



Brain CT and MRI medical image fusion using convolutional neural networks and a dual-channel spiking cortical model

Ruichao Hou¹ · Dongming Zhou¹ · Rencan Nie¹ · Dong Liu¹ · Xiaoli Ruan¹

Received: 12 June 2018 / Accepted: 17 November 2018 / Published online: 23 November 2018
© International Federation for Medical and Biological Engineering 2018

Abstract

The aim of medical image fusion is to improve the clinical diagnosis accuracy, so the fused image is generated by preserving salient features and details of the source images. This paper designs a novel fusion scheme for CT and MRI medical images based on convolutional neural networks (CNNs) and a dual-channel spiking cortical model (DCSCM). Firstly, non-subsampled shearlet transform (NSST) is utilized to decompose the source image into a low-frequency coefficient and a series of high-frequency coefficients. Secondly, the low-frequency coefficient is fused by the CNN framework, where weight map is generated by a series of feature maps and an adaptive selection rule, and then the high-frequency coefficients are fused by DCSCM, where the modified average gradient of the high-frequency coefficients is adopted as the input stimulus of DCSCM. Finally, the fused image is reconstructed by inverse NSST. Experimental results indicate that the proposed scheme performs well in both subjective visual performance and objective evaluation and has superiorities in detail retention and visual effect over other current typical ones.

Keywords Image fusion · Non-subsampled shearlet transform · Convolutional neural networks · Dual-channel spiking cortical model

1 Introduction

With the rapid development of sensor and computer science technology, medical imaging has been playing an essential role in various clinical applications including medical diagnosis, surgical navigation, and treatment planning, which is a critical tool for the doctors to diagnose the diseases accurately [1].

Commonly, medical images are generated by different imaging mechanisms, which are focused on specific tissue or organ information, such as X-ray, computed tomography (CT), and magnetic resonance imaging (MRI). The CT images are used for the precise localization of dense structures like bones and implants, the MRI images can provide enough soft-tissue details with high-resolution anatomical information [2]. The main task of image fusion is to generate a single comprehensive image containing the unique characteristics of

multimodal medical images, which can help doctors to make accurate decisions for various diagnoses [3].

Over the last few years, multiscale transform (MST) methods applied to image fusion have been studied extensively. Among the conventional tools of MST, we can mention discrete wavelet transform (DWT) [4], Laplacian pyramid (LAP) [5], and contourlet transform (CT) [6]. To achieve better frequency selectivity and regularity than CT and remove pseudo-Gibbs phenomena along the edges to some extent, non-subsampled contourlet transform (NSCT) was proposed by Da Cunha et al. [7]. In comparison with other decomposition methods, NSCT requires a larger amount of computation. To reduce the computational complexity of NSCT, non-subsampled shearlet transform (NSST) was proposed by Zhang et al. [8]; NSST has the shift-invariance of non-subsampled processes and inherits the main properties from shearlet and wavelet, such as the characteristics of anisotropy and computing speed. Therefore, NSST has a significant advantage in obtaining more details for image fusion.

In addition to choose the excellent decomposition method, how to generate a robust weight map is also the key step of image fusion. In conventional transform domain fusion methods, the weight maps were generated by the simple

✉ Dongming Zhou
zhoudm@ynu.edu.cn

¹ Information College, Yunnan University, Kunming 650504, China

fusion rules such as weighted-average or choose-max. This kind of fusion rules does not consider the relationship between pixels which reduce the contrast of fused image and lose saliency information on a certain degree. To get the better fusion performance, the methods based on pulse-coupled neural network (PCNN) have also become a research hotspot. PCNN owns some superior characteristics, such as coupling and pulse synchronization which can be used to generate fused coefficients [9]. X. Jin et al. [10] proposed an image fusion method based on NSST and PCNN. K. J. He et al. [11] introduced a fusion method which combines focus-region-level partition and PCNN. However, PCNN has a large number of parameters which are always set as constants by human experience leading to the lack of universality. To address these problems, a modified neural network model called spiking cortical model (SCM) was proposed by Hou et al. [12], which devised a novel scheme based on SCM and NSST and overcame the shortcoming of parameters setting, and utilized the saliency map to fuse low-frequency coefficients. However, this algorithm has a certain limitation that saliency detection method only achieves outstanding performance on visible and infrared images.

In recent years, deep learning has gained many breakthroughs in various computer vision and image processing problems, such as image segmentation, super resolution restoration, classification, saliency detection, and so on [13]. Y. Liu et al. [14] proposed a novel multi-focus image fusion scheme using convolutional neural networks (CNNs); they used CNNs to classify the focus region and get a decision map. The fused image was generated by combining decision map and source images. Although deep learning achieves good performance, the limitation of this method is that it is just suitable for multi-focus image fusion. Then Y. Liu et al. [1] extended the CNNs model to medical image fusion which acquired a good effect.

In this paper, we propose a novel medical image fusion scheme based on deep learning framework and improved artificial neural network. In the beginning, NSST decomposes the source image into a low-frequency coefficient and a series of high-frequency coefficients. The main task in this paper is to design robust activity level measurements and weight assignment strategies, so the CNNs are used to encode a direct mapping from the source image to weight map. To enhance adaptability of the algorithm for different images, we design an effective weight assignment rule. Then the low-frequency coefficients are fused by deep learning framework. For high-frequency coefficients, we proposed a dual-channel SCM (DCSCM) to fuse the decomposed coefficients. Finally, the fused image is obtained via inverse NSST. Experimental results show that the proposed method does well in the fusion of medical images and can preserve not only the dense structures information of the CT image but also the soft-tissue information of the

MRI image; thus, the result contains rich details and has a good visual effect.

The remaining sections of this paper are summarized as follows. Section 2 reviews the theory of related algorithms and describes the image fusion strategies and steps in detail. Experimental results are given in Section 3. Section 4 shows the discussion about the experimental results. Some conclusions are summarized in Section 5.

2 Methods

In this section, we briefly review the theory of NSST, CNNs, and DCSCM, which are essential components of the proposed method.

2.1 Non-subsampled shearlet transform

NSST, which was proposed by Zhang et al. [8], is an extension of the wavelet in multidimensional space and combines the non-subsampled Laplacian pyramid (NSLP) filter with shearlet transform to provide the multiscale decomposition. The shearlet transform (ST) is close to optimal sparse representation; the synthetic expansion of affine system is described as follows:

$$A_{AB}(\psi) = \left\{ \psi_{j,l,k}(x) = |\det A|^{j/2} \psi(B^l A^j x - k) : j, l \in \mathbb{Z}, k \in \mathbb{Z}^2 \right\} \quad (1)$$

where $\psi_{j,l,k}$ is expressed as a composite wavelet, A denotes the anisotropy matrix for multiscale decomposition, B is a shear matrix for directional analysis, and j, l and k are scale, the direction of decomposition and shift parameter, respectively. When $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, the composite wavelet transformed into shearlet, the structure of the frequency tiling by the shearlet is shown in Figs. 1 and 2, which show three-level multiscale and multidirectional decomposition of the NSST.

The NSST decomposition is divided into two major steps:

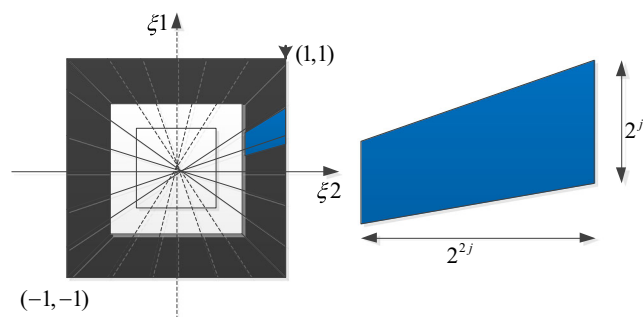
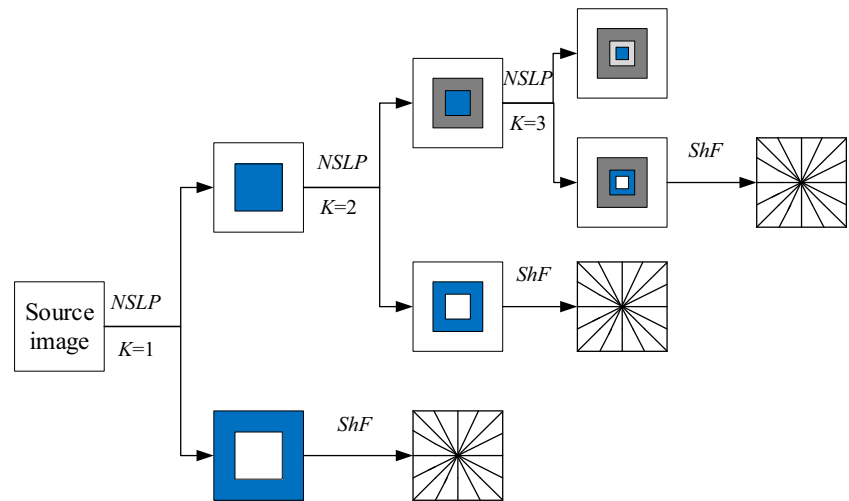


