



Article

Multi-Modality Medical Image Fusion Using Convolutional Neural Network and Contrast Pyramid

Kunpeng Wang ^{1,2}, Mingyao Zheng ³, Hongyan Wei ³, Guanqiu Qi ⁴ and Yuanyuan Li ^{3,*}

¹ School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China; wkphnzk@163.com

² Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, Mianyang 621010, China

³ College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; ZMYzhengmingyao@126.com (M.Z.); weihy12@126.com (H.W.)

⁴ Computer Information Systems Department, State University of New York at Buffalo State, Buffalo, NY 14222, USA; qig@buffalostate.edu

* Correspondence: liyy@cqupt.edu.cn

Received: 7 March 2020; Accepted: 8 April 2020; Published: 11 April 2020



Abstract: Medical image fusion techniques can fuse medical images from different morphologies to make the medical diagnosis more reliable and accurate, which play an increasingly important role in many clinical applications. To obtain a fused image with high visual quality and clear structure details, this paper proposes a convolutional neural network (CNN) based medical image fusion algorithm. The proposed algorithm uses the trained Siamese convolutional network to fuse the pixel activity information of source images to realize the generation of weight map. Meanwhile, a contrast pyramid is implemented to decompose the source image. According to different spatial frequency bands and a weighted fusion operator, source images are integrated. The results of comparative experiments show that the proposed fusion algorithm can effectively preserve the detailed structure information of source images and achieve good human visual effects.

Keywords: medical image fusion; convolutional neural network; image pyramid; multi-scale decomposition

1. Introduction

In the clinical diagnosis of modern medicine, various types of medical images play an indispensable role and provide great help for the diagnosis of diseases. To obtain sufficient information for accurate diagnosis, doctors generally need to combine multiple different types of medical images from the same position to diagnose the patient's condition, which often causes great inconvenience. If multiple types of medical images are only analyzed by doctor's space concepts and speculations, the analysis accuracy is subjectively affected, even parts of image information may be neglected. Image fusion techniques provide an effective way to solve these issues [1]. As the variety of medical imaging devices increases, the obtained medical images from different modalities contain complementary as well as redundant information. Medical image fusion techniques can fuse multi-modality medical images for more reliable and accurate medical diagnosis [2,3].

This paper proposes a CNN-based medical image fusion method. First, CNN-based model generates a weight map for any-size source image. Then, Gaussian pyramid decomposition is performed on the generated weight map, and the contrast image pyramid decomposition is applied to source images for obtaining the corresponding multi-scale sub-resolution images. Next, a weighted fusion operator based on the measurement of regional characteristics is used to set different

thresholds for the top layer and the remaining layers of sub-decomposed images to obtain the fused sub-decomposed images. Finally, the fused image is obtained by the reconstruction of contrast pyramid. This paper has three main contributions as follows:

- (1) In training process of CNN, source images can be directly mapped to the weight map. Thus, it can also achieve the measurement of activity level and weight distribution in an optimal way to overcome the difficulties in design by learning network parameters in the training process.
- (2) Human visual system is sensitive to the changes of image contrast. Thus, this paper proposes a multi-scale contrast pyramid decomposition based image fusion solution, which can selectively highlight the contrast information of fused image to achieve better human visual effects.
- (3) The proposed solution uses a weighted fusion operator based on the measurement of regional characteristics. In the same decomposition layer, the fusion operators applied to different local regions may be different. Thus, the complementary and redundant information of fused image can be fully explored to achieve a better fusion effect and highlight important detailed features.

The remainder of this paper is organized as follows. Section 2 discusses the related works of medical image fusion. Section 3 demonstrates the proposed CNN-based medical image fusion solution in detail. Section 4 presents the comparative experiments and compares corresponding results. Section 5 concludes this paper.

2. Related Works

Researchers have proposed many medical image fusion methods in recent years [4,5]. Mainstream medical image fusion methods include decomposition-based and learning-based image fusion methods [6,7]. As a commonly used decomposition-based medical image fusion method, multi-scale transform (MST) generally has three steps in the fusion process: decomposition, fusion, and reconstruction. Pyramid-based method, wavelet, and multi-scale geometric analysis (MAG) based method are commonly used in MST [8]. In MAG-based methods, nonsubsampling contourlet transform (NSCT) [9,10] and nonsubsampling shearlet transform (NSST) [11] based methods have high efficiency in image representation. In addition to image transformation, the analysis of high- and low-frequency coefficients is also a key issue of MST-based fusion methods. Traditionally, the activity level of high-frequency coefficient is usually based on its absolute value. It is calculated in a pixel- or window-based way, and then uses a simple fusion rule, such as the selection of the maximum or weighted average, to obtain the fused coefficient. Averaging the coefficients of different source images was the most popular low-frequency fusion strategy in early research. In recent years, more advanced image transformations and more complex fusion strategies have been developed [12–17]. Liu proposed an integrated sparse representation (SR)- and MST-based medical image fusion framework [18]. Zhu proposed an NSCT based multi-modality decomposition method for medical images, which uses the phase consistency and local Laplacian energy to fuse high- and low-pass sub-bands, respectively [9]. Yin proposed a multi-modality medical image fusion method in NSST domain, which introduced pulse coupled neural network (PCNN) for image fusion [19]. To improve the fusion quality of multi-modality images, a novel multi-sensor image fusion framework based on NSST and PCNN was proposed by Li [20].

In the past decade, learning-based methods have been widely used in medical image fusion. Especially, SR- and deep learning-based fusion methods are most widely used [21,22]. In the early stage, SR-based fusion methods used a standard sparse coding model based on a single image component and local image blocks [23–25]. In the original spatial domain, source images were segmented into a set of overlapping image blocks for sparse coding. Most existing SR-based fusion methods attempt to improve their performances in the following ways: adding detailed constraints [5], designing more efficient dictionary learning strategy [26], using multiple sub-dictionaries in representation [27,28], etc. As an SR-based model, Kim proposed a dictionary learning method based on joint image block clustering for multi-modality image fusion. Zhu proposed a medical image fusion method based on

cartoon-texture decomposition (CTD), and used an SR-based fusion strategy to fuse the decomposed coefficients [29]. Liu proposed an adaptive sparse representation (ASR) model for simultaneous image fusion and denoising [28]. All the above-mentioned methods propose complex fusion rules or different SR-based models. However, these specific rules cannot be applicable to every type of medical image fusion [27].

With the rapid development of artificial intelligence, deep learning-based image fusion methods have become a hot research topic [30–33]. As a main representative of artificial intelligence, deep learning is developed on the basis of traditional artificial neural networks. It can learn data characteristics autonomously, establish a human-like learning mechanism by simulating the neural network of human brain, and then analyze and learn the related data, such as images and texts [34,35]. CNN as a classical deep learning model can achieve the encoding of direct mapping from source images to weight map during the training process [29,36]. Thus, both activity-level measurements and weight distribution can be achieved together in an optimal way by learning network parameters. In addition, CNN's local connection and weight sharing feature can further improve the performance of image fusion algorithms, while reducing the complexity of entire network and the number of weights. At present, CNN plays an increasingly important role in medical image fusion. Xia integrated multi-scale transform and CNN into a multi-modality medical image fusion framework, which uses the deep stacked neural network to divide source images into high- and low-frequency components to do corresponding image fusion [37]. Liu proposed a CNN-based multi-modality medical image fusion algorithm, which applies image pyramids to the medical image fusion process in a multi-scale manner [38].

The calculation of weight map, which fuses the pixel activity information from different sources, is one of the most critical issues in existing deep learning based image fusion [38]. Most existing fusion methods use a two-step solution that contains activity-level measurement and weight assignment. In traditional transform-domain fusion methods, the absolute value of decomposition coefficient is used to measure its activity first. Then, the fusion rule, such as “choose-max” or “weighted-average”, is used to select the maximum or weighted average [39]. According to the obtained measurements, the corresponding weights are finally assigned to different sources. To improve the fusion performance, many complicated decomposition methods and detailed weight assignment strategies have been proposed in recent years [28,40–45]. However, it is not easy to design an ideal activity level measurement or weight assignment strategy, which can consider all key issues [37].

