SPECT and MRI image fusion results. Two close-ups are provided in each set for better comparisons. (a1) CT. (a2) MRI. (a3) GF. (a4) JPCD. (a5) CSR. (a6) LP-SR. (a7) SR-JD. (a8) CSMCA. (a9) PAPCNN. (a10) LP-CNN. (a11) Scheme1. (a12) Scheme2. (b1) CT. (b2) MRI. (b3) GF. (b4) JPCD. (b5) CSR. (b6) LP-SR. (b7) SR-JD. (b8) CSMCA. (b9) PAPCNN. (b10) LP-CNN. (b11) Scheme1. (b12) Scheme2. (c1) CT. (c2) MRI. (c3) GF. (c4) JPCD. (c5) CSR. (c6) LP-SR. (c7) SR-JD. (c8) CSMCA. (c9) PAPCNN. (c10) LP-CNN. (c11) Scheme1. (c12) Scheme2.

within a body. The fusion of SPECT and MRI images aims to facilitate the observation of soft tissue and functional information. In this experiment, we considered the SPECT image as a color image. Thus, YUV transform is performed on the SPECT image to produce a three-dimensional color space including a luminance component and two chrominance components (U and V), considering human perception. Then, the luminance component is extracted as a gray image that continues to participate in the fusion process with the MRI image. Finally, the fused luminance component, original U component, and V component are integrated by performing YUV inverse transform to obtain the fused image.

Fig. 8 shows three sets of fusion results of SPECT and MRI images. Obviously, the fusion performance of the GF, JPCD,

TABLE 1. Quantitative indices of fusion results.

Source images	Index	GF	JPCD	CSR	LP-SR	SR-JD	CSMCA	PAPCNN	LP-CNN	Scheme1	Scheme2
CT-MRI	Q_{FMI}	0.868	0.8602	0.8665	0.8705	0.8695	**0.8734(1)**	0.8644	0.8686	0.8708(3)	0.8715(2)
	Q_G	0.5267	0.4991	0.5163	0.5511(2)	0.4901	0.5225	0.4924	0.5268	**0.5517(1)**	0.5431(3)
	$VIFF$	0.2854	0.3405	0.3041	0.3825(3)	0.3124	0.3348	**0.3978(1)**	0.3699	0.385(2)	0.3774
	SD	67.7208	77.4667	63.7807	82.2876	73.0474	70.1369	84.4019(2)	82.7266	83.0496(3)	**85.4193(1)**
MR-T1-MR-T2	Q_{FMI}	0.8677	0.8511	0.8623	0.8708(2)	0.8701	0.8661	0.8572	0.866	0.8704(3)	**0.8711(1)**
	Q_G	**0.6295(1)**	0.5382	0.5712	0.6206	0.6059	0.6131	0.5867	0.6023	0.6221(3)	0.6225(2)
	$VIFF$	0.4396	0.4419	0.4397	0.5157	0.4284	0.4669	**0.5319(1)**	0.5027	0.5259(2)	0.5161(3)
	SD	71.0051	77.7311	63.8339	76.5089	72.3048	71.2275	77.9385(3)	75.3774	78.3882(2)	**78.4171(1)**
SPECT-MRI	Q_{FMI}	0.8659	0.8458	0.8654	0.8667	0.8625	0.8677	0.8684	0.873(3)	0.8753(2)	**0.879(1)**
	Q_G	0.6392	0.5332	0.6225	0.6394	0.6108	0.6472(3)	0.6398	0.6325	0.6699(2)	**0.6793(1)**
	$VIFF$	0.5139	0.4616	0.4764	0.5546	0.5221	0.5292	0.5771(3)	0.5847(2)	**0.5854(1)**	0.5329
	SD	52.9718	50.3807	49.6997	55.3902	52.7833	53.6961	**62.5575(1)**	61.9258	61.9673(3)	62.3352(2)

CSR, LP-SR, SR-JD, and CSMCA methods in preserving color fidelity are relatively low, causing serious visual inconsistency problem [see the right region in Fig. 8 (a3)-(a8)]; such condition is unsuitable for medical diagnosis. The PAPCNN and LP-CNN methods generally perform well in extracting functional information from the SPECT image and also exhibit good visual effects; however, a few defects still exist in terms of color distortion [see the right regions in Fig. 8 (b9) and (c10)]. Additionally, some undesired artifacts exist in the region with a high pixel value [see the white region in Fig. 8 (a9) and (a10)]. Scheme 1 also performs well in color preservation. However, a few artifacts are introduced into the fused image [see the right region in Fig. 8 (a11)]. Scheme 2 has the best performance in preserving color fidelity [see the right enlarged region in Fig. 8 (a12)]; however, partial textures are lost [see Fig. 8 (a12) and (c12)].