Fig. 1 The structure of the frequency tiling by the shearlet

Fig. 2 Three level multiscale and multidirectional decomposition of the NSST



- (I) Multiscale decomposition. $(k + 1)$ sub-bands as same size as the source image can be obtained by using the k -class non-subsampled pyramid filter, including a low-frequency map and a series of high-frequency maps.
- (II) The direction of localization. In pseudo polarization grid coordinates, standard shearlet is calculated by Meyer window function, which requires the subsampled operation to obtain the shift-invariance. However, NSST direction of localization uses the modified shearlet filter, which can map from the pseudo polarization to the Cartesian coordinate system avoiding the next sampling operation via Fourier inverse transform, so NSST has the characteristic of the shift-invariance.

2.2 Convolutional neural networks

Recently, CNNs have shown impressive performance across various artificial intelligence tasks. While CNNs have achieved state-of-the-art results in many high-level computer vision tasks like classification, object detection, scene understanding, and much more, their performance on low-level image processing problems such as filtering and image fusion is not studied extensively [15].

CNNs become a new type of an artificial neural network model, which are combining artificial neural network and deep learning network. The convolution layer is the key to construct the convolutional neural network which is defined as follows.

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} * w_{ij}^l + b_j^l \right) \tag{2}$$

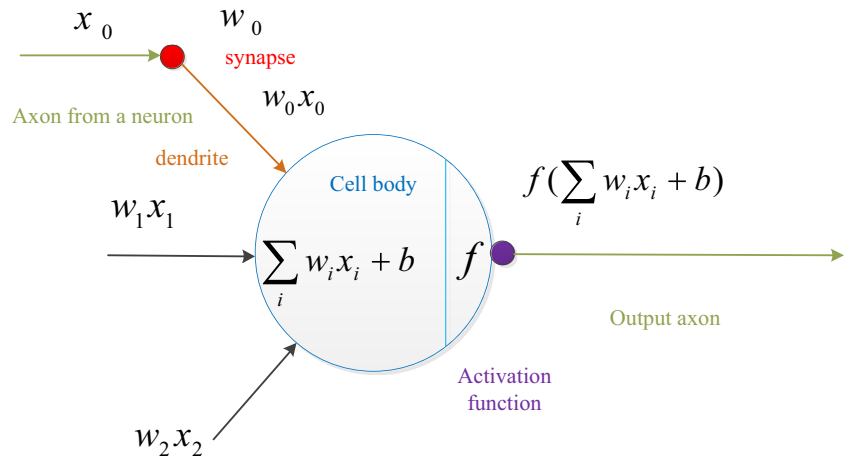
where l denotes l -th layer, w denotes the convolution kernel, M_j is the receptive field of the input layer, b is the bias, $*$ denotes the convolution operation, and f represents the activation function.

The structure of neuron is shown in Fig. 3. Firstly, each neuron connected via synapses and neurons capture input signals from its dendrites, then the dendrites would transmit the signals to the cell body, eventually along the axons to produce the output signal. The proper activation function is an important part of neural network, Eq. (2) can be rewritten as Eq. (3) by incorporating the rectified linear unit (ReLU) activation function [16].

$$x_j^l = \max \left(\sum_{i \in M_j} x_i^{l-1} * w_{ij}^l + b_j^l, 0 \right) \tag{3}$$

VGG network increases the depth of network to achieve a better performance. Gatys et al. [17] proposed a novel image style transfer technology based on VGG network. Firstly, the VGG-19 network was used to extract the deep feature in multi-layers from content and style images, then the loss function of style and content was defined, and the generated image was achieved after a certain number of iterative training, which fuses the style and content of the source images respectively. There is no doubt that VGG network is an effective feature extractor that contains different information in each layer, and the structure of VGG-19 neural network is shown in Fig. 4. Inspired by the above work, the fixed VGG network in our paper which is trained on ImageNet (ILSVRC2012) to extract the multi-layer deep features of source medical images. Specifically, the training is carried out by optimizing the multinomial logistic regression objective using mini-batch gradient descent with momentum and the learning is stopped after 74 epochs [18]. Different from the style transfer and classification tasks, we do not need neural network to reconstruct images or output the probability of classification, but just utilize deep features of ReLU_1_1, ReLU_2_1, and ReLU_3_1 activation layer from the VGG-19 network to design the robust weight maps, and more details are described in Section 2.4.1.

Fig. 3 The structure of neuron



2.3 Dual-channel spiking cortical model

Conventional SCM was presented by K Zhan et al. [19] and has the simple structure and fewer parameters. It consists of multiple neurons, and each neuron contains three main function units: receptive field, modulation field, and pulse generator. Moreover, it does not need to learn or train, and it can extract the useful information from the complex background. We modified the conventional SCM into dual-channel SCM (DCSCM) which can enhance its ability to extract details in the dark region. The mathematical expressions of the model are as follows:

$$U_{ij}^1(n) = fU_{ij}^1(n-1) + S_{ij}^1 \sum_{kl} W_{kl} Y_{kl}(n-1) \quad (4)$$

$$U_{ij}^2(n) = fU_{ij}^2(n-1) + S_{ij}^2 \sum_{kl} W_{kl} Y_{kl}(n-1) \quad (5)$$

$$U_{ij}(n) = \max(U_{ij}^1(n), U_{ij}^2(n)) \quad (6)$$

$$E_{ij}(n) = gE_{ij}(n-1) + V_{\theta} Y_{ij}(n-1) \quad (7)$$

$$X_{ij}(n) = \frac{1}{1 + e^{(E_{ij} - U_{ij})}} \quad (8)$$

$$Y_{ij}(n) = \begin{cases} 1, & \text{if } X_{ij}(n) > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where n denotes the iteration times, (i, j) is the location of the image pixel, $S_{ij}(n)$ is the input excitation signal, and the superscript 1 and 2 represent channel 1 and channel 2 respectively. $U_{ij}(n)$ refers to the internal active state of the neuron,

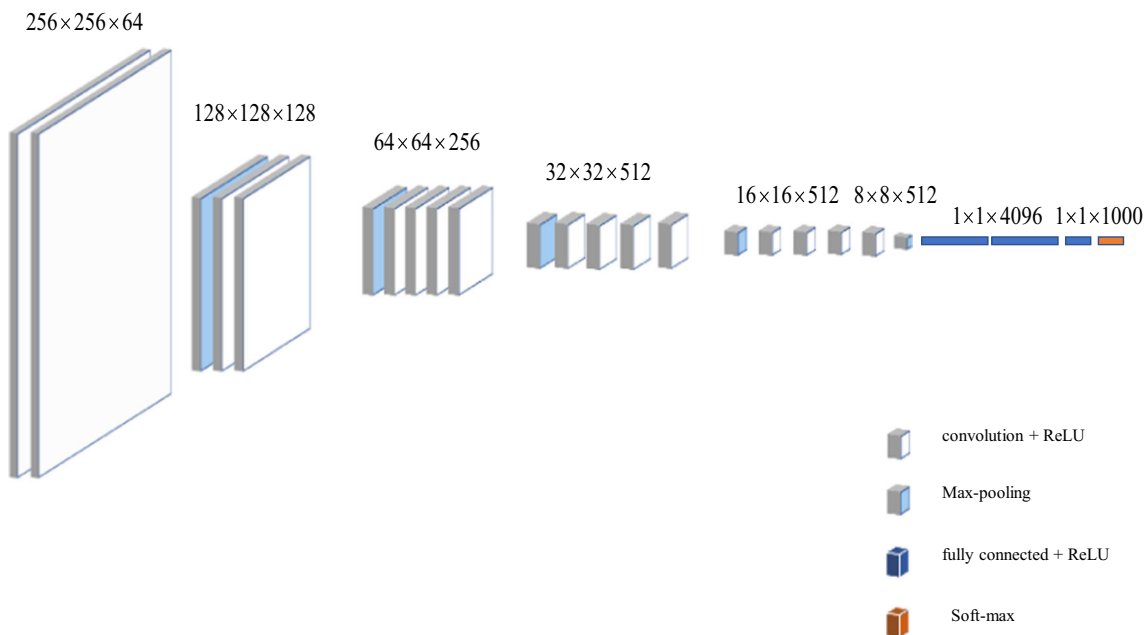


Fig. 4 The structure of VGG-19 neural network

and it depends on the maximum of $U_{ij}^1(n)$ and $U_{ij}^2(n)$, W_{kl} is the weighted coefficient matrix of linking between neurons, $E_{ij}(n)$ is the dynamic threshold, V_θ is the threshold of amplification factor, $Y_{ij}(n)$ is the output signal of the neuron at n th iteration, and f and g are the internal active and dynamic threshold signal decay coefficients, respectively.

