IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2021, accepted April 17, 2021, date of publication April 27, 2021, date of current version May 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075953

Multimodal Medical Image Fusion Based on Gabor Representation Combination of Multi-CNN and Fuzzy Neural Network

LIFANG WANG, JIN ZHANG, YANG LIU, JIA MI, AND JIONG ZHANG
College of Big Data, North University of China, Shanxi 030051, China

Corresponding author: Jin Zhang (1648283554@qq.com)

This work was supported by the National Natural Science Foundation of China under Grant 201901D111152 (Project name: the research and exploration of texture and edge information in the fusion of medical images of cancer and brain tumor).

ABSTRACT Aiming at the current multimodal medical image fusion methods that cannot fully characterize the complex textures and edge information of the lesion in the fused image, a method based on Gabor representation of multi-CNN combination and fuzzy neural network is proposed. This method first filters the CT and MR image sets through a set of Gabor filter banks with different proportions and directions to obtain different Gabor representations pairs of CT and MR, each pair of different Gabor representations is used to train the corresponding CNN to generate a G-CNN and multiple G-CNN form a G-CNN group, namely G-CNNs; then when fusing CT and MR images, CT and MR are represented by Gabors to get Gabor representation pairs firstly, each Gabor representation pair is put into the corresponding trained G-CNN for preliminary fusion, then use the fuzzy neural network to fuse multiple outputs of the G-CNNs to obtain the final fused image. Compared with the nine recent state-of-the-art multimodal fusion methods, the average mutual information of the three groups of experiments has increased by 13%, 10.3%, and 10% respectively; the average spatial frequency has increased by 10.3%, 20%, and 10.7%; the average standard deviation has increased respectively 12.4%, 10.8%, 14.4%; the average edge retention information increased by 33.5%, 22%, and 43%. The experimental results show that the proposed fusion method is significantly better than the other comparative fusion methods in objective evaluation and visual quality. It has the best performance on the four indicators and can better integrate the rich texture features and the clear edge information of the source images into the final fused image, which improves the quality of multimodal medical image fusion, and effectively assists doctors in disease diagnosis.

INDEX TERMS Medical image fusion, G-CNNs, Gabor representation, convolutional neural network, fuzzy neural network.

I. INTRODUCTION

With the development of medical imaging technology, medical images are becoming more and more important in clinical diagnosis. Medical images of different modalities have their own advantages and limitations. For example, computed tomography (CT) images are very clear for bone imaging and can provide a good reference for lesion location, but the contrast of soft tissues is relatively low; magnetic resonance imaging (MR) images can clearly reflect the structure of soft tissues, but the imaging of bone tissue information is very flawed [1]. Therefore, a single mode of medical image cannot fully and accurately reflect the pathological tissue information. In clinical practice, doctors generally need to check multiple different modalities of medical images of the same part to diagnose the patient's condition, which often causes the diagnosis complexity and error. Multimodal medical image fusion aims to extract useful information from the multimodal source images as much as possible to integrate a fused image which is able to help doctors diagnose and treat diseases accurately, quickly and comprehensively [2].

When diagnosing brain tumor diseases, due to the invasive growth of malignant brain tumor, brain tumors images often have insufficient texture and unclear boundaries. Therefore the associated medical images of the brain tumor urgently need to be fused together to demonstrate the rich textures

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

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

features and edge information of the lesion area. Different texture features can reflect the pathological detailed conditions of the lesion area at different times, and the edge information can help to distinguish the tumor boundary and other lesions. However, the existing traditional multimodal fusion methods cannot fully characterize the complex texture features and edge information in the fused image. Traditional image fusion methods generally include two categories: spatial domain based methods and transform domain based methods. The fusion methods based on the spatial domain [3], such as principal component analysis and intensity color saturation, commonly cause the spectral and spatial distortion of the fused image, which is not conducive to the doctor's observation of the lesion area. In transform domain, fusion methods can be classified as follows: pyramid-based methods, and wavelet transform methods. Based on the pyramid transformation [4], the laplacian pyramid, the gaussian pyramid, the contrast pyramid, and the morphological pyramid, these methods fail to involve spatial direction selectivity in the decomposition process, resulting in the block effect, and meanwhile produce many artifacts in the edge of the fused image affecting the fusion result; wavelet transform fusion methods [5], such as discrete wavelet transform, redundant wavelet transform and multi-wavelet transform, only capture limited directional information, obtain limited information on edges and texture areas, and cannot clearly characterize the edges of the images.

In recent years, with the rise of deep learning, Convolutional Neural Network (CNN) as an important branch of deep learning has stronger feature extraction capabilities than traditional methods, and is more suitable for image fusion [6]–[8]. Liu *et al.* [9] propose a multi-focus image fusion method based on Convolutional Neural Network (CNN), which uses CNN to classify the focus area and obtain a decision map. The decision graph is combined with the source images to generate a fused image. However, in the CNN model, the final output feature dimensions of convolution and downsampling are low, resulting in the loss of fused image information. Zhanga *et al.* [10] propose a general image fusion framework FCNN based on a fully convolutional neural network. This framework uses a fully convolutional neural network to solve the problem of information loss, but the convolution kernel setting of its convolutional layer is too simple, which causes the extracted features cannot well represent the texture information of the lesion, making it difficult to use the fused image for clinical diagnosis. Recently, ZhenChao *et al.* propose a method that apply improved fuzzy neural network to medical image fusion. This method can effectively solve the uncertainty and ambiguity of the lesion during the fusion process. However it still has the weak capacity of containing full texture features and edge details of the fused image lesion. Alenezi and Salari [11] propose a method that combines Gabor filter and PCNN to fuse multimodal medical images. Taking advantage of the good performance of Gabor filter in maintaining texture features and edge details, the texture features and edges of the fused images are improved, but the information content of the fused image is insufficient.

In consideration of the advantages and inadvantages of all the above mentioned methods, in order to obtain a fused image with rich textures, clear edges and sufficient information content, this paper proposes a method based on Gabor representation of multi-CNN groups (G-CNNs) combining with fuzzy neural network for multimodal medical images Fusion. The proposed method mainly consists of two parts: 1) Constructing the G-CNN groups (G-CNNs); 2) Fusing multimodal medical images with the G-CNNs and fuzzy neural networks. In the first part, different Gabor representation pairs of CT and MR are obtained through a set of Gabor filters with different proportions and directions, and then the corresponding CNNs are trained using different Gabor representations pair respectively to generate multiple G-CNN. Then we combine the multiple G-CNN and get G-CNNs. In the second part, when CT and MR are fused, the same as the first part, source images through a Gabor bank get Gabor representation pairs, then put these into corresponding trained G-CNNS to get multiple preliminary fusion outputs. The fuzzy neural network is used to fuse the diverse outputs of the G-CNNs to get the final fused image. Contributions of this paper are as followings:

1) The method uses Gabor expression: $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ as CNN input enhancing and retaining the texture features and edge details of source images.

2) Considering the advantage of CNN extracting effective information from complex backgrounds, we successfully extract the deep features of $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ of source images by G-CNNs and get multiple preliminary fusion outputs through G-CNNs.

3) In view of the capacity of the fuzzy neural networks solving the problems of ambiguity and uncertainty in the fusion process, we leverage the fuzzy neural network to fuse the diverse outputs of G-CNNs and obtain the final fused image which retains the rich texture features and clear edge information of medical source images CT and MR, improves the visual characteristics and reinforces the information content.

II. RELATED WORK

A. GABOR REPRESENTATION

Gabor is a direction-sensitive linear filter, mainly used for edge detection and texture analysis [12]. A 2-dimensional Gabor filter is the product of a sine plane wave and a Gaussian kernel function [13]. The filter has a real component (Equation 1) and an imaginary component (Equation 2) that represent the orthogonal direction. These two components can form a complex number (Equation 3), or they can also be used separately.

$$g(x, y : \lambda, \theta, \phi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 + y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \phi\right) \quad (1)$$

$$g(x, y : \lambda, \theta, \phi, \sigma, \gamma)$$

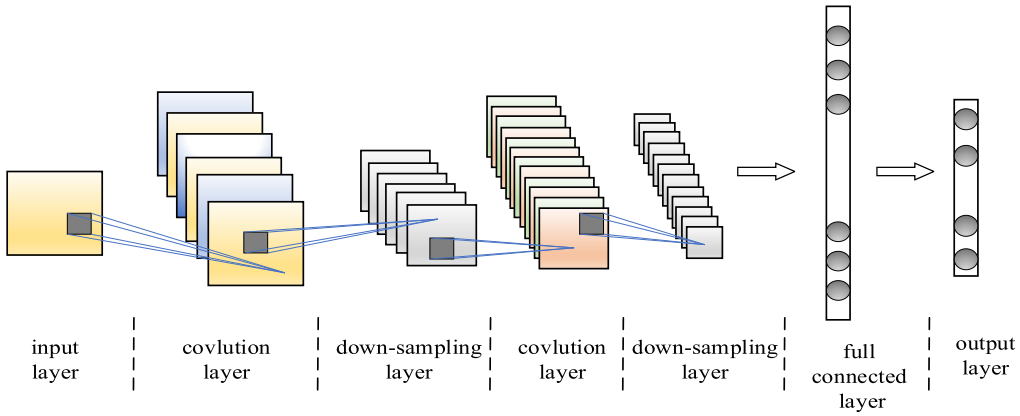


FIGURE 1. Typical CNN structure diagram.