2) OBJECTIVE EVALUATION

The objective evaluation results of the eight fusion methods and the two proposed schemes on three types of medical image fusion are reported in Table 1, wherein the average value of each method on all the experimentally used images in each fusion problem is listed. For all the ten methods, the highest score of each metric is highlighted in bold and the scores ranking second to third places are underlined. Table 1 is visually presented in Fig. 9. Table 1 and Fig. 9 clearly show that Scheme 1 always ranks in the top three places for almost all the objective indexes, except for the metric *SD* in the SPECT-MRI image fusion problem. Also, Scheme 2 always ranks in the top three places for most of the objective indexes, except for the metric *VIFF* in the CT-MRI and SPECT-MRI image fusion problems. Compared with the eight other fusion methods, the two fusion schemes are not always the best, but they achieve stable performance on all four metrics. Therefore, our fusion framework performs comparably or even better than the compared methods.

In particular, compared with Scheme 1, Scheme 2 exhibits obvious advantages on the metrics Q_{FMI} and *SD*, indicating that Scheme 2 obtains better performance in preserving global features and energy of source images. For metric Q_G, Scheme 2 outperforms Scheme 1 in MR-T1-MR-T2 and SPECT-MRI image fusion problems. However, the metric *VIFF* of Scheme 2 is lower than that of Scheme 1 for all the three fusion problems.

The LP-SR method also demonstrates high objective performance in CT-MRI and MR-T1-MR-T2 fusion problems, but it shows inferior performance in the SPECT-MRI fusion problem. Also, the LP-CNN method achieves higher performance in the SPECT-MRI fusion problem and inferior performance in the two other medical image fusion problems.

The PAPCNN method performs well on the metric *VIFF*, this result also confirms that it achieves better visual effects. However, it exhibits poor performance on the metrics Q_{FMI} and Q_G because it focuses on the extraction of structural details and cannot always successfully identify the important salient features. Scheme 2 outperforms the PAPCNN method on the aforementioned metrics on all three image fusion methods, and shows a slight advantage on the metric *SD* in the CT-MRI and SPECT-MRI image fusion problems, indicating that Scheme 2 achieves better objective evaluation.