3. The Proposed Medical Image Fusion Solution

As shown in Figure 1, the proposed medical image fusion framework has three main steps. First, it uses Siamese network model to generate the same-size weight map W for any-size source image A and B , respectively. Then, Gaussian pyramid decomposition is applied to the generated weight map W to obtain corresponding multi-scale sub-decomposed image G_W , which is used to determine the fusion operator in coefficient fusion process. $G_{W,l=N}^{l,k}$ and $G_{W,0 \leq l < N}^{l,k}$ are the top layer and the remaining layers of sub-decomposed image. It applies the contrast pyramid to the decomposition of source image A and B . The multi-scale sub-decomposed images C_A and C_B are obtained for the subsequent coefficient fusion process. $C_{A,l=N}^{l,k}$ and $C_{B,l=N}^{l,k}$ are the top layer of sub-decomposed image C_A and C_B , respectively. $C_{A,0 \leq l < N}^{l,k}$ and $C_{B,0 \leq l < N}^{l,k}$ are used to represent the remaining layers of sub-decomposed image C_A and C_B , respectively. Finally, different thresholds are set for the top layer and the remaining layers of sub-decomposed images, respectively. A weighted fusion operator based on the measurement of regional characteristics is used to fuse the different regions in the same decomposition layer to obtain the fused sub-decomposed image C_F . The final fused image F is obtained by the reconstruction of contrast pyramid.

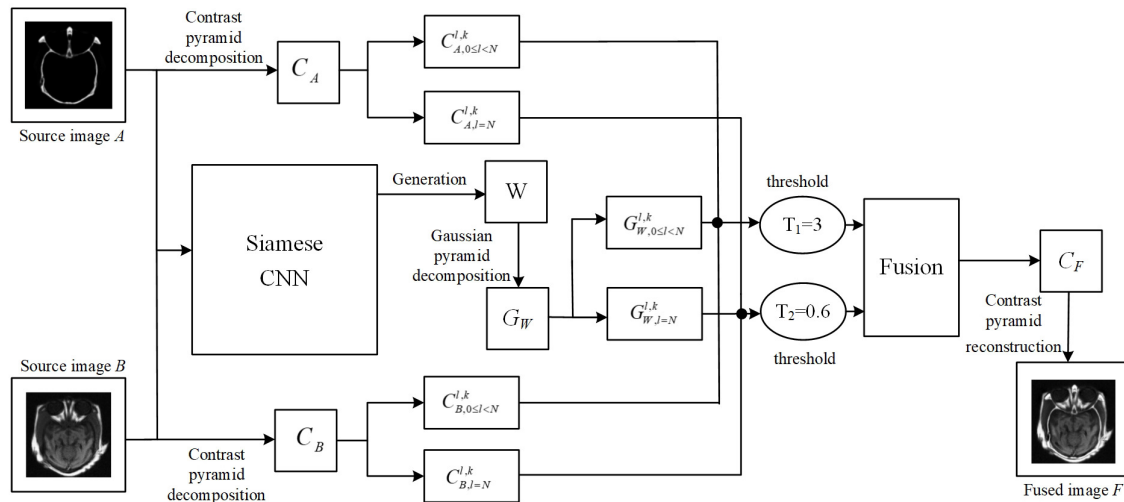


Figure 1. The proposed medical image fusion framework.

3.1. Generation of CNN-Based Weight Map

3.1.1. Network Construction

To obtain the weight map of pixel activity information from multiple source images, the proposed method uses CNN to achieve the measurement of optimal pixel activity level and weight distribution. This paper uses siamese network to improve the efficiency of CNN training. Siamese network has two branches. Each branch contains three convolutional layers and one max-pooling layer. The first two layers are convolutional layers. The first layer is used for the simple feature extraction of input image. In the second layer, the number of feature maps increases. The features of output map in the upper convolutional layer are extracted. The third layer is a max-pooling layer. It removes unimportant samples from feature map to further reduce the number of parameters. As a convolution layer, the fourth layer extracts more complex features from the output map of the pooling layer. To reduce memory consumption, it uses a lightweight network structure to reduce the training complexity. Specifically, the feature map of each branch's final output is concatenated first. Then, the concatenated ones are directly connected to a two-dimensional vector by a fully connected layer. To predict the probability distribution of different characteristics, the two-dimensional vector obtained by mapping is sent to a bi-directional softmax layer, and then classified by probability value. This paper uses the siamese network training architecture shown in Figure 2.

To achieve the classification in CNN network, this paper uses softmax classifier to obtain the classification probability by Equation (1).

$$f(p_i) = \frac{e^{p_i}}{\sum_{j=1}^n e^{p_j}} \quad (1)$$

If one p_i is larger than all the other p , then its mapping component is close to 1, and the others are close to 0, which normalizes all input vectors. The batch size is set to 128, thus the softmax loss function is obtained as Equation (2).

$$L = \sum_{i=0}^{batchsize} -\log f(p_i) \quad (2)$$

Taking the softmax loss function as the optimization goal, stochastic gradient descent is used to minimize the loss function. As the initial parameter settings, the momentum and weight decay are set to 0.9 and 0.0005, respectively. Thus, Equations (3) and (4) are used to update the weights.

$$v_{i+1} = 0.9 \cdot v_i - 0.0005 \cdot \alpha \cdot w_i - \alpha \cdot \frac{\partial L}{\partial w_i} \quad (3)$$

$$w_{i+1} = w_i + v_{i+1} \quad (4)$$

where v_i is the dynamic variable, w_i is weight after i th iteration, α is the learning rate, L represents the loss function, and $\frac{\partial L}{\partial w_i}$ is the loss derivative of weight w_i .

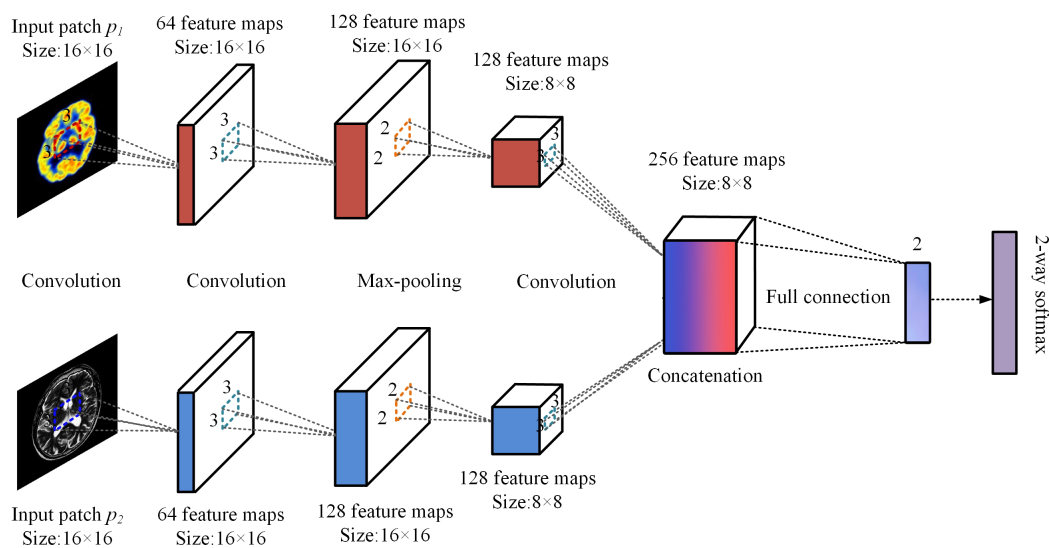


Figure 2. Siamese network training architecture.

