A Multi-Focus Image Fusion Method Based on Brushlet and CNN

Haonan Yu[1]

Wollongong Joint Institute, Central China Normal University, Wuhan, Hubei, China

yh20212022@163.com

Abstract. The fusion of images in the transform domain using convolutional neural networks method can improve the fusion effect, but if the training sample set input to the CNN model is not selected properly, the fused image will show "pseudo-edge", "artificial texture" and other phenomena. In this paper, we propose a CNN image fusion algorithm based on Brushlet energy, which performs non-down sampling contour wave transform on the original image to obtain high and low frequency coefficient maps, uses Brushlet to bilayer decompose the coefficient maps to obtain complex coefficients, obtains the coefficient map chunk energy values by real and imaginary energy solving method, and uses them as the input sample set of CNN model for processing, the CNN model The output is the final decision map for fusion, which can be applied to each high and low frequency coefficient map of NSCT to achieve more accurate image fusion. The experimental results show that the method proposed in this paper has some improvement over other algorithms in both subjective human eye perception effect and quantitative objective evaluation index.

Keywords: image fusion; brushlet complex energy; convolutional neural network

1 Introduction

Due to the limited focusing ability of imaging devices such as digital cameras, the same scene is often photographed with only one focus, i.e., at the focal point, the image quality is good, while in other unfocused areas there is a blurring phenomenon, so to complete a clear image, it is necessary to focus different areas of the same scene several times to process them separately and use image fusion technology [1] to fuse each of the acquired, with images with different focusing areas are fused and processed to obtain a clear image, which can be provided to the human eye for better perception and understanding, and to computers for various analysis and processing. Currently, there are numerous applications of digital image fusion techniques in the fields of medical images [2-3], remote sensing images [4-6], and infrared and visible images [7-9].

[1] Haonan Yu, first year graduate student at CCNU, currently focus on image fusion, deep reinforcement learning.

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 514–525, 2022.
https://doi.org/10.2991/978-94-6463-108-1_59

Image fusion techniques can be broadly classified into three levels, pixel-level, feature-level, and decision-level, according to the different fusion information selection, among which pixel-level based fusion techniques are more widely used because they take into account the integrity of information and are especially well characterized for image texture and detail parts [10-11]. Currently, the pixel-level fusion techniques can be subdivided into two schemes: fusion in the null domain and fusion in the transform domain. Among them, fusion in the null domain is mostly achieved by choosing fusion algorithms such as calculating the energy of the region for weighting or parametric minimization; while the key to achieving fusion in the transform domain is the image multiscale geometric transformation method.

For example, paper [12] proposed to achieve image fusion in the contourlet domain, using the Contourlet Transform (CT) that can separate the image into multiple high and low frequency components, which can better distinguish the details and energy concentrated parts of the image. But the fusion has certain "artifacts" phenomenon due to the lack of translation invariance of CT. The paper [13] proposed an image fusion algorithm based on Non-Subsampled Contourlet Transform (NSCT) and Pulse-Coupled Neural Network (PCNN), which can better achieve image fusion, but the complexity of the algorithm is high, and there is a certain misclassification region. The image fusion algorithm based on Shearlet transform (ST) and PCNN proposed in the literature [14] has the phenomenon of "artificial texture" when using PCNN due to the spectral overlap in the ST segmentation process. In the Brushlet domain [15], the fusion effect is effectively improved by calculating the complex energy, taking into account the texture continuity of the image, but some image data are discarded in the low-frequency region, so the fusion rules for the low-frequency part can be improved.

In the past few years, deep learning has penetrated various fields of image processing [16-18], all of which have achieved good research results, and within the field of image fusion, CNN-based image fusion algorithms are widely mentioned. Among them, the paper [19] used CNN to directly perform feature extraction and discriminative classification of null domain pixels, ignoring the information of high and low frequencies, so the fusion effect was not satisfactory, and an improved algorithm was proposed in the paper [20], which used a CNN model to simultaneously generate horizontal activity measurements and corresponding fusion rules, by the laws of human eye vision, which effectively improved the fusion performance in both subjective and objective evaluation. However, the high complexity of CNN and insufficient samples are also one of the factors that affect the fusion effect.