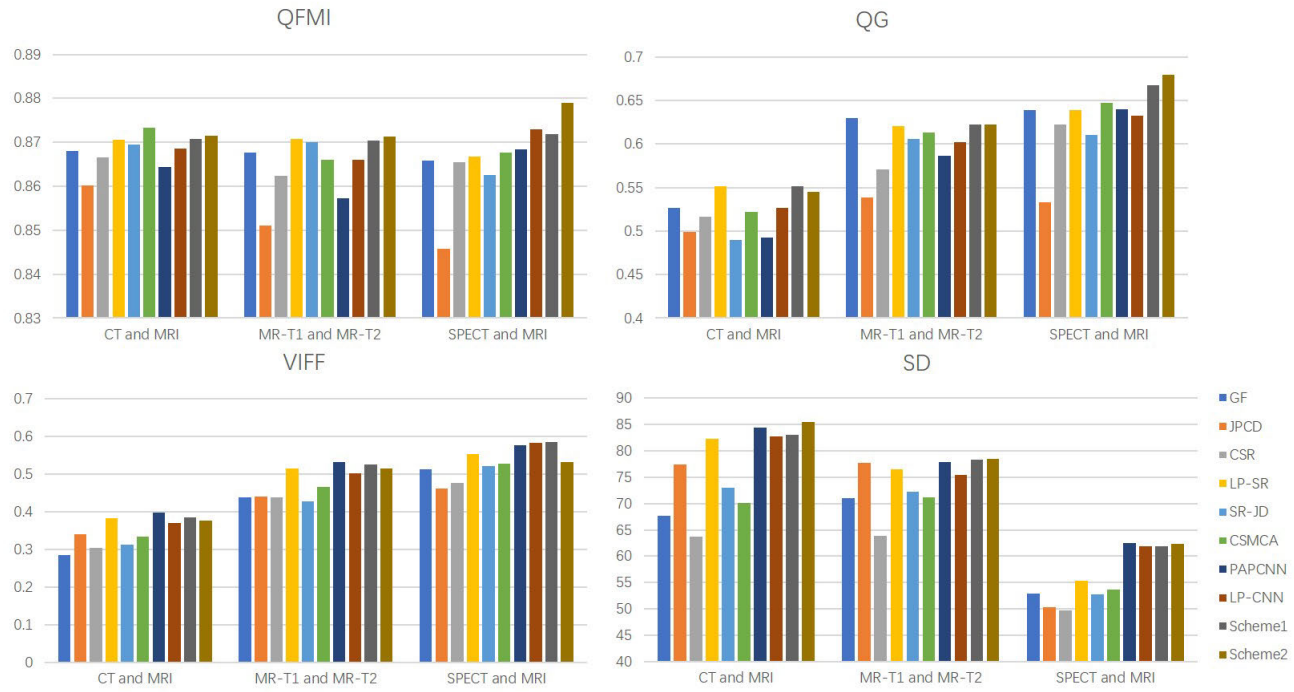


FIGURE 9. Quantitative indices of fusion results.

F. DISCUSSIONS

Table 1 clearly shows that Scheme 1 and Scheme 2 achieve better robustness in three different types of medical image fusion problems compared with the other fusion methods. Additionally, Scheme 2 exhibits higher performance over metrics Q_{FMI}, Q_G, and SD, implying that this scheme appears more comprehensive in extracting salient edge features and image energy. Therefore, the images fused by Scheme 2 appear more natural and are in accord with human visual perception. By combining the objective evaluation with the visual perception of the two schemes, Scheme 2 can be considered better than Scheme 1. Moreover, Figs. 6-8 show that the proposed framework can preserve the bone regions well and nearly all the abnormal details in the same region of the complementary images are filtered out. This fusion result is also in accord with the good performance in preserving salient information and less meaningless details.

In comparison to Scheme 1, Scheme 2 holds a small pyramid decomposition level and eliminates the normalization phase in the dictionary construction and sparse coding stages, in which the mean value of atoms or patches is subtracted, causing the sparse coding stage to merely process the structural details of low-frequency layers in terms of pixels. As a consequence, Scheme 1 pays more attention to the extraction of salient information and details, whereas Scheme 2 focuses more on the preservation of global features. Therefore, we can see from Figs. 6-8 that some artifacts exist in the fused images obtained by Scheme 1 since this scheme cannot identify the meaningful and significant features, resulting in unsatisfactory fusion results. It also can be expected is that Scheme 2 blurs some details, and this result agrees with the inferior visual effect caused by the ignorance of salient details.

Another notable problem is that the two schemes are easily prone to the loss of textures, especially for Scheme 2. Fig. 10 shows a typical example used in experiments for this defect. Clearly, the GF, LP-SR, SR-JD, CSMCA, LP-CNN methods preserve most of the texture details from the MR-T2 image, whereas the other fusion methods including Scheme 1 and Scheme 2 achieve poor performance in extracting the textures, which can be seen from left close-up in Fig. 10. The edge information seems to be strengthened by JPCD method, but it still loses textures. The PAPCNN method introduces a few meaningless details of MR-T1 image. Scheme 1 and Scheme 2 have flaws of textures loss when referring to the MR-T2 image [see the left close-up in Fig. 10 (k) and the right close-up in Fig. 10 (l)]. This is mainly because almost all the texture details in the MRI image are retained in the detail layer. Since the Laplacian operator followed by the Gaussian filtering provides superior performance in highlighting edges or gradients. These textures and noises are processed using a guided filter-based scheme, which can be essentially considered the extraction of edge features and the smoothing of textures with noises. Consequently, the areas with high gradient are preserved and the textures are blurred. Hence, our framework can efficiently protect bone features from the introduction of meaningless details at the sacrifice of a few textures.