$$= \exp(-\frac{x'^2+\gamma^2+y'^2}{2\sigma^2})\sin(2\pi\frac{x'}{\lambda}+\phi) \quad (2)$$

$$g(x, y : \lambda, \theta, \phi, \sigma, \gamma)$$

$$= \exp(-\frac{x'^2+\gamma^2+y'^2}{2\sigma^2})\exp(i(2\pi\frac{x'}{\lambda}+\phi)) \quad (3)$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, λ, θ, and ϕ are the parameters of the sine part. λ denotes the wavelength of the sine curve; θ indicates the rotation of the image; ϕ represents the phase shift of the sine wave, that is, how much the pattern needs to shift from the center; σ means the standard deviation of the Gaussian part; γ is the aspect ratio of the pattern.

B. CONVOLUTIONAL NEURAL NETWORK

CNN is a typical deep learning model, which attempts to learn hierarchical representations of image data with different levels of abstraction [14]. A typical CNN structure is shown in Figure 1, which consists of an input layer, a convolution layer, a down-sampling layer, a full connected layer, and an output layer.

In CNN, each convolutional layer transforms the input image into a certain number of feature maps through a convolution operation. Generally, the nonlinear activation function immediately follows the convolution operation. The convolution operation and the nonlinear Relu activation function in CNN [15] can be expressed as:

$$y^j = \max(0, b^j + \sum_i k^{ij} * x^i) \quad (4)$$

In formula (4), x^i and y^i respectively represent the i^{th} input feature map and the j^{th} output feature map of the convolutional layer. k^{ij} denotes the convolution kernel between x^i and y^i. b^j is the deviation. $*$ indicates a convolution operation. When there are M input maps and N output maps, the layer will contain N convolution kernels of size $q \times q \times M$ ($q \times q$ is the size of the receptive field).

The down-sampling layer, also called the pooling layer, is commonly located in the middle of the continuous convolutional layer, which can simplify the output of the convolutional layer and reduce the feature dimension, network parameters and computational complexity. The pooling layer includes maximum pooling and average pooling. The most commonly used is maximum pooling, which can be defined as:

$$y^i_{h,w} = \max_{0\le a,b\le p}(x^i_{h\times p+a, w\times p+b}) \quad (5)$$

In formula (5), $y^i_{h,w}$ denotes the neuron at position (h, w) in the i^{th} output map of the pooling layer, and the maximum value is specified in the local area of $p \times p$ in the x^i input map i^{th}, and a, b are the pooling range in the $p \times p$ local area.

The full connected layer is usually at the end of CNN. After feature extraction of input data through the convolutional layer and pooling layer, the full connected layer carries out nonlinear combination of extracted features to obtain output.

C. FUZZY NEURAL NETWORK

Fuzzy Neural Network (FNN) is the combination of Fuzzy logic and neural Network. It can combine the uncertainty and interpretability of Fuzzy logic with the learning ability of neural Network. It has advantages in solving uncertainty, data imprecision and nonlinear problems and can deal with large-scale fuzzy application problems [16]. Therefore, researchers propose image fusion methods based on fuzzy neural network [17], [18]. These methods can solve the problems of ambiguity and uncertainty in the fusion process, have good ability of noise removal and information retention, and can better represent the edge of the image.

The type and number of the layers in the fuzzy neural network model can be set according to the complexity of the problem. A typical model structure hides one or two layers, which are responsible for the process of fuzzy reasoning, rule making, fuzzification or defuzzification of data [19], [20]. The first layer is the fuzzification process, and the last layer is the final result of the model. Some hidden layers in the middle are the creation of fuzzy rules, the processing of model response, the matching degree of output membership function and input data, and the resolution of fuzzy rules, etc.

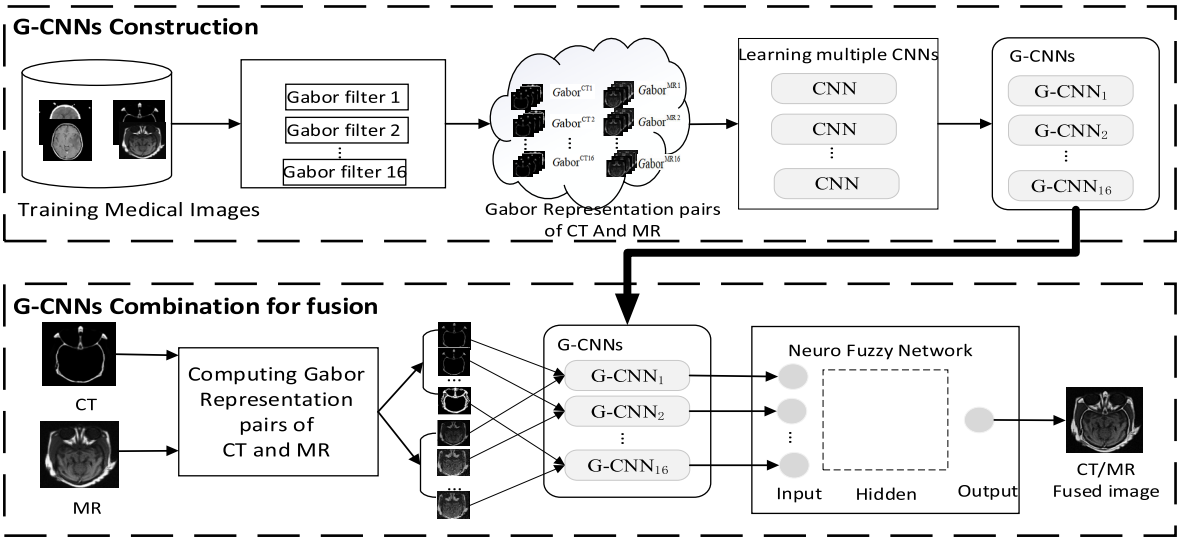


FIGURE 2. Multimodal medical image fusion process based on G-CNNS and fuzzy neural network.

D. MOTIVATION OF THIS WORK

Medical image fusion, as a characteristic application field of information fusion, makes full use of multimodal images to obtain complementary information, which makes clinical diagnosis and treatment more accurate. Although the existing multimodal medical fusion methods have been improved and developed, these methods still cannot fully represent the complex textures and edge information of the lesion in the fused image. Whereas, the processing of texture and edge information largely determines the quality of the fused image. Considering Gabor is very suitable for edge detection and texture analysis, the parameters of the filter can be set according to the source images to establish Gabor filters of different scales and directions, which can be used in image filtering to get the full texture and edge features of the image in different scales and directions. Meanwhile Convolutional Neural Network (CNN) has brought new development to the field of image fusion. The convolutional layer plays a role in feature extraction, which is usually able to extract more extensive and richer features than the traditional manual feature extractor. At the same time, there are problems of ambiguity and uncertainty in the fusion process. The fuzzy neural network can be used to solve this problem. Therefore we use Gabor group of different scales and directions to represent the texture and edge features of resource images. Then we utilize G-CNNs to extract deep features of the textures and edges of Gabor representation pairs and get multiple preliminary fused images of Gabor representation pairs. Finally we leverage the fuzzy neural network to fuse multiple outputs of G-CNNs.

III. MULTIMODEL MEDICAL IMAGE FUSION BASED ON THE COMBINATION OF G-CNNs AND FUZZY NEURAL NETWORK

In order to fully characterize the texture details of the lesions in the multimodal medical images and make the edges clearer, we propose a method combining Gabor-CNNs (G-CNNs) with fuzzy neural networks to fuse multimodal medical images, as shown in Figure 2. The method consists of two parts: (1) The first part is to build "G-CNNs" group (G-CNNs), namely "G-CNNs Construction" as shown in the upper section of Figure 2. In order to fully represent texture and edge features of source images, we use 16 pair Gabor filters with different scales and directions to filter CT and MR source images, respectively. And 16 different Gabor representation pairs of CT and MR are expressed as: $Gabor^{CT_i}$ and $Gabor^{MR_i}$, $i = 1, \ldots, 16$. The representations have multiple detail texture and edge information in different directions and scales to enhance the texture features of the source images, and then each pair of Gabor: $Gabor^{CT_i}$ and $Gabor^{MR_i}$, $i = 1, \ldots, 16$ is used to train the corresponding 16 CNN, and 16 trained CNN form G-CNNs. Since trained CNN in different texture directions and scales, each G-CNN obtains different texture characterization information. (2) The second part is the fusion of CT and MR medical images based on G-CNNs, as shown in the lower part "G-CNNs Combination for Fusion" of Figure 2. CT and MR images to be fused are represented by Gabor filter to obtain the Gabor representation pairs of $Gabor^{CT_i}$ and $Gabor^{MR_i}$ respectively, and each pair of Gabor representation $Gabor^{CT_i}$ and $Gabor^{MR_i}$ is put into the corresponding trained G-CNN$_i$, i $= 1 \ldots\ldots 16$ for preliminary fusion. Then, multiple outputs of the G-CNNs are fused by the Fuzzy Neural Network to obtain the final fused image.