To address the above shortcomings, a CNN image fusion algorithm based on Brushlet energy is proposed in this paper. After the image is decomposed by NSCT, the image is further decomposed on each high and low frequency coefficient map using Brushlet to get the corresponding complex coefficients and decomposed into 16*16 sub-blocks, and the energy is solved for the sub-blocks to get each energy block, which is input to the CNN model for training, and the output decision map is obtained, based on which the fusion of source images can be completed in each high and low frequency region, and then the final fused image is obtained by inversion of NSCT. The final fused image is then obtained by inversion of NSCT. The advantages of this method are mainly reflected in the following three aspects.

(1) The high and low frequency coefficient maps of the source image obtained by NSCT decomposition are free of spectral overlap, which can effectively characterize the detailed information in the source image and ensure the accuracy of sub-block energy calculation.

(2) The high-frequency directional coefficient map and low-frequency coefficient map of the NSCT decomposition using Brushlet can achieve a good characterization of the energy concentration degree.

(3) The energy block sample set has good focusing ability after being trained by CNN model, so the obtained decision map can achieve accurate segmentation of the fusion boundary.

2 Algorithm Architecture

2.1 A. Basic principles of NSCT

NSCT is a modified form of contour wave transform. Compared with CT, NSCT adopts Nonsubsampled Laplacian Pyramid (NLP) structure and Nonsubsampled Directional Filter Banks (NDFB). Therefore, the high and low frequency coefficient maps obtained by decomposing the image using NSCT do not have spectral overlap and have better frequency selectivity and regularity than the contour transform. Figure 1 shows the decomposition framework of NSCT.

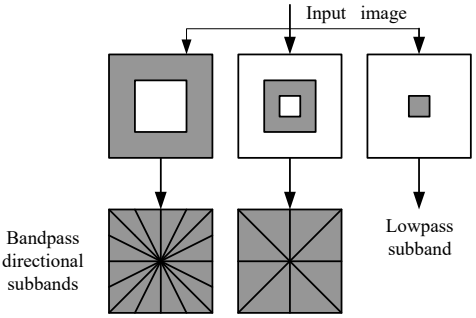


Fig. 1. Framework diagram of NSCT scale and orientation decomposition

Brushlet has a multi-layer decomposition result that enables an excellent decomposition of the frequency domain. the main difference between Brushlet basis and wavelet packet is the arbitrary tiling of the time-frequency plane and the perfect localization of single frequencies within one coefficient. In addition, the coefficients obtained after the image is decomposed by Brushlet are complex-valued functions, and this complex-valued information facilitates the energy characterization. To further elaborate the direction selectivity of Brushlet, this paper takes a no-copyright image as an example, as shown in Figure 2(a). The one-layer decomposition of Brushlet is to divide the frequency domain into four quadrants, and the corresponding direction can be characterized as $\frac{\pi}{4} + k\frac{\pi}{2}, k = 0,1,2,3$ (Figure 2(b)), and the two-layer decomposition of Brushlet is to further decompose the coefficient map corresponding to the four quadrants into 16 coefficients

based on the one-layer decomposition (Figure 2(c)), as seen from the figure, the multi-layer decomposition is the process of completing the refinement of the directional coefficient map of the previous layer, and it is generally appropriate to choose two-layer decomposition.