3.1.2. Networking Training

It selects a high-quality multi-modality medical image set from <http://www.med.harvard.edu/aanlib/home.html> as training samples. It applies a Gaussian filter to each image to obtain corresponding five different-level fuzzy versions. Specifically, a Gaussian filter with a standard deviation of 2 and a cutoff value of 7×7 is used. Gaussian filter is used to blur the original image to obtain the first blurred image. In the following Gaussian filtering, the previous output image is used as the next input image. For instance, the output image of first Gaussian filtering is used as the input image of the second Gaussian filtering. Then, for each blurred and clear image, it randomly samples 20 pairs of 16×16 image blocks. p_c and p_b represent a pair of clear and blurred image blocks. When $p_1 = p_c$ and $p_2 = p_b$, it is defined as a positive example (marked as 1), where p_1 and p_2 are the inputs for the first and second branch, respectively. Oppositely, when $p_1 = p_b$ and $p_2 = p_c$, it is defined as a negative example (marked as 0). Therefore, the training set is ultimately composed of positive and negative examples. After the sample is generated, the weight of each convolutional layer is initialized by using Xavier algorithm, which adaptively determines the initialization scale based on the number of input and output neurons. The deviation of each layer is initialized to 0. The inclination rates of all layers are equal, and their initial values are set to 0.0001. When the loss reaches a steady state, the inclination rates are manually reduced to 10% of previous values. After about ten iterations, it can complete the network training.

3.1.3. The Generation of Weight Map W

In the image testing and fusion process, to process any-size source images, it converts the fully connected layer into two equivalent convolutional layers of equal kernel size. When the conversion is completed, any-size image A and B to be fused can be processed as a whole to generate a dense prediction map S . Every prediction S_i is a two-dimensional vector, and the value of each dimension is between 0 and 1. If one dimension is larger than another, this dimension can be normalized to 1, and the other one is set to 0. It simplifies the weight of corresponding image block with an output dimension value of 1. For two adjacent predictions in S , the steps of corresponding image blocks overlap. For overlapping areas, the weights are averaged. The output is the average weight of the overlapping image blocks. In the above way, it is possible to input any-size image A and B into the network, and generate the corresponding same-size weight map W .

3.2. Pyramid Decomposition

This paper uses both contrast pyramid and Gaussian pyramid to decompose source images. It builds the contrast pyramid first. Then, when the Gaussian pyramid is established, G^0 is the zeroth layer (bottom layer), and the l 'th layer G^l can be constructed in the following manner. As shown in Equation (5), it convolves G^{l-1} by a window function $w(m, n)$ with low-pass characteristics first, and then downsamples the convolutional result by the interlaced every other row and column.

$$G^l = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) G^{l-1}(2i + m, 2j + n), \quad 0 < l \leq N, 0 \leq i < C_l, 0 \leq j < R_l \quad (5)$$

where $w(m, n)$ is the window function, C_l and R_l are the number of columns and the number of rows in the l 'th-layer sub-image of the Gaussian pyramid, respectively, and N is the total number of the pyramid layers.

- (1) Separability: $w(m, n) = w(m)w(n)$, $m \in [-2, 2]$, $n \in [-2, 2]$;
- (2) Normalization: $\sum_{m=-2}^2 w(m) = 1$;
- (3) Symmetry: $w(n) = w(-n)$; and
- (4) Equal contribution of odd and even terms: $w(-2) + w(2) + w(0) = w(-1) + w(1)$.

According to the above constraints, it can construct $w(0) = 3/8$, $w(1) = w(-1) = 1/4$, and $w(2) = w(-2) = 1/16$. Then, according to Constraint 1, it can get the window function $w(m, n)$ by calculation, as shown in Equation (6).

$$w = \frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}. \quad (6)$$

At this point, the image Gaussian pyramid is constructed by G^0, G^1, \dots, G^N .

After the construction of a Gaussian pyramid image by halving the size of each layer one by one, the interpolation method is used to interpolate and expand the Gaussian pyramid. Thus, the expanded l th-layer image G^l and the $l - 1$ th-layer image G^{l-1} have the same size and the operation is shown as follows:

$$G_*^l(i, j) = 4 \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) G^l \left[\frac{m+i}{2}, \frac{n+j}{2} \right], \quad 0 < l \leq N, 0 < i < C_l, 0 < j < R_l \quad (7)$$

$$G^l \left[\frac{m+i}{2}, \frac{n+j}{2} \right] = \begin{cases} G^l \left(\frac{m+i}{2}, \frac{n+j}{2} \right), & \text{when } \frac{m+i}{2}, \frac{n+j}{2} \text{ are integer} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where G_*^l is an expansion version of image Gaussian pyramid G^l . According to the above formulas, an expansion sequence is obtained by interpolating and expanding each layer of Gaussian pyramid, respectively.

According to the above formulas, an expansion sequence is obtained by interpolating and expanding each layer of Gaussian pyramid respectively. The decomposition of image contrast is shown as Equation (9).

$$\begin{aligned} C^l &= \frac{G^l}{G_*^l} - I, & N > l \geq 0 \\ C^N &= G^N, & l = N \end{aligned} \quad (9)$$

where C^l is the contrast pyramid, G^l is the Gaussian pyramid, I is the image decomposed by contrast pyramid, and l is the decomposition level, which composes the contrast pyramid C^0, C^1, \dots, C^N of source image.

Source image A and B are decomposed into corresponding sub-images by contrast pyramid, respectively. For the weight map generated in CNN network, it is decomposed into sub-images by Gaussian pyramid. Different thresholds are set for the top layer and the remaining layers of the obtained sub-images respectively in the fusion processing.

3.3. Fusion Rules

In the fusion process, to obtain better visual characteristics, richer details, and outstanding fusion effects, this paper adopts new fusion rules and the weighted average fusion operators based on regional characteristics. The fusion rules and operators are shown as follows:

- (1) After the contrast pyramid decomposition, it calculates the energy E_A^l and E_B^l of corresponding local regions in each decomposition level l of source image A and B , respectively.

$$\begin{aligned} E_A^l(x, y) &= \sum_m \sum_n C_A^l(x+m, y+n)^2 \\ E_B^l(x, y) &= \sum_m \sum_n C_B^l(x+m, y+n)^2 \end{aligned} \quad (10)$$

where $E^l(x, y)$ represents the local area energy centered at (x, y) on the l th layer of contrast pyramid, C^l is the l th-layer image of contrast pyramid, and m and n represent the size of local area.

- (2) Calculate the similarity of corresponding local regions in two source images.

$$M^l(x, y) = \frac{2 \sum_m \sum_n C_A^l(x+m, y+n) C_B^l(x+m, y+n)}{E_A^l(x, y) E_B^l(x, y)} \quad (11)$$

where E_A^l and E_B^l are calculated by Equation (10). The range of similarity is $[-1, 1]$, and a value close to 1 indicates high similarity.

- (3) Determine the fusion operators. Define a similarity threshold T (when $0 \leq l < N, T_1 = 3$; when $l = N, T_2 = 0.6$). When $M^l(x, y) < T$, it obtains:

$$\left. \begin{aligned} &\text{when } E_A^l(x, y) \geq E_B^l(x, y), C_F^l(x, y) = C_A^l(x, y); \\ &\text{when } E_A^l(x, y) < E_B^l(x, y), C_F^l(x, y) = C_B^l(x, y); \end{aligned} \right\} \quad (12)$$

when $M^l(x, y) \geq T$, weight map W based weighted mean model is:

$$\left. \begin{array}{l} \text{when } E_A^l(x, y) \geq E_B^l(x, y), \\ C_F^l(x, y) = W_{\max}^l(x, y) C_A^l(x, y) + W_{\min}^l(x, y) C_B^l(x, y); \\ \text{when } E_A^l(x, y) < E_B^l(x, y), \\ C_F^l(x, y) = W_{\min}^l(x, y) C_A^l(x, y) + W_{\max}^l(x, y) C_B^l(x, y); \end{array} \right\} \quad (13)$$

where C_F^l is the l th layer of sub-image after fusion.

$$\begin{aligned} W_{\min}^l(x, y) &= G_W^l(x, y) \\ W_{\max}^l(x, y) &= 1 - W_{\min}^l(x, y) \end{aligned} \quad (14)$$

Finally, the integration strategy can be summarized as a whole by Equation (15).

$$C_F^l(x, y) = \begin{cases} C_A^l(x, y), & \text{if } M^l(x, y) < T \& E_A^l(x, y) \geq E_B^l(x, y); \\ C_B^l(x, y), & \text{if } M^l(x, y) < T \& E_A^l(x, y) < E_B^l(x, y); \\ W_{\max}^l(x, y) C_A^l(x, y) + W_{\min}^l(x, y) C_B^l(x, y), & \text{if } M^l(x, y) \geq T \& E_A^l(x, y) \geq E_B^l(x, y); \\ W_{\min}^l(x, y) C_A^l(x, y) + W_{\max}^l(x, y) C_B^l(x, y), & \text{if } M^l(x, y) \geq T \& E_A^l(x, y) < E_B^l(x, y); \end{cases} \quad (15)$$

According to the above algorithm, when the similarity between the corresponding local regions of source image A and B is less than threshold T , it means that the “energy” difference of two local regions is large. At this time, the central pixel of the region with a larger “energy” is selected as the central pixel of corresponding region in the fused image. Conversely, when the similarity is greater than or equal to threshold T , it means that the “energy” of the region is similar in two source images. At this time, the weighted fusion operator is used to determine the contrast or gray value of the central pixel of the region in the fused image.