In order to show the difference within ignition range, the sigmoid function is used to improve the neuron output signal [20], as shown in Eq. (8), $X_{ij}(n)$ denotes the pixel pulse ignition output amplitude, as $X_{ij}(n) > 0.5$, the neuron produces a pulse, which is called one firing time, the signal is captured by the linking matrix W_{kl} , and the adjacent neurons achieve synchronization pulse release at the spatial position. $T_{ij}(n)$ expresses the neuron firing times matrix after n th iteration, the structure of the basic SCM neuron is shown in Fig. 5, and the mathematical expression is described as follows.

$$T_{ij}(n) = \begin{cases} n, & Y_{ij}(n) = 1 \\ T_{ij}(n-1) + Y_{ij}(n), & otherwise \end{cases} \quad (10)$$

2.4 Fusion strategies and specific steps

2.4.1 Low-frequency coefficient fusion strategy

Commonly the part of low-frequency contains the main components of the source image. On the contrary, the high-frequency coefficients preserve the more details of the source image. The low-frequency coefficients of the source images are fused by the simple weighted averaging or maximum value-based strategies, which do not consider the relationship between pixels. To address the limitation of traditional fusion strategies, we use the VGG-19 model to extract the multi-layer features of source images, the weight maps are generated by

adaptive selection rule. Finally, the low-frequency coefficients are fused by the source images and weight maps. The low-frequency coefficient fusion framework is shown in Fig. 6.

We introduce this fusion strategy in details, $f_k^{n,m}$ denotes the feature map of k th source image at n th layer, and m is the dimension of the feature map, $m = 64 \times 2^{n-1}$, $k=2$, where F_n indicates the layer in VGG-19 network, $n \in \{1, 2, 3\}$ represents the ReLU_1_1, ReLU_2_1, and ReLU_3_1 activation layer respectively. $A_k^n(i, j)$ is the activity level map which is generated by l_1 -norm at position (i, j) , to make the fusion method robust to misregistration; the block-based average operator is used to calculate the final activity level map $\hat{A}_k^n(i, j)$, where r denotes the block size, to preserve more details, $r = 1$.

$$f_k^{n,m} = F_n(I_k) \quad (11)$$

$$A_k^n(x, y) = \left\| f_k^{n,m}(x, y) \right\|_1 \quad (12)$$

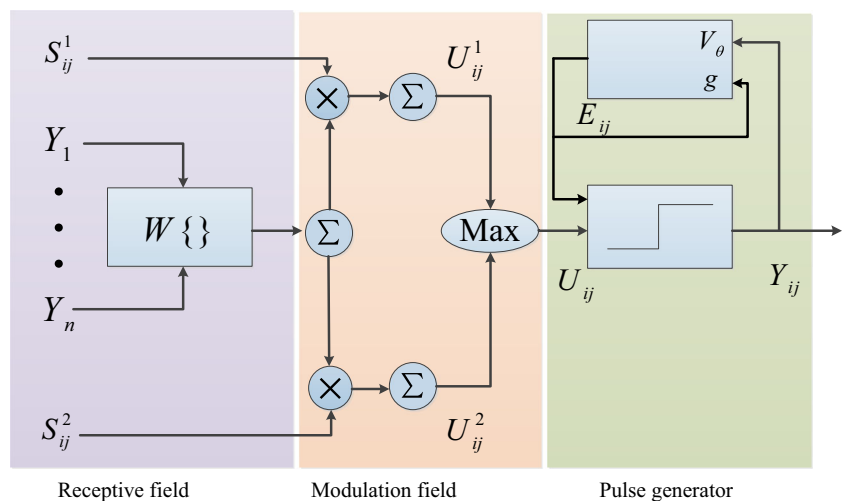
$$\hat{A}_k^n(i, j) = \frac{\sum_{\beta=-r}^r \sum_{\theta=-r}^r A_k^n(i + \beta, j + \theta)}{(2r + 1)^2} \quad (13)$$

Then we design an adaptive selection rule to make the weight maps more robust, $t_n(i, j)$ is the ratio of two activity level maps by Eq. (13), $W_1^n(i, j)$ is the weight map which denotes that if $t_n(i, j)$ tends to zero, the L_1 coefficient has a greater weight, and the same goes for $W_2^n(i, j)$. The VGG-19 has pooling operation; it is necessary to carry out the upsampling operator and resize the weight maps size consistent with source images size.

$$t_n(i, j) = \frac{\hat{A}_1^n(i, j)}{\hat{A}_2^n(i, j)} \quad (14)$$

$$W_1^n(i, j) = \frac{t_n^3(i, j)}{1 + t_n^3(i, j)} \quad (15)$$

Fig. 5 The structure of the basic SCM neuron



$$W_2^n(i, j) = \frac{1}{1 + I_n^3(i, j)} \tag{16}$$

$$\hat{W}_k^n(i + p, j + q) = W_k^n(i, j) \tag{17}$$

$$p, q \in \{0, 1, \dots, (2^{n-1} - 1)\} \tag{18}$$

Finally, we carry out the maximum value rule for the initial fused coefficients of the three layers, so as to merge them into the final fused coefficient; the specific expressions are as follows.

$$L_{Fused}^n(i, j) = L_1 \times \hat{W}_1^n(i, j) + L_2 \times \hat{W}_2^n(i, j) \tag{19}$$

$$L_{Fused}(i, j) = \max [L_{fused}^n(i, j), n \in \{1, 2, 3\}] \tag{20}$$

2.4.2 High-frequency coefficient fusion strategy

The existing high-frequency fusion strategies contain the largest absolute value, regional energy [21], variance, and gradient [22], but these strategies cannot extract detail information from the image adequately while only considering the individual pixels or regional characteristics. The gray value of a single pixel is used as the excitation of the neural network; this may lose image edges and texture features. Two diagonal gradient changes are added on the basis of the conventional method; it can be utilized to stimulate the DCSCM.

Suppose $H(i, j)$ denotes the high-frequency coefficients at the location (i, j) , and modified average gradient (MAG) is measured using slipping windows (the size is 3×3) of the coefficients, then MAG in each coefficient is used to motivate the neuron, and it is defined as follows:

$$MAG(i, j) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(\frac{\nabla H_h(i, j) + \nabla H_v(i, j) + \nabla H_{md}(i, j) + \nabla H_{vd}(i, j)}{2} \right)^{1/2} \tag{21}$$

$$\nabla H_h(i, j) = [H(i, j) - H(i, j - 1)]^2 \tag{22}$$

$$\nabla H_v(i, j) = [H(i, j) - H(i - 1, j)]^2 \tag{23}$$

$$\nabla H_{md}(i, j) = [H(i, j) - H(i - 1, j - 1)]^2 \tag{24}$$

$$\nabla H_{vd}(i, j) = [H(i, j) - H(i - 1, j + 1)]^2 \tag{25}$$

where $\nabla H_h(i, j)$, $\nabla H_v(i, j)$, $\nabla H_{md}(i, j)$, and $\nabla H_{vd}(i, j)$ denote the gradient changes in the horizontal, vertical, main diagonal, and oblique diagonal directions, respectively. N and M are the size of the slipping window.

2.4.3 Fusion steps

Assume that the CT and MRI images have been matched and treated with uniform size accurately.

The image fusion scheme is shown in Fig. 7.

The steps of the image fusion method are narrated as follows.

- Step 1. Decompose the CT and MRI images using NSST to obtain their low-frequency coefficients $\{L_1, L_2\}$ and a series of high-frequency coefficients $\{H_1^{l,k}, H_2^{l,k}\}$ at each K -scale and l -direction, where $1 \leq k \leq K$.
- Step 2. The deep learning framework is used to fuse the low-frequency coefficients. According to Eqs. (11)–(20), the weight maps are generated by the VGG-19 network which can choose the low-frequency coefficients adaptively.