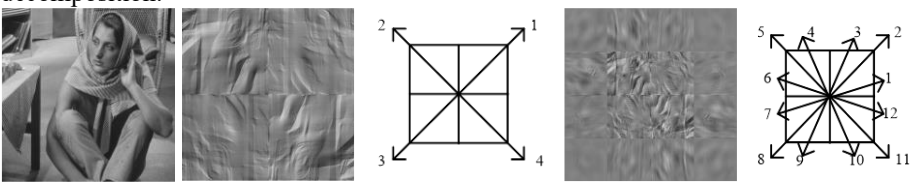


Fig. 2. Brushlet second scale decomposition diagram and related directions example

Let *f* represent the coefficient of the Brushlet decomposition, *freal* represent the real part and *fimg* represent the imaginary part of the complex number, then the energy identity is expressed by E_f as equation (1) follows:

$$E_f = \sum_{n=1}^{N}\sum_{m=1}^{M}|f(m,n)| = \sum_{n=1}^{N}\sum_{m=1}^{M}\left[\left(freal(m,n)\right)^2 + \left(fimg(m,n)\right)^2\right]^{\frac{1}{2}} \quad (1)$$

where M and N represent the size of *f*.

3 CNN fusion rules based on NSCT domain

In this paper, two images are selected for fusion, and in practice the fusion mechanism for multiple focused images can be arranged two by two and fused sequentially. The basic framework of fusion is shown in Figure 3 below. The algorithm in this paper has four main parts: NSCT decomposition, Brushlet decomposition, and energy calculation, CNN training to obtain the decision map, and final fusion.

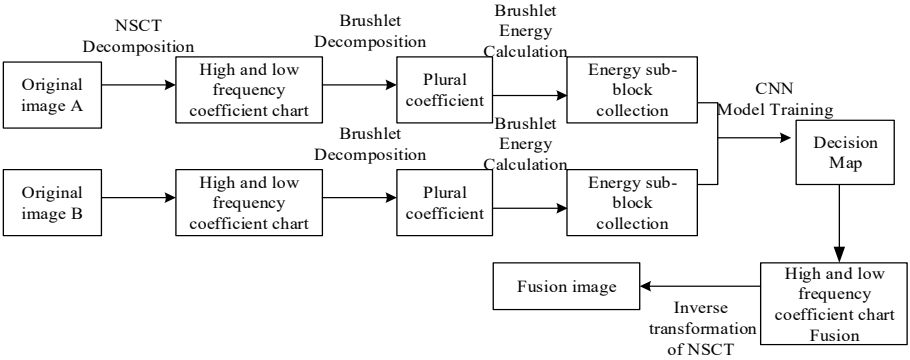


Fig. 3. Fusion framework diagram

After the NSCT decomposition of the two source images, a low-frequency coefficient map and a series of high-frequency directional coefficient maps are obtained. To extract the focus region of each coefficient map, Brushlet is selected to further carry out two-

layer decomposition on the coefficient map, and the complex coefficient map is decomposed into 16*16 subblocks, and the energy of these subblocks is calculated by using the complex coefficients to determine the set of energy block samples of each coefficient map, which is input into the CNN model for training to obtain the decision map corresponding to all coefficient maps, and then used to achieve fusion in the NSCT domain and obtain the fused image by NSCT inverse transform.

3.1 A. NSCT domain sample set construction

Let the source image IM_1, IM_2 be implemented by NSCT with multi-scale multi-directional filtering, which can produce low-frequency coefficient maps $C1_k^j\ C2_k^j$, where $j = 1, 2 \cdots J$ characterizes the j-th scale decomposition layer and $k = 1, 2 \cdots K$ represents the kth directional coefficient map corresponding to that scale decomposition layer. A two-level decomposition is performed using $C1_k^j, C2_k^j$ to obtain the coefficient matrices of the real and imaginary parts, characterized as $C1real_k^j$, $C1img_k^j$, $C2real_k^j$, $C2img_k^j$, respectively.

Taking $C1real_k^j$ as an example, the 16*16 block region is divided as $\left\{yBreal_d^k(1), yBreal_d^k(2), \cdots, yBreal_d^k(P)\right\}$ and a total of P sub-blocks are assumed to be obtained, and the same principle of sub-block division is used for $C1img_k^j$, $C2real_k^j$, $C2img_k^j$. The algorithm for solving the energy of the real and imaginary parts solves the energy of each block within $C1_k^j, C2_k^j$, respectively, as shown in equation (2)

$$E_k^j(p) = \sum_{m \in M, n \in N}\left(C1real_k^j(p)(m,n)^2 + C1img_k^j(p)(m,n)^2\right) \tag{2}$$

where M, N is the size of the subregion.