Since the central pixel with large local energy represents a distinct feature of source image, the local image features generally do not only depend on a certain pixel. Therefore, the weighted fusion operator based on region characteristics is used, which is more reasonable than other determination methods of fused pixel based on the simple selection or the weight of an independent pixel.

Finally, the decomposed sub-image C_F^l obtained after fusion is inversely transformed by contrast pyramid, which is also called image reconstruction. According to Equation (16), the accurate image reconstruction by contrast pyramid can be obtained.

$$\begin{aligned} G^l &= (C^l + F) \odot G_*^l, & N > l \geq 0 \\ G^N &= C^N, & l = N \end{aligned} \quad (16)$$

where \odot denotes Hadamard product (also known as the element-wise multiplication).

The fused image F can be obtained by calculating the above-mentioned image reconstruction formula. Algorithm 1 shows the main steps of the proposed medical image fusion solution.

Algorithm 1 Proposed NSST-based multi-sensor image fusion framework.**Input:**

source image A and B ;

Parameters: pyramid decomposition level l , the number of pyramid levels N , similarity threshold T

Output:

the fused image F

- 1: It inputs two any-size source images A and B to the trained siamese network.
- 2: It generates a dense prediction map S , where each prediction has two dimensions.
- 3: **for** any prediction S_i **do**
- 4: It does normalization processing to obtain a corresponding image block weight with a dimension value of 1.
- 5: **end for**
- 6: **for** an overlapping region of two adjacent predictions S_j and S_{j+1} **do**
- 7: It does the averaging process to obtain the mean value of the overlapping image block weights.
- 8: It outputs the same size weight map W as source image.
- 9: **end for**
- 10: **for** each source image A , B , and weight map W **do**
- 11: It does pyramid decomposition respectively to obtain a contrast sub-images C_A , C_B and a Gaussian sub-image G_W .
- 12: **for** each decomposition level l obtained by the contrast pyramid decomposition of source image **do**
- 13: It calculates the energy $E_{A,B}^l(x,y)$ of its corresponding local area.
- 14: It determines the similarity of fusion mode $M^l(x,y)$.
- 15: It defines a similarity threshold T (when $0 \leq l < N$, $T_1=3$; when $l = N$, $T_2=0.6$) to determine the strategy of coefficient fusion.
- 16: **end for**
- 17: **end for**
- 18: The fused image F is obtained by the inverse pyramid transform of sub-image C_F^l after fusion.

4. Comparative Experiments and Analysis

4.1. Experiment Results and Analysis

The following comparative experiments were used to prove that the proposed CNN-based algorithm has good performance in medical image fusion. Eight different image fusion methods were used to fuse MR-CT, MR-T1-MR-T2, MR-PET, and MR-SPECT images, respectively. These eight methods are MST-SR [18], NSCT-PC [9], NSST-PCNN [19], ASR [28], CT [29], KIM [26], CNN-LIU [38], and the proposed solution.

Figure 3 shows the results of MR-CT image fusion experiments. In Figure 3c, the fused image obtained by MST-SR method has a general visualization performance, and the image contrast is high by analyzing the partially enlarged image. As shown in Figure 3d,e, the fused images by NSCT-PC and NSST-PCNN have a high brightness. According to the partial enlargements marked in green and red dashed frame, both methods have the poor performance in the preservation of image details. In Figure 3f,g, the fused images obtained by ASR and CT have low brightness. According to the analysis of details, the detailed information of image edge is not obvious, which is not good for human eye observation. As shown in Figure 3h, the fused image obtained by KIM has low sharpness and poor visual effect. Comparing Figure 3i,j, as well as the partially magnified images, it is difficult to visually distinguish the quality of the fused image obtained by CNN-LIU and the proposed method.

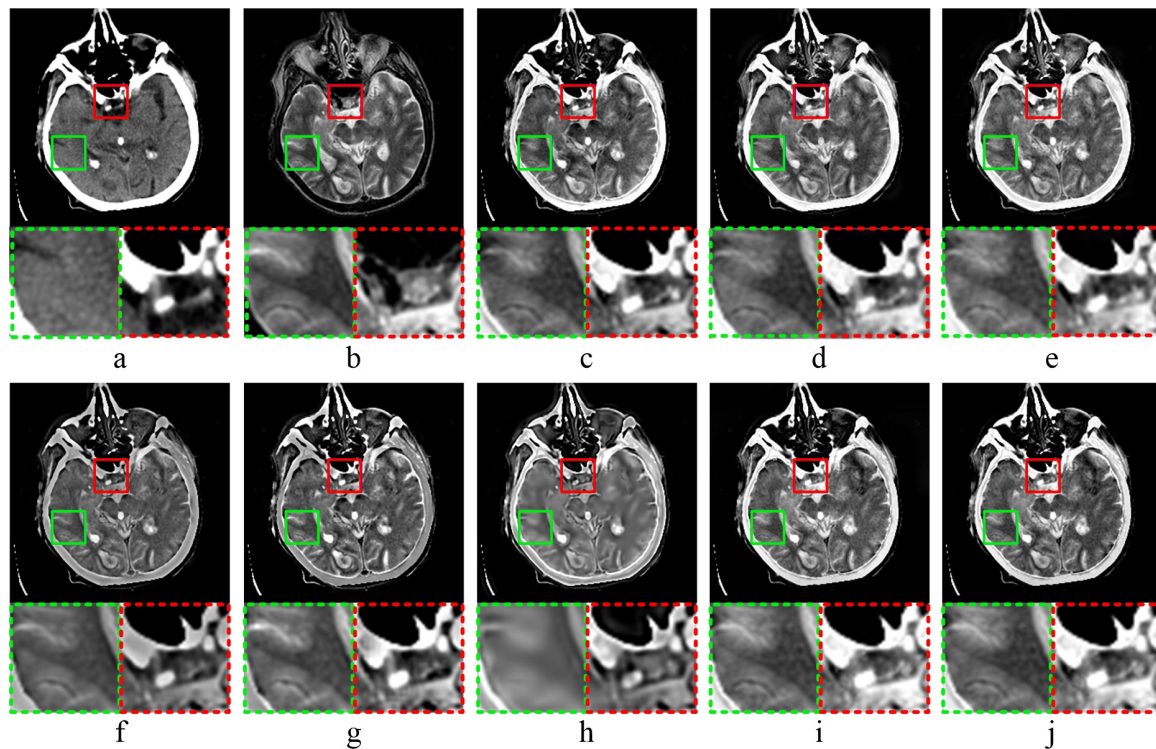


Figure 3. MR-CT image fusion experiments: (a,b) source images; and (c–j) the fused image obtained by MST-SR, NSCT-PC, NSST-PCNN, ASR, CT, KIM, CNN-LIU, and the proposed method, respectively. Two partially enlarged images marked in green and red dashed frames correspond to the regions surrounded by green and red frames in the fused image.