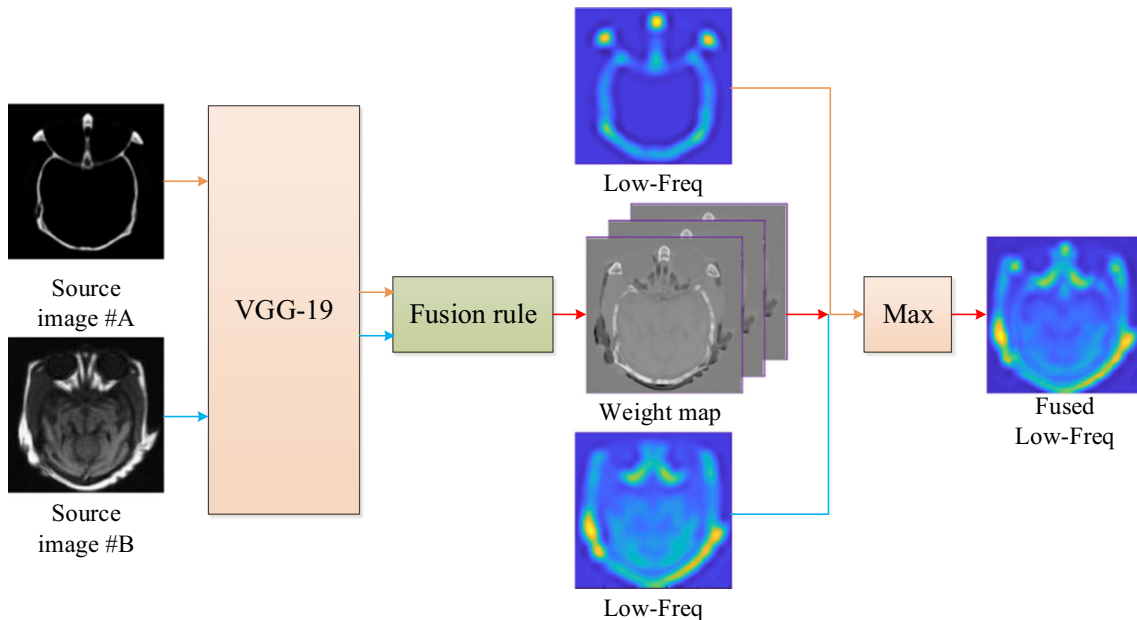


Fig. 6 The low-frequency coefficient fusion framework

Step 3. DCSCM is utilized to deal with the high-frequency coefficients. Let the MAG maps be the feedback inputs of the DCSCM.

- (a) Calculate the MAG_1 and MAG_2 maps according to Eq. (21), and all coefficients are normalized.
- (b) Set the initial values as follows: $U_{ij}(0) = T_{ij}(0) = E_{ij}(0) = 0$. In the initial state, all the neurons are inactivated, so $Y_{ij}(0) = 0$.
- (c) Calculate $U_{ij}(n)$, $E_{ij}(n)$, and $Y_{ij}(n)$ by Eq. (6), Eq. (7), and Eq. (9), respectively, and then compute the neuron’s firing times $T_{ij}(n)$ according to Eq. (10). The fusion coefficients are selected according to $U_{ij}(n)$, N is the maximum number of iterations, and the rule is described as follows:

$$H_{Fused}^K(i, j) = \begin{cases} H_1^K(i, j), & U_{ij}(N) = U_{ij}^1(N) \\ H_2^K(i, j), & U_{ij}(N) = U_{ij}^2(N) \end{cases} \quad (26)$$

Step 4. Perform the inverse NSST of the low-frequency and the high-frequency coefficients to obtain the fused image.

3 Results

The simulation experiments were conducted by MATLAB2017b and Python 2.7 software on PC with Intel

E5 2670 2.6 GHz CPU, 16 GB RAM, GTX1080ti GPU. We take several groups of accurate matching of CT image and MRI image to test, and a pair of CT and MRI images belongs to the same patient, as shown in Figs. 8, 9, 10, 11, and 12. The source medical images were collected from <http://www.med.harvard.edu/AANLIB/>.

3.1 Experiment parameters setting

In this section, extensive experiments on CT and MRI medical images are performed to verify the effectiveness of the proposed method. Our fusion method is compared with three representative conventional fusion methods and two state-of-the-art fusion methods including wavelet-based method (DWT) [4], Laplacian pyramid (LAP) [5], multiscale transform-based method (NSST-SCM) [23], sparse representation-based method (SR) [24], and guided filter-based fusion method (GFF) [25]. In the experiments, DWT uses “db2” as the filter; NSST uses a non-subsampling pyramid “maxflat” filter, and its decomposition directions are set as [2–4]. The fusion rule for low-frequency coefficient is averaging while the high-frequency coefficients are fused using absolute maximum choosing rule. According to [23] and artificial experience, the parameters of the SCM and DCSCM are set as follows:

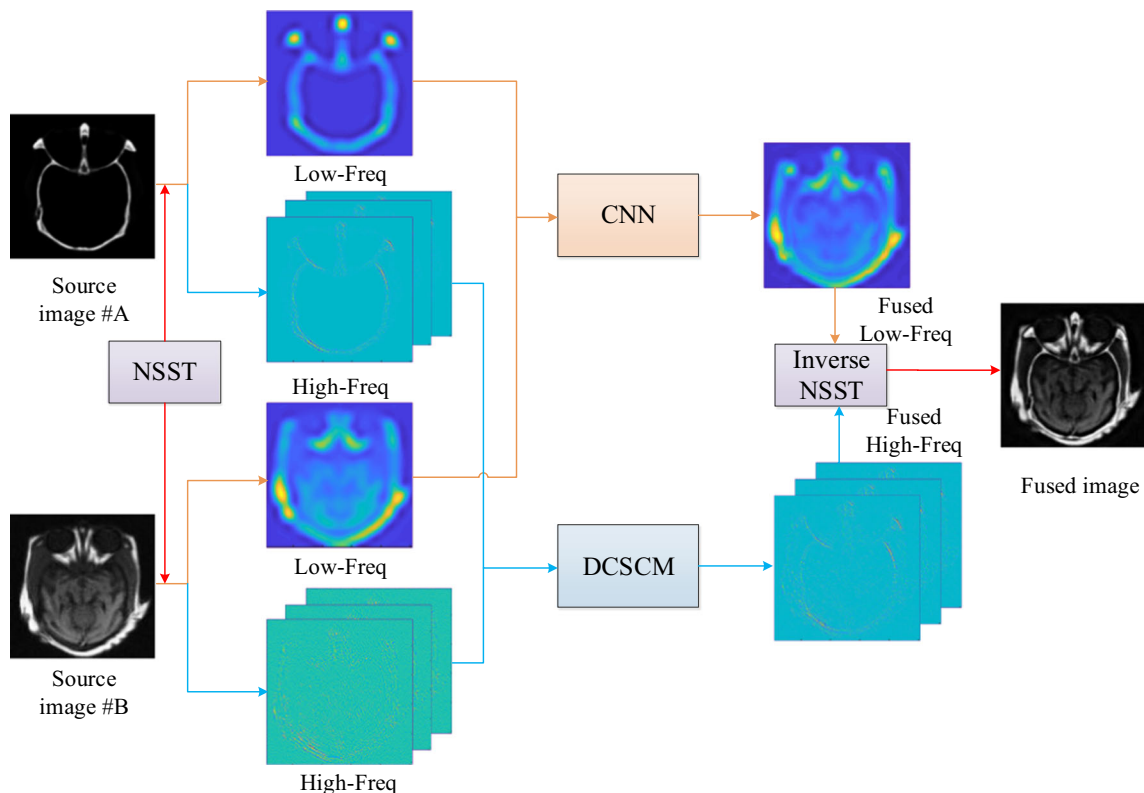


Fig. 7 Schematic diagram of the proposed image fusion framework

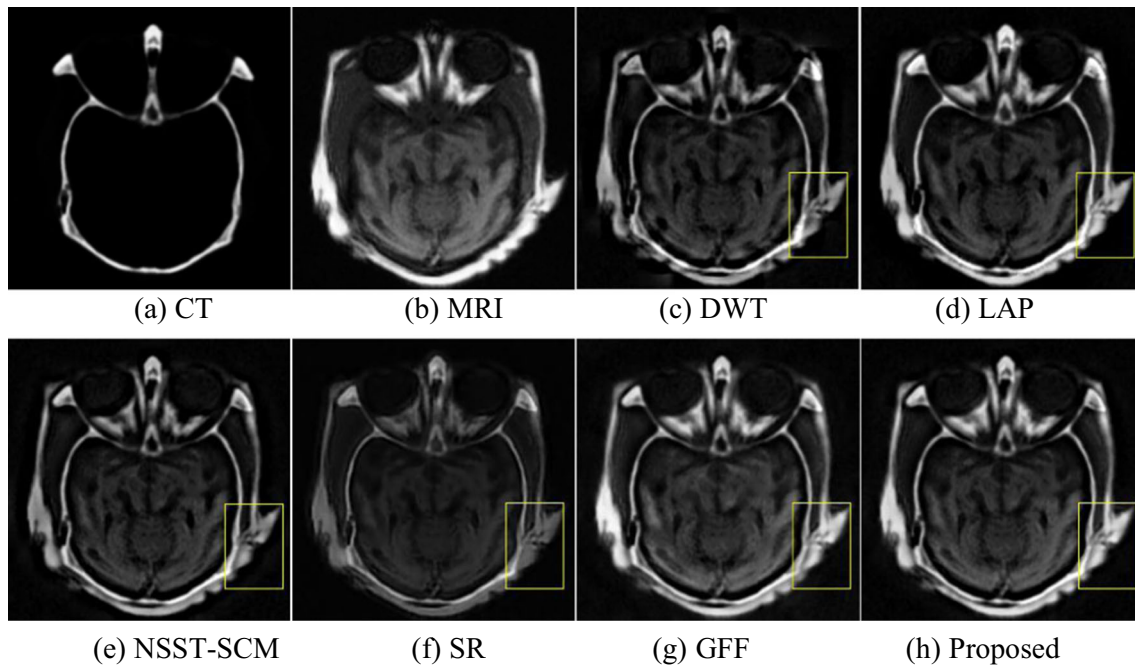


Fig. 8 Fusion results of the “Data-1” image set with different methods. **a** CT, **b** MRI, **c** DWT, **d** LAP, **e** NSST-SCM, **f** SR, **g** GFF, **h** proposed

$$f = 0.2, g = 0.6, V_{\theta} = 20, W$$

$$= \begin{bmatrix} 0.1091 & 0.1409 & 0.1091 \\ 0.1409 & 0 & 0.1409 \\ 0.1091 & 0.1409 & 0.1091 \end{bmatrix}.$$

3.2 Subjective evaluations

For convenience, five pairs of CT and MRI images respectively called “Data-1,” “Data-2,” “Data-3,” “Data-4,” and “Data-5” are selected as representative results to demonstrate the performance of the proposed fusion method. All of them cover 256 gray levels and have the same size 256×256 .

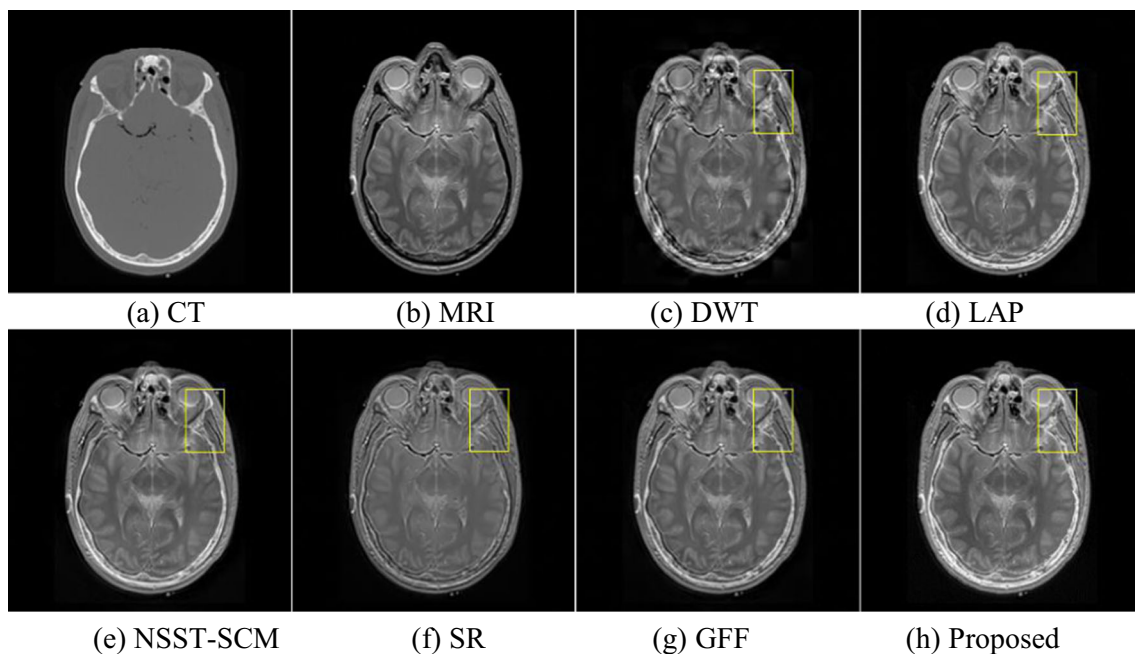


Fig. 9 Fusion results of the “Data-2” image set with different methods. **a** CT, **b** MRI, **c** DWT, **d** LAP, **e** NSST-SCM, **f** SR, **g** GFF, **h** proposed

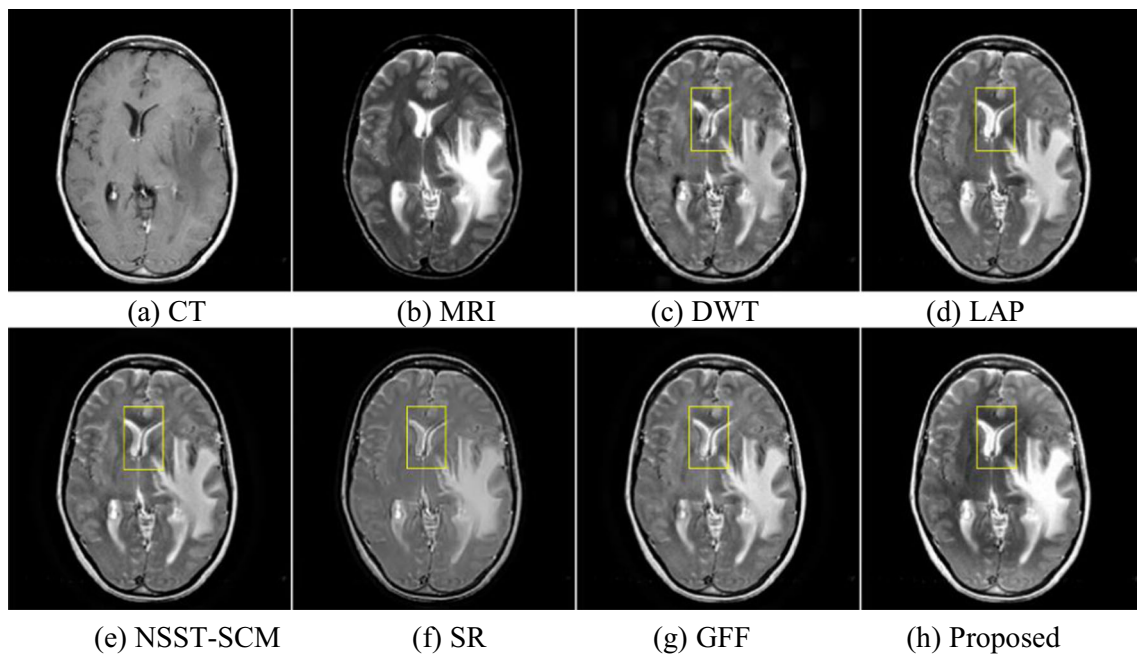


Fig. 10 Fusion results of the “Data-3” image set with different methods. **a** CT, **b** MRI, **c** DWT, **d** LAP, **e** NSST-SCM, **f** SR, **g** GFF, **h** proposed

The fusion results based on different methods for “Data-1” image set are shown in Fig. 8. The CT and MRI images are respectively shown in Fig. 8a and b. The fusion results obtained from DWT, LAP, NSST, SR, and the proposed method are represented in Fig. 8c–h, respectively. The fusion results mainly retain both the bone structures of CT image and soft tissues of MRI image. However, there are slight differences in contrast and detail preservation.

