IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received September 15, 2019, accepted October 6, 2019, date of current version October 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947378

Multi-Focus Image Fusion Based on Residual Network in Non-Subsampled Shearlet Domain

SHUAIQI LIU[1,2], JIE WANG[1,2], YUCONG LU[1,2], SHAOHAI HU[3], XIAOLE MA[3], AND YIFEI WU[4]

[1]College of Electronic and Information Engineering, Hebei University, Baoding 071000, China
[2]Machine Vision Engineering Research Center of Hebei Province, Baoding 071000, China
[3]Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[4]Department of Electrical and Computer Engineering, University of California at San Diego, San Diego, CA 92037, USA

Corresponding authors: Xiaole Ma (maxiaole@bjtu.edu.cn) and Yifei Wu (y3wu@eng.ucsd.edu)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572063 and Grant 61401308, in part by the Natural Science Foundation of Hebei Province under Grant F2016201142 and Grant F2018210148, in part by the Opening Foundation of Machine Vision Engineering Research Center of Hebei Province under Grant 2018HBMV02, in part by the Science Research Project of Hebei Province under Grant QN2016085, and in part by the Natural Science Foundation of Hebei University under Grant 2014-303.

ABSTRACT In order to obtain a panoramic image which is clearer, and has more layers and texture features, we propose an innovative multi-focus image fusion algorithm by combining with non-subsampled shearlet transform (NSST) and residual network (ResNet). First, NSST decomposes a pair of input images to produce subband coefficients of different frequencies for subsequent feature processing. Then, ResNet is applied to fuse the low frequency subband coefficients, and improved gradient sum of Laplace energy (IGSML) perform high frequency feature information processing. Finally, the inverse NSST is performed on the fused coefficients of different frequencies to obtain the final fused image. In our method, we fully consider the low frequency global features and high frequency detail information in image by using NSST. For low-frequency coefficients fusion, we can also obtain the spatial information features of low-frequency coefficient images by using ResNet, which has a deep network structure. IGSML can use different directional gradients to process high-frequency subband coefficients of different levels and directions, which is more conducive to the fusion of the coefficients. The experiment results show that the proposed method has been improved in the structural features and edge texture in the fusion images.

INDEX TERMS Image fusion, multi-focus image fusion, NSST, ResNet.

I. INTRODUCTION

In the field of digital image processing, different imaging devices acquire different information from the same scene. As in the optical lens, the acquired image is not an all-focus image since the limited of the lens depth range. The optical image is only clear in the part of the scene that is focused in the lens range, and the rest is a blurred defocused image. Typically, image fusion is often used to produce good results which is superior to the original image quality [1], [2]. The fused image contains more scene information, which is more suitable for imaging features of the human eye and is also convenient for later computer processing. Therefore, the process of multi-focus image fusion can be considered as a tool for producing high quality result images [3], [4].

In the development of multi-focus image fusion, there are two types of fusion methods, namely spatial domain fusion and transform domain fusion [5]. However, the most important aspect is the design of fusion rules in the image fusion processing. The image fusion methods based on transform domain are a popular and widely in this fields. In transform domain-based fusion algorithm, the multi-scale decomposition of the original images is mainly applied by multiscale transform (MST), and image fusion is performed by using different fusion rules for image coefficients at different scales. In image fusion based on transform domain, the performance of the algorithm is mainly dependent on the choice of transform domain and the design of fusion rules. Commonly, transform domain algorithms include gradient pyramid (GP) [6], discrete wavelet transform (DWT) [7], double tree complex

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

wavelet transform (DTCWT) [8], shearlet transform (ST) [5], non-subsampled contourlet transform (NSCT) [9], discrete cosine transform (DCT) [10] and high-order singular value decomposition based method (HOSVD) [11] and others. After selecting the corresponding transform, the fusion processing is turn to the design of fusion rules in multi-focus image fusion algorithms based on transform domain.

The fusion rules are different in low frequency and high frequency. Usually, the weighted average fusion rules are used to fuse the low frequency coefficients, while the larger sum of energy for high frequency coefficient is applied to low frequency coefficients fusion. In [5], complex-shearlet is used to decompose the source images. The authors use weighted average fusion rules based on guided filtering to fuse the low frequency coefficients, and use the larger sum modified Laplacian (SML) with guided filtering to fuse the high frequency coefficients. In [6], image fusion is performed on a multiresolution gradient map representation domain of image signal information. The authors use weighted average fusion rules called arithmetic combinations to fuse the low frequency coefficients, and use pixel-based select max approach to fuse the high frequency coefficients. In [7], after the source images are decomposed by DWT, two different window-based fusion rules named maximum sharpness focus measure and maximum neighboring energy are separately employed to combine the low frequency and high frequency coefficients. In [8], a different image fusion strategy by various fusion rules are innovatively combined in Q-shift DTCWT is presented in this work. In [9], the fusion rules are similar to the one in [8]. In [10], the fusion rule called larger spatial frequencies is used to fuse the DCT coefficients. In [11], the generated coefficients are fused by the multi-level fusion strategy of the sigmoid function. In the transform domain, the sharpness measurements of the source images are measured by transforming coefficients of different scales and directions in the domain. Then, the final desired fusion image can be obtained by the final fusion coefficients through inverse transform. The spatial image fusion algorithms usually perform image fusions by measuring the spatial definition of the source images.

Image fusion algorithms based on spatial domain include: weighted mean fusion algorithm [12], principal component analysis based fusion algorithm [13], image fusion algorithm for improving Laplacian energy [14], pulse coupled neural network based fusion algorithm [15], image gradient based fusion algorithm [16], surface area focusing criterion based fusion algorithm [17], non-local mean filtering based fusion algorithm [18], graphic-based visual saliency based fusion algorithm [19], self-similarity and depth information based fusion algorithm [20], the structure saliency based fusion algorithm [21] and so on. Image fusion algorithms for spatial domain are relatively simple and easy to implement. However, traditional spatial image fusion algorithms may produce artificial texture and also has more serious blocking artifacts. And, the in-depth and continuous improvement of spatial domain research has improved the quality of fused images. In particular, the development of deep learning theory has made the spatial domain algorithm based on deep learning achieve good results. Compared with the traditional spatial multi-focus image fusion methods, the deep learning algorithm has a great development prospect in optimizing the fusion image result processing [22].

In recent years, with the development of deep learning theory meanwhile get good research results in related fields. The proposal of deep learning (DL) [22] is derived from artificial neural networks. The so-called ''deep'' is a perceptron with multiple hidden layers. Deep learning can effectively combine low-level features. Therefore, CNN has also been applied to fusion process. In [23], a fusion rule use CNN is proposed. It treats the generation of the fused map as a classification problem. In [24], a multi-focus image fusion algorithm, which uses image segmentation based on multi-scale CNN to generate fusion decision map, generate high quality fused image. In [25], pixel convolutional neural network for multi-focus image fusion (P-CNN) is proposed. This algorithm can select the different focus degree pixels from the neighborhood information of source images. In addition, the P-CNN can also set precise labels according to different focal length levels for image classification processing to form accurate focus information feature maps. Yang *et al.* [26] proposed multi-level features convolutional neural network for multi-focus image fusion (MLFCNN). In this method, all the features learned from the previous layer are passed to the next layer, and 1 × 1 convolution block is added to each path between the upper layer and the next layer to reduce redundancy. This method firstly inputs the input images into a pre-trained MLFCNN model to obtain an initial focus map. Then, the initial focus map is refined by the morphological opening and closing operation. Finally, Gaussian filtering is performed to obtain the final fused map. The final fused image is generated by using weighted sums of fused map.

These fused algorithms can effectively integrate the judgment of the focus area and design of the fusion rule based on a large amount of image learning. Achieving clear results with higher quality. However, in the training of CNN, the number of feature extraction layers is small, the accuracy of extracting and identifying the focus block is low, and the extraction of image edges and texture features is not rich enough. In turn, the total amount of information and visual effects of the fusion results are affected. Moreover, these methods do not divide the frequencies of the image in the processing of image fusion. This is obviously not in line with the human eye to observe the image features.

In order to improve the disadvantages of the above algorithms, we propose an innovative algorithm by combining with NSST and ResNet. First of all, we use ResNet to deal with the shortcomings of insufficient image feature extraction based on CNN image fusion algorithm. He *et al.* [27] proposed ResNet, and the structure of ResNet has high training efficiency and the model accuracy has been greatly improved. Therefore, we use ResNet for image fusion in our algorithm. Similar to CNN-based image fusion, focus measurement and fusion decision map are generated by training ResNet model, which can overcome the difficulties of manual operation and extract more complete information features. NSST can sparsely represent image [28]–[30]. And the NSST decomposed subband image is the same size as the input image. Then, in order to make the process result has more global contents and detail edge structure of the source images, NSST is used to divide the frequency of the original images, and separately fuse different coefficients. The low-resolution images produced by the image after NSST decomposition contain more overall information content and the high-resolution images contain more texture detail features. In this paper, ResNet is applied to fuse low frequency coefficients, while IGSML is applied to fuse high frequency coefficients. ResNet can better extract the hierarchical features of the low frequency images when processing the low frequency images containing the contour information, which can contain more overall information of the source images after image fusion. The high-pass coefficients mainly contain detail texture, and IGSML can fully compare the information characteristics of different scales and directions, which can contain better detail feature of the original images after image fusion.

The structure of this paper is similar to [15], however, the idea of the proposed method is very different to the method in [15]. In our algorithm, the source images are decomposed to low and high frequency, and different fusion rules are applied to low and high frequency. The ResNet-based based image fusion rule is applied to low frequency coefficients, which can obtain a more complete image overall contour. However, the high frequency coefficients of NSST contain the details and edges of the image. We using IGSML to exact the gradient and energy processing of high-frequency coefficients. The proposed algorithm considers the visual structure features of the image and extraction of deep detail features, and it has good robustness, suppresses the generation of artificial texture, and improves the visual clarity of the image. The core idea of the method in [15] is dual-channel-SCM model, which is designed and applied to the fusion process of NSST decomposition coefficients. And through the difference images between the original fused image and the original images, further fusion is obtained to get the final fusion result. The fusion of the high and low pass coefficients of the decomposition is not carried out by different fusion rules. And the fusion processing is directly performed by the dual channel-SCM model, which produces ideal results. Therefore, the two papers have certain similarities in structure, but they have their own innovative parts.

Compared with the pure deep learning image fusion algorithms, we use NSST to process the images and fully considers the overall and local features of the images. At the same time, the optimized ResNet is introduced to perform deep feature extraction without increasing computational complexity. So, the proposed algorithm considers the visual structure features of the image and extraction of deep detail features, and it has good robustness, suppresses the generation of artificial texture, and improves the visual clarity of the image.

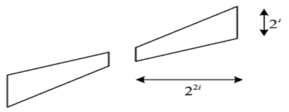


FIGURE 1. The trapezoidal pair frame of shearlet.

The chapter of the paper is divided as following. The second section introduces the principle of NSST. The third section introduces the fusion rules based on the ResNet-50 layer model. It mainly includes the general ResNet network model structure, the ResNet-50 network model used in this paper and training process of the model. The fourth section gives the ResNet fusion rules based on the NSST, which mainly includes the fusion rules of different frequencies in NSST domain, and also gives all the steps of the proposed image fusion algorithm. The fifth part gives the experimental results and evaluation indicators of different fusion methods. The sixth part summaizes the innovations and shortcomings of our method.

II. NON-SUBSAMPLED SHEARLET TRANSFORM

Shearlet [28] has good directionality and realizes multiscale geometric transform with relatively easy process. Synthetic wavelet technology is the theoretical basis of shearlet transform. When the dimension n=2, the shearlet functions are generated by affine transform as following (1), as shown at the bottom of this page.

The function expression is obtained by a series of different spatial transform such as scaling, shearing and translation of the basis function of the shearlet basis function. It can well represent the morphological features of the image and reduce artifacts. And $\psi \in L^2(R^2)$, $\mathbf{A}, \mathbf{B}$ are 2×2 invertible matrices and $|\det \mathbf{B}| = 1$. The dilations $\mathbf{A}^j$ are telescopic transform matrices, while the matrices $\mathbf{B}^l$ are related to geometric transform of the preservative region. Normally $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$ or $A = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ represents the anisotropic dilation matrix, and $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ or $B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ represents shear matrix. From FIGURE 1, we can know that the bias function of shearlet $\hat{\psi}_{j,l,\mathbf{k}}$ is supported on a pair of trapezoids of approximate size $2^{2j} \times 2^j$, and the slope of the trapezoids is $l2^{-j}$ [31], [32].

Therefore, the continuous Shearlet transform of f is expressed as the following.

$$SH_{\psi} = \left\langle f, \psi_{j,l,k} \right\rangle, \tag{2}$$

where $j \geq 0$, $l = -2^j, 2^j - 1$, $\boldsymbol{k} \in Z^2$. The transform diagram of shearlet is shown in FIGURE 2.

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

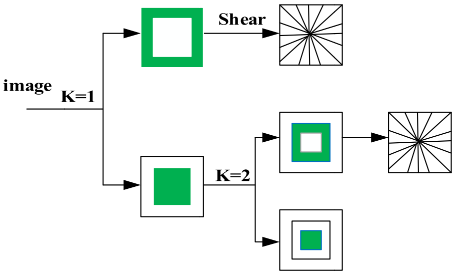


FIGURE 2. Shearlet transform diagram.

NSST [29], [30] is an extension of the shearlet transform, which makes it have good frequency domain characteristics in the process of processing images. NSST [34] is realized through different scales and directions. First of all, the image is decomposed by non-subsampled pyramids (NSP). It can be concluded by scale decomposition that NSP can generate k high frequency sub-band coefficients and one low frequency sub-band coefficient after k-th scale decomposition. In the second part, multi-directional transform of the decomposed different frequency sub-images is achieved by applying an improved shearing filter (SF). Because the NSST transform can well overcome the down-sampling operation of the image, it has the translation invariance of the image transform. When performing related image processing operations, it is possible to have more information features of the input images in the fused result, which will greatly improve the overall fused image effect. So, NSST is widely used in image process.

In this paper, to get better computational efficiency and extracting image detail features, NSP is used to for source images' decomposition. And the scale of decomposition is two. SF is used for directional decomposition. The high frequency in the first scale has four directions, while the second scale has eight directions.

NSCT also has multi-scale and multi-directional decomposition ability, and also has the characteristics of anisotropy and translation invariance. However, the speed of NSCT is slowly. And direction decomposition is not flexible enough to implement in NSCT.

In order to compare the performance of the fused method base on NSCT and NSST, we replace the NSST by NSCT in our method. And we set the decomposition parameters of the NSCT to be the same as NSST. In FIGURE 3, the experiment gives the image fusion results of a pair of original images after NSCT-based fused method and NSST-based fused method. In the fusion result, it can be found that the fused image by NSST-based method has clear detail features than NSCT-based method.

TABLE 1 gives the mutual information of the two fused images. The result show that NSST perform a little better than NSCT. However, from TABLE 1, we find that NSST-based fused method has high computational efficiency, and NSCT is time consuming. So, we use NSST to process images, which can achieve better performance and high computational efficiency.

TABLE 1. Comparesion of NSCT and NSST.

The fused methods	The mutual information	Time(s)
NSCT-based method	9.0235	170.0352
NSST-based method	**9.1525**	**58.2231**

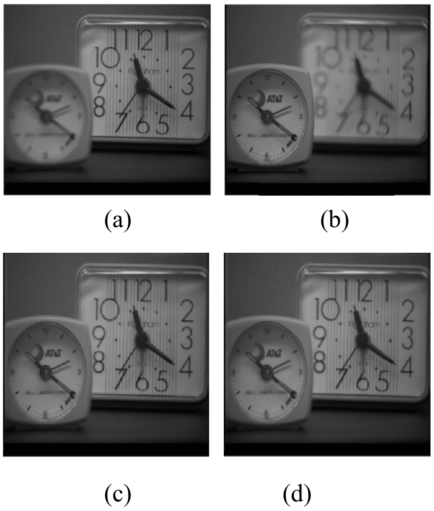


FIGURE 3. (a)-(b) are the original images and (c)-(d) are NSCT-based and NSST-based fusion results.

III. FUSION RULES BASED ON RESNET-50 MODEL

A. RESNET BLOCK MODEL

Traditional convolutional networks or fully connected networks have more or less information loss during information transmission. At the same time, there may be gradient disappearance or gradient explosion, which may lead to the network layers being too many to train the network model. ResNet has dealt with the above problems to a certain extent by connecting the network mapping. ResNet is widely used in the field of image processing, such as license plate recognition [33], speckle suppression [34] and image classification [35]. But it is rarely used in image fusion. ResNet can extract deep image features and has good image feature extraction performance. Therefore, in this paper, we extend it to the field of multi-focus image fusion.

The main idea of ResNet is to add a direct connection channel to the network, which is the idea of the highway network. The idea of ResNet is very similar to the highway network. It allows the original input information to be passed directly to the later layers, which removing the same body part, thereby highlighting minor changes. Therefore, while increasing the depth of the network hierarchy, there will be no introduction of additional parameters and no increase in computational difficulty [27]. Assuming that the target map to be learned by a sub-module of the neural network is $H(x)$, this mapping function may be complicated and difficult to fit. The idea of residual learning is to employ the stack nonlinear layer to fit another mapping relationship: $F(x) = H(x) - x$, then the actual mapping relationship can be expressed as $F(x) + x$. That is, in the residual neural network, a sub-module consists of two parts: an identity map $x \to x$ and a

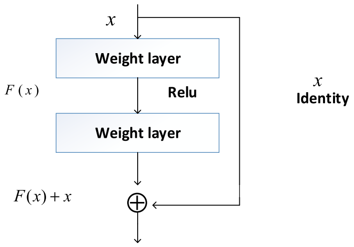


FIGURE 4. Residual block structure.

nonlinear map $F(x)$. FIGURE 4 is a schematic diagram of a sub-module for residual learning.

When the identity map $x \rightarrow x$ is the optimal map, the learning algorithm can easily set all the parameters of the nonlinear map $F(x)$ to 0. This is obviously much easier than making $F(x)$ to fit an identity map $x \rightarrow x$. The specific network structure is as follows.

$$y = F(x, \{W_i\}) + x, \tag{3}$$

where x and y are the input and output vectors of one layer in ResNet. The function $F(x, \{W_i\})$ represents the residual map to be learned.

In order to build a deeper neural network, the cost of stacking multiple residual learning modules shown in FIGURE 3 is still a bit large. In [27], a residual learning module structure called bottleneck is also proposed. The bottleneck structure consists of two 1*1 convolutional layer and a k^*k convolutional layer. Usually the value of k is 3. Suppose the input data of this module has 256 channels. Then the first 1*1 convolutional layer is to reduce the input data dimension to 64. This operation can effectively reduce the parameters and achieve cross-channel information fusion to some extent. The role of the second 1*1 convolutional layer is to raise the data dimension to 256, which ensures that the data dimensions of the two operations are the same in the $F(x)+x$. This bottleneck block is the key to enabling the residual neural network to reach hundreds or even thousands of layers.

B. RESNET-50 MODEL

Through the deep ResNet model, we compare the subjective, objective and time factors, and consider that the network hierarchy of ResNet-50 layer model is the most suitable for image fusion. This structure combines a plurality of shallow residual blocks stacked together. Among them, the structure diagram of a shallow residual block is presented in FIGURE 5.

Among them, each residual block is composed of convolution layer, normalization, resizing and activation function. In FIGURE 5 the residual block is composed of four convolutional layers. Each convolution layer core has sizes of 1*1, 1*1, 3*3, and 1*1, and the corresponding number of filters is 256, 64, 64, and 256. Note that the identity mapping part is processed by the convolutional layer and the normalization layer. The purpose is to ensure that the dimensions and sizes of the identity mapping are consistent with the dimensions and sizes of the features obtained by the residual function.

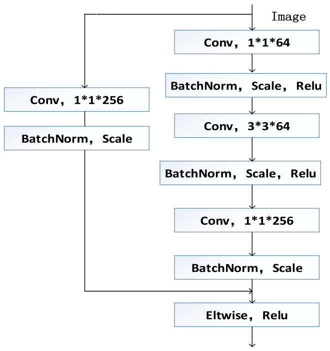


FIGURE 5. The structure diagram of residual block in our algorithm.

Res2a-b is made up of two residual blocks of Res2a and Res2b. Res2a and Res2b are similar to the residual structure shown in Figure 5, and the difference is just that the identity mapping part is directly connected by a line, and there is no corresponding convolution operation. Res3a-c, Res4a-e and Res5a-b are residual block groups similar to Res2a-b. During the training of the network, the number of filters used in each convolution layer is different, and the number of feature maps generated is also different, while the number of features and the number of filters is the same. In each residual structure, the number of feature maps is shown in FIGURE 7.

The feature map obtained by the two branches trained by the last residual block is completely connected, and finally the 2-dimensional feature vector is obtained. After changing the network structure, the part is the two full connected processing, which is convenient for input images of any size to produce dense fractional graphs [36]–[39]. We consider Res2a-b and Res2c in Figure 6 as a whole, that is, the Res2 feature vector output layer, and display it in the feature map of FIGURE 7. The Res3, Res4, and Res5 layers are similar to the Res2 layer. It is also an eigenvector output layer.

C. MODEL TRAINING

Since image mergence can be regarded as a two-class problem, that is, the algorithm should distinguish the corresponding positions belong to the fuzzy part or the clear part. Therefore, the ResNet-50 training can be easily completed. We use the Matconvnet toolbox to complete the training process of the model. The training and verification images used in this paper are part of the image dataset in ImageNet. Due to lacking of organized open multi-focus image training set, in this paper, some data sets in ImageNet are directly processed by Gaussian blurring to produce images of different degree of focus, which are used for model training.

The specific training process, such as the pre-processing of the image (including the size of image block pair is 16*16, the measurement range of the image block in different degrees of focus is 0-1) is similar in the literature [40]. In this paper, the standard deviation of the Gaussian filter is 2, and the clear image is filtered to produce a blurred image. The first blurred

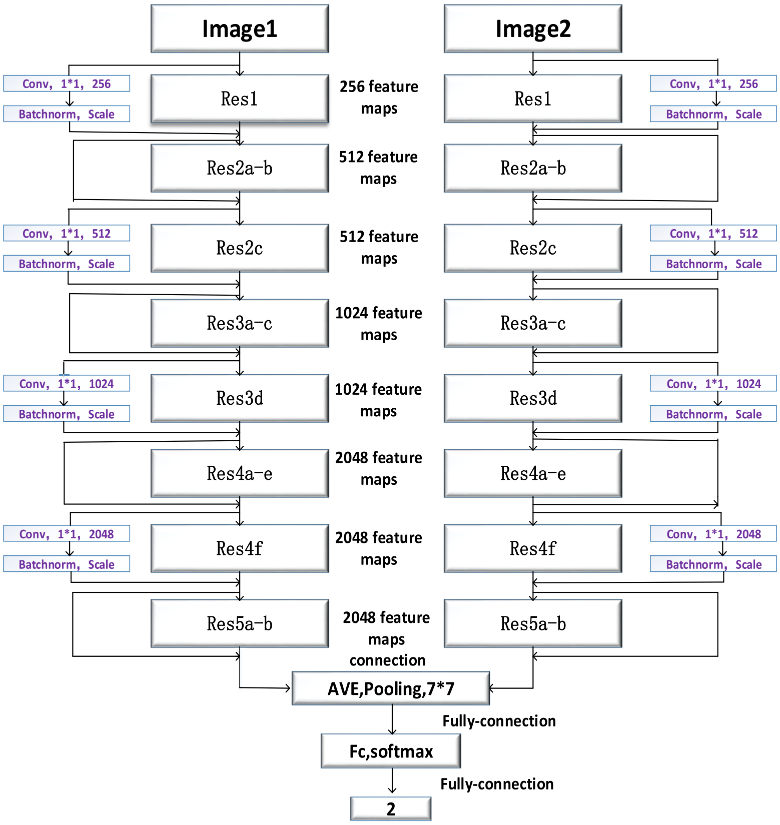


FIGURE 6. ResNet-50 model.

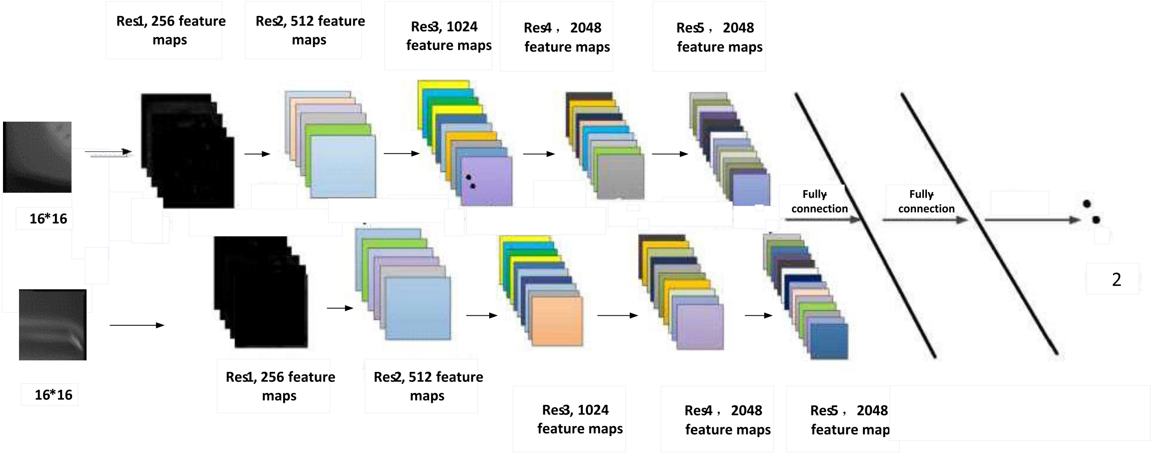


FIGURE 7. The training process of Resnet-50 model.

image is then filtered again to produce another blurred image that is different from the first blur, and so on. Taking into account the time and the accuracy of training results, we used four images with different degrees of blur as the training and verification image sets.

In the process of training, our fusion decision map is generated by labels 0 and 1 marked on the image set. We define a and b as the inputs to the two branches, labeled 1 if the a and b branches correspond to clear and blurred image blocks, respectively. Conversely, if a and b correspond to obscured and clear image blocks, they are marked as 0. Finally, 50,000 image pairs labeled 1 and 50,000 image pairs labeled 0 are selected as training sets.

In the process of training, the ResNet-50 layer model is selected as the training network model. As a training target, the softmax function achieves the desired value by several iterations of updating the weight parameters of the training network. The determination of the initial value during the training is adaptively selected by the number of neurons. In order to show the training results more clearly, the feature map during the training process is given in FIGURE 6. For a pair of input image blocks, each residual block can generate a corresponding set of feature extraction maps to extract information features of different details. Further, the focus information features of the original image can be better represented. Using the last two parts of the full connection operation, we can get the focus score map of the image. The full connection operation uses a convolution kernel of 1*1 size, which can realize the processing of input images of any size, and is more suitable for practical requirements.

IV. FUSION RULES FOR RESNET BASED ON NSST CHANGE DOMAIN

In this paper, the NSST decomposition of the original images can obtain high and low frequency images representing different levels of information features. Since the low frequency coefficients of the images contain the main features of the image, there is strong contour information with the source images. We use the Resnet-50 based network model for the fusion of low frequency coefficients. ResNet with deep structure can fully extract the global hierarchical features of the image and perform accurate classification. Since the human visual perception system is sensitive to the information characteristics of the texture, edge and direction of the image, while it is not very sensitive to the response of a single pixel in the image. The high frequency images generated by the NSST contain more detailed information, and the gradient can well show the degree of detail change of the images. In this paper, the gradient changes of the main diagonal and the sub-diagonal are added to the gradient changes of the original rows and columns to form gradients in four directions. Therefore, the texture features of the high frequency images can be better extracted. SML is a regional energy function that can better represent the details of image edge, but it is not directional. In order to make the fused image have better spatial continuity, we combine the improved gradient and SML to form the IGSML operator for high-frequency image fusion. Let A and B denote two source images with different focus field, and F is the fusion image. After the source images A and B decomposing by NSST, the corresponding high frequency coefficients are $S_A^{l,d}(i,j)$, $S_B^{l,d}(i,j)$ and the low frequency coefficients are $S_A^{0,d}(i,j)$ and $S_B^{0,d}(i,j)$. For multiple images fusion, we can obtain a final high-resolution fused image by means of two-in-one integration.

A. LOW FREQUENCY COEFFICIENTS FUSION

First, the low frequency coefficients $S_A^{0,d}(i,j)$ and $S_B^{0,d}(i,j)$ are used as inputs into the ResNet-50, thereby obtaining low frequency coefficient score map *map*(i,j) of image A and B. In order to display the source images information more clearly, when *map*(i,j) is greater than 0.5, the value is uniformly set to 1, and otherwise take 0. Then, the fusion decision map $Z(i,j)$ is obtained as following.

$$Z(i,j) = \begin{cases} 1 & m(i,j) > 0.5 \\ 0 & other, \end{cases} \tag{4}$$

According to the weighted average fusion rule of the corresponding pixel, the low frequency fusion coefficients of images A and B are obtained as following.

$$S_F^{0,d}(i,j) = Z(i,j)S_A^{0,d}(i,j) + (1 - Z(i,j))\, S_B^{0,d}(i,j) \tag{5}$$

B. HIGH FREQUENCY COFFICIENTS FUSION

For the high frequency coefficients $S_A^{l,d}(i,j)$ and $S_B^{l,d}(i,j)$, we get the fusion coefficients by using fusion strategy with larger IGSML value. The gradient can show the details of the image. The gradient only considers the first-order difference features of the row and column directions, there are fewer information features. So, in our paper, we improve the gradient with increasing the direction information of the main diagonal and the sub-diagonal. It can obtain more detail direction information, which make the feature representation more specific and comprehensive. The solution formula for the improved gradient is as follows:

$$G(i,j) = \sqrt{f_x^2 + f_y^2 + f_{xy}^2 + f_{-xy}^2}, \tag{6}$$

where f_x^2, f_y^2, f_{xy}^2 and f_{-xy}^2 represent the first-order difference formulas of the row, column, main diagonal, and sub-diagonal, respectively. Their calculation formula is as follows.

$$f_x = I(i,j) - I(i-1,j), \tag{7}$$
$$f_y = I(i,j) - I(i,j-1), \tag{8}$$
$$f_{xy} = I(i,j) - I(i+1,j+1), \tag{9}$$
$$f_{-xy} = I(i,j) - I(i+1,j-1), \tag{10}$$

The SML can be calculated as follows (11), as shown at the bottom of the next page.

We calculate gradient and SML by using the sliding window of 3 × 3, and we can get the local average gradient $G(i,j)$ and the SML $s(i,j)$. The regional gradient and SML are normalized as the weighting factors g_A, g_B, s_A and s_B. Then, the high frequency coefficients are adaptively weighted, and we can obtain fused coefficients as following.

$$\begin{cases} S_F^{l,d}(i,j) = \mu S_A^{l,d}(i,j) + (1-\mu)\, S_B^{l,d}(i,j) \\ \mu = (g_A + s_A)\big/(g_A + s_A + g_B + s_B), \end{cases} \tag{12}$$

where S represents the pixel value of the different high frequency coefficients, μ is the weighting coefficient, and A, B represent the original images. g_A and g_B are normalized gradients which can be calculated as following.

$$g_A = \frac{G_A}{\max(G_A) - \min(G_A)}, \tag{13}$$
$$g_B = \frac{G_B}{\max(G_B) - \min(G_B)}, \tag{14}$$

And s_A and s_B are normalized SML, which can be calculated as following.

$$s_A = \frac{s_A}{\max(s_A) - \min(s_A)}, \tag{15}$$
$$s_B = \frac{s_B}{\max(s_B) - \min(s_B)}, \tag{16}$$

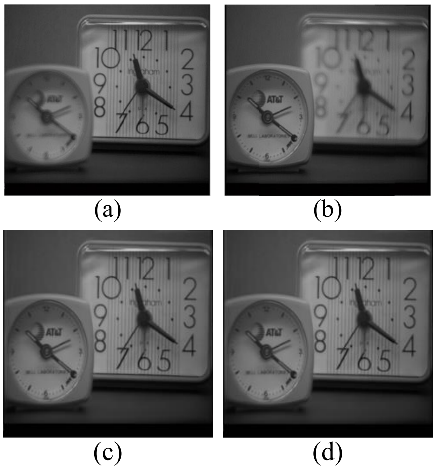


FIGURE 8. (a)-(b) are the original images and (c)-(d) are SML-based fused image and IGSML-based fused image.

TABLE 2. Evaluation indicators for SML- and IGSM-based method.

The fusion methods	Q_Y	Q_{PC}
SML-based	0.9598	0.7327
IGSML-based	**0.9685**	**0.7421**

Combining the advantages of local average gradient and SML, IGSML can obtain the different directional gradients and details information of the high frequency coefficients. To compare the performance of IGSML and SML, we replace IGSML by SML in our method. Figure 8 gives the fused images by IGSML-based fusion method and SML-based fusion method. TABLE 2 shows the objective indicators (which is explained in the next section).

From the subjective visual features of FIGURE 8, it can be found that the results of IGSML and SML all has good visual effect. However, the objective indicators in TABLE 2 show that IGSML-based method is better than SML-based method, which contributed to the using of IGSML.

According to the above processing, we can get the final fused image by inverse NSST applied to the fused high and low frequency coefficients. FIGURE 9 shows flow chart of the proposed fusion algorithm.

V. EXPERIMENT AND ANALYSIS

A. EXPERIMENT SETTINGS

In order to effectively evaluate the fusion performance of the proposed algorithm in different degrees of focus images, we test our method in the commonly used multi-focus images (shown in FIGURE 10). And the performance indicators from both subjective and objective aspects are compared with the other nine representative multi-focus fusion algorithms, such as multi-focus image fusion based on sparse representation (SR) proposed in [41], image fusion with guided filtering (GFF) proposed in [42], multi-focus image fusion in DCT domain by using variance and energy of Laplacian and correlation coefficient for visual sensor networks (DCT) proposed in [43], image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain (NSCT-PCNN) proposed in [44], multi-focus image fusion base on multi-scale weighted gradient (MWG) proposed in [45], multi-focus image fusion with a deep convolutional neural network (CNN) proposed in [23] and multi-focus image fusion based on pixel convolutional neural network (P-CNN) proposed in [23], multi-focus image fusion based on dual-SCM in NSST domain (NSST-SCM) proposed in [15], boundary aware multi-focus image fusion using deep neural network (BADNN) [4]. The computing environment of all the algorithms in this paper is UltraLAB Alpha600 super graphics workstation, the basic configuration is CPU E5-4627, and the memory is 16G*16. All algorithms were run by using MATLAB 2014A. The parameters used by each fusion algorithm are the same as those in the original papers.

We also make subjective and objective evaluations of the fused images generated by different algorithms. In the objective evaluation of images, we use five kinds of objective evaluation indicators including normalized mutual information (MI) [46], image similarity based measures (Q_Y) [47], phase congruency-based fusion metric (Q_{PC}) [48], edge information similarity measurement ($Q^{AB/F}$) [49] and fusion measure of human perception (Q_{CB}) [50]. MI shows how much information of the original images are in the merged results, Q_Y indicates the similarity between the merged image and the input images, Q_{PC} indicates the measurement of the phase similarity of each frequency domain component of the image, $Q^{AB/F}$ shows the amount of information of the source images owned by the fusion result and Q_{CB} is a measurement of human perception, mainly using human visual features to measure. Among the five objective evaluation indicators, the larger the evaluation value is, the better the results obtained.

B. COMPARISON OF FUSION RESULTS BY DIFFERENT FUSION ALGORITHMS

First, we compare the fusion effects of different fusion algorithms from the perspective of visual effects. In FIGURE 11, experimental results of different fusion methods with "Clock" as the original images are given. In order to more clearly see the difference between the features of each fused image and the different fusion results, the difference images between the different fusion results and the original

$$s(i,j) = \sum_{(i,j)\in(-3,3)} [|2I(i,j) - I(i-1,j) - I(i+1,j)| + |2I(i,j) - I(i,j-1) - I(i,j+1)|]^2, \tag{11}$$

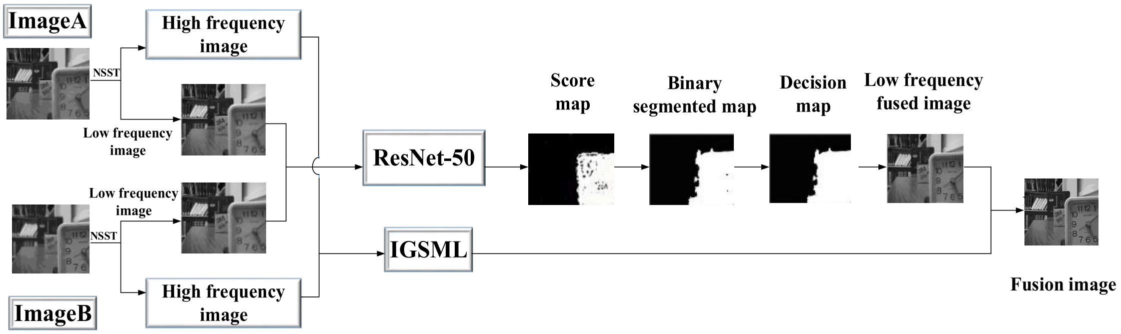


FIGURE 9. Flow chart of the proposed fusion algorithm.

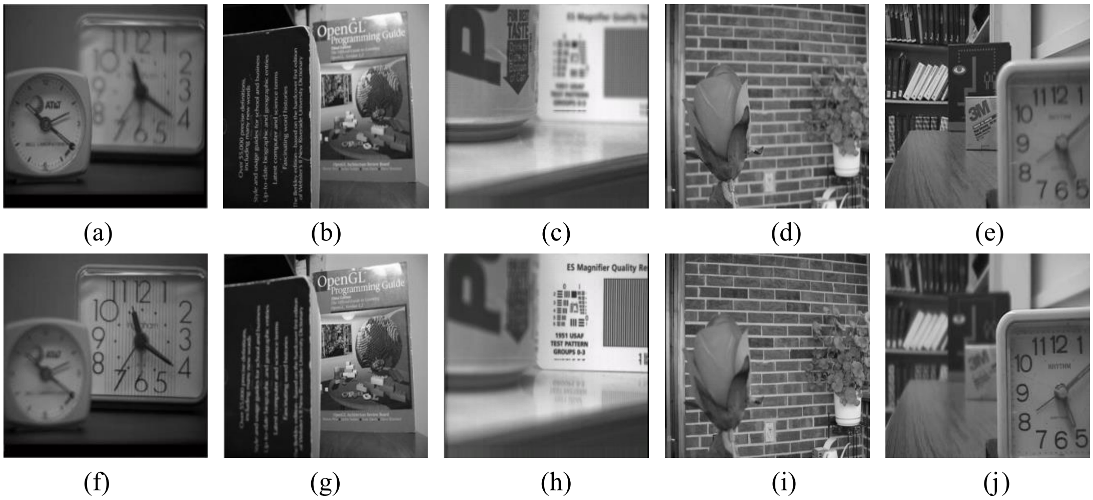


FIGURE 10. 5 pairs of multi-focus source images.

images are also given in FIGURE 11. It can be seen from the difference images between the left-focused image and right-focused image that there are shadows in the images of FIGURES 11 (a1) and (a2), which are most obvious in the upper right corner of the small clock near the number "9". It indicating that the SR-based algorithm does not display the detailed features of the image well. FIGURES 11(b1-d1) and (b2-d2) have poor continuity blocks and artifacts. It is shown that GFF, DCT, NSCT-PCNN produce block effects with poor continuity. FIGURE 11 (e1-g1) and FIGURE 11(e2-g2) have more artifacts and isolated points at the boundary, which indicating that MWG, CNN and P-CNN are not ideal in the boundary part, and the boundaries of the images are not well extracted. There are block area on the right side of the FIGURE(h2). It can be clearly seen from the FIGURE 11(i1,i2) that there are more edge textures. It shows that BADNN cannot extract focus information very well. Our algorithm can extract the features of the image focus areas well, and make the merged result clearer and more visually appealing to the human eye.

In TABLE 3, five different objective evaluation indicators for each fusion algorithm are given to show the fusion effects of different fusion algorithms. At the same time, we also give the running time of each method. Among the values of the given evaluation indicators, large values indicate that the fused algorithm has good experimental result. TABLE 3 shows that the proposed algorithm has the best values in the evaluation indexes of MI, Q_{PC} and $Q^{AB/F}$. In the Q_Y evaluation index, the GFF algorithm has the best effect, indicating that the adaptive dual-channel network used in NSST-SCM can better extract the similarity characteristics of the image. Among the Q_{CB} indicator, P-CNN has good visual perception effect, and the network structure of P-CNN can process image feature information well. Among the two indicators of Q_Y and Q_{CB}, the performance of the proposed algorithm is not perfect, but it is not the worst in the compared algorithms, and has certain room for improvement. Compare to SR, CNN, NSCT-PCNN and BADNN, the proposed algorithm is efficient. The performance of the proposed method and NSST-SCM looks similar, that is, both algorithms have coefficient decomposition of the original images, and the calculation efficiency is similar. Compared with the direct processing of different coefficients by using the dual-channel SCM algorithm, our algorithm deals with the high and low frequency coefficients of the images separately. And, for low frequency coefficients fusion, ResNet in our method can extract the deep features of the source image, so the final result of our method has better global and detailed information, and thus has better

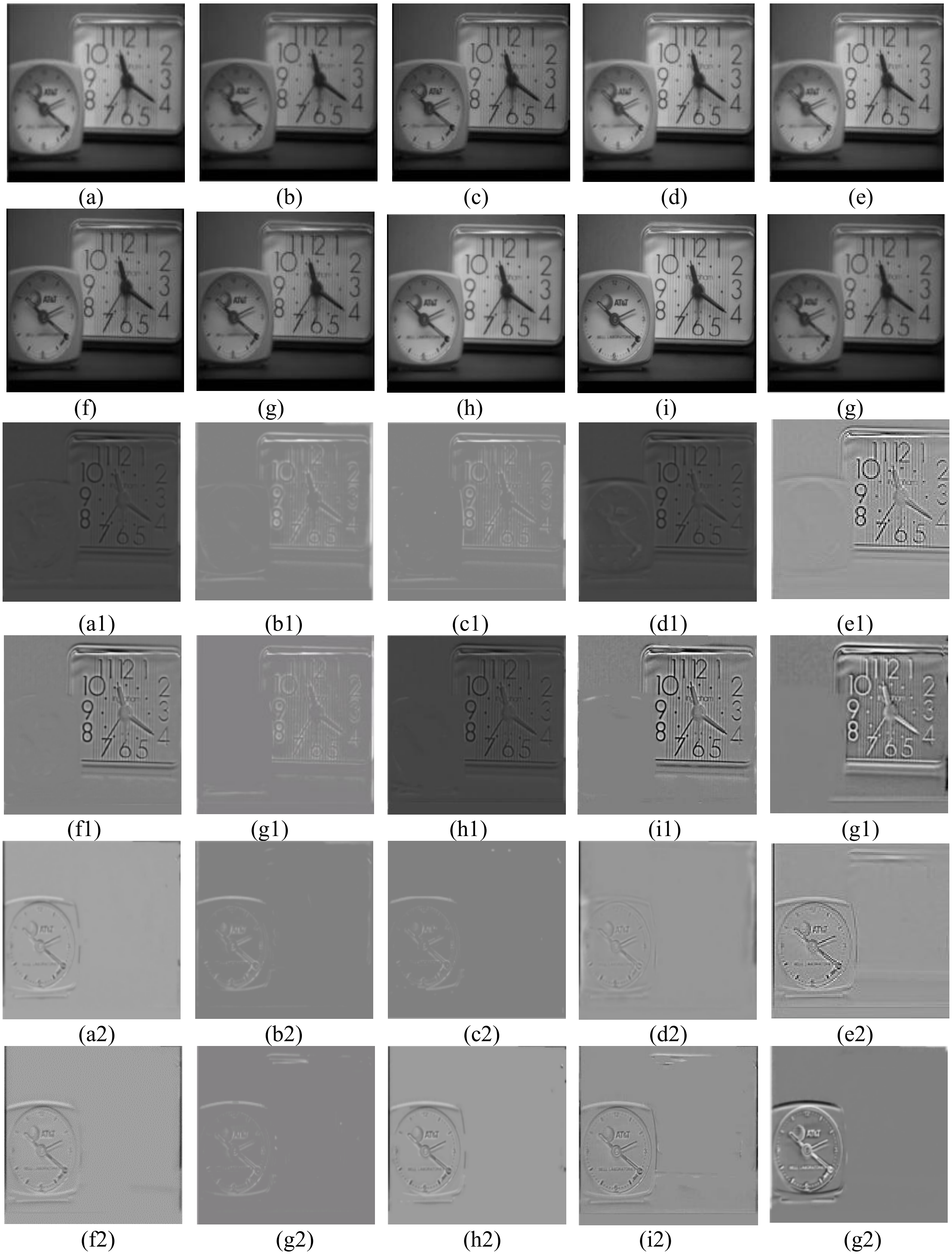


FIGURE 11. The results of the fusion of the Clock images. (a-i) are fusion results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet. (a1-j1), (a2-j2) are difference images with the left focused image FIGURE 10(a) and the right focused image FIGURE 10(f), respectively.

image fusion performance. In the comparison of objective evaluation indicators, it demonstrates that the comprehensive performance of the algorithm is the best.

FIGURE 12 shows the effects of eight fusion algorithms on the fusion of "Book" multi-focus images. There is a fuzzy artifact in the corner of the book on the left side of FIGURE 12(a), indicating that the SR-based fusion method cannot extract image detail features better. In FIGURE 12(b-d), there are distortions on the edge of the images, which indicates that the algorithm based on GFF, DCT and NSCT-PCNN cannot extract the overall information of the images well, and the visual effects are not ideal.

TABLE 3. Objective evaluation indicators of different fusion algorithms in figure 11.

Fusion methods	MI	Q_Y	Q_{PC}	$Q^{AB/F}$	Q_{CB}	Time(s)
SR	8.3412	0.9545	0.7156	0.7252	0.7532	80.9925
GFF	8.6614	0.9894	0.7120	0.7232	0.7738	0.7778
DCT	8.3253	0.9673	0.7121	0.6982	0.7601	0.2341
NSCT-PCNN	7.4598	0.9653	0.7188	0.6880	0.7634	255.3342
MWG	8.7212	0.9723	0.7201	0.7279	0.7658	50.3328
CNN	9.1012	0.9785	0.7016	0.7345	0.7811	161.2210
P-CNN	9.1496	0.9844	0.7315	0.7244	**0.7812**	**0.0633**
NSST-SCM	9.0997	**0.9895**	0.7389	0.7589	0.7798	60.7580
BADNN	8.3214	0.9425	0.7033	0.7127	0.7352	3120.11
NSST-ResNet	**9.1525**	0.9685	**0.7421**	**0.7599**	0.7808	58.2231

In FIGURE 12(e-g), there are isolated points and discontinuous small pieces at the connecting edges of the two books. The images produced in the MWG, CNN and P-CNN are discontinuous. The block area of the image can be clearly seen from FIGURE 12 (i2). It shows that BADNN has poor coherence when processing images. Compared with Figure FIGURE 12(h), FIGURE 12(j) shows that the merged image obtained by the proposed algorithm is clear in both the overall information and the edge detail textures. In order to see the difference between the different algorithms more clearly, the difference images of FIGURE 12 (a1-j2) are given. We found that the proposed method can extract complete image information at the edge or at the combination of focus and defocus, and present clear results.

TABLE 4 demonstrates that the proposed algorithm possesses the best results among the four objective evaluation indicators of MI, Q_Y, Q_{PC} and Q_{CB}. In particular, it shows outstanding performance in the MI and phase consistency metrics of the image, which indicates that the proposed algorithm has an absolute advantage in retaining the original image details, at the same time, the fusion results are more consistent and accord with visual effects. Although the proposed algorithm is slightly worse than the GFF, NSST-SCM and P-CNN algorithms in the comparison of $Q^{AB/F}$ evaluation values, its other indicators are much higher. So, the proposed algorithm is more competitive in image fusion.

FIGURE 13 shows the effects of eight fusion algorithms on the fusion of ''Soda'' images. FIGURE 13 shows that the proposed algorithm has the best visual effect on both the detail information and the overall definition. At the same time, it was obviously discovered that the fusion image obtained by the proposed algorithm is more prominent in hierarchical features. The others fusion algorithms have some misclassified regions at the edge of the image and at the junction of focus and defocus. In particular, there are clear outline images on the right-side portion in FIGURES 13(a2) and (d2), which indicating that the fused images cannot sufficiently extract the focus feature information of the images in the right parts. It means that SR and NSCT-PCNN have some disadvantages in extracting and retaining image information features. And there are clear shadows of letter ''P'' on in the left-side portion in FIGURE 13(b1), FIGURE 13(f1), FIGURE 13(g1) and FIGURE 13(i1), which means that GFF, CNN, P-CNN and BADNN cannot extract the image information features of left-focus part of source images. There is a break of the straight line in FIGURE 13(c2), which means that DCT cannot express the image well. Compare to MWG and NSST-SCM, the difference images show that our method has a better visual effect.

TABLE 5 gives the objective evaluation indicators for all algorithms in FIGURE 13. TABLE 5 indicates that our method is more better than other fusion methods in the evaluation index values of MI, Q_{PC} and $Q^{AB/F}$, which indicating that the proposed algorithm has advantages in suppressing image artifacts. Among the evaluation values of Q_Y and Q_{CB}, the proposed algorithm is slightly smaller than the MWG and P-CNN. In general, the proposed algorithm produces a satisfactory fusion result.

FIGURE 14 shows the effects of eight fusion algorithms on the fusion of ''Flower'' multi-focus images. The fusion characteristics of different algorithms can be clearly found from the difference images. FIGURE 14 (a1, b1) show that SR does not extract the detailed features of the source images very well. There are obvious block effects in the images in FIGURES 14 (b1-c1) and (b2-c2), which shows that the GFF and DCT algorithms have poor continuity in the information extraction process, and need to improve the spatial continuity of the images. In FIGURE 14(d), the characteristics of the reserved original images are less, and the artificial texture is generated, that is, NSCT-PCNN cannot extract the information structures of the source images well. In FIGURE 14(e1,e2), the overall feature processing is very good. However, the details of the intersection of the petals on the right side are not obvious. So, MWG should improve the detail processing. In FIGURE 14(f1,f2), there are artifacts and individual block regions around the edges, which indicating the fusion result of CNN has no advantage in the extraction and retention of the source images. FIGURE 14(i2) has a large number of block regions and artifacts. In FIGURE 14(g1), the information is not fully expressed, indicating that there is information loss in P-CNN. FIGURE 14(h) has a clear fusion result, but the fused image obtained by our algorithm is more information coherent and clear contour features. Our algorithm can better extract the useful information from the source images and generate the

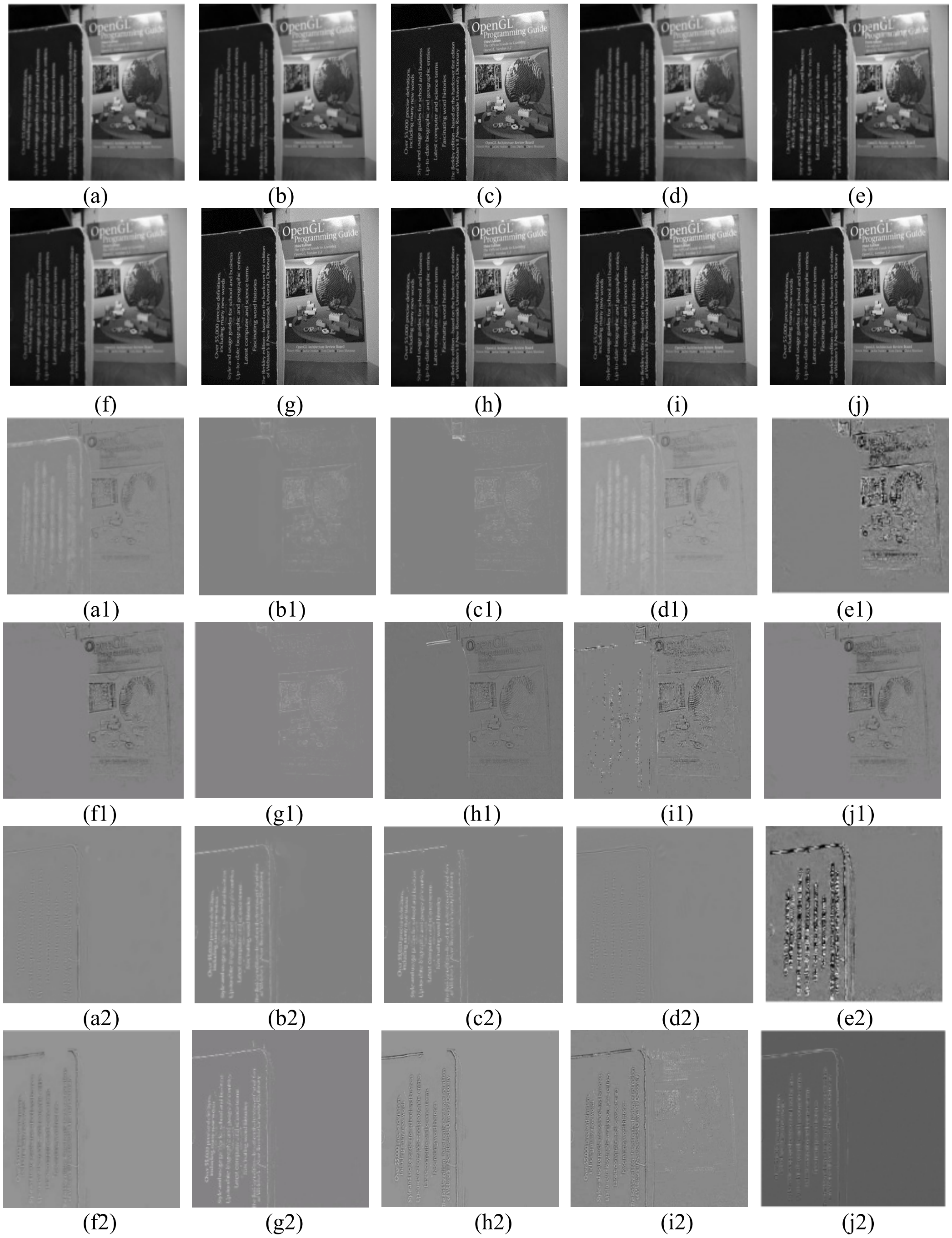


FIGURE 12. The results of the fusion of the Book images. (a-i) are fusion results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet. (a1-j1), (a2-j2) are difference images with the left focus image FIGURE 10(b) and the right focus image FIGURE 10(g), respectively.

best visual effect image, which is better applied to the image field.

TABLE 6 gives the objective evaluation indicators for all algorithms in FIGURE 14. TABLE 6 demonstrates that our method possesses the best result in the evaluated values, especially MI and Q_Y. It indicates that the proposed algorithm has great advantages in image detail information processing and overall structural composition. So, the entire fused image has better spatial continuity, and the information of each layer of the image can be more fully utilized, which make the resulting image more conducive to computer vision processing.

FIGURE 15 shows the effects of eight fusion algorithms on the fusion of "Desk" images. In FIGURE 15 (a1-c2), there are discontinuous blocks, and the overall information characteristics of the image are inconsistent, which means that it has block effects in SR, GFF and DCT. In FIGURE 15 (d1-e2), there are over smoothing phenomena at the intersection of

TABLE 4. Objective evaluation indicators of different fusion algorithms in figure 12.

Fusion methods	MI	Q_Y	Q_{PC}	$Q^{AB/F}$	Q_{CB}	Time(s)
SR	8.0287	0.9053	0.6802	0.7682	0.7425	89.2236
GFF	8.1356	0.9214	0.6825	0.8012	0.7642	0.7520
DCT	8.3562	0.9708	0.6932	0.7561	0.7514	0.2214
NSCT-PCNN	8.4631	0.9212	0.7142	0.7810	0.7521	262.8897
MWG	8.6414	0.9574	0.7385	0.7481	0.7623	55.6314
CNN	8.8608	0.9778	0.6829	0.7677	0.7724	174.3321
P-CNN	8.8610	0.9732	0.9576	0.8011	0.7779	**0.0433**
NSST-SCM	8.8527	0.9794	0.8998	**0.8089**	0.7662	58.5660
BADNN	7.9822	0.8934	0.6972	0.7152	0.7165	612.33
NSST-ResNet	**8.9882**	**0.9804**	**0.9632**	0.7914	**0.7782**	52.7787

TABLE 5. Objective evaluation indicators of different fusion algorithms in figure 13.

Fusion methods	MI	Q_Y	Q_{PC}	$Q^{AB/F}$	Q_{CB}	Time(s)
SR	8.0287	0.8941	0.7514	0.7661	0.7658	78.8852
GFF	8.3194	0.8995	0.7538	0.7914	0.7785	0.6532
DCT	8.7562	0.9538	0.7608	0.7502	0.7752	0.2148
NSCT-PCNN	8.4631	0.9026	0.7572	0.7309	0.7434	244.5211
MWG	8.2414	**0.9652**	0.7482	0.7558	0.7884	61.7741
CNN	9.1025	0.9525	0.7559	0.7824	0.7964	162.3370
P-CNN	9.2308	0.9406	0.8123	0.7877	**0.7997**	**0.0358**
NSST-SCM	8.9995	0.9617	0.8055	0.7941	0.7596	53.6652
BADNN	8.3255	0.8951	0.7602	0.7426	0.7605	3121.57
NSST-Resnet	**9.3241**	0.9358	**0.8203**	**0.7956**	0.7692	49.858

TABLE 6. Objective evaluation indicators of different fusion algorithms in figure 14.

Fusion methods	MI	Q_Y	Q_{PC}	$Q^{AB/F}$	Q_{CB}	Time(s)
SR	7.8841	0.9485	0.7289	0.7002	0.7684	74.6658
GFF	7.9852	0.9534	0.7365	0.7212	0.7752	0.6532
DCT	7.4753	0.9647	0.7548	0.7258	0.7855	0.2148
NSCT-PCNN	6.2315	0.9632	0.7485	0.6885	0.8042	184.3227
MWG	8.2141	0.9789	0.7495	0.6868	0.8074	99.8542
CNN	8.4532	0.9608	0.7037	0.7356	0.8136	162.3370
P-CNN	8.5874	0.9806	**0.7896**	0.7394	0.8149	**0.0299**
NSST-SCM	8.5441	0.9889	0.7985	**0.7423**	0.8147	63.6652
BADNN	7.9557	0.9568	0.7531	0.6982	0.7984	3141.54
NSST-Resnet	**8.5937**	**0.9891**	0.7623	0.7254	**0.8227**	57.4417

the clocks, which showing that the NSCT-PCNN and MWG cannot highlight the content presented by the image very well. In FIGURE 15(g), there is a misjudgment between the focus and the defocus area. It shows that the P-CNN algorithm has poor ability to handle details. In FIGURE 15(i2), there are clear defocus areas. It shows that the BADNN algorithm can't extract and fuse image features very well. Comparing FIGURE 15(f), (h) and (j), it can be found that the fused image produced by our algorithm looks clear and comfortable. Compared with other algorithms, the proposed algorithm can effectively preserve the overall and detail features of the image and reduce the artificial texture and fuzzy artifacts.

TABLE 7 gives objective evaluation indicators for all algorithms in FIGURE 15. TABLE 7 demonstrates that our algorithm possesses good evaluation effect on all indicators. NSST-SCM has the best evaluation value in Q_{PC} and $Q^{AB/F}$. However, the proposed algorithm is slightly smaller than the best evaluation value. Since the test image has more edges and corners, and there are different degrees of angularity at the edge of each object. So, it is a high demand for detailed processing of images. Although the proposed algorithm does not have the best results in every evaluation indicator, it has three good objective values and has satisfactory results in detail processing.

FIGURE 16 shows an enlarged view of a local of the fused images by different fusion algorithms in FIGURE 15. The upper left corner area of the different fusion result image is selected as the enlarged area in test images "Clock". It can be seen from the enlarged area that the fusion results of the local amplification regions of different methods are different. FIGURE 16 (a-c) have obvious blurring, which indicates that SR, GFF and DCT extraction source images information is incomplete. There are obvious discontinuous regions near the edge of the images in FIGURE 16 (d-f), which meaning that the image blocks are not accurately classified by the NSCT-PCNN, MWG, and CNN. FIGURE 16 (g) has severe man-made texture information around edge, which indicates that P-CNN cannot effectively perform edge feature extraction and fusion. In FIGURES 16 (h, i), it can be seen that there is blurring near the corner of the "Clock". It shows that the results of NSST-SCM and BADNN are less consistent. In the enlarged image area of FIGURE 14, it can be found that there is a partially blurred area at the junction of the clock and other objects. In the enlarged image area of FIGURE 16, it can be found that there are partially blurred areas at the

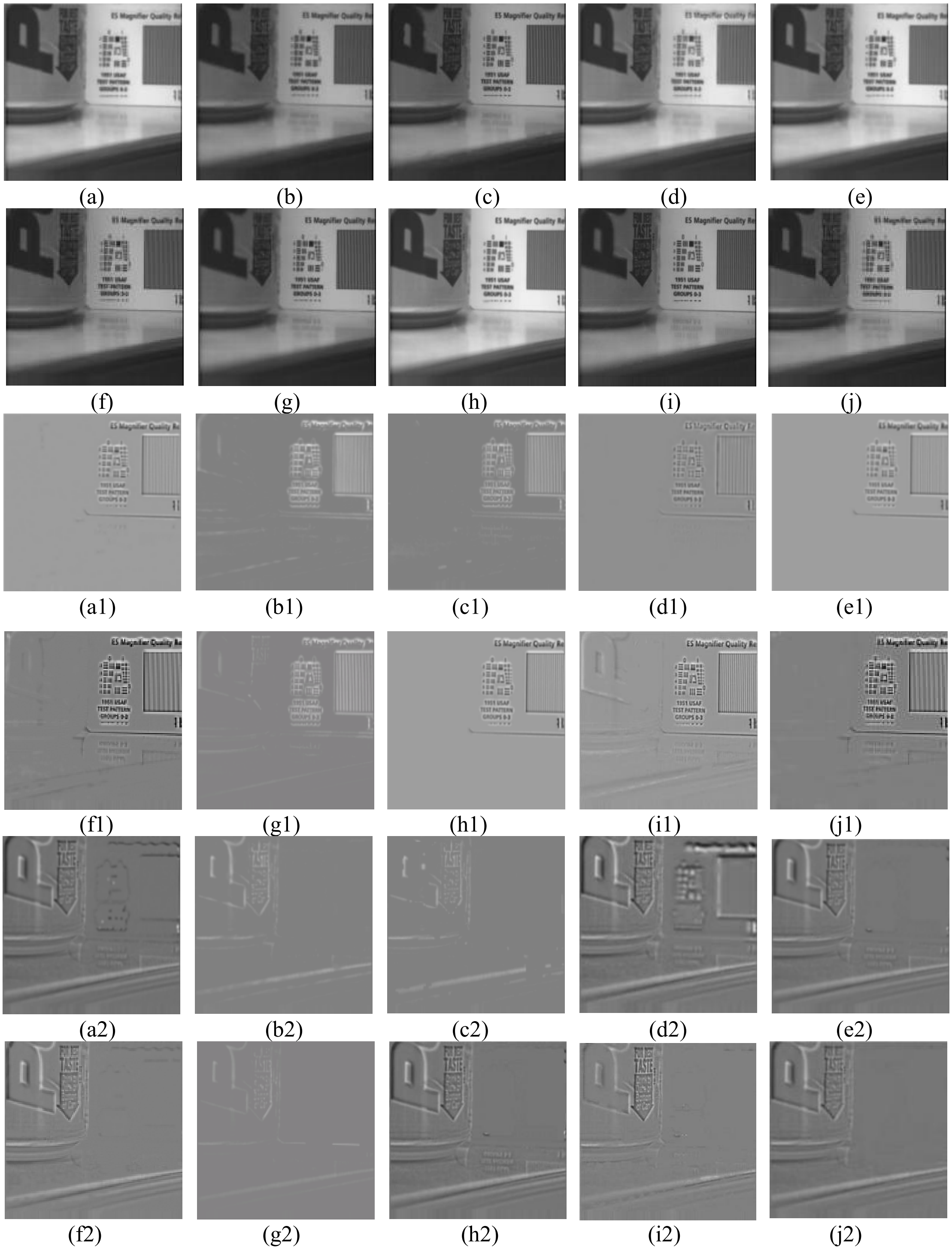


FIGURE 13. The results of the fusion of the Soda images. (a-i) are fusion results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet. (a1-j1), (a2-j2) are difference images with the left focused image FIGURE 10(c) and the right focused image FIGURE 10(h), respectively.

junction of the timepiece and other objects. It can be seen that each algorithm has some shortcomings in the processing of small features at the edge of different object edges. However, the proposed algorithm has the least degree of blur and the clearest texture information, and has better spatial continuity in the edge part.

C. COLOR IMAGE EXPERIMENT RESULTS

In FIGURE 17, five pairs of color images are presented to verify the fusion results of different algorithms. These five pairs of color images have a complex background. From different fusion results, the proposed algorithm has some advantages in the fusion image clarity. In FIGURE 17.2(a),

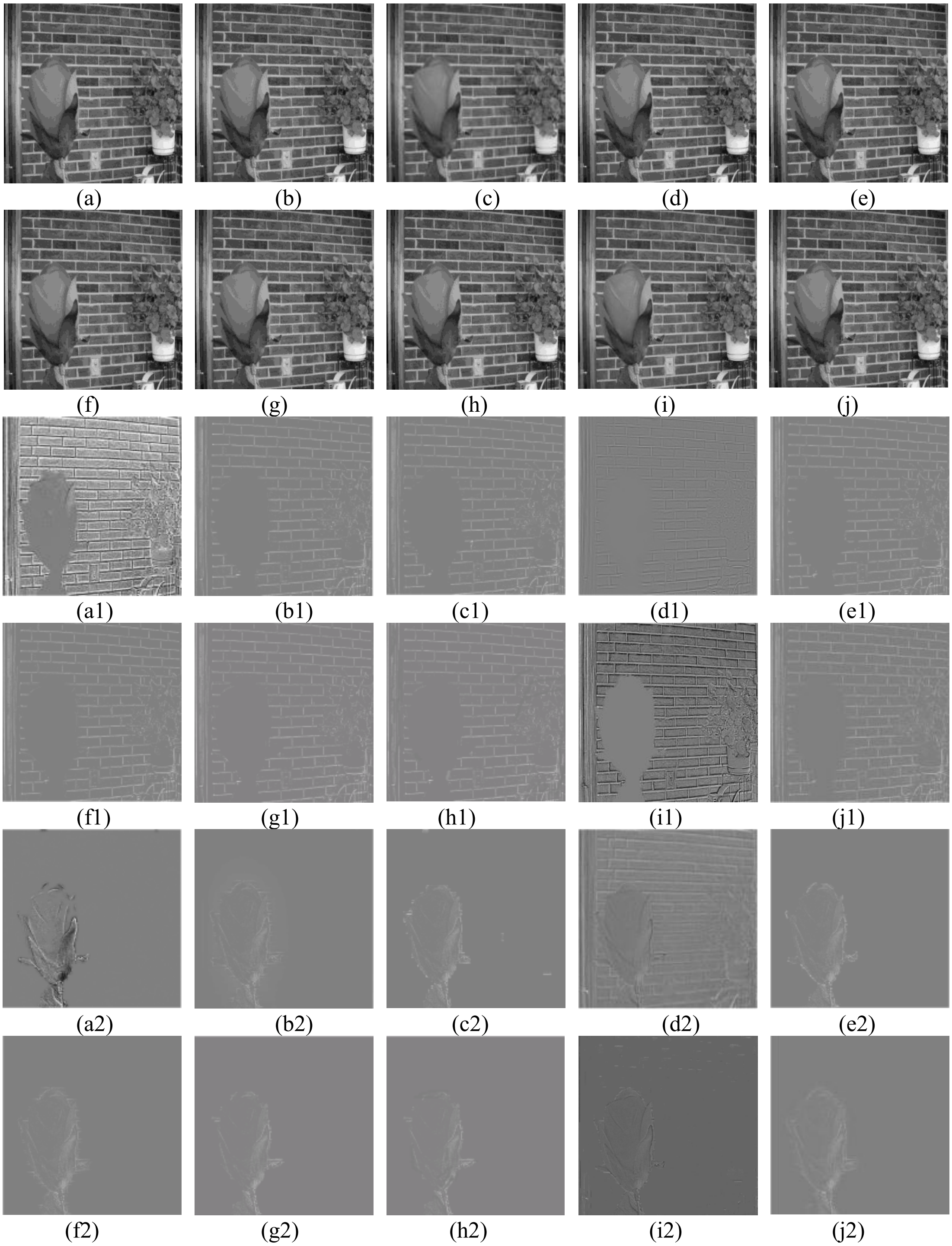
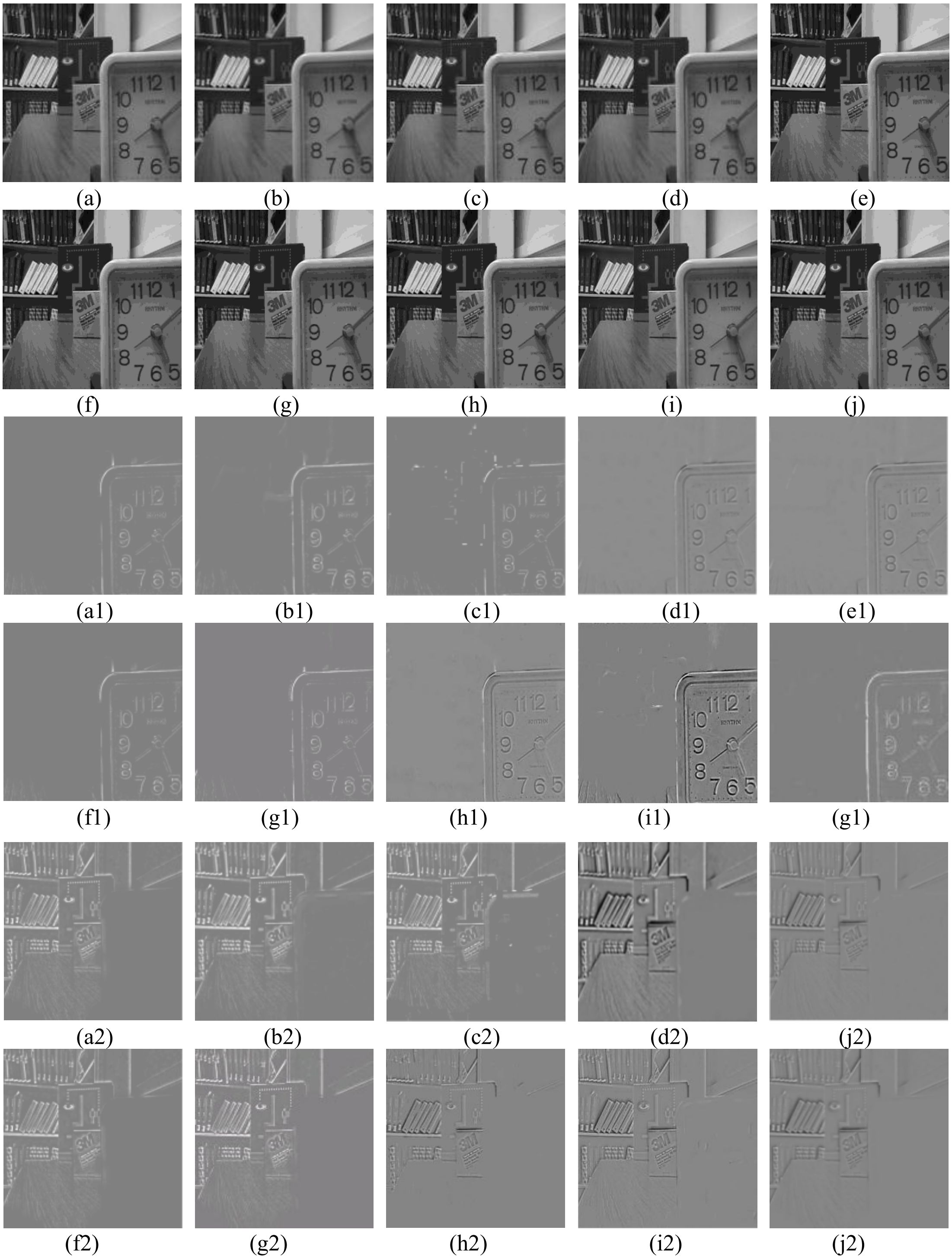


FIGURE 14. The results of the fusion of the Flower images. (a-i) are fusion results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet. (a1-j1), (a2-j2) are difference images with the left focused image FIGURE 10(d) and the right focused image FIGURE 10(i), respectively.

the "small flag" in the distance is blur, which means that SR cannot extract all the complete information, making the fusion result unclear. In FIGURES 17.4(b) and 17.5(b), artifacts appear in the "floor" and "hair" sections, which indicating that the GFF cannot preserve the sources image information well. There are small areas with discontinuities in FIGURES 17.1(c), 17.3(c), 17.2(d) and 17.3(d), which shows that DCT obtains the error selection by the scale transformation to generate the error map, or the scale transform process has errors. The decision map obtained by NSCT-PCNN through scale transform and neural network is not proficient in retaining the source images detail information features, and it also is easy to ignore small detail features. The edges are blurring in FIGURE 17.2(e), and parts of the sources image


(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(a1) (b1) (c1) (d1) (e1)

(f1) (g1) (h1) (i1) (g1)

(a2) (b2) (c2) (d2) (j2)

(f2) (g2) (h2) (i2) (j2)

FIGURE 15. The results of the fusion of the Desk images. (a-i) are fusion results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet. (a1-j1), (a2-j2) are difference images with the left focused image FIGURE 10(e) and the right focused image FIGURE 10(g), respectively.

information is lost, which shows that MWG has poor image features and details preserver ability. Connection part the of ‘‘glasses and the sea’’ in FIGURE 17.1(f) has a defocused area, and local part near the ‘‘floor’’ in FIGURE 17.5(g) has a defocused part of the defocus. They show that the fusion images generated by CNN and P-CNN also fail to correctly classify all the scattered focus regions. There are many shadows in the window of FIGURE 17.4(i), which shows that the fusion results based on NSST-SCM are clear, but the resolution is lower relative to our algorithm. The last picture of each line shows that our innovative method can better preserve the details of the source images in terms of

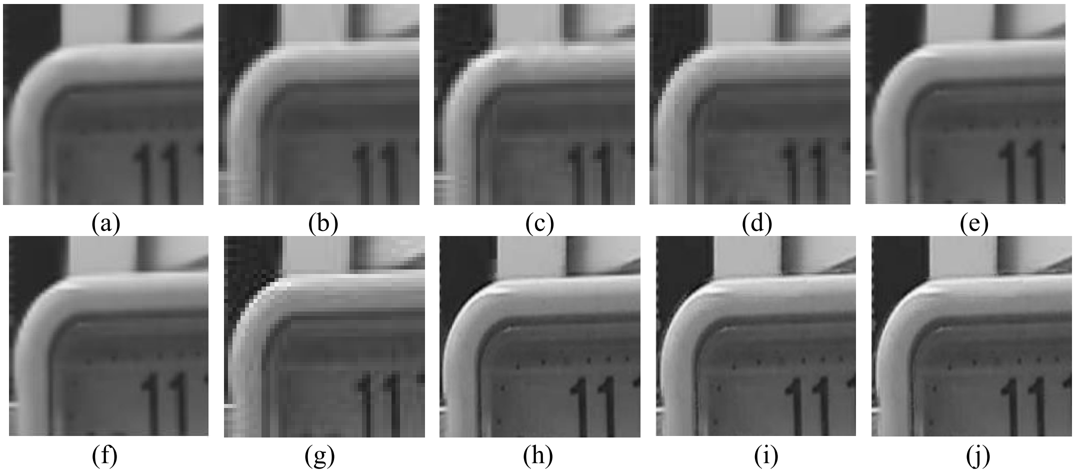


FIGURE 16. An enlarged view of the fusion results of the Desk images. (a-j) are enlarged images of the fusion results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet.

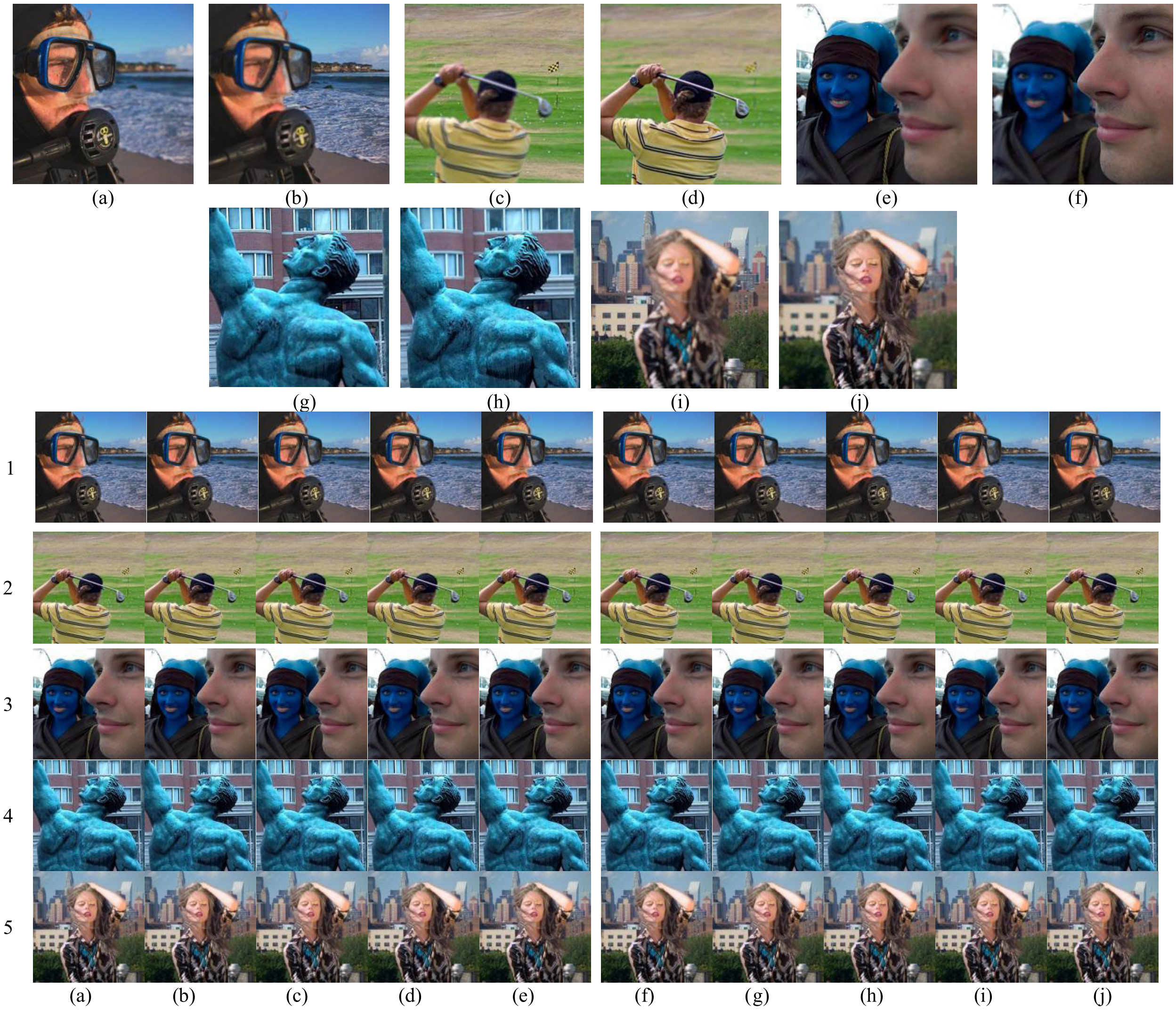


FIGURE 17. (a-j) are the original multi-focus image pairs from Lytro multi-focus dataset. The numbers (1-5) are the fusion results of the source images using different methods. From left to right are the fused results based on SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN, NSST-ResNet.

TABLE 7. Objective evaluation indicators of different fusion algorithms in figure 15.

Fusion methods	MI	Q_Y	Q_{PC}	$Q^{AB/F}$	Q_{CB}	Time(s)
SR	7.8841	0.9485	0.7289	0.7002	0.7684	74.6658
GFF	7.9852	0.9534	0.7365	0.7212	0.7752	0.6532
DCT	7.4753	0.9647	0.7548	0.7258	0.7855	0.2148
NSCT-PCNN	6.2315	0.9632	0.7485	0.6885	0.8042	184.3227
MWG	8.2141	0.9789	0.7495	0.6868	0.8074	99.8542
CNN	8.4532	0.9608	0.7037	0.7356	0.8136	162.3370
P-CNN	8.5874	0.9806	**0.7896**	0.7394	0.8149	**0.0299**
NSST-SCM	8.5441	0.9889	0.7985	**0.7423**	0.8147	63.6652
BADNN	7.9557	0.9568	0.7531	0.6982	0.7984	3141.54
NSST-Resnet	**8.5937**	**0.9891**	0.7623	0.7254	**0.8227**	57.4417

TABLE 8. Objective evaluation indicators of different fusion algorithms in figure 17.

	Fusion methods	MI	Q_Y	Q_{PC}	$Q^{AB/F}$	Q_{CB}
FIGURE 15.1	SR	7.9882	0.9238	0.8124	0.7253	0.7479
	GFF	8.7841	0.9641	0.8223	0.7559	0.7649
	DCT	8.1324	0.9774	0.8173	0.7225	0.7562
	NSCT-PCNN	8.3329	0.9726	0.8253	0.7158	0.7583
	MWG	8.6898	0.9427	0.8369	0.7542	0.7687
	CNN	8.9102	0.9528	0.7934	0.7565	0.7854
	P-CNN	**9.0128**	0.9587	0.8199	0.7461	0.7862
	NSST-SCM	8.9951	0.9782	0.8377	0.7557	0.7809
	BADNN	8.2563	0.9668	0.8134	0.7206	0.7598
	NSST-ResNet	8.7735	**0.9893**	**0.8415**	**0.7591**	**0.7905**
FIGURE 15.2	SR	7.8441	0.8698	0.7314	0.7189	0.7603
	GFF	7.9253	0.8702	0.7589	0.7208	0.7821
	DCT	8.2436	0.9774	0.7509	0.7225	0.7642
	NSCT-PCNN	8.3412	0.9713	0.7502	0.7381	0.7614
	MWG	7.4024	0.9674	0.7634	0.7497	0.7785
	CNN	7.5363	0.8837	0.7938	**0.7532**	0.7983
	P-CNN	9.0738	0.8675	**0.8052**	0.7428	**0.7989**
	NSST-CSM	8.9928	0.8996	0.7806	0.7584	0.7904
	BADNN	7.9844	0.8357	0.7308	0.7054	0.7531
	NSST-ResNet	**9.1147**	**0.9823**	0.7915	0.7334	0.7896
FIGURE 15.3	SR	7.6675	0.9328	0.7613	0.7401	0.7586
	GFF	7.6938	0.9742	0.7642	0.7463	0.7792
	DCT	7.8557	0.9756	0.7658	0.7415	0.7735
	NSCT-PCNN	7.8893	0.9789	0.7694	0.7423	0.7694
	MWG	8.1725	0.9842	0.7742	0.7438	0.7815
	CNN	8.5782	0.9867	0.7931	0.7506	**0.8142**
	P-CNN	**8.6991**	0.9884	0.8036	0.7529	0.8126
	NSST-CSM	8.4582	0.9807	0.8005	0.7487	0.7995
	BADNN	7.5674	0.8966	0.7598	0.7285	0.7804
	NSST-ResNet	8.4829	**0.9892**	**0.8147**	**0.7587**	0.8079
FIGURE 15.4	SR	8.0165	0.9438	0.7931	0.7204	0.7614
	GFF	8.1017	0.9388	0.8115	0.7235	0.7802
	DCT	8.1726	0.9374	0.8209	0.7247	0.7635
	NSCT-PCNN	8.2319	0.9584	0.8152	0.7298	0.7683
	MWG	8.2573	0.9682	**0.8329**	0.7341	0.7792
	CNN	8.9142	0.9418	0.8214	0.7529	0.8086
	P-CNN	8.9481	0.9435	0.8226	0.7498	0.8125
	NSST-SCM	8.7726	0.9480	0.8205	**0.7567**	0.8096
	BADNN	8.1874	0.9025	0.7996	0.7258	0.7754
	NSST-ResNet	**8.9735**	**0.9766**	0.8302	0.7385	**0.8239**
FIGURE 15.5	SR	8.0347	0.9438	0.7148	0.7428	0.7496
	GFF	8.0629	0.9625	0.7169	0.7496	0.7752
	DCT	8.2728	0.9762	0.7173	0.7413	0.7731
	NSCT-PCNN	8.2936	0.9743	0.7205	0.7402	0.7689
	MWG	8.3374	0.9829	0.7247	0.7531	0.7739
	CNN	8.7659	0.9846	0.7326	0.7608	0.7983
	P-CNN	8.8842	0.9854	0.7391	0.7619	0.8086
	NSST-SCM	8.8797	0.9752	0.7405	**0.7621**	**0.8245**
	BADNN	8.1447	0.9352	0.6855	0.7154	0.7142
	NSST-ResNet	**8.8907**	**0.9871**	**0.7412**	0.7539	0.8142

the overall fusion process. At the same time, the focus area is effectively used in the fusion result, so that the image has a clearer level, which is easier for human eyes to distinguish. In the objective evaluation values in TABLE 8, it can be found that the proposed algorithm still has good evaluation values for the processing of the fused image.

Based on the two commonly used objective evaluation indicators *MI* and $Q^{AB/F}$, we add three objective evaluation indexes to evaluate the quality of the fusion image more comprehensively. It can evaluate the fusion results from multiple aspects of the image. TABLE 8 is an objective evaluation of the different fusion results in FIGURE 17, and the part of the bold font is the portion with better result values. Due to the different information of different image, the evaluation results of the final results are different after different algorithms. It