The results of MR-T1-MR-T2 image fusion experiments are shown in Figure 4. Comparing the fused result (Figure 4c) with source image (Figure 4a,b), the fused image obtained by MST-SR method has a low similarity to source image (Figure 4a), and does not well retain the detailed structure information of source image (Figure 4a). As shown in Figure 4d, the fused image obtained by NSCT-PC method is too smooth in some areas, and the detailed image texture is not sufficiently obvious. In Figure 4e, the fused image obtained by NSST-PCNN method has high brightness, and does not well preserve the detailed features of source images. ASR method obtains the fused image with low contrast and a lot of noises, as shown in Figure 4f. The fused image shown in Figure 4g was obtained by CT method, and has high edge brightness, which weakens the detailed texture information of image edges. According to Figure 4h, the fused image obtained by KIM method has low sharpness, and is blurred. As shown in Figure 4i,j, CNN-LIU and the proposed method reach the almost same visual performance of human eyes.

Figures 5 and 6 show the results of MR-PET image fusion experiments. In both Figures 5c and 6c, the fused images obtained by MST-SR method have high darkness, which is not conducive to human visual observation. According to Figure 5d,e, as well as the partially magnified areas, the fused images obtained by NSCT-PC and NSST-PCNN method have high brightness, and the detailed image information is not clear. Figures 5f,g and 6f,g show that the fused images obtained by ASR and CT methods have low brightness. It means that these two methods have poor performance in the preservation of image details. Comparing Figure 5i,j, the fused image of KIM method has low sharpness, which means the image is blurred. As shown in Figure 5i, the fused image obtained by CNN-LIU method has a low contrast, and the detailed edge information is not obvious. Both Figures 5j and 6j show that the proposed fusion method can preserve the detailed information of source images well, which is conducive to the observation of medical images and the diagnosis of diseases.

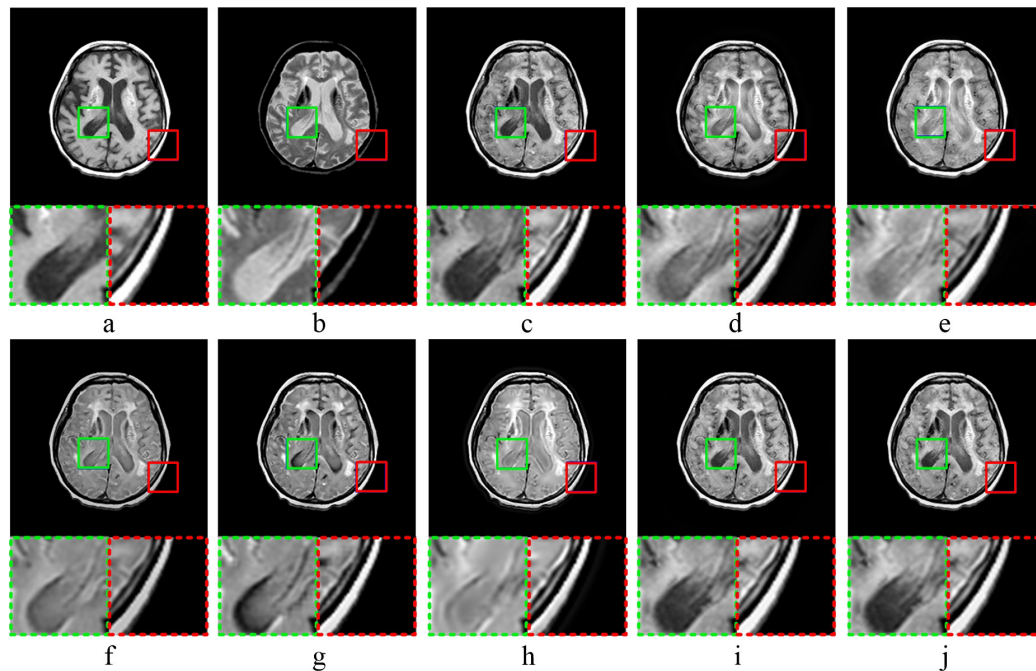


Figure 4. MR-T1 CMR-T2 image fusion experiments: (a,b) source images; and (c–j) the fused image obtained by MST-SR, NSCT-PC, NSST-PCNN, ASR, CT, KIM, CNN-LIU, and the proposed method, respectively. Two partially enlarged images marked in green and red dashed frames correspond to the regions surrounded by green and red frames in the fused image.

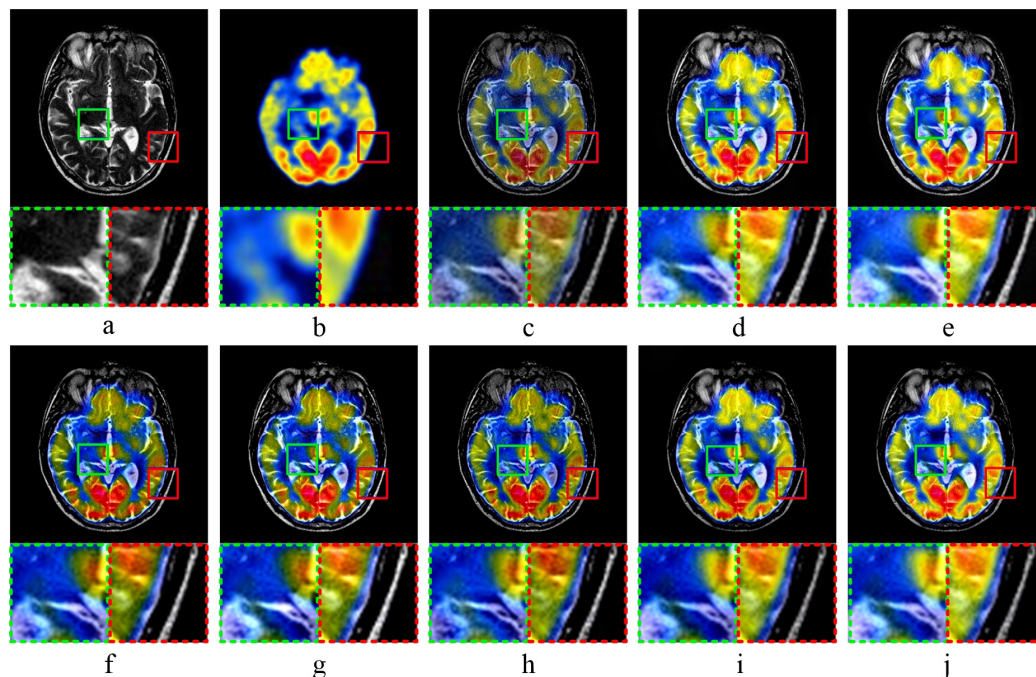


Figure 5. MR-PET image fusion Experiment 1: (a,b) source images; and (c–j) the fused image obtained by MST-SR, NSCT-PC, NSST-PCNN, ASR, CT, KIM, CNN-LIU, and the proposed method, respectively. Two partially enlarged images marked in green and red dashed frames correspond to the regions surrounded by green and red frames in the fused image.

Figures 7 and 8 show the results of MR-SPECT image fusion experiments. In Figure 7c, the fused image of MST-SR method has a low contrast and unclear edge details. As shown in Figure 8d,e, some edge regions are too smooth in the fused images obtained by NSCT-PC and NSST-PCNN methods, and the edge details are not clear. In Figures 7g and 8g, the images obtained by CT method have the high

contrast, and CT method performs poorly on the detail retention of source images. The fused images shown in Figures 7f,h and 8f,h, which were obtained by ASR and KIM method, respectively, have the low brightness and poor visualization performance. As shown in Figures 7i,j and 8i,j, the fused images obtained by both CCN-LIU and the proposed method have the high brightness and good visualization performance. Comparing all the fused results in Figures 7 and 8, the fused images obtained by the proposed fusion method have the high similarity with source images, which can preserve the detailed structures of source images well and achieve good fusion performance.