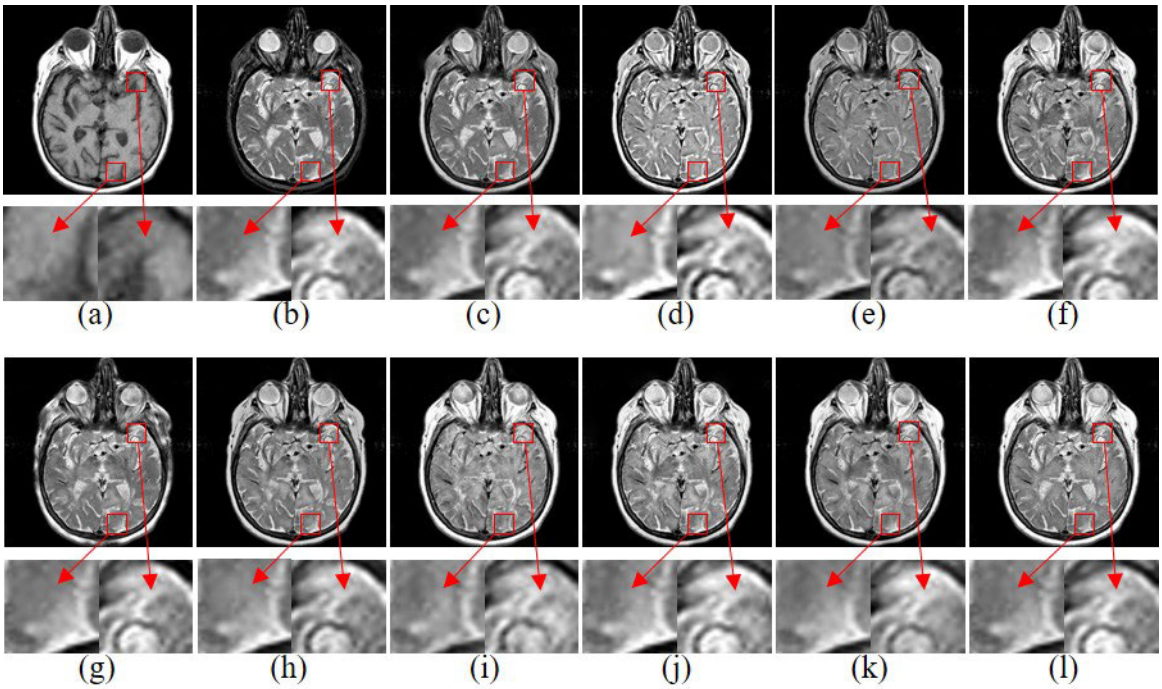


FIGURE 10. A typical example of the defect of our fusion framework. (a)-(l) refer to the source images and fusion results by different methods: (a) MR-T1 image; (b) MR-T2 image; (c) GF; (d) JPCD; (e) CSR; (f) LP-SR; (g) SR-JD; (h) CSMCA; (i) PAPCNN; (j) LP-CNN; (k) Scheme1; (l) Scheme2.

TABLE 2. Average time-consumption of different methods when fusing two images of size 256 × 256 pixels (Units:Seconds).

Methods	GF	JCPD	CSR	LP-SR	SR-JD	CSMCA	PAPCNN	LP-CNN	Scheme1	Scheme2
Times	0.06	47.93	29.49	0.09	531.14	101.14	9.71	13.57	322.53	334.46

In particular, Scheme 2 focuses on preserving global features and thus appears more serious than Scheme 1 in terms of losing details, which in accord with a lower score over metric *VIFF*. To address this disadvantage, applying effective contrast enhancement schemes to the detail layer may be helpful for enhancing texture details.

G. COMPUTATIONAL EFFICIENCY

The analysis of the computational efficiency of different fusion methods is compared in this section. The experimental environment is MATLAB R2014b on a computer equipped with an Intel(R) Core (TM) i7-8750H CPU (2.20 GHz) and 8GB RAM installed on a Win 10 64-bit operating system. All the source images in the fusion experiments are used to compute the average running time. Table 2 provides the average running time of ten different methods. As shown in Table 2, the GF and LP-SR methods require less running time than the other fusion methods. The PAPCNN and LP-CNN methods also exhibit good computational efficiency due to their simplicity of implementation. The JPCD method uses a novel online dictionary learning method and takes less time to form a fused image compared with the proposed method. The SR-JD method has the lowest computational efficiency because of its time-consuming joint dictionary construction method. The CSMCA method achieves competitive performance at the expense of running time. In particular, the proposed method practically uses a pre-trained dictionary that is required to construct a dictionary before the fusion process. It requires more running time, which is primarily attributed to the time-consuming dictionary learning process, in which the sliding window scheme with the sliding length of one pixel is used to divide source images into patches. In conclusion, the proposed method produces superior fused images at the cost of running time.