For all high frequency coefficient maps, the energy values of sub-regions in the same direction and at the same position are combined. The calculation is shown in the following equation (3):

$$Esum_k(p) = \sum_{j \in J} E_k^j(p) \tag{3}$$

Let $Esum^{(1)}{}_k(p)$ and $Esum^{(2)}{}_k(p)$ come from IM_1 and IM_2, respectively. Comparing the magnitude of these two energy values can reflect the focused region of the source image, i.e., a larger energy value corresponds to the focused region (clear region) and a smaller energy value for the unfocused region (blurred region). This is fed into the CNN model as a training sample set for optimization.

3.2 CNN network structure

In this paper, the CNN network structure is divided into 5 layers as shown in Figure 4.

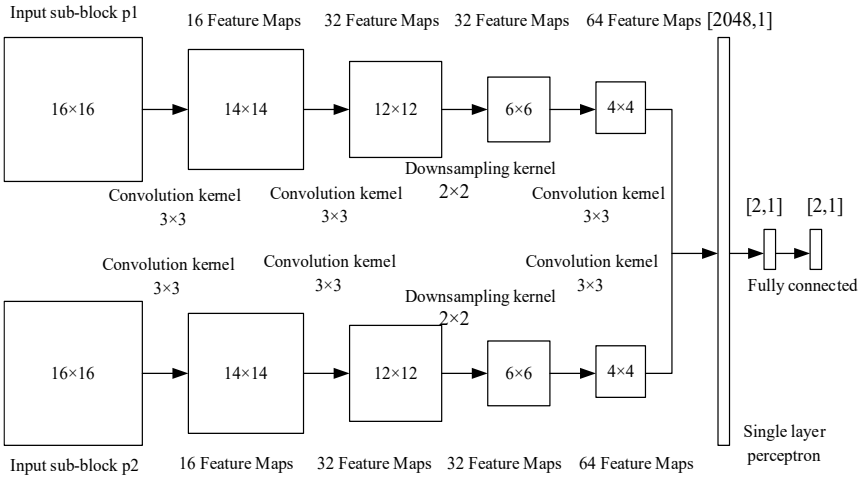


Fig. 4. Schematic diagram of the CNN network structure used by the algorithm in this paper

The network consists of an input layer, three convolutional layers, and a maximum pooling layer. The sample set first enters the input layer. The second layer is the convolutional layer, in which the sample set needs to be convolved with a convolutional kernel, and here a convolutional kernel of 3*3 and amplitude 1 is chosen to process the sample set, and 16 feature maps are obtained, corresponding to a size of 14*14. The third layer is the convolutional layer, and the choice of convolutional kernel is exactly the same as the second layer, but the convolution can obtain 32 feature maps of size 12*12. The fourth layer is the maximum pooling layer, and the kernel is chosen to perform the maximum pooling operation with the 32 feature maps obtained in the previous layer, where the size of the sampling kernel is chosen to be 2*2, and the magnitude is twice that of the convolution kernel, and the output of the pooling operation is still 32 feature maps, but its corresponding size becomes 6*6. The last layer is the convolution layer, and the choice of the convolution kernel is the same as the second and third layers The pooling layer is then used to convolve the 32 feature maps into 64 feature maps of size 4*4. The 64 feature maps of each branch are then connected to form a feature vector F1 of length 2048. The fully connected operation is then used to obtain F2 of length 2. The Softmax function is called again to calculate and finally obtain a feature vector F3 of size [2, 1], F3 Each element value of the vector corresponds to the probability of each class.