A. GABOR REPRESENTATION OF MEDICAL IMAGES

Gabor representation of medical images can enhance the texture features of the source images and obtain more edge details, so that the fused images contain more texture features and edge information about the relevant lesion sites with Gabor representation. In order to obtain Gabor

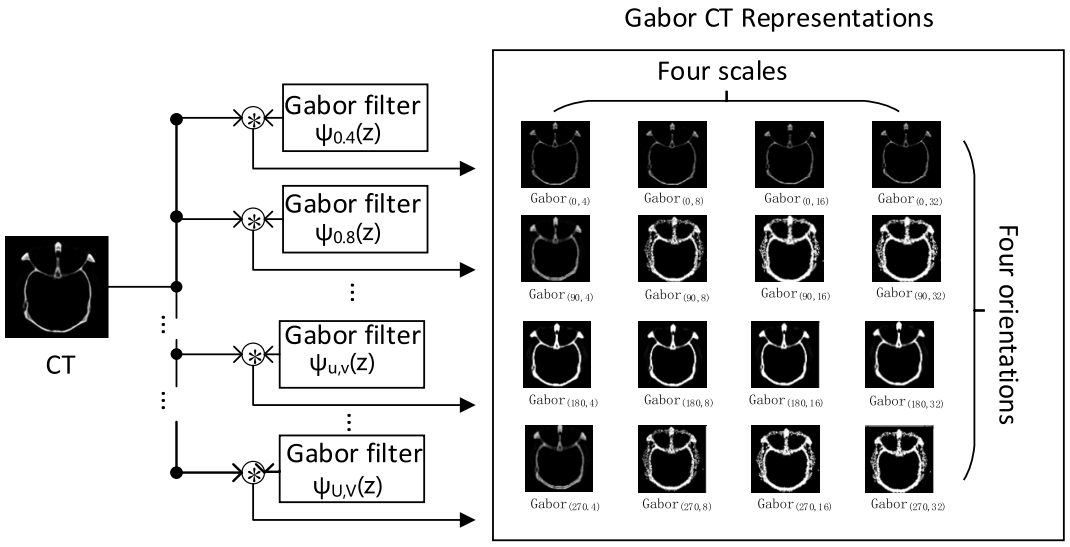


FIGURE 3. Gabor representation of CT image.

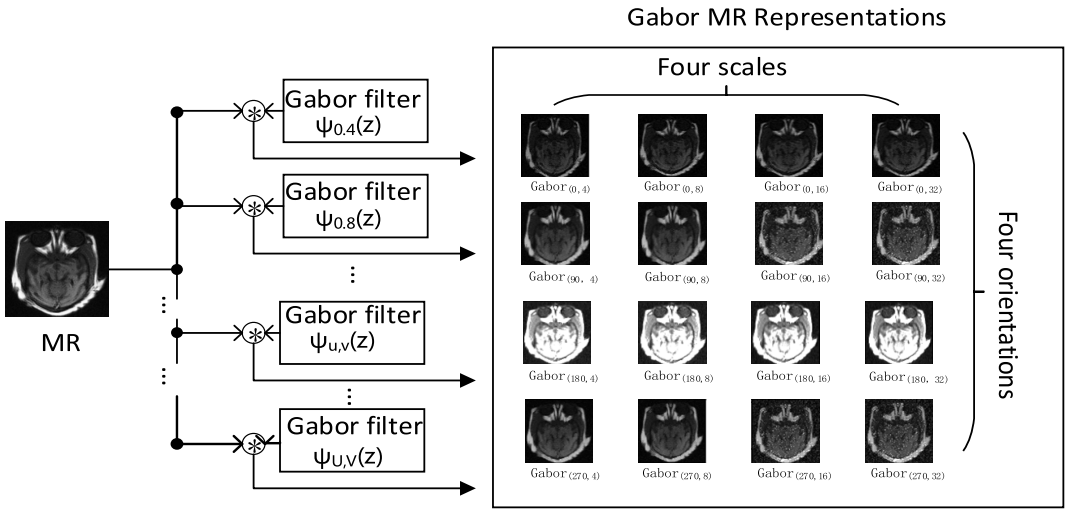


FIGURE 4. Gabor representation of MR image.

representations of CT and MR of the given medical images, we need design a set of Gabor filter banks with different scales and orientations. CT and MR images are both two-dimensional gray images, so a set of two-dimensional Gabor filters is defined:

$$\psi_{\mathrm{u},v}(z) = \frac{\left\|k_{u,v}\right\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2\|z\|^2/2\sigma^2)} \times \left[\mathrm{e}^{ik_{u,v}z} - e^{-\sigma^2/2}\right] \tag{6}$$

In equation (6), u and v represent the direction and scale of the filter bank. $z = (x, y)$ denotes the position of the pixel. $\mathrm{k}_{u,v} = k_v e^{i\varphi_u}$, $k_v = k_{\max}/f^v$, $\varphi_u = \pi u/8$, $k_{\max}$ is the maximum frequencies, f indicates the spacing factor between filters in the frequency domain, σ means the standard deviation of the Gaussian part. According to Equation (6), the filter can be set with different u directions and v scales.

Gabor representations of CT and MR with specific u direction and v scale are obtained by the following operations:

$$Gabor_{(u,v)}^{CT_i}(z) = I(z) * \psi_{u,v}(z) \quad \text{for } 0 \le \mathrm{u} \le U-1, \; 0 \le \mathrm{v} \le V-1 \tag{7}$$

$$Gabor_{(u,v)}^{MR_i}(z) = I(z) * \psi_{u,v}(z) \quad \text{for}, 0 \le \mathrm{u} \le U-1, \; 0 \le \mathrm{v} \le V-1 \quad i = 1, \ldots, 16 \tag{8}$$

In formula (7) and (8), $*$ represents the convolution operation, $G\mathrm{abor}_{(\mathrm{u,v})}^{\mathrm{CT_i}}(z)$ and $G\mathrm{abor}_{(\mathrm{u,v})}^{\mathrm{MR_i}}(z)$ are the Gabor representation pair of CT and MR of the given source images (the direction is u and the scale is v). (Here, the representation of each pair is different due to different filters.)The directions of the Gabor filters are set as 0°, 90°, 180° and 270°, and the scales are set as 4, 8, 16, and 32, respectively. In this way, the 4 directions and 4 scale filters are combined in pairs to produce 16 filters which form a filter bank. Figure 3 and Figure 4

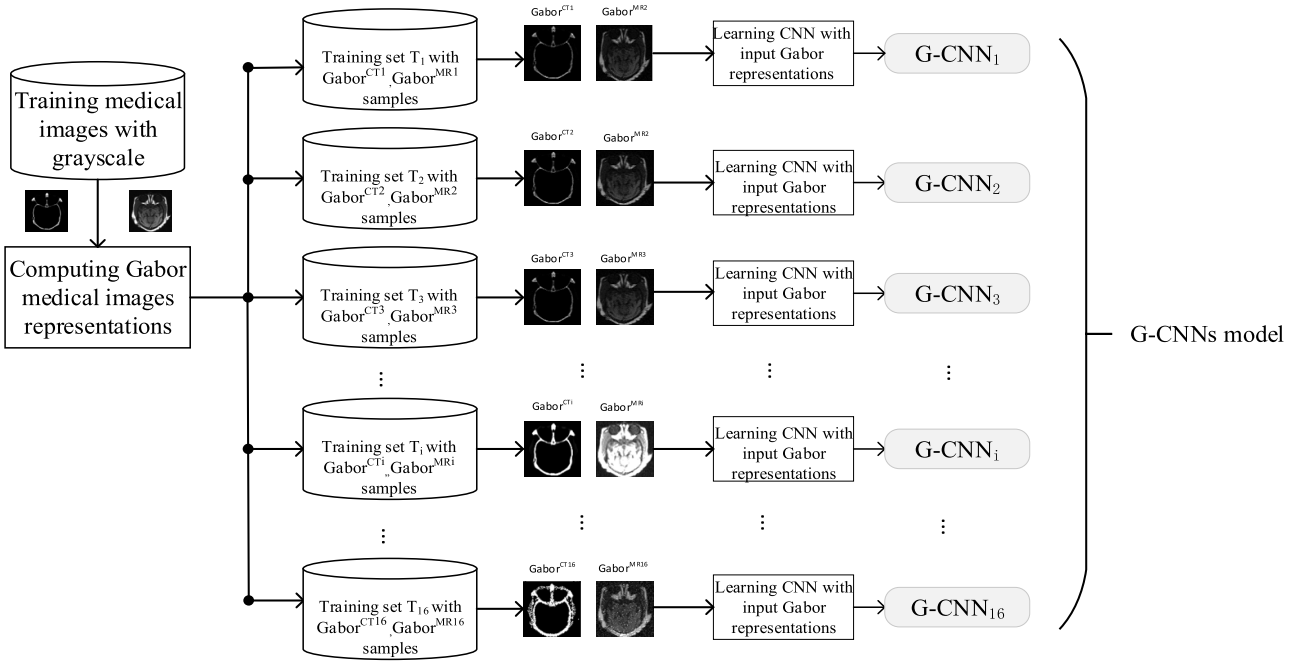


FIGURE 5. Generation process of G-CNNS structure.