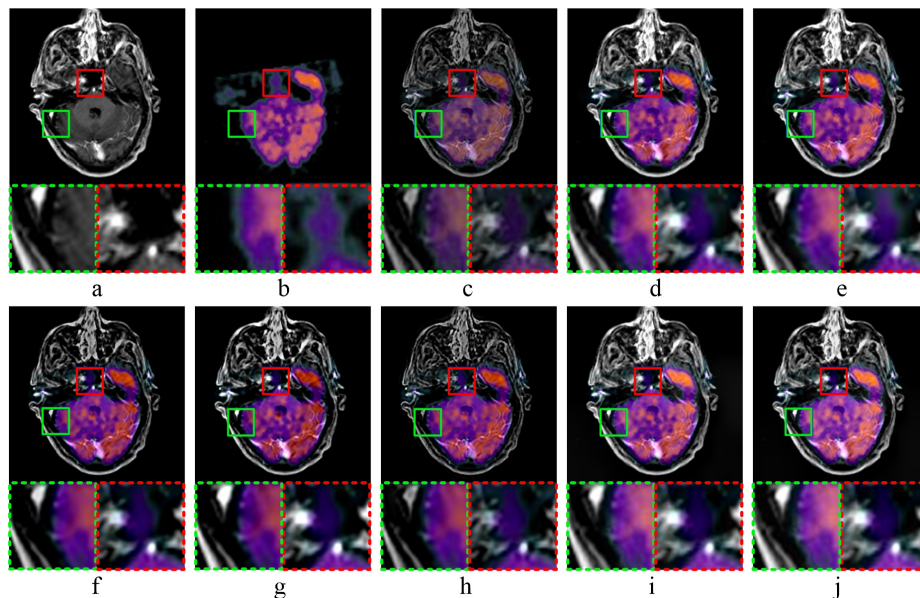


Figure 6. MR-PET image fusion Experiment 2: (a,b) source images; and (c–j) the fused image obtained by MST-SR, NSCT-PC, NSST-PCNN, ASR, CT, KIM, CNN-LIU, and the proposed method, respectively. Two partially enlarged images marked in green and red dashed frames correspond to the regions surrounded by green and red frames in the fused image.

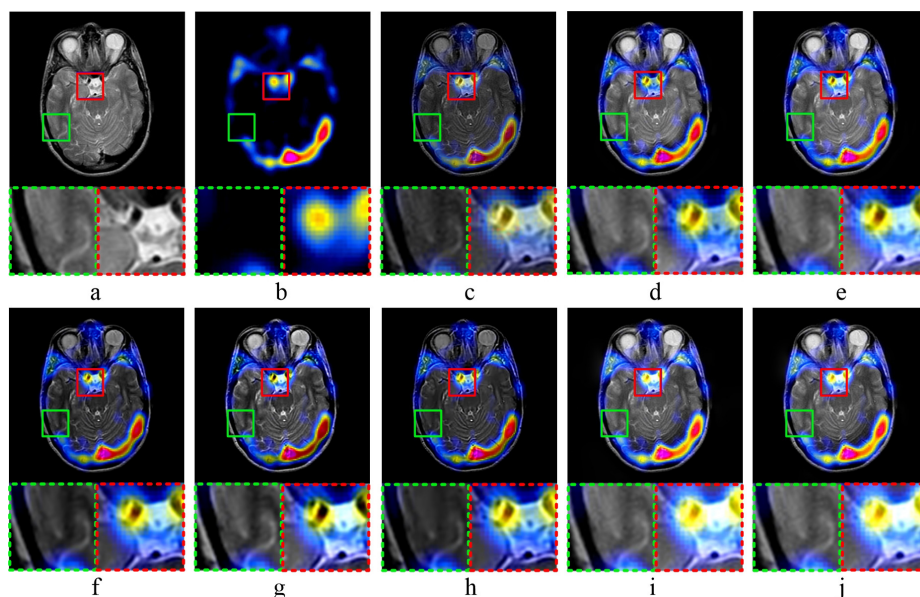


Figure 7. MR-SPECT image fusion Experiment 1: (a,b) source images; and (c–j) the fused image obtained by MST-SR, NSCT-PC, NSST-PCNN, ASR, CT, KIM, CNN-LIU, and the proposed method, respectively. Two partially enlarged images marked in green and red dashed frames correspond to the regions surrounded by green and red frames in the fused image.

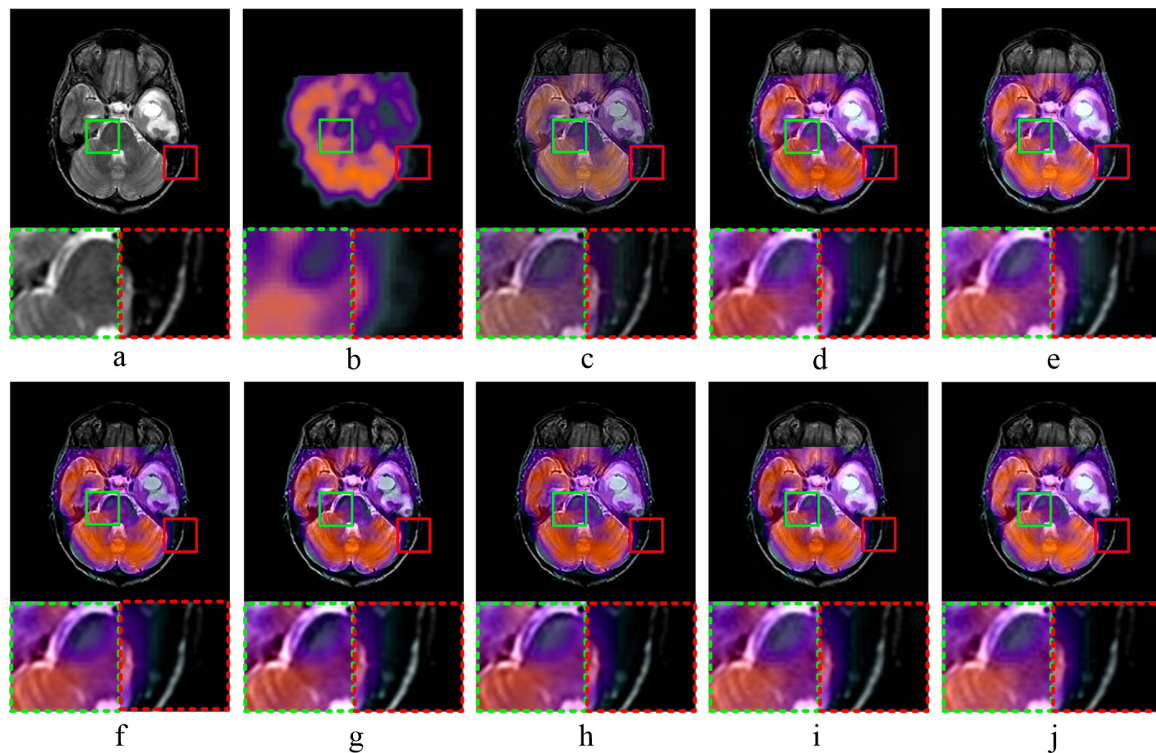


Figure 8. MR-SPECT image fusion Experiments 2: (a,b) source images; and (c–j) the fused image obtained by MST-SR, NSCT-PC, NSST-PCNN, ASR, CT, KIM, CNN-LIU, and the proposed method, respectively. Two partially enlarged images marked in green and red dashed frames correspond to the regions surrounded by green and red frames in the fused image.

4.2. Evaluation of Objective Metrics

For image fusion, a single evaluation metric lacks objectivity. Therefore, it is necessary to do a comprehensive analysis by using multiple evaluation metrics. In this study, four objective evaluation metrics, namely Q^{TE} [46,47], $Q^{AB/F}$ [29,48], Q^{MI} [47], and Q^{VIF} [29,49], were used to evaluate the performances of different fusion methods. Q^{TE} is the Tsallis entropy of the fused image. The entropy value represents the amount of average information contained in the fused image. $Q^{AB/F}$ as a gradient-based quality indicator is mainly used to measure the edge information of fused images. Q^{MI} is the mutual information indicator, which is used to measure the amount of information contained in the fused image. Q^{VIF} is the information ratio between the fused image and source images to evaluate the human visualization performance of the fused image. The objective evaluation results of medical image fusion are shown in Figure 9. Among all the fusion results, the proposed method achieves good performance in all four objective evaluations. It confirms that the proposed method can preserve the detailed structure information of source images well and realize good human visual effects.