To show the difference of comparison methods more directly, we marked the experimental results with a yellow rectangle. As shown in Fig. 8c and f, the fused results have very low contrast in the yellow marked region. Although the fused images using LAP and NSST shown in Fig. 8d preserve more information of CT image, they lose some details of MRI image. In terms of visual effects, the performance of the GFF is similar to our proposed

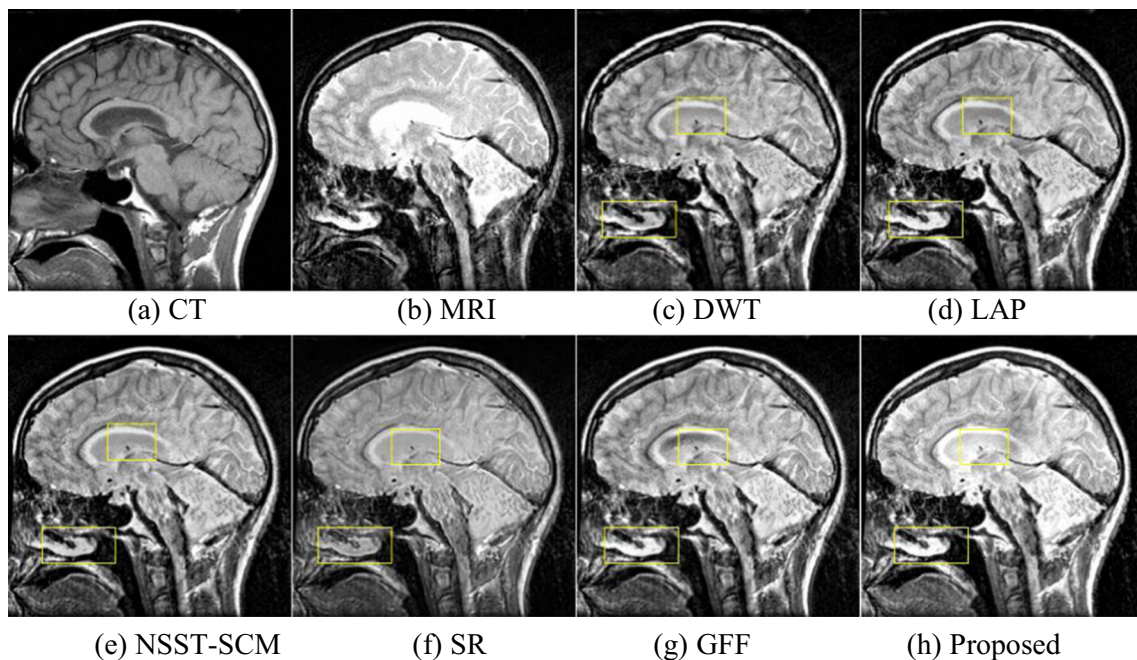


Fig. 11 Fusion results of the “Data-4” image set with different methods. **a** CT, **b** MRI, **c** DWT, **d** LAP, **e** NSST-SCM, **f** SR, **g** GFF, **h** proposed

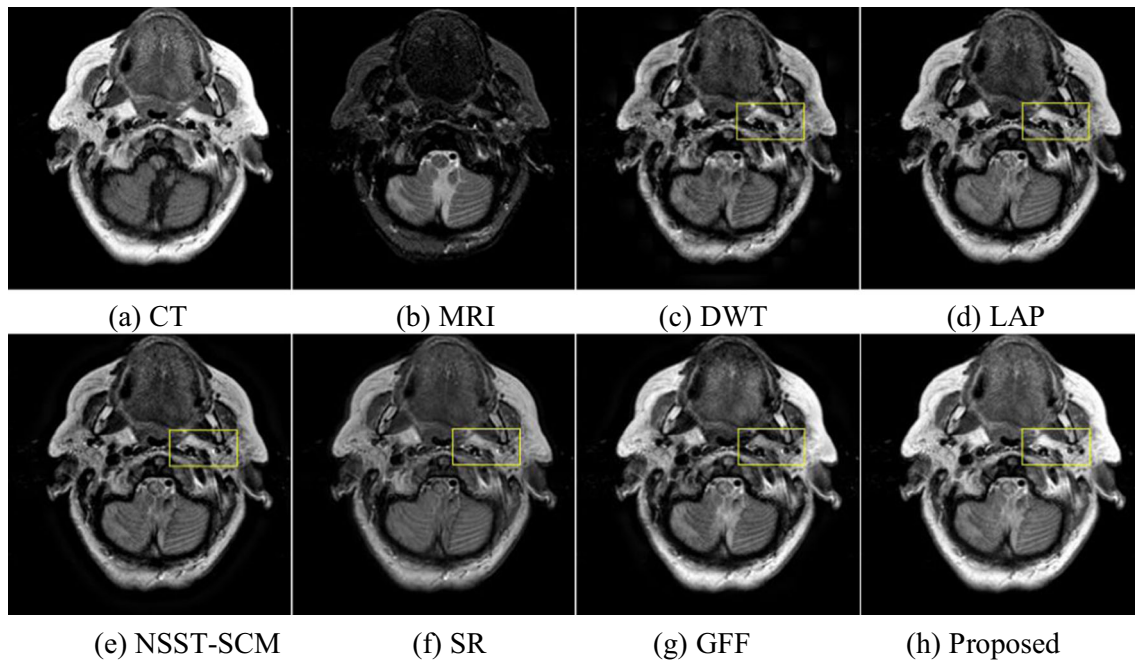


Fig. 12 Fusion results of the “Data-5” image set with different methods. **a** CT, **b** MRI, **c** DWT, **d** LAP, **e** NSST-SCM, **f** SR, **g** GFF, **h** proposed

method which can fully retain the details of the source images. The next medical image set “Data-2” is shown in Fig. 9. The result of using DWT is the lack of bone structure information and has a bad visual effect; the problem of low contrast also exists in Fig. 9f. There were no significant differences in the other three groups of results. Figure 10 shows the medical image set “Data-3,” compared with other comparison algorithms; the fusion result of the proposed method has high contrast and fully retains the soft tissue information as shown in Fig. 10h. The DWT, NSST-SCM, and SR methods lose the details of source images, and bone structures information is insufficient in Figs. 11 and 12. The obtained results by the proposed method have sharp edges, more details, and enhanced contrast.

3.3 Quantitative comparison

In addition to subjective visual evaluation, quantitative evaluation metric is an important tool to measure fusion performance. In this paper, those quantitative evaluation metrics include mutual information (MI) [26], mean structural similarity ($MSSIM$) [27], standard deviation (SD) [10], spatial frequency (SF) [10], image entropy (IE) [10], and margin information retention ($Q^{AB/F}$) [28] which are used to evaluate the different fusion methods.

1) MI shows the correlation between two events. The MI of two discrete random variables U and V can be defined as follows:

$$MI(U, V) = \sum_{v \in V} \sum_{u \in U} p(u, v) \log_2 \frac{p(u, v)}{p(u)p(v)} \quad (27)$$

where $p(u, v)$ is the joint probability distribution of U and V , $p(u)$ and $p(v)$ are the marginal probability distribution of U and V , respectively. The sum of mutual information between the fused image and two source images can be calculated to denote the difference of fusion quality, and then the mutual information metric can be described as follows:

$$MI_F^{AB} = MI(A, F) + MI(B, F) \quad (28)$$

Equation (28) reflects a total amount of information that fused image $F(i, j)$ contains both source image $A(i, j)$ and source image $B(i, j)$. The higher score of MI is, the richer the information is obtained from the source images.

2) SD is a measure of the dispersion degree of a set of image data averages. The standard deviation of the fused image is calculated as.

$$SD = \sqrt{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (F(i, j) - \mu)^2} \quad (29)$$

where $F(i, j)$ is the pixel value of the fused image at the location (i, j) , and μ is the mean value. The metric indicates the clarity of the fused image; the higher this score is, the higher the image quality is.

3) SF is composed of row frequency (RF) and column frequency (CF) and is described as follows.

$$SF = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (RF + CF) \tag{30}$$

$$RF = [C(i, j) - C(i, j-1)]^2 \tag{31}$$

$$CF = [C(i, j) - C(i-1, j)]^2 \tag{32}$$

where $C(i, j)$ denotes the pixel value of the image, in which the size of the image is $M \times N$. The higher score of this metric represents the fused image with higher resolution.

4) IE represents the amount of information in the fused image and the gray distribution of an image is $P = \{P_1, P_2, \dots, P_n\}$, P_l denotes the ratio of the pixel number of gray value l and the total pixels of the image, and n is the total number of gray level. It can be acquired by Eq. (33).

$$IE = - \sum_{l=0}^L P(l) \log_2 P(l) \tag{33}$$

where $P(l)$ expresses the probability density of L , and L represents the gray level of an image. The higher score of IE is, the more information the fused image contains.