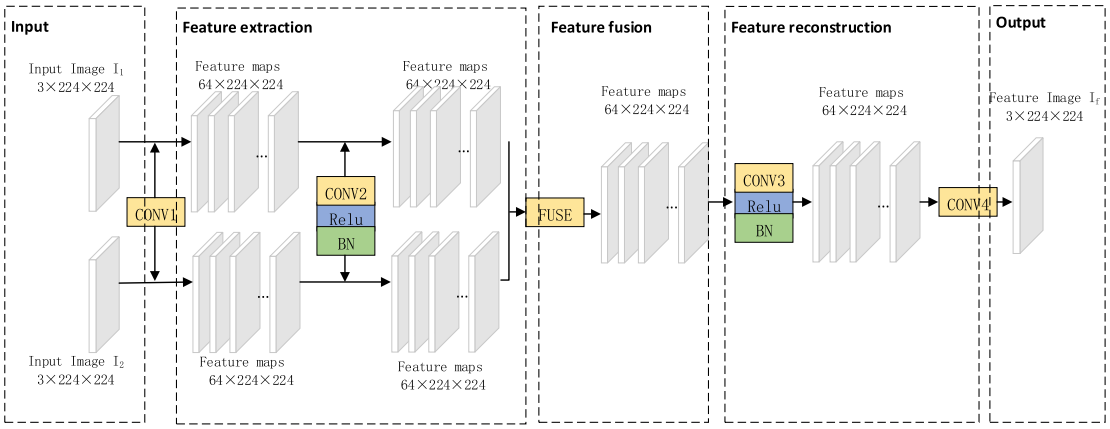


FIGURE 6. G-CNN structure diagram.

show the Gabor representation process of CT and MR images. The images CT and MR are convolved by the Gabor filter bank to obtain 16 pairs of $Gabor^{CT_i}$ and $Gabor^{MR_i} i = 1, \ldots, 16$ representations in different directions and scales. These representations characterize the features of the lesion from different directions and different scales to enhance the key information, especially for textures and edges.

B. G-CNNs STRUCTURE

As shown in Figure 5, The medical images with grayscale are represented by 16 Gabor filters with different scales and directions. Hence we get different training sets:T_1-T_{16} and each training set includes: $Gabor^{CT_i}$ and $Gabor^{MR_i}(i = 1, \ldots, 16)$. The different training sets T: $Gabor^{CT_i}$ and $Gabor^{MR_i}(i = 1, \ldots, 16)$ trains the corresponding CNN, and we obtain corresponding $\text{G-CNN}_\text{i}(\text{i} = 1, 2, \ldots 16)$. Each G-CNN_i represents the texture and edge features of medical images in a direction and scale. 16 $\text{G-CNN}_\text{i}(i = 1, 2, \ldots 16)$ form a G-CNN group, namely G-CNNs.

In order to realize the fusion of $Gabor^{CT}$ and $Gabor^{MR}$ images, the model structure of G-CNN is designed as shown in Figure 6. The structure consists of three parts: feature extraction, feature fusion and image reconstruction.

Since the down-sampling feature map inevitably loses the source information of the input images, affecting the quality of the image, thus there is no down-sampling operation in the G-CNN structure. The size of the output feature map at each layer is consistent with the size of the input $Gabor^{CT}$ and $Gabor^{MR}$ images. The network structure consists of four convolutional layers CONV1, CONV2, CONV3 and

CONV4, as shown in Figure 6. Firstly, two convolution layers CONV1 and CONV2 are used to extract the information features of the images; secondly, the feature fusion module FUSE is used to fuse the convolution features of the input images; finally, the fused features are reconstructed through two convolution layers CONV3 and CONV4 to obtain a preliminary fused image of $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ images.

1) FEATURE EXTRACTION

Input $3 \times 224 \times 224$ $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ images into the G-CNN, and two convolution layers are used to extract depth features from $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ images. The first convolutional layer of the advanced ResNet101 pre-trained on ImageNet is used as the first convolutional layer CONV1. CONV1 contains 64 convolution kernels with the size of 7×7, and the step size and filling parameters are set to 1 and 3 respectively. In order to extract deeper features and adjust the convolution features of CONV1, the second convolution layer CONV2 is added. The core number and core size of CONV2 are set to 64 and 3×3 respectively, and the step size and filling parameters of CONV2 are set to 1. To overcome the over-fitting problem and accelerate the network convergence, CONV2 sets up the ReLU activation function [14] and the batch normalization function (BN) [21]. Through the convolution operation of the two convolution layers, 64 feature maps of 224×224 $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ images are obtained respectively.

2) FEATURE FUSION

In the fusion stage, the pixel-level fusion rules are used to fuse the convolution features of multiple inputs, as expressed in equation (9). Generally speaking, there are three commonly used fusion rules, namely element-level maximum, element-level sum and element-level average.

$$\hat{f}(x, y) = \underset{i}{fuse}(f_{i,C_2}(\mathrm{x}, y)), \quad 1 \leq \mathrm{i} \leq N \tag{9}$$

In formula (9), N represents the number of input images, $N \geq 2$, f_{i,c_2} means the feature map of the fourth input image extracted by CONV2. $\hat{f}$ denotes the fusion feature map generated by the feature fusion module and *fuse* indicates the element-level fusion rule. This paper uses the element maximum fusion rule to fuse the feature maps of the $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ images.

3) IMAGE RECONSTRUCTION

After feature fusion, two convolution layers CONV3 and CONV4 are utilized to reconstruct the fused image from the fused convolution features. CONV3 parameter settings are the same as CONV2, which can adjust the fusion convolution features. Therefore, the core number and core size of CONV3 are set to 64 and 3×3 respectively, and the ReLU activation function and batch normalization function of CONV3 are same as CONV2. Finally, CONV4 uses the element weighted average to reconstruct the feature map. The core number and core size of CONV4 are set to 3 and 1×1 respectively, and the step size and filling parameters are set as 0.

After the convolution operation of the two convolution layers of image reconstruction, $3 \times 224 \times 224$ $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ fused image is obtained. The size of the feature map in the whole process is consistent with the size of the source images input in the model, and the information of source images are better maintained. The fused image of $G\text{abor}^{\text{CT}}$ and $G\text{abor}^{\text{MR}}$ has rich texture features and clear edges, but there are still problems such as uncertain and fuzzy boundaries of the diffuse lesion area of malignant brain tumors.

C. FUSION OF G-CNNs BASED ON FUZZY NEURAL NETWORK

The G-CNNs structure is obtained through Gabor representation of medical images in Section A and training G-CNNs in Section B. When the CT and MR images are fused, as shown in Figure 7, the CT and MR images are first passed through the filter bank to obtain different Gabor representation pairs $G\text{abor}^{\text{CT}_i}$ and $G\text{abor}^{\text{MR}_i} (i = 1, \ldots, 16)$,and the corresponding Gabor representation pairs $G\text{abor}^{\text{CT}_i}$ and $G\text{abor}^{\text{MR}_i}$ are put into the corresponding G-CNNs for preliminary fusion, and the fused images of $G\text{abor}^{\text{CT}_i}$ and $G\text{abor}^{\text{MR}_i}$ are obtained. But these fused images have defects such as uncertain and blurred boundaries in the area of the diffuse lesion of malignant brain tumors. Fuzzy neural network can effectively solve the defects in the fusion process through fuzzy sets and membership functions. Therefore, this paper uses fuzzy neural network to fuse the preliminary fused images by G-CNNs one more time to obtain the final fused image.

The Fuzzy Neural Network structure is shown in the blue dotted box in Figure 7. It consists of five layers:Input, Fuzzy partition, Front combination, Inference and Output.