3.3 CNN training model

As the sample data is input to the CNN training model, it is set to the activation regions RA and RB, and the parameters are optimized by using two propagation stages (forward and inverse) of the convolutional neural network. forward propagation in the convolutional layer can be described by equation (4). Where the i-th input feature map is denoted by X_i, K_{ij} is the convolutional kernel corresponding to the i, j output feature maps, and b_j is the bias value for the j-th output feature map. n, f represent the number of input feature maps and Relu activation function.

$$X_j^l = f\left(\sum_{i=1}^{n} X_i^{l-1} \times K_{ij}^l + b_j^l\right) \tag{4}$$

The maximum pooling layer has the same number of input and output feature maps for forward propagation, but the size is reduced, and the output map corresponds to a multiplicative bias β and an additive bias b, as shown in equation (5), where down() represents the down sampling operator.

$$X_j^l = f\left(\beta_j^l down\left(X_i^{l-1}\right) + b_j^l\right) \tag{5}$$

The difference between the CNN prediction and the true value can be measured by introducing a squared error loss function, and the difference can be further reduced by iterative training, which is the purpose of training. There are N labeled samples $\{z^1, y^1\}$, $\{z^2, y^2\}$, … $\{z^n, y^n\}$ with the "one-of-c" labeling format. The expression of the squared error cost function E^n of z^n is shown in equation (6) for each independent sample z^n with a total of c classes. Where t_k^n and y_k^n characterize the predicted and true probability values of the nth sample belonging to the k-th class.

$$E^n = \frac{1}{2}\sum_{k=1}^{c}(t_k^n - y_k^n)^2 \tag{6}$$

The error of the full sample training set is the superposition of each sample error, as shown in (7)

$$E^N = \sum_{n=1}^{N} E^n = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}(t_k^n - y_k^n)^2 \tag{7}$$

All the parameters to be optimized form a vector group, denoted by W, and W* represents its optimal parameters, and if we want to obtain W*, we just need to minimize E^N. As shown in equation (8)

$$W^* = \arg\min_{W} E^N \tag{8}$$

Considering the existence of a large number of parameters in EN, the optimal solution will form an NP-hard problem, so the gradient descent method is chosen to perform two derivations in the specific solution to complete the parameter optimization, as shown in equation (9). The output is the optimization result.

$$W^{(k)} = W^{(k-1)} - \alpha \frac{\partial E^N}{\partial W}\Big|_{W-W^{(k)}} \tag{9}$$

According to the optimization results of CNN output for the fusion of NSCT high-frequency layers, the coefficient map of the same direction, the low-frequency coefficient map of NSCT is also fused in the same way, and the fusion effect is finally obtained.

4 Results and discussion

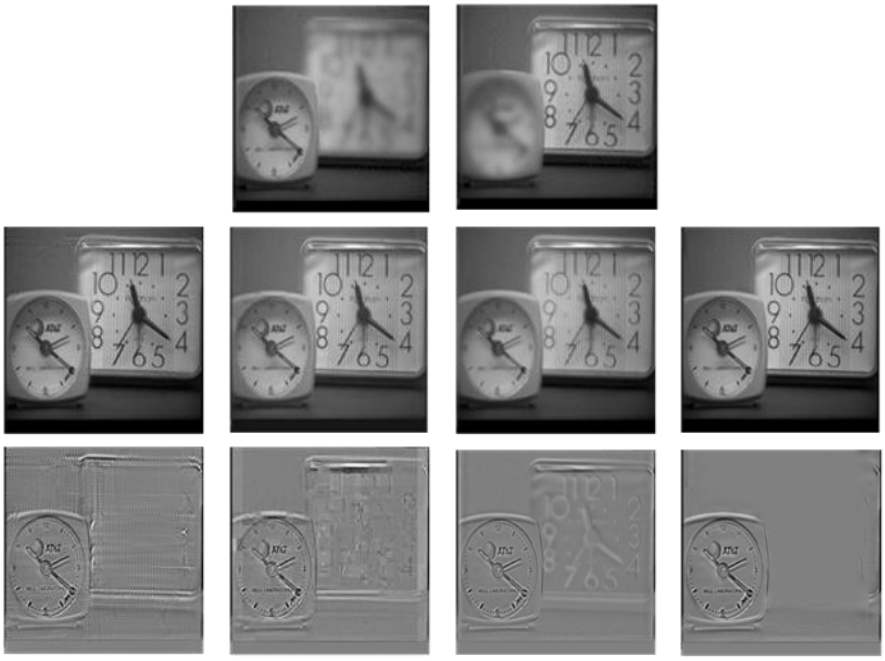
To verify the effectiveness of the algorithm in image fusion, the natural images Clock1/2, Bird1/2, Fighter1/2, and medical images CT&MIR were selected as the simulation test images. The simulation was done on a PC with Win10 flagship operating system, 16.00 GB of memory, 2.61 GHz CPU, and the fused images were aligned. The experimental software environment is MATLAB 2016a. The selection of suitable objective evaluation criteria can effectively verify the merits of the image fusion algorithm, and in this paper, two evaluation criteria, "mutual information MI" [21] and $Q^{AB/F}$ [22], are chosen to verify the effectiveness of the algorithm. MI is a qualitative measure of the amount of information contained in the original image. The $Q^{AB/F}$ is a measure of how much information is contained in the fused image from the significant edges of the image. To effectively test whether the algorithm in this paper can improve the fusion effect in image fusion, we have compared with the algorithms proposed in paper [13], paper [19], and paper [20] with respect to MI and $Q^{AB/F}$. Table 1 lists the objective criteria obtained by the different fusion algorithms.