5) $MSSIM$ is an effective measure of similarity of two images, which is calculated as follows.

$$MSSIM = \frac{SSIM(A, F) + SSIM(B, F)}{2} \tag{34}$$

where $SSIM(A, F)$ and $SSIM(B, F)$ are correlation coefficients between the CT image and the fused image, the MRI image and the fused image respectively. $SSIM(i, j)$ is defined as follows.

Table 1 Quantitative assessments comparison of different methods

Dataset	Methods	MI	IE	SF	SD	$MSSIM$	$Q^{AB/F}$
Data-1	DWT	1.9221	6.1766	17.1339	44.7173	0.5246	0.6073
	LAP	2.4316	6.0799	18.1079	52.8910	0.5518	0.7184
	NSST-SCM	2.0570	6.2015	17.0573	44.1620	0.5366	0.6816
	SR	2.5735	5.8054	11.6897	30.8241	0.5122	0.5756
	GFF	2.3191	6.5253	16.9774	52.8964	0.5634	0.7210
	Proposed	2.6023	6.5997	17.6021	53.0515	0.5676	0.7276
Data-2	DWT	3.1957	5.1935	22.1574	55.7300	0.7915	0.5051
	LAP	3.3457	4.8935	21.9313	59.9259	0.8146	0.5888
	NSST-SCM	3.3489	5.1244	20.9562	54.5607	0.8160	0.5887
	SR	3.4259	4.9400	17.7694	51.4073	0.8248	0.5178
	GFF	3.7904	5.2091	20.2566	55.6818	0.8207	0.6028
	Proposed	3.8849	5.2231	22.6706	62.3212	0.8273	0.6108
Data-3	DWT	3.1289	4.8668	25.1158	66.5339	0.7489	0.5473
	LAP	3.3880	4.3915	25.9915	69.6022	0.7775	0.6042
	NSST-SCM	3.2093	4.8812	24.5271	65.8981	0.7733	0.5971
	SR	3.1845	4.5228	20.1904	61.5043	0.7640	0.5157
	GFF	3.3467	5.0519	24.3932	69.6355	0.7762	0.6119
	Proposed	3.8130	4.6208	25.6074	73.2812	0.7669	0.6534
Data-4	DWT	3.0829	7.4137	35.9431	76.8087	0.6263	0.4699
	LAP	3.2691	7.3196	37.0391	79.8434	0.6462	0.5171
	NSST-SCM	3.2365	7.4479	34.6064	79.4965	0.6628	0.5349
	SR	3.3354	7.5617	28.9810	69.8406	0.6532	0.4964
	GFF	3.5241	7.7616	34.3087	75.3693	0.6602	0.5510
	Proposed	3.6284	7.5901	35.5445	88.1087	0.6683	0.5516
Data-5	DWT	2.8048	5.3647	22.2882	55.7298	0.7098	0.4573
	LAP	3.1872	4.8391	23.0648	61.1195	0.7448	0.5214
	NSST-SCM	2.9475	5.4499	21.4789	53.7913	0.7311	0.5226
	SR	3.1968	5.2472	17.5857	51.7154	0.7427	0.4823
	GFF	3.2308	5.7828	21.5629	66.9815	0.7342	0.5330
	Proposed	4.4628	5.5307	23.0713	74.3093	0.7456	0.5990

$$SSIM(i, j) = \frac{(2\mu_i\mu_j + C_1)(2\sigma_{ij} + C_2)}{(\mu_i^2 + \mu_j^2 + C_1)(\sigma_i^2 + \sigma_j^2 + C_2)} \quad (35)$$

where μ_i , σ_j , and σ_{ij} express the mean, standard deviation, and cross-correlation, respectively. C_1 and C_2 are used to ensure stability when the mean value and the variance are close to zero. The rotationally symmetric Gaussian window with standard deviation 1.5 was selected in *MSSIM*. The higher score of *MSSIM* is, the smaller the distortion of the fused image is.

6) $Q^{AB/F}$ represents the transformation degree of edge information of the fused image and the source image. It is defined as follows.

$$Q^{AB/F} = \frac{\sum_{i=1}^N \sum_{j=1}^M (Q^{AF}(i, j)w^A(i, j) + Q^{BF}(i, j)w^B(i, j))}{\sum_i \sum_j (w^A(i, j) + w^B(i, j))} \quad (36)$$

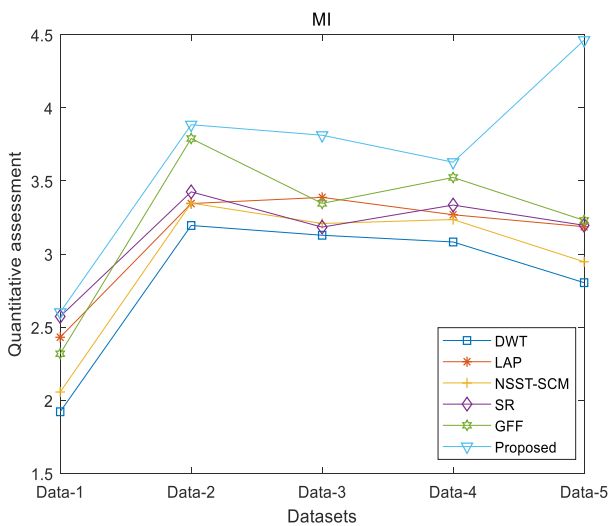
where $Q^{AF}(i, j) = Q_g^{AF}(i, j)Q_o^{AF}(i, j)$, $Q_g^{AF}(i, j)$, and $Q_o^{AF}(i, j)$ are the edge strength and orientation preservation value at the location (i, j) , respectively. N and M are the size of the image, and $Q^{BF}(i, j)$ is similar to $Q^{AF}(i, j)$; $w^A(i, j)$ and $w^B(i, j)$ reflect the weight of $Q^{AF}(i, j)$ and $Q^{BF}(i, j)$ respectively. If the $Q^{AB/F}$ gets the value higher and closer to unity, it means that the fused image is produced with less edge information loss.

The objective quantitative assessments based on image quality metrics are shown in Table 1. The best results are in italic. The proposed method achieves the best metric in terms

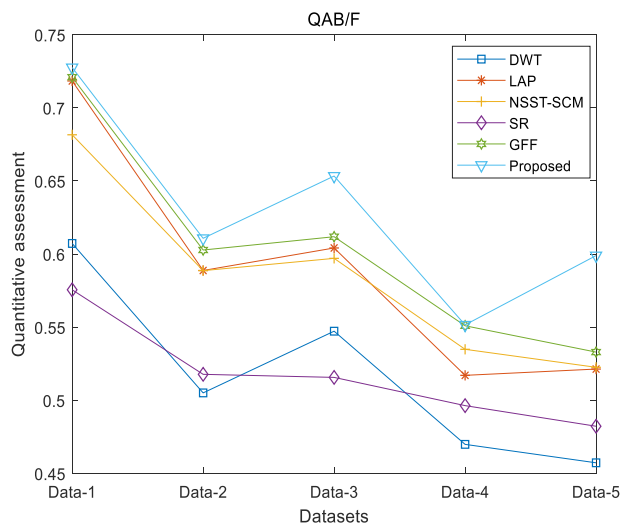
of *MI*, *SD*, *MSSIM*, and $Q^{AB/F}$, and remaining metrics have subtle different between each other which demonstrate that the proposed method has a high level of competence in preserving edge details and saliency information. Moreover, the two important relative evaluation metrics that *MI* and $Q^{AB/F}$ will be represented as line chart intuitively, as shown in Fig. 13.