In the first input layer, the sixteen neurons $x_1, x_2, \ldots, x_{16}$ are the specific pixel values represented by the same positions of the 16 Gabor fused images obtained by G-CNNs.

In the second fuzzy partition layer, because the gray levels of different tissues in medical images are different, the pixels are divided into five fuzzy sets according to the pixel value: very dark, dark, normal, bright and very bright. The membership function is used to represent the fuzzy set, here it is represented by a Gaussian function. The Gaussian function is shown in formula (10):

$$\mu_{\mathrm{i}}^{j}(x_i) = \exp\left(-\frac{1}{2}\left(\frac{x_i - c_i^j}{\sigma_i^j}\right)^2\right) \tag{10}$$

In formula (10), c and σ are the center and width of the Gaussian function, $i = 1, 2, \ldots 16, j = 1, 2, \ldots 5$. Different fuzzy sets correspond to different membership functions. The logical sequence of the five fuzzy sets is shown in Figure 8. The five different membership functions correspond to ''very dark'', ''dark'', ''normal'', ''bright'' and ''very bright'' in turn. The abscissa represents the image pixel value, and the ordinate denotes the output value of the membership

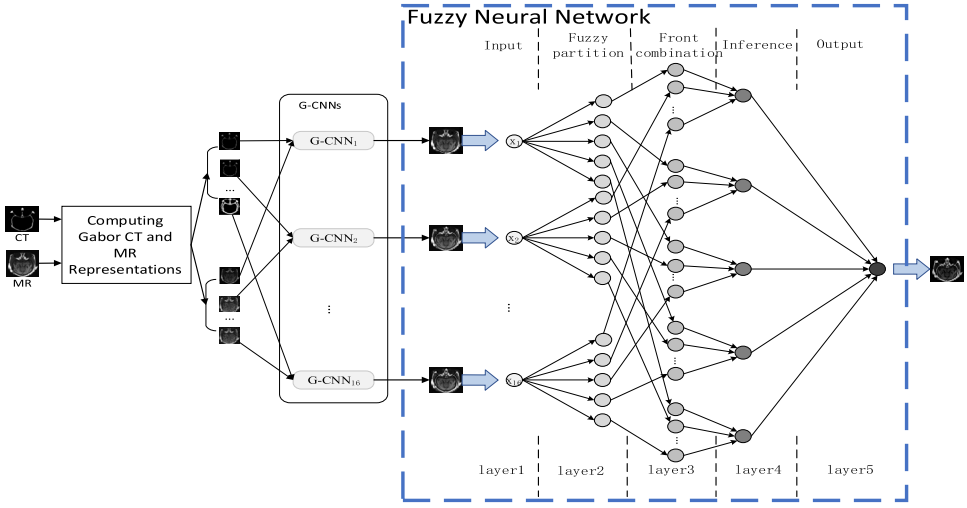


FIGURE 7. Fusion of G-CNNS base on fuzzy neural network.

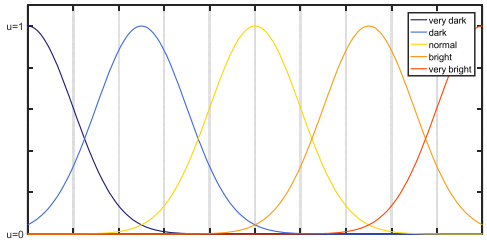


FIGURE 8. Fuzzy set corresponding to membership function.

function μ. The specific process is: for a specific pixel value, the five membership functions have different output μ values, and the fuzzy set corresponding to the largest μ value is selected as the fuzzy set of this pixel value.

The third layer is the front combination layer, which is the premise of forming rules, and arranges the nodes of the same fuzzy set together; the fourth layer is the inference layer, which constitutes rule-based inference, and its nodes are formed by the 16 nodes representing the same fuzzy set in the third layer, and the rules are shown in Figure 9. Five fuzzy rules are set according to the divided five fuzzy sets.

In the fifth output layer, the neuron represents the pixel value of the fused image based on the same position as the input image. In order to obtain the fused image, set the connection weighting factor V between the inference layer and the output layer as follows:

$$V_l = p_0^l + p_1^l x_1 + p_2^l x_2 + \ldots + p_{16}^l x_{16} \tag{11}$$

$p_0, p_1 \ldots p_{16}$ are the weighted coefficient obtained by linear regression in Equation (11) $p_0 = 1, P_1 = P_2 = \ldots = P_{16} = 1/16$, $l = 1, 2, 3, \ldots 5$. Then, the pixel values of 16 Gabor fused images are calculated by the deblurring function, as shown below:

$$\mathrm{y} = \frac{\sum_{l=1}^{L} v_l R_l}{\sum_{l=1}^{L} R_l} \tag{12}$$

In formula (12), L = 5, R_l is the rule output value.

Through the five-layer fuzzy neural network,16 Gabor fused images obtained by G-CNNs are finally fused into a fused image, which solves the problem of fuzzy feature uncertainty in the fusion process, and finally the fused image of CT/MR is obtained with clear edges and sufficient textures and rich content especially for the diffuse lesion of malignant brain tumors.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The hardware platform of this experiment: IntelCorei7-3700 CPU, RAM 16G, NVIDIA 1080Ti GPU 12GB. Software platform is Windows10 64 bit; environment framework is PyTorch 3.6.0.

A. DATA SET AND EXPERIMENTAL ENVIROMENT

In this experiment, high quality CT and MR images with high resolution, rich textures and complex details are selected from the brain images of normal brain and tumor diseases of Harvard Medical School [22] and the Key Laboratory of Biomedical Imaging and Image Big Data of Shanxi Province. After the Gabor representation of the data sets, for each image pair of $G\mathrm{abor}^{\mathrm{CT}}$ and $G\mathrm{abor}^{\mathrm{MR}}$ is first cropped at a random ratio of 0.5 to 1, adjusted to 224 $\times$ 224, and then randomly flipped in the vertical and horizontal directions to double the data set. Finally, the processed source images are used as the input of G-CNN, the corresponding real fused image is used as the label image.

When deep learning algorithms are used, the model parameters need to be optimized with appropriate loss functions to obtain prediction results similar to the real situation. The goal of G-CNN model proposed in this paper is to get a preliminary fused image from multiple input images. The mean square error (MSE), or fundamental loss function B_{loss}, is a basic and commonly used loss function utilized to regularize a model's predictions to approximate the real output. However, only the model with mean square error should be regularized

When x_1 belongs to very dark,x_2 belongs to very dark,...x_{16} belongs to very dark,
then rule output: $R_1 = \mu_1^1(x_1) \times \mu_2^1(x_2) \times ... \times \mu_{16}^1(x_{16})$;
When x_1 belongs to dark,x_2 belongs to dark,...x_{16} belongs to dark,
then rule output: $R_2 = \mu_1^2(x_1) \times \mu_2^2(x_2) \times ... \times \mu_{16}^2(x_{16})$;
When x_1 belongs to normal ,x_2 belongs to normal,...x_{16} belongs to normal,
then rule output: $R_3 = \mu_1^3(x_1) \times \mu_2^3(x_2) \times ... \times \mu_{16}^3(x_{16})$;
When x_1 belongs to bright,x_2 belongs to bright,...x_{16} belongs to bright,
then rule output: $R_4 = \mu_1^4(x_1) \times \mu_2^4(x_2) \times ... \times \mu_{16}^4(x_{16})$;
When x_1 belongs to very bright,x_2 belongs to very bright,...x_{16} belongs to bright,
then rule output: $R_5 = \mu_1^5(x_1) \times \mu_2^5(x_2) \times ... \times \mu_{16}^5(x_{16})$;

FIGURE 9. Five fusion rules of fuzzy neural network system.

to obtain a smooth fused image. In order to solve this problem, the perceptual loss function P_{loss} is usually introduced to regularize the network so as to generate an image with greater structural similarity to the real fused image.

The experiment is divided into pre-training and fine training. First, during the pre-training period, the mini-batch training method is adopted, the batch-size is set to 64, the data set is input into the network in batches, the basic loss function B_{loss} of 5000 iterations is regularized by the loss formula (13). The momentum of the batch normalization function is linearly reduced from 0.99 to 0. Then during the fine training period, we freeze the batch normalization function parameters in order to further regularize the network to generate fused image with greater structural similarity to the real image. On the basis of the basic loss function, the perceptual loss function P_{loss} equation (14) is added as the comprehensive loss function T_{loss} equation (15) to finely train the model. The comprehensive loss function T_{loss} performs the parameters of the convolutional layer 60000 iterations adjustment. In order to ensure the stability of training and accelerate the convergence speed of the model during the fine training process, the batch-size is set to 32. In the setting of learning rate, the basic learning rate during pre-training and fine training is set to 0.01 which is gradually reduced to 0 according to the ''multiple'' learning rate policy.

$$B_{loss} = \frac{1}{H_g W_g} \sum_{i,x,y} [I_P^i(x, y) - I_g^i(x, y)]^2 \tag{13}$$

In formula (13), I_p and I_g respectively indicate the predicted fused image and the real fused image. i represents the channel index of the image. H_g and W_g denote the height and width of the real fused image.

$$P_{loss} = \frac{1}{C_f H_f W_f} \sum_{i,x,y} [f_P^i(x, y) - f_g^i(x, y)]^2 \tag{14}$$

In formula (14), f_p and f_g respectively are the feature maps of the predicted fused image and the real fused image. i indicates the channel index of the feature map. C_f, H_f and W_f respectively denote the number of channels, height and width of the feature map.

$$T_{loss} = w_1 B_{loss} + w_2 P_{loss} \tag{15}$$

In formula (15), w_1 and w_2 respectively represent the weight coefficients of basic loss and perceptual loss. In this paper, both w_1 and w_2 are set to 1.