Table 1. Values of the fusion evaluation indexes in the different methods

Images	Fusion methods				
	Criteria	Paper [13]	Paper [19]	Paper [20]	This
Clock1/2	MI	6.6298	7.1031	7.4120	8.4013
	$Q^{AB/F}$	0.6900	0.7204	0.7301	0.7502
Bird1/2	MI	5.6002	6.1925	6.3497	7.5107
	$Q^{AB/F}$	0.5597	0.6142	0.6407	0.7109
Fighter1/2	MI	6.9489	7.0078	7.4204	7.5621
	$Q^{AB/F}$	0.6699	0.6930	0.7746	0.8124
CT&MIR	MI	3.0109	4.0362	4.2007	4.3648
	$Q^{AB/F}$	0.6075	0.6285	0.6769	0.7375

As seen from Table 1, under the measurement of mutual information, this paper's algorithm in Clock1/2, Bird1/2, Fighter1/2, and CT&MIR are better than those of the other three methods. Among them, the MI value of Clock1/2 fusion reaches 8.4013, which indicates that the algorithm in this paper has a better fusion effect for the images with clearer fused edges. On top, the fusion index of the algorithm in this paper is above 0.7 for each image, among which the fusion index reaches 0.8124 for the Fighter1/2 image, which indicates that the fused image contains more information about the original image edges and has a better degree of detail retention. Therefore, the fusion method proposed in this paper can guarantee the amount of information while preserving the details and texture structure of the original image to a greater extent, which meets the requirement of high clarity of the fused image.

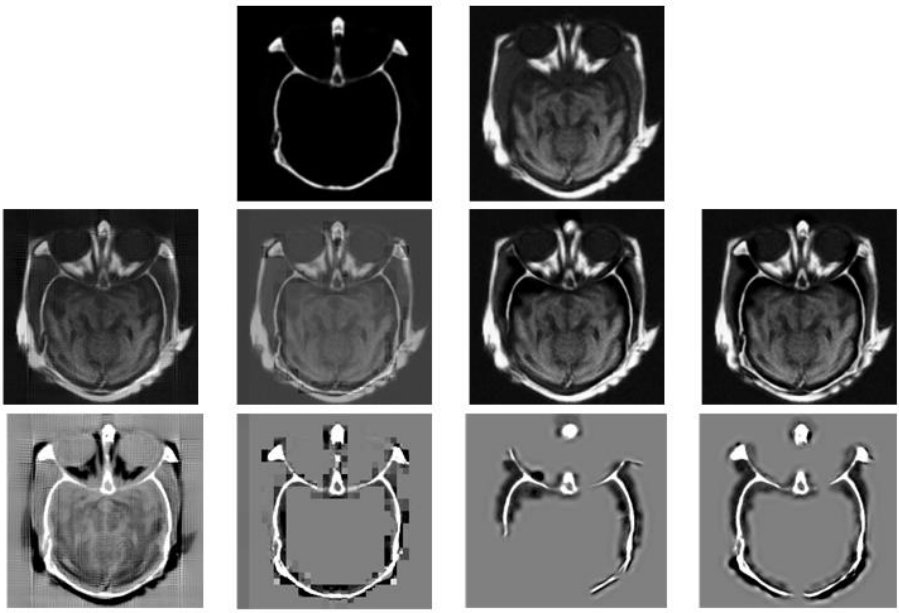
In this paper, three types of multifocal images (size and 256 gray levels) are selected for testing. One of the fusion results using the source image clock is shown in Figure 5.