Table 1 shows the values of four objective metrics for eight fusion methods. The proposed method achieves the highest Q^{TE} value. Comparing with the seven other fusion methods, the fused image obtained by the proposed method has the highest Tsallis entropy, and contains more information than the others. According to the analysis of $Q^{AB/F}$, the fused images obtained by NSCT-PC, NSST-PCNN, CNN-LIU, and the proposed method have high $Q^{AB/F}$, which means these fused images perform well in the preservation of edge details. The fused image obtained by KIM has low $Q^{AB/F}$, which indicates that KIM does not have good performance in the preservation of edge information. For Q^{MI} , the proposed method is a little bit higher than the others. It means more information of source images is retained in the fused image, and the preservation ability of source image details is strong. The proposed method has the highest Q^{VIF} . Comparing with CNN-LIU, the proposed method has a

higher information ratio between the fused image and source images, and achieves a better human visual effect as well.

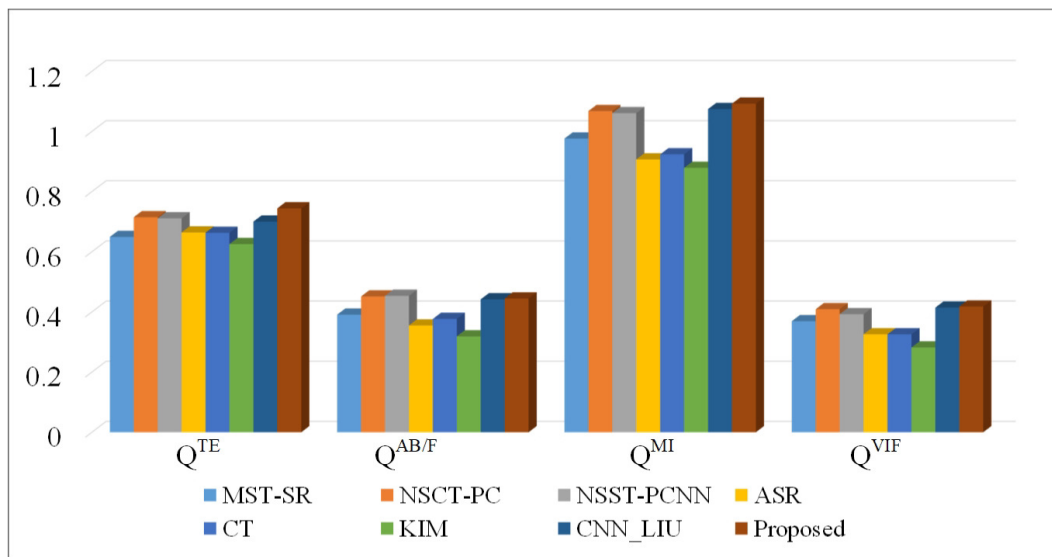


Figure 9. Objective evaluation results of eight fusion methods.

Table 1. Objective evaluations of medical image fusion comparative experiments.

	Q^{TE}	$Q^{AB/F}$	Q^{MI}	Q^{VIF}	Average Processing Time
MST-SR	0.6495	0.3911	0.9764	0.3693	15.0541
NSCT-PC	0.7150	0.4515	1.0681	0.4092	3.7743
NSST-PCNN	0.7113	0.4537	1.0610	0.3924	6.1595
ASR	0.6643	0.3550	0.9072	0.3258	35.2493
CT	0.6631	0.3769	0.9240	0.3258	14.5846
KIM	0.6257	0.3188	0.8792	0.2820	59.1929
CNN-LIU	0.7003	0.4421	1.0745	0.4145	14.5846
Proposed	0.7445	0.4449	1.0925	0.4181	12.8667

4.3. Threshold Discussion

In this study, a similarity threshold T was defined to fuse the multi-scale sub-decomposed images. For the top layer of sub-decomposed images, the threshold was set to 0.6. For the remaining layers of sub-decomposed images, the threshold was set to 3. Table 2 shows the values of five objective metrics for the fusion framework with different thresholds. According to Q^{MI} , the proposed method is a little bit lower than others. However, for Q^{TE} , $Q^{AB/F}$, and Q^{VIF} , the proposed method is higher than the others. It means more average information and edge information is contained in the fused image, and it has a higher information ratio between the fused image and source images. In addition, these three methods have close values in terms of time consumption. Overall, the proposed method performs better on five objective metrics.

Table 2. Objective evaluations of the fusion framework with different thresholds.

	Q^{TE}	$Q^{AB/F}$	Q^{MI}	Q^{VIF}	Average Processing Time
Threshold = 0.6	0.6973	0.4248	1.1165	0.4151	12.7256
Threshold = 3	0.7289	0.4255	1.2540	0.4068	12.6652
Proposed	0.7445	0.4449	1.0925	0.4181	12.8667

5. Conclusions

This paper proposes a CNN-based medical image fusion solution. The proposed method implements the measurement of activity level and weight distribution by CNN training to generate a weight map including the integrated pixel activity information. To obtain better visual effects, the multi-scale decomposition method based on contrast pyramid is used to fuse corresponding image components in different spatial frequency bands. Meanwhile, the complementary and redundant information of fused images is explored by the local similarity strategy in adaptive fusion mode. Comparative experiment results show that the fused images by proposed method have high visual quality and objective indicators. In the future, we will continue to explore the great potential of deep learning techniques and apply them to other types of multi-modality image fusion, such as infrared-visible and multi-focus image fusion.

Author Contributions: K.W. designed the proposed algorithm and wrote the paper; M.Z. and H.W. participated in the algorithm design, algorithm programming, and testing the proposed method; G.Q. participated in the algorithm design and paper writing processes; and Y.L. participated in the algorithm design, provided technical support, and revised the paper. All authors have read and approved the final manuscript.

Funding: This work was jointly supported by the National Natural Science Foundation of China under Grants 61803061 and 61906026; the National Nuclear Energy Development Project of China (Grant No. 18zg6103); Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN201800603); Chongqing Natural Science Foundation Grant cstc2018jcyjAX0167; the Common Key Technology Innovation Special of Key Industries of Chongqing science and Technology Commission under Grant Nos. cstc2017zdcy-zdyfX0067, cstc2017zdcy-zdyfX0055, and cstc2018jszx-cyzd0634; and the Artificial Intelligence Technology Innovation Significant Theme Special Project of Chongqing science and Technology Commission under Grant Nos. cstc2017rgzn-zdyfX0014 and cstc2017rgzn-zdyfX0035.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ganasala, P.; Kumar, V. Feature-Motivated Simplified Adaptive PCNN-Based Medical Image Fusion Algorithm in NSST Domain. *J. Digit. Imaging* **2016**, *29*, 73–85. [[CrossRef](#)] [[PubMed](#)]
2. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [[CrossRef](#)]
3. James, A.P.; Dasarathy, B.V. Medical image fusion: A survey of the state of the art. *Inf. Fusion* **2014**, *19*, 4–19. [[CrossRef](#)]
4. Zhu, Z.; Chai, Y.; Yin, H.; Li, Y.; Liu, Z. A novel dictionary learning approach for multi-modality medical image fusion. *Neurocomputing* **2016**, *214*, 471–482. [[CrossRef](#)]
5. Li, H.; He, X.; Tao, D.; Tang, Y.; Wang, R. Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recognit.* **2018**, *79*, 130–146. [[CrossRef](#)]
6. Li, Y.; Sun, Y.; Zheng, M.; Huang, X.; Qi, G.; Hu, H.; Zhu, Z. A novel multi-exposure image fusion method based on adaptive patch structure. *Entropy* **2018**, *20*, 935. [[CrossRef](#)]
7. Qi, G.; Wang, J.; Zhang, Q.; Zeng, F.; Zhu, Z. An Integrated Dictionary-Learning Entropy-Based Medical Image Fusion Framework. *Future Internet* **2017**, *9*, 61. [[CrossRef](#)]
8. Shen, J.; Zhao, Y.; Yan, S.; Li, X. Exposure Fusion Using Boosting Laplacian Pyramid. *IEEE Trans. Cybern.* **2014**, *44*, 1579–1590. [[CrossRef](#)]
9. Zhu, Z.; Zheng, M.; Qi, G.; Wang, D.; Xiang, Y. A Phase Congruency and Local Laplacian Energy Based Multi-Modality Medical Image Fusion Method in NSCT Domain. *IEEE Access* **2019**, *7*, 20811–20824. [[CrossRef](#)]
10. Li, Y.; Sun, Y.; Huang, X.; Qi, G.; Zheng, M.; Zhu, Z. An Image Fusion Method Based on Sparse Representation and Sum Modified-Laplacian in NSCT Domain. *Entropy* **2018**, *20*, 522. [[CrossRef](#)]
11. Xu, L.; Gao, G.; Feng, D. Multi-focus image fusion based on non-subsampled shearlet transform. *IET Image Process.* **2013**, *7*, 633–639. [[CrossRef](#)]
12. Qu, X.B.; Yan, J.W.; Xiao, H.Z.; Zhu, Z.Q. Image Fusion Algorithm Based on Spatial Frequency-Motivated Pulse Coupled Neural Networks in Nonsubsampled Contourlet Transform Domain. *Acta Autom. Sin.* **2008**, *34*, 1508–1514. [[CrossRef](#)]