B. ANALYSIS OF EXPERIMENTAL RESULTS

In order to verify the effectiveness of this method, three groups of registered multimodal brain medical images are selected for a comparative experiment. The size of the three groups of image sets is 256 × 256 pixels and the gray level is 256. The data sets include CT/MR images of cerebrovascular disease, CT/MR images of cerebral infarction and CT/MR images of brain tumor. The proposed fusion methods are compared with the recent state-of-the art fusion methods, which are as follows: 1) the wavelet based fusion method (DWT) [5]; 2) the method based on the laplace pyramid (LAP) [4]; 3) multi-scale transform base approach (NSST) [23]; 4) fusion method based on the combination of multi-scale transformation and PCNN(NSST-PCNN) [24]; 5)the method based on the sparse representation (SR) [25]; 6)the method based on the guided filter (GFF); 7) the method based on the convolutional neural network based approach(CNN) [9]; 8) the method based on the full convolutional neural network (FCNN) [10]; 9) method based on the combination of Gabor and PCNN (GPCNN) [11].

In order to objectively evaluate the performance of different image fusion methods, the quantitative performance of the proposed method on multimodal medical image data sets is evaluated by four commonly used indexes which are mutual information (MI), spatial frequency (SF), standard deviation (SD) and edge information retention ($Q^{AB/F}$). The higher the value of MI, the richer the information obtained from the source image; the larger the value of SD, the higher the image quality; the higher the index of SF, the better the resolution of the fused image; the larger the value of $Q^{AB/F}$, the smaller the loss of edge information of the fused image.

Figure 10 shows the fusion results of CT and MR images of cerebrovascular disease. Fusion results should be as detailed

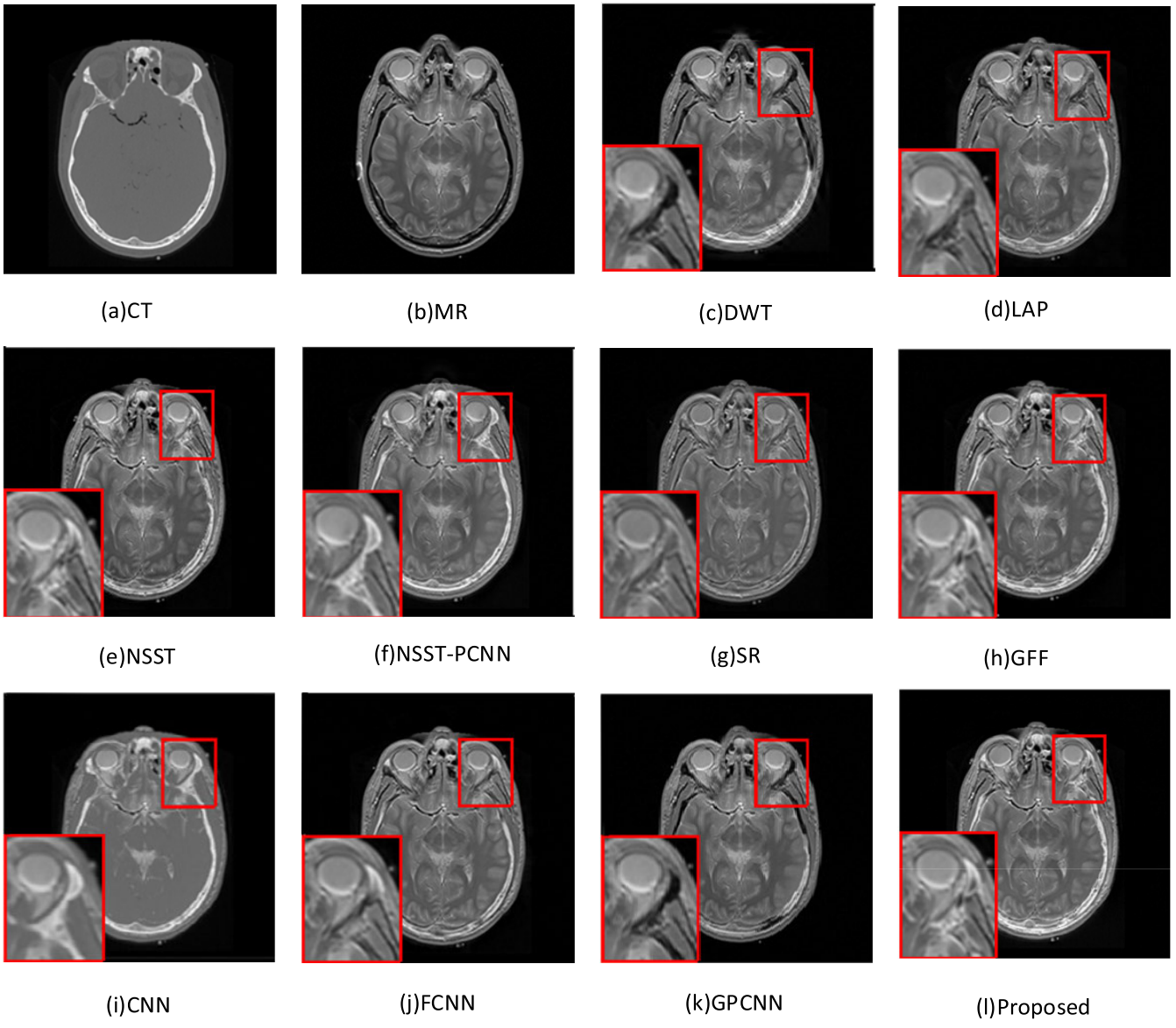


FIGURE 10. CT/MR fusion results of cerebrovascular diseases.

as possible at the vascular lesion site to facilitate the diagnosis and observation of cortical structures. Among them, images a and b represent CT and MR source images of cerebrovascular diseases respectively, and images c-l are respectively the fusion results of DWT, LAP, NSST, NSST-PCNN, SR, GFF, CNN, FCNN, GPCNN and the proposed method. To more visually show the differences in comparison methods, the experimental results are marked with red rectangles. From the red marked area of the fusion result, the fusion result obtained by DWT method lacks edge information and has poor visual effect. The fusion result obtained by LAP method loses a lot of edges of the MR image, leading to blurred image. The fused images by NSST and NSST-PCNN have higher brightness, and the edge information of the lesion site are not clear. The image brightness obtained by SR method is weak. The CNN method has the problem of low contrast of edge information, and the fused image edge details obtained by FCNN and GPCNN methods are not clear enough. The fusion effects of the GFF and the proposed method are better compared with the other methods.

From the quantitative data in Table 1, it can be seen that DWT, LAP, NSST, NSST-PCNN, SR, GFF, CNN, FCNN, GPCNN are lower in index *MI* and index $Q^{AB/F}$, indicating that these methods have lower edge information preservation capabilities compared with the proposed method. And that the proposed method can retain the complete information of the edge of the lesion, the fusion quality is better, which is conducive to the accurate diagnosis

TABLE 1. Comparison of Brain cerebrovascular disease CT/MR fusion results.

Fusion method	The evaluation index			
	MI	*SF*	*SD*	$Q^{AB/F}$
DWT	3.1957	17.9674	51.4073	0.5428
LAP	3.3457	21.9373	59.9259	0.6430
NSST	3.3487	20.8564	54.5502	0.6310
NSST-PCNN	3.3489	20.9562	54.5607	0.6258
SR	3.4259	21.6523	55.7300	0.6670
GEF	3.7904	20.2566	55.6818	0.7626
CNN	3.6876	20.1354	54.8635	0.6750
FCNN	3.6986	20.2653	55.8691	0.6920
GPCNN	3.6952	20.6520	56.1532	0.6956
Proposed	**3.8849**	**22.6706**	**62.3212**	**0.8652**

Note: Bold font indicates optimal results.

TABLE 2. Comparison of Brain cerebral infarction CT/MR fusion results.