V. CONCLUSION

A two-scale multimodal medical image framework based on guided filtering and sparse representation is presented in this study. The proposed method utilizes a guided filter for two-scale decomposition to improve edge information. Then, the LP-SR and guided filtering-based rules are applied to merge the base and detail layers, respectively.

To enhance the sparse representation performance for the pyramid-decomposed low-frequency layer, a spatial degraded dictionary is learned from patches of source images by using a selection-based rule. The fused image is finally obtained by integrating the fused base and detail layers. In particular, two schemes under this framework are proposed to identify the best scheme on the basis of preserving salient features and retaining less meaningless details. Extensive experiments on three different types of medical image pairs are conducted. Compared with state-of-the-art fusion methods, the proposed fusion framework shows strong robustness and achieves competitive performance in the visual effect and objective evaluation of three types of medical images because it extracts accurately salient information and less abnormal details. We theoretically analyzed the significant difference between the two schemes and verified by conducting comparison experiments, which indicates Scheme 2 is better than Scheme 1 in terms of both the visual effect and objective evaluation. However, these two schemes perform well at the expense of losing textures due to the inherent denoising process that is applied in the detail layer. In the future, we will explore effective contrast enhancement methods to improve the texture details. Another notable problem is that the presented framework is time-consuming due to the dictionary learning process. Therefore, some online dictionary learning scheme for medical images can be designed.