13. Bhatnagar, G.; Wu, J.; Liu, Z. Directive Contrast Based Multimodal Medical Image Fusion in NSCT Domain. *IEEE Trans. Multimed.* **2013**, *15*, 1014–1024. [[CrossRef](#)]
14. Das, S.; Kundu, M.K. A Neuro-Fuzzy Approach for Medical Image Fusion. *IEEE. Trans. Biomed. Eng.* **2013**, *60*, 3347–3353. [[CrossRef](#)]
15. Liu, Z.; Yin, H.; Chai, Y.; Yang, S.X. A novel approach for multimodal medical image fusion. *Expert Syst. Appl.* **2014**, *41*, 7425–7435. [[CrossRef](#)]
16. Wang, L.; Li, B.; Tian, L. Multimodal Medical Volumetric Data Fusion Using 3-D Discrete Shearlet Transform and Global-to-Local Rule. *IEEE. Trans. Biomed. Eng.* **2014**, *61*, 197–206. [[CrossRef](#)]
17. Yang, Y.; Que, Y.; Huang, S.; Lin, P. Multimodal Sensor Medical Image Fusion Based on Type-2 Fuzzy Logic in NSCT Domain. *IEEE Sens. J.* **2016**, *16*, 3735–3745. [[CrossRef](#)]
18. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [[CrossRef](#)]
19. Yin, M.; Liu, X.; Liu, Y.; Chen, X. Medical Image Fusion with Parameter-Adaptive Pulse Coupled Neural Network in Nonsampled Shearlet Transform Domain. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 49–64. [[CrossRef](#)]
20. Yin, L.; Zheng, M.; Qi, G.; Zhu, Z.; Jin, F.; Sim, J. A Novel Image Fusion Framework Based on Sparse Representation and Pulse Coupled Neural Network. *IEEE Access* **2019**, *7*, 98290–98305. [[CrossRef](#)]
21. Zhang, Q.; Liu, Y.; Blum, R.S.; Han, J.; Tao, D. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Inf. Fusion* **2018**, *40*, 57–75. [[CrossRef](#)]
22. Qi, G.; Zhang, Q.; Zeng, F.; Wang, J.; Zhu, Z. Multi-focus image fusion via morphological similarity-based dictionary construction and sparse representation. *CAAI TIT.* **2018**, *3*, 83–94. [[CrossRef](#)]
23. Wang, K.; Qi, G.; Zhu, Z.; Cai, Y. A Novel Geometric Dictionary Construction Approach for Sparse Representation Based Image Fusion. *Entropy* **2017**, *19*, 306. [[CrossRef](#)]
24. Yang, B.; Li, S. Multifocus Image Fusion and Restoration with Sparse Representation. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 884–892. [[CrossRef](#)]
25. Yang, B.; Li, S. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Inf. Fusion* **2012**, *13*, 10–19. [[CrossRef](#)]
26. Kim, M.; Han, D.K.; Ko, H. Joint patch clustering-based dictionary learning for multimodal image fusion. *Inf. Fusion* **2016**, *27*, 198–214. [[CrossRef](#)]
27. Li, S.; Yin, H.; Fang, L. Group-Sparse Representation with Dictionary Learning for Medical Image Denoising and Fusion. *IEEE. Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [[CrossRef](#)]
28. Liu, Y.; Wang, Z. Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Process.* **2015**, *9*, 347–357. [[CrossRef](#)]
29. Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; Qi, G. A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inf. Sci.* **2018**, *432*, 516–529. [[CrossRef](#)]
30. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)]
31. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
32. Zhu, Z.; Qi, G.; Li, Y.; Wei, H.; Liu, Y. Image Dehazing by An Artificial Image Fusion Method based on Adaptive Structure Decomposition. *IEEE Sens. J.* **2020**, *42*, 1–11. [[CrossRef](#)]
33. Qi, G.; Chang, L.; Luo, Y.; Chen, Y.; Zhu, Z.; Wang, S. A Precise Multi-Exposure Image Fusion Method Based on Low-level Features. *Sensors* **2020**, *20*, 1597. [[CrossRef](#)]
34. Qi, G.; Zhu, Z.; Erqinhu, K.; Chen, Y.; Chai, Y.; Sun, J. Fault-diagnosis for reciprocating compressors using big data and machine learning. *Simul. Model. Pract. Theory* **2018**, *80*, 104–127. [[CrossRef](#)]
35. Li, D.; Dong, Y. Deep Learning: Methods and Applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [[CrossRef](#)]
36. Qi, G.; Wang, H.; Haner, M.; Weng, C.; Chen, S.; Zhu, Z. Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation. *CAAI TIT.* **2019**, *4*, 80–91. [[CrossRef](#)]
37. Xia, K.J.; Yin, H.S.; Wang, J.Q. A novel improved deep convolutional neural network model for medical image fusion. *Cluster Comput.* **2018**, *22*, 1515–1527. [[CrossRef](#)]

38. Liu, Y.; Chen, X.; Cheng, J.; Peng, H. A medical image fusion method based on convolutional neural networks. In Proceedings of the 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017. [[CrossRef](#)]
39. Li, H.; Li, X.; Yu, Z.; Mao, C. Multifocus image fusion by combining with mixed-order structure tensors and multiscale neighborhood. *Inf. Sci.* **2016**, *349–350*, 25–49. [[CrossRef](#)]
40. Shen, R.; Cheng, I.; Basu, A. Cross-Scale Coefficient Selection for Volumetric Medical Image Fusion. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1069–1079. [[CrossRef](#)]
41. Singh, R.; Khare, A. Fusion of multimodal medical images using Daubechies complex wavelet transform-A multiresolution approach. *Inf. Fusion* **2014**, *19*, 49–60. [[CrossRef](#)]
42. Zhu, Z.; Qi, G.; Chai, Y.; Li, P. A Geometric Dictionary Learning Based Approach for Fluorescence Spectroscopy Image Fusion. *Appl. Sci.* **2017**, *7*, 161. [[CrossRef](#)]
43. Bhatnagar, G.; Wu, Q.M.J.; Liu, Z. A new contrast based multimodal medical image fusion framework. *Neurocomputing* **2015**, *157*, 143–152. [[CrossRef](#)]
44. Li, H.; Yu, Z.; Mao, C. Fractional differential and variational method for image fusion and super-resolution. *Neurocomputing* **2016**, *171*, 138–148. [[CrossRef](#)]
45. Li, H.; Liu, X.; Yu, Z.; Zhang, Y. Performance improvement scheme of multifocus image fusion derived by difference images. *Signal Process.* **2016**, *128*, 474–493. [[CrossRef](#)]
46. Cvejic, N.; Canagarajah, C.; Bull, D. Image fusion metric based on mutual information and Tsallis entropy. *Electron. Lett.* **2006**, *42*, 626–627. [[CrossRef](#)]
47. Liu, Z.; Blasch, E.; Xue, Z.; Zhao, J.; Laganiere, R.; Wu, W. Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 94–109. [[CrossRef](#)]
48. Petrović, V. Subjective tests for image fusion evaluation and objective metric validation. *Inf. Fusion* **2007**, *8*, 208–216. [[CrossRef](#)]
49. Sheikh, H.; Bovik, A. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).