	The evaluation index			
	MI	*SF*	*SD*	$Q^{AB/F}$
DWT	3.0829	16.5365	53.5426	0.5125
LAP	3.2691	16.8562	60.5241	0.5632
NSST	3.2456	18.6032	62.4562	0.5631
NSST-PCNN	3.2365	18.5634	62.4358	0.5524
SR	3.3354	20.5691	55.4350	0.5340
GEF	3.5241	21.4652	64.7562	0.5862
CNN	3.5042	20.9653	63.4520	0.5620
FCNN	3.4965	20.4563	62.8652	0.5576
GPCNN	3.6027	22.3021	64.9852	0.6231
Proposed	**3.6285**	**22.5620**	**65.7424**	**0.6690**

Note: Bold font indicates optimal results.

TABLE 3. Comparison of Brain tumor disease CT/MR fusion results.

融合方法	评价指标			
	MI	*SF*	*SD*	$Q^{AB/F}$
DWT	3.0221	17.0573	50.8241	0.5756
LAP	3.2316	17.1079	52.8910	0.6123
NSST	3.0524	17.1520	54.6352	0.6021
NSST-PCNN	3.0570	17.1339	54.1620	0.6005
SR	3.0735	15.6897	54.7173	0.6125
GFF	3.3191	16.9774	62.8964	0.7034
CNN	3.2410	16.0254	60.4531	0.6952
FCNN	3.2245	16.1564	60.7531	0.7012
GPCNN	3.3211	17.1230	62.9420	0.7042
Proposed	**3.6023**	**18.6021**	**63.0515**	**0.7865**

Note: Bold font indicates optimal results

and observation of the boundary information of the lesion area.

Figure 11 shows the fusion results of CT/MR images of cerebral infarction. The results of fusion should fully characterize the texture features of the lesion site so that doctors can accurately determine the site and extent of cerebral infarction. Images a and b represent CT and MR source images of cerebra infarction diseases respectively, and images c-l are respectively the fusion results of DWT, LAP, NSST, NSST-PCNN, SR, GFF, CNN, FCNN, GPCNN and the proposed method. The similarity between the result obtained by DWT with the source images is low, and the texture features of the lesion in the source images cannot be well retained. The image obtained by LAP method has low contrast and high noise, which causes the texture features in the lesion area partially blurred. The fusion results obtained by NSST and NSI-PCNN have high brightness in the lesion area, which weaken the texture information in the image. The fused image obtained by SR method has low brightness and insufficient texture features. The texture features of GFF fused images in the lesion area is not adequately represented. The contrast of the texture features of the fused images obtained by the CNN method is low. The fused image obtained by FCNN loses part of the texture features of the source MR image. The fusion effects of GPCNN and the proposed method are good. It can be seen from Table 2 that the proposed fusion method on the four indicators is higher than other fusion methods. From the fusion results and quantitative indicators, the method of this paper can fully and clearly retain the texture characteristics of the cerebral infarction lesions, the fusion quality is good which is beneficial to the clinical diagnosis.

Figure 12 shows the fusion results of CT/MR images of brain tumor disease. The fusion results should retain the rich texture features and clear edge information of the brain tumor site as much as possible to help doctors accurately determine

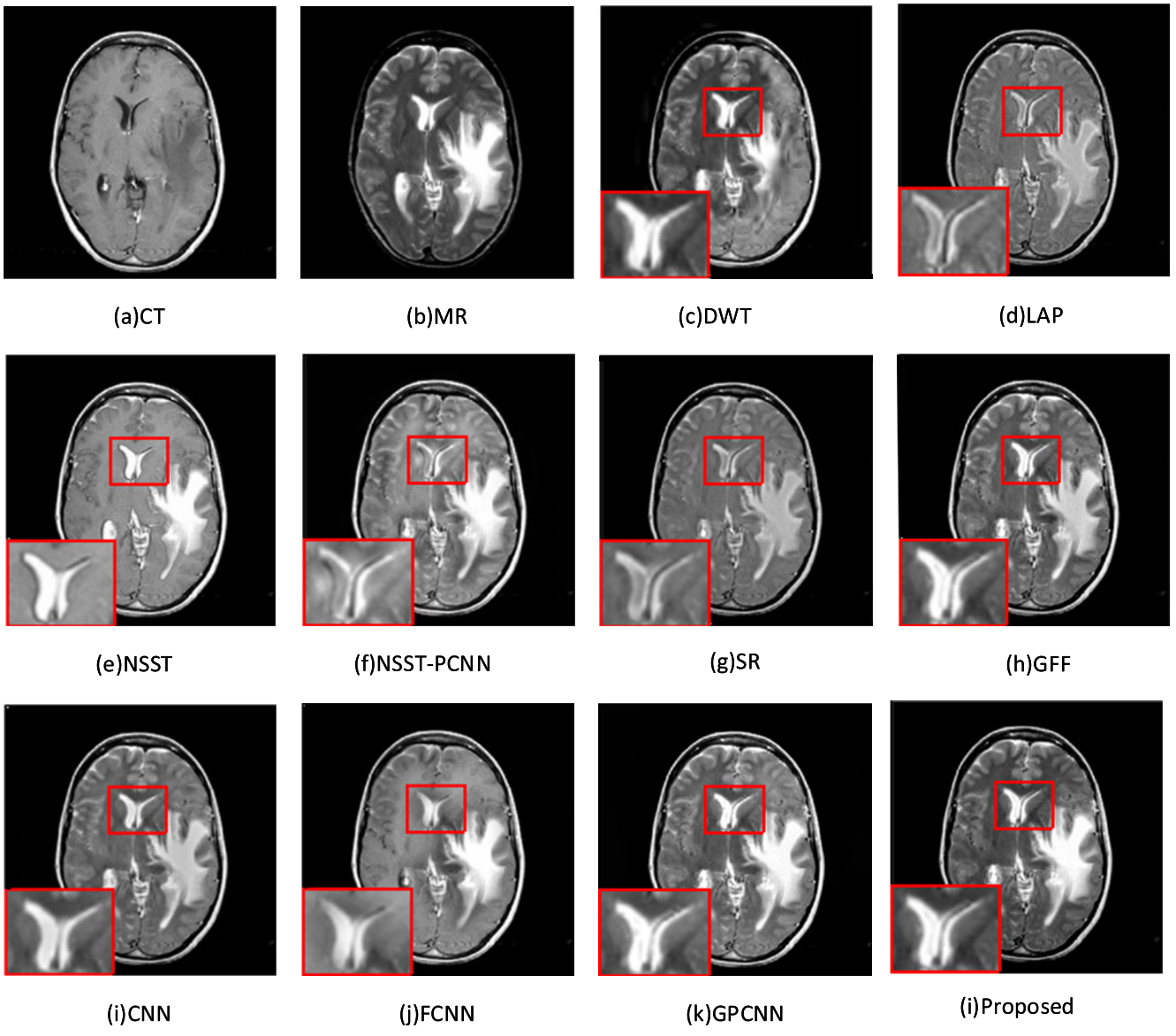


FIGURE 11. CT/MR fusion results of cerebral infarction.

the grade and boundary of the brain tumor, to facilitate the diagnosis to remove the brain tumor in surgery. Images a and b represent CT and MR source images of brain tumor disease respectively. Images c-l are respectively the fusion results of DWT, LAP, NSST, NSST-PCNN, SR, GFF, CNN, FCNN, GPCNN and proposed method. The fused image obtained by DWT method has low edge clarity and fuzzy texture features in the red marked area. The fused image obtained by LAP retains more CT image information, but loses some details of MR image, resulting in partial loss of texture features and edge details of the lesion. The fused images obtained by NSST and NSST-PCNN have higher brightness, which weaken the texture features and edge details of the source images. The fused image obtained by SR has low brightness, and texture features and edge information are not clear enough. The fusion results obtained by CNN and FCNN have very low contrast, and the texture features and edge details of the brain tumor sites are not obvious. In terms of visual effect, the performance of GPCNN and GFF are similar to the method proposed in this paper, but the lesion area is somewhat smooth, the texture features are not sufficient, and the edge details are not clear. From the quantitative indicators in Table 3, it can be seen that the comparison methods are relatively low in index MI and index $Q^{AB/F}$, and it shows that there are defects in the retention of texture features and edge details, which are consistent with subjective vision. The method presented in this paper is superior to the comparison methods in preserving texture features and edge details.