REFERENCES

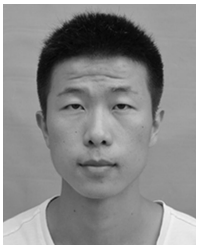
[1] X. Liu, W. Mei, and H. Du, "Structure tensor and nonsubsampled shearlet transform based algorithm for CT and MRI image fusion," *Neurocomputing*, vol. 235, pp. 131–139, Apr. 2017.
[2] J. Du, W. Li, and H. Tan, "Intrinsic image decomposition-based grey and pseudo-color medical image fusion," *IEEE Access*, vol. 7, pp. 56443–56456, May 2019.
[3] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.
[4] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
[5] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, May 1995.
[6] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel-and region-based image fusion with complex wavelets," *Inf. Fusion*, vol. 8, no. 2, pp. 119–130, Apr. 2007.
[7] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, Apr. 2007.
[8] Q. Zhang and B.-L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, Jul. 2009.
[9] M. Yin, W. Liu, X. Zhao, Y. Yin, and Y. Guo, "A novel image fusion algorithm based on nonsubsampled shearlet transform," *Optik*, vol. 125, no. 10, pp. 2274–2282, May 2014.
[10] X. Jin, G. Chen, J. Hou, Q. Jiang, D. Zhou, and S. Yao, "Multimodal sensor medical image fusion based on nonsubsampled shearlet transform and S-PCNNs in HSV space," *Signal Process.*, vol. 153, pp. 379–395, Dec. 2018.
[11] M. Nejati, S. Samavi, N. Karimi, S. M. R. Soroushmehr, S. Shirani, I. Roosta, and K. Najarian, "Surface area-based focus criterion for multi-focus image fusion," *Inf. Fusion*, vol. 36, pp. 284–295, Jul. 2017.
[12] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.
[13] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[14] W. Li, L. Jia, and J. Du, "Multi-modal sensor medical image fusion based on multiple salient features with guided image filter," *IEEE Access*, vol. 7, pp. 173019–173033, 2019.
[15] S. Li, B. Yang, and J. Du, "Multifocus image fusion by combining curvelet and wavelet transform," *IEEE Access*, vol. 29, no. 9, pp. 1295–1301, Jul. 2008.
[16] S. Li and B. Yang, "Hybrid multiresolution method for multisensor multimodal image fusion," *IEEE Sensors J.*, vol. 10, no. 9, pp. 1519–1526, Sep. 2010.
[17] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.
[18] 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 Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.
[19] L. Jian, X. Yang, Z. Zhou, K. Zhou, and K. Liu, "Multi-scale image fusion through rolling guidance filter," *Future Gener. Comput. Syst.*, vol. 83, pp. 310–325, Jun. 2018.
[20] T. Ma, J. Ma, B. Fang, F. Hu, S. Quan, and H. Du, "Multi-scale decomposition based fusion of infrared and visible image via total variation and saliency analysis," *Infr. Phys. Technol.*, vol. 92, pp. 154–162, Aug. 2018.
[21] X. Meng, J. Li, H. Shen, L. Zhang, and H. Zhang, "Pansharpening with a guided filter based on three-layer decomposition," *Sensors*, vol. 16, no. 7, pp. 1068–1082, Jul. 2016.
[22] J. Zhu, W. Jin, L. Li, Z. Han, and X. Wang, "Multiscale infrared and visible image fusion using gradient domain guided image filtering," *Infr. Phys. Technol.*, vol. 89, pp. 8–19, Mar. 2018.
[23] K. Padmavathi, C. S. Asha, and V. K. Maya, "A novel medical image fusion by combining TV-L1 decomposed textures based on adaptive weighting scheme," *Eng. Sci. Technol., Int. J.*, vol. 23, no. 1, pp. 225–239, Feb. 2020.
[24] Y. Liu, D. Zhou, R. Nie, R. Hou, Z. Ding, Y. Guo, and J. Zhou, "Robust spiking cortical model and total-variational decomposition for multimodal medical image fusion," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 101996.
[25] C. Liu, X. Wang, and J. Mao, "Research on multi-focus image fusion algorithm based on total variation and quad-tree decomposition," *Multimedia Tools Appl.*, vol. 79, nos. 15–16, pp. 10475–10488, Apr. 2020.
[26] C. Xing, M. Wang, C. Dong, C. Duan, and Z. Wang, "Using Taylor expansion and convolutional sparse representation for image fusion," *Neurocomputing*, vol. 402, pp. 437–455, Aug. 2020.
[27] F. Luo, H. Huang, Y. Duan, J. Liu, and Y. Liao, "Local geometric structure feature for dimensionality reduction of hyperspectral imagery," *Remote Sens.*, vol. 9, no. 8, pp. 790–812, Aug. 2017.
[28] F. Feng, W. Li, Q. Du, and B. Zhang, "Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity," *Remote Sens.*, vol. 9, no. 4, pp. 323–336, Mar. 2017.
[29] S. Maqsood and U. Javed, "Multi-modal medical image fusion based on two-scale image decomposition and sparse representation," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101810.
[30] J. Du and W. Li, "Two-scale image decomposition based image fusion using structure tensor," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 2, pp. 271–284, Jun. 2020.
[31] 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, Feb. 2019.
[32] F. Shabanzade, M. Khateri, and Z. Liu, "MR and PET image fusion using nonparametric Bayesian joint dictionary learning," *IEEE Sensors Lett.*, vol. 3, no. 7, pp. 1–4, Jul. 2019.
[33] F. Zhou, X. Li, M. Zhou, Y. Chen, and H. Tan, "A new dictionary construction based multimodal medical image fusion framework," *Entropy*, vol. 21, no. 3, p. 267, Mar. 2019.
[34] N. Aishwarya, H. Thangammal, and C. Bennila, "A novel multimodal medical image fusion using sparse representation and modified spatial frequency," *Int. J. Imag. Syst. Technol.*, vol. 28, no. 3, pp. 175–185, Sep. 2018.
[35] J.-J. Zong and T.-S. Qiu, "Medical image fusion based on sparse representation of classified image patches," *Biomed. Signal Process. Control*, vol. 34, pp. 195–205, Apr. 2017.
[36] Q. Zhang, Y. Liu, R. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Inf. Fusion*, vol. 40, pp. 57–75, Mar. 2018.