Row 1 from left to right: Clock A, Clock B
Row 2 (Fusion image): Reference [13], [19], [20], and this paper
Row 3 (Difference image with the original image): Reference [13], [19], [20], and this paper

Fig. 5. Alarm clock image fusion effect diagram

This paper also applies the proposed method to medical image fusion. CT images show the structure of bone and MRI images show areas of soft tissue. In clinical applications, physicians need to see the location of bones and tissues to determine pathology and aid in diagnosis. Therefore, a hybrid image, which includes as much CT and MRI information as possible, is usually required in practice. The fusion results are shown in Figure 6.



Row 1 from left to right: CT, MRI
Row 2 (Fusion image): Reference [13], [19], [20], and this paper
Row 3 (Difference image with the original image): Reference [13], [19], [20], and this paper

Fig. 6. Medical image fusion rendering diagram

In Figure 5, the visual inspection of the fused images from the fusion method of the paper [13] is not satisfactory. The reason for this is that the differences between the multifocal images are small and there are often transition regions due to pixel blurring. Using only regional features as a fusion strategy, the selection of coefficients is not accurate. The paper [19] uses CNN for feature extraction and discriminative classification of null domain pixels directly, ignoring the information of high and low frequencies, so the fusion effect is not satisfactory. The paper [20] uses CNN model to simultaneously generate horizontal activity measurements and corresponding fusion rules, which is consistent with the laws of human eye vision and has fusion performance in subjective and objective evaluation, but CNN compound samples are insufficient to exist pseudo-edge phenomenon. The algorithm proposed in this paper can accurately segment the boundary and outperforms other fusion methods in terms of visual effect.

From Figure 6, the fused image of paper [13] is relatively blurred and artificial texture exists; the fused image of paper [19] produces a pseudo-edge blending phenomenon. And the fusion image constructed by the method of paper [20] is not accurate enough due to the loss of some details of the CT image. The algorithm in this paper can fully fuse the effective information of two medical images, so the fusion effect is better than other methods.

5 Conclusion

In this paper, a new scheme of joint Brushlet energy and CNN model training is proposed. The novelty of this method is that the complex coefficients obtained by Brushlet decomposition can accurately characterize the energy of coefficient sub-blocks, which are used as the input sample set of CNN model, and after training by CNN model, they have good classification ability and can obtain accurate decision maps for fusing high and low frequency coefficient maps. The experimental results show that the method in this paper has certain advantages over other algorithms, both in terms of subjective human eye perception effect and quantitative objective evaluation index.

The CNN model used in this paper is an unsupervised learning fusion network, which does not need to construct a huge label training set, but only needs to use Brushlet chunking energy to construct training samples, so the network can not only be used for fusion of natural and medical images but also can be extended to the fusion of infrared and visible images, the fusion of remote sensing images, etc. The fused images can be further applied to target detection, tracking, etc.

Due to the training method of CNN model used in this paper, the complexity of the algorithm is high. Future work considers the Fast RCNN model to improve the convergence speed and reduce the complexity of the algorithm. In addition, the mapping relationship between the source image pair and the fused image can be further explored, and there is still room for improvement of the algorithm in this paper by optimizing the network structure and loss function, etc.