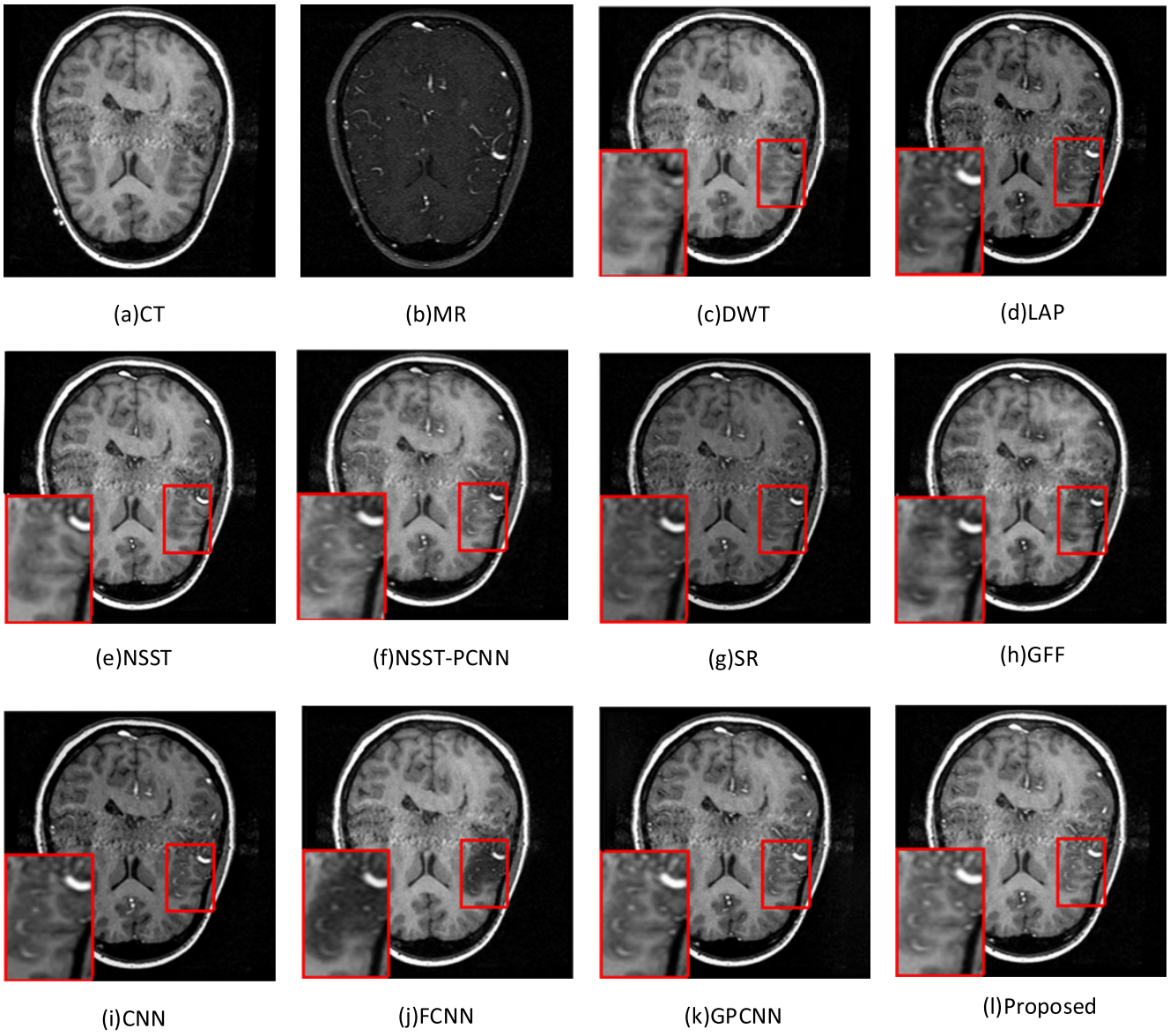


FIGURE 12. CT/MR fusion results of brain tumor disease.

V. CONCLUSION

Aiming at the fact that the existing multimodal medical image fusion methods cannot fully characterize the texture features and edge information of the lesion, an image fusion method based on the combination of G-CNNs and fuzzy neural networks is proposed. This method takes advantage of the idea of integration in machine learning, the medical images are represented by a Gabor filter bank with different directions and scales, which are used as training sets to train the corresponding 16 CNN, and G-CNNs is obtained. When CT and MR images are fused, the Gabor representations are performed first, and the corresponding representations are put into the corresponding 16 G-CNN to obtain multiple preliminary fused Gabor images, and then the multiple preliminary fused Gabor images by G-CNNs are finally fused together through the fuzzy neural network. According to comparison experiments, it can be seen that the proposed method is significantly better than the comparison methods in characterizing the texture features and edge information of the lesion in the medical image. The fused image based on this method is more suitable for assisting doctors to diagnose and treat diseases more accurately.

Although the experimental results verify that the method has a good improvement in the texture and edge details of the fused image, there are still some problems: the scale and direction parameters of Gabor filter bank need to be set manually and the training of CNN network structure has

high energy consumption, which will be the focuses of our follow-up research.

REFERENCES

[1] L. Chang, X. Feng, X. Zhu, R. Zhang, R. He, and C. Xu, "CT and MRI image fusion based on multiscale decomposition method and hybrid approach," *IET Image Process.*, vol. 13, no. 1, pp. 83–88, Jan. 2019.

[2] B. Meher, S. Agrawal, R. Panda, and A. Abraham, "A survey on region based image fusion methods," *Inf. Fusion*, vol. 48, pp. 119–132, Aug. 2019.

[3] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jan. 2017.

[4] J. Du, W. Li, K. Lu, and B. Xiao, "An overview of multi-modal medical image fusion," *Neurocomputing*, vol. 215, pp. 3–20, Nov. 2016.

[5] X. Xu, Y. Wang, and S. Chen, "Medical image fusion using discrete fractional wavelet transform," *Biomed. Signal Process. Control*, vol. 27, pp. 103–111, May 2016.

[6] 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.

[7] Z. Wang, X. Li, H. Duan, X. Zhang, and H. Wang, "Multifocus image fusion using convolutional neural networks in the discrete wavelet transform domain," *Multimedia Tools Appl.*, vol. 78, no. 24, pp. 34483–34512, Dec. 2019.

[8] 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.

[9] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method base on convolutional neural networks," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Xi'an, China, Aug. 2017, pp. 1–7.

[10] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020.

[11] F. Alenezi and E. Salari, "Novel technique for improved texture and information content of fused medical images," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2018, pp. 348–353.

[12] S. QiuHu and F. Zhang, "Multi-modality medical image fusion base on separable dictionary learning and Gabor filtering," *Signal Process., Image Commun.*, vol. 83, Apr. 2020, Art. no. 115758.

[13] S. Molaei and M. E. S. A. Abadi, "Maintaining filter structure:A Gabor-based convolutional neural network for image analysis," *Appl. Soft Comput. J.*, vol. 88, May 2020, Art. no. 105960.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jun. 2017, pp. 2881–2890.

[15] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 694–711.

[16] S. Das and M. K. Kundu, "NSCT-based multimodal medical image fusion using pulse-coupled neural network and modified spatial frequency," *Med. Biol. Eng. Comput.*, vol. 50, no. 10, pp. 1105–1114, Oct. 2012.

[17] P. V. de Campos Souza, "Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature," *Appl. Soft Comput.*, vol. 92, Jul. 2020, Art. no. 106275.

[18] T.-L. Nguyen, S. Kavuri, and M. Lee, "A multimodal convolutional neuro-fuzzy network for emotion understanding of movie clips," *Neural Netw.*, vol. 118, pp. 208–219, Oct. 2019.

[19] Z. Chao, D. Kim, and H.-J. Kim, "Multi-modality image fusion based on enhanced fuzzy radial basis function neural networks," *Phys. Medica*, vol. 48, pp. 11–20, Apr. 2018.

[20] K. V. Shihabudheen and G. N. Pillai, "Recent advances in neuro-fuzzy system: A survey," *Knowl.-Based Syst.*, vol. 152, pp. 136–162, Jul. 2018.

[21] S. Iofffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[22] *The Whole Brain Altas of Harvard Medical School*. [Online]. Available: http://www.med.rvad.edu

[23] 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.

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

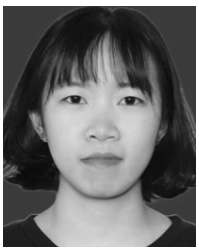
[25] S. Polinati and R. Dhuli, "Structural and functional medical image fusion using an adaptive Fourier analysis," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 23645–23668, Sep. 2020.

LIFANG WANG was born in 1977. She received the bachelor's degree from Shaanxi Normal University, in 2000, and the master's and Ph.D. degrees from the North University of China, in 2007 and 2014, respectively.

She is currently an Associate Professor. As the first author, she has published more than 20 scientific research articles, including eight articles included in EI. Her main research interests include machine vision, big data processing, and medical processing.

JIN ZHANG was born in 1996. She received the bachelor's degree from the North University of China, in 2019. She is currently pursuing the M.S. degree.

Her main research interests include medical image fusion and machine learning.



YANG LIU was born in 1997. She is currently pursuing the M.S. degree. Her main research interest includes medical image fusion.

JIA MI was born in 1998. She is currently pursuing the M.S. degree. Her main research interest includes medical image fusion.

JIONG ZHANG was born in 1998. He is currently pursuing the M.S. degree. His main research interest includes medical image fusion.