4 Discussions

To summarize the experimental results, the fused images based on SR and DWT method look unsatisfactory due to low contrast and bone structure information loss. The visual effect of results based on LAP, NSST-SCM, and GFF has improved, but texture and edge in yellow marked region are not preserved fully. By contrast, the proposed method achieves clear and high contrast fused results by retaining salient features, which synthesizes soft tissues and bone structure information to the maximum extent. Next, we discuss the quantitative assessments of the image in detail. Among the six metrics, such as *IE*, *SF*, and *SD* reflect the internal features of a single image, which can measure the quality of the fused image commonly. The *IE* represents the information entropy of the fused image. The *SF* reflects the clarity of the image. The *SD* describes the contrast of the fused image. The bigger the *SD* is, the more dispersed the distribution of gray level in the image is, and the greater the contrast is, the better the visualization of the fused image is. However, some results contain more redundant information, which also lead to increase the score of those three metrics. To make more comprehensive and objective analysis, we introduce the other three metrics including *MI*, *MSSIM*, and $Q^{AB/F}$ that can better reflect the



(a) The comparison line chart for MI



(b) The comparison line chart for $Q^{AB/F}$

Fig. 13 Objective evaluation results based on Figs. 8, 9, 10, 11, and 12. a The comparison line chart for MI, b the comparison line chart for $Q^{AB/F}$

internal relationship between the source images and the fused image. The MI describes the similarity of the image intensity distributions of the corresponding image pair and estimates how much information is obtained from source images. The higher score of MI indicates the richer information and details obtained from CT and MRI images which also assures more activity and clarity level in the fused images. The $MSSIM$ represents the degree of distortion of the fused image. In addition, the $Q^{AB/F}$ measures the amount of edge information transferred from the source images to the fused images. This metric is essential for medical image fusion that the higher score means the more the edge details such as bone structure and texture are fused, which is helpful for accurate pathological analysis. According to the quantitative analysis of the experimental results, the proposed method achieves the best performance in the MI , SD , and $Q^{AB/F}$ metrics especially, which also means that the fused image has a high contrast and less distortion and contains enough dense structures, soft tissues information, texture, and edge details. The results based on the proposed method are more suitable for assisting doctors in the accurate diagnosis diseases. At the present stage, we adopt the deep learning scheme to extract multi-layer features and generate the low-frequency fusion weight which achieves good performance, but the selection of the deep features still depends on the artificial designed rule. In the future, we will continue to study the relationship between multi-model features and try to build an unsupervised deep fusion model and make the fusion framework more robust and adaptive.

5 Conclusions

In this paper, a novel fusion method for CT and MRI medical images is proposed by integrating CNNs and DCSCM in NSST domain. In the proposed method, the NSST provides both the multiscale and direction analysis of the source images. The CNNs based an activity level measurement that is used to fuse the low-frequency coefficients. In the term of high-frequency coefficients, the modified average gradient is utilized as the external incentive of DCSCM. Extensive contrast experiments have been carried out on different pairs of CT and MR images which can verify the superiority of the proposed method in both visual effects and objective evaluation. Moreover, the results demonstrate this fusion scheme has application prospect in the field of medical image fusion.

Acknowledgements The authors thank the editors and the anonymous reviewers for their careful works and valuable suggestions for this study.

Funding information This work was supported by the National Natural Science Foundation of China under Grants 61463052 and 61365001 and Yunnan Province University Key Laboratory Construction Plan Funding, China.

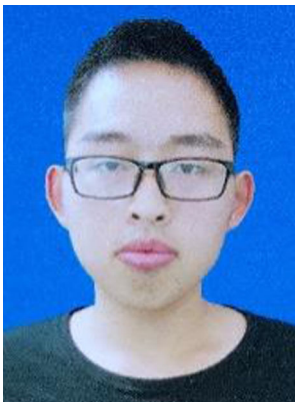
Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interests.

References

1. Y Liu, X Chen, J Cheng et al. A medical image fusion method based on convolutional neural networks, International Conference on Information Fusion. IEEE, 1–7(2017)
2. Gaurav Bhatnagar QM, Wu J, Liu Z (2013) Directive contrast based multimodal medical image fusion in NSCT domain. IEEE Trans Multimedia 15(5):1014–1024
3. Liu X, Mei W, Du H (2017) Structure tensor and nonsubsampling shearlet transform based algorithm for CT and MRI image fusion. Neurocomputing 235:131–139
4. Vijayarajan R, Muttan S (2015) Discrete wavelet transform based principal component averaging fusion for medical images. AEU-Int J Electron C 69(6):896–902
5. Toet A (1989) Image fusion by a ratio of low-pass pyramid. Pattern Recogn Lett 9(4):245–253
6. Do Minh N, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans Image Process 14(12):2091–2106
7. Da Cunha AL, Zhou J, Do MN (2006) The nonsubsampling contourlet transform: theory, design, and applications. IEEE Trans Image Process 15(10):3089–3101
8. Zhang T, Zhou Q, Feng H et al (2007) Fusion of infrared and visible light images based on nonsubsampling shearlet transform. Int J Infrared Millimeter Waves 26(6):476–480
9. Johnson JL, Padgett ML (1999) PCNN models and applications. IEEE Trans Neural Netw 17(3):480–498
10. Jin X, Nie RC, Zhou DM et al (2016) Multifocus color image fusion based on NSST and PCNN. J Sens 8359602. <https://doi.org/10.1155/2016/8359602>
11. He K, Zhou D, Zhang X et al (2018) Multi-focus image fusion combining focus-region-level partition and pulse-coupled neural network. Soft Comput 4:1–15. <https://doi.org/10.1007/s00500-018-3118-9>
12. Hou R, Nie R, Zhou D et al (2018) Infrared and visible images fusion using visual saliency and optimized spiking cortical model in non-subsampling shearlet transform domain. Multimed Tools Appl 1:1–24. <https://doi.org/10.1007/s11042-018-6099-x>
13. Liu Y, Chen X, Wang Z et al (2018) Deep learning for pixel-level image fusion: recent advances and future prospects. Inform Fusion 42:158–173
14. Liu Y, Chen X, Peng H et al (2017) Multi-focus image fusion with a deep convolutional neural network. Inform Fusion 36:191–207
15. Prabhakar KR, Srikar VS, Babu RV (2017) DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image Pairs. IEEE International Conference on Computer Vision. IEEE Computer Society, 4724–4732
16. Ide H, Kurita T (2017) Improvement of learning for CNN with ReLU activation by sparse regularization. International Joint Conference on Neural Networks IEEE, 2684–2691
17. Gatys LA, Ecker AS, Bethge M (2016) Image style transfer using convolutional neural networks. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2414–2423
18. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. Comput Sci

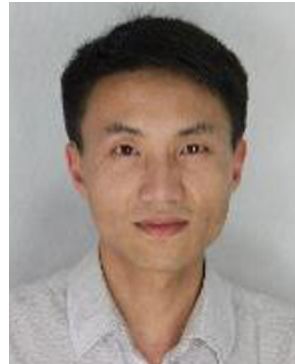
19. Zhan K, Zhang H, Ma Y (2009) New spiking cortical model for invariant texture retrieval and image processing. *IEEE Trans Neural Netw* 20(12):1980–1986
20. Yoshifusa I (1991) Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. *Neural Netw* 4(3):385–394
21. Zhang Y (2007) Adaptive region-based image fusion using energy evaluation model for fusion decision. *SIViP* 1.3:215–223
22. Liu X, Mei W, Du H (2016) Multimodality medical image fusion algorithm based on gradient minimization smoothing filter and pulse coupled neural network. *Biomed Signal Process Control* 30: 140–148
23. Huang Z, Ding M, Zhang X (2017) Medical image fusion based on non-subsampled shearlet transform and spiking cortical model. *J Med Imag Health In* 7.1:229–234
24. Kim M, Han DK, Ko H (2010) Joint patch clustering-based dictionary learning for multimodal image fusion. *ACM T Sensor Network* 6.3:20
25. Li S, Kang X, Hu J (2013) Image fusion with guided filtering. *IEEE Trans Image Process* 22.7:2864–2875
26. Hossny M, Nahavandi S, Creighton D (2008) Comments on “Information measure for performance of image fusion”. *Electron Lett* 44(18):1066–1067
27. Wang Z, Bovik AC, Sheikh HR et al (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
28. Vladimir P, Costas X (2004) Evaluation of image fusion performance with visible differences. 8th European Conference on Computer Vision, ECCV 2004, Lecture Notes in Computer Science, 3023: 380–391



Ruichao Hou is currently pursuing a master’s degree at Yunnan University, China. He received his B.S. degree in electrical engineering and automation in 2016. His current research interests include neural networks theory and its application, image processing, computer vision, and deep learning.



Dongming Zhou received the B.S. and M.S. degrees in automatic control engineering from Huazhong University of Science and Technology in 1985 and 1988, respectively, and the Ph.D. degree in electronic circuits and systems from Fudan University in 2004. He is currently a professor at Yunnan University, China. His current research interests include neural network theory and applications, artificial intelligence control theory, and image processing.



Rencan Nie received the B.S., M.S., and Ph.D. degrees in telecommunications engineering from Yunnan University, China, in 2004, 2007, and 2014, respectively. He is currently an associate professor at Yunnan University. His current research interests include pulse-coupled neural networks’ theory and applications, image processing, and machine learning.



Dong Liu is currently pursuing a master’s degree at Yunnan University, China. He received his B.S. degree in electrical automation from Tianjin Polytechnic University, China, in 2016. His current research interests include image processing and image fusion.



Xiaoli Ruan is currently pursuing a Ph.D. degree at Yunnan University, China. Her current research interests include biomedical signal processing, neural networks theory and its application, and deep learning.