References

1. Kaur H, Koundal D, Kadyan V. Image fusion techniques: a survey[J]. Archives of Computational Methods in Engineering, 2021, 28(7): 4425-4447.
2. Li X, Guo X, Han P, et al. Laplacian redecomposition for multimodal medical image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(9): 6880-6890.
3. Palani U, Vasanthi D, Begam M S R. Enhancement of medical image fusion using image processing[J]. Journal of Innovative Image Processing (JIIP), 2020, 2(04): 165-174.
4. Shao Z, Cai J. Remote sensing image fusion with deep convolutional neural network[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2018, 11(5): 1656-1669.
5. Belgiu M, Stein A. Spatiotemporal image fusion in remote sensing[J]. Remote sensing, 2019, 11(7): 818.
6. Yang Y, Wan W, Huang S, et al. Remote sensing image fusion based on adaptive IHS and multiscale guided filter[J]. IEEE Access, 2016, 4: 4573-4582.
7. Jee S H, Kang M G. Image fusion method for a single sensor based multispectral filter array containing a near infra-red channel[J]. Electronic Imaging, 2016, 2016(15): 1-5.
8. Hadi B, Majeed A. Transform infra-red image using discrete wavelet function[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019, 571(1): 012114.
9. Asif M, Sharooq M, Rohith L, et al. Multiresolution Palm Image Fusion Approach for Image Enhancement[J]. European Journal of Molecular & Clinical Medicine, 2020, 7(4): 2432-2437.

10. Martinez J, Pistonesi S, Maciel M C, et al. Multi-scale fidelity measure for image fusion quality assessment[J]. Information Fusion, 2019, 50: 197-211.
11. Gattim N K, Rajesh V, Partheepan R, et al. Multimodal image fusion using curvelet and genetic algorithm[J]. 2017.
12. Shabanzade F, Ghassemian H. Combination of wavelet and contourlet transforms for PET and MRI image fusion[C]//2017 artificial intelligence and signal processing conference (AISP). IEEE, 2017: 178-183.
13. Ding S, Zhao X, Xu H, et al. NSCT-PCNN image fusion based on image gradient motivation[J]. IET Computer Vision, 2018, 12(4): 377-383.
14. Yin M, Liu X, Liu Y, et al. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain[J]. IEEE Transactions on Instrumentation and Measurement, 2018, 68(1): 49-64.
15. Pai Z. Image Fusion Based on Joint Nonsubsampled Contourlet and Overcomplete Brushlet Transforms[J]. International Journal on Smart Sensing & Intelligent Systems, 2016, 9(4): 2186-2203.
16. Zhang Y, Liu Y, Sun P, et al. IFCNN: A general image fusion framework based on convolutional neural network [J]. Information Fusion, 2020, 54: 99-118.
17. Amin-Naji M, Aghagolzadeh A, Ezoji M. Ensemble of CNN for multi-focus image fusion[J]. Information fusion, 2019, 51: 201-214.
18. Dian R, Li S, Kang X. Regularizing hyperspectral and multispectral image fusion by CNN denoiser[J]. IEEE transactions on neural networks and learning systems, 2020, 32(3): 1124-1135.
19. Bhavana D, Kumar K K, Rajesh V, et al. Deep learning for pixel-level image fusion using CNN [J]. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(6): 49-56.
20. Liu Y, Chen X, Peng H, et al. Multi-focus image fusion with a deep convolutional neural network [J]. Information Fusion, 2017, 36: 191-207.
21. Qu G H, Zhang D L, Yan P. “Information measure for performance of image fusion”. Electronics Letters, Vol 38(7):313-315, 2002.
22. Petrovic V, Xydeas C. “On the effects of sensor noise in pixel-level image fusion performance”. In: Proceedings of the 3rd International Conference on Image Fusion. Paris, France: IEEE, 14-19, 2000.