[37] Z. Zhu, Y. Chai, H. Yin, Y. Li, and Z. Liu, "A novel dictionary learning approach for multi-modality medical image fusion," *Neurocomputing*, vol. 214, pp. 471–482, Nov. 2016.
[38] M. Kim, D. K. Han, and H. Ko, "Joint patch clustering-based dictionary learning for multimodal image fusion," *Inf. Fusion*, vol. 27, pp. 198–214, Jan. 2016.
[39] H. Yin, Y. Li, Y. Chai, Z. Liu, and Z. Zhu, "A novel sparse-representation-based multi-focus image fusion approach," *Neurocomputing*, vol. 216, pp. 216–229, Dec. 2016.
[40] Q. Hu, S. Hu, and F. Zhang, "Multi-modality medical image fusion based on separable dictionary learning and Gabor filtering," *Signal Process., Image Commun.*, vol. 83, Apr. 2020, Art. no. 115758.
[41] 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 Recognit.*, vol. 79, pp. 130–146, Jul. 2018.
[42] 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 Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1082–1102, Apr. 2020.
[43] H. Li, X. He, Z. Yu, and J. Luo, "Noise-robust image fusion with low-rank sparse decomposition guided by external patch prior," *Inf. Sci.*, vol. 523, pp. 14–37, Jun. 2020.
[44] Y. Zhang, M. Yang, N. Li, and Z. Yu, "Analysis-synthesis dictionary pair learning and patch saliency measure for image fusion," *Signal Process.*, vol. 167, Feb. 2020, Art. no. 107327.
[45] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.
[46] K.-J. Xia, H.-S. Yin, and J.-Q. Wang, "A novel improved deep convolutional neural network model for medical image fusion," *Cluster Comput.*, vol. 22, no. S1, pp. 1515–1527, Jan. 2019.
[47] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.
[48] J. Huang, Z. Le, Y. Ma, F. Fan, H. Zhang, and L. Yang, "MGMDcGAN: Medical image fusion using multi-generator multi-discriminator conditional generative adversarial network," *IEEE Access*, vol. 8, pp. 55145–55157, 2020.
[49] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–7.
[50] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A unified densely connected network for image fusion," in *Proc. 34th Conf. Artif. Intell. (AAAI)*, Apr. 2020, pp. 12484–12491.
[51] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity," in *Proc. 34th Conf. Artif. Intell. (AAAI)*, Apr. 2020, pp. 12797–12804.
[52] M. Aharon, M. Elad, and A. Bruckstein, "*K*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[53] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Jan. 1993.
[54] S. Singh and R. S. Anand, "Multimodal medical image sensor fusion model using sparse K-SVD dictionary learning in nonsubsampled shearlet domain," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 593–607, Feb. 2020.
[55] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 733–740.
[56] Q. Wang, W. Zheng, and R. Piramuthu, "GraB: Visual saliency via novel graph model and background priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 535–543.
[57] X. Li, L. Zhao, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling, and J. Wang, "DeepSaliency: Multi-task deep neural network model for salient object detection," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3919–3930, Aug. 2016.
[58] K. A. Johnson and J. A. Becker. (2011). [Online]. Available:. [Online]. Available: http://www.med.harvard.edu/aanlib/
[59] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.
[60] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 485–489, Mar. 2019.
[61] M. Haghighat and M. A. Razian, "Fast-FMI: Non-reference image fusion metric," in *Proc. IEEE 8th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Oct. 2014, pp. 424–426.
[62] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Mar. 2000.
[63] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Electron. Lett.*, vol. 14, no. 2, pp. 127–135, Apr. 2013.
[64] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.

CHUNYANG PEI (Graduate Student Member, IEEE) received the B.S. degree in automation from the Shandong University of Science and Technology, Tai'an, China, in 2018. He is currently pursuing the M.S. degree in control science and engineering with the Jiangxi University of Science and Technology, Ganzhou, China. His current research interests include image fusion, sparse representation, and image processing.

KUANGANG FAN (Member, IEEE) was born in Linyi, China, in 1981. He received the B.S., M.S., and Ph.D. degrees in instrumentation science from Jilin University, in June 2006, June 2008, and June 2011, respectively.

From 2012 to 2014, he held a postdoctoral position at the State Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. From 2015 to 2016, he was a Visiting Scholar with the School of Electronics and Computer Engineering, Peking University Shenzhen Graduate School. From 2018 to 2019, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, UC Davis, Davis, CA, USA. He is currently an Associate Professor of electrical and computer engineering with the Jiangxi University of Science and Technology, China. He has published over 30 refereed articles and book chapters, and holds more than 30 invention patents. His research contributions cover a broad range of signal processing and control engineering, including blind channel estimation and equalization, source separation, parameter estimation, and adaptive control.



WENSHUAI WANG (Student Member, IEEE) received the B.S. degree in measurement and control technology and instrument from the Jiangxi University of Science and Technology, Ganzhou, China, in 2019, where he is currently pursuing the M.S. degree in instruments science and technology. His current research interests include digital signal processing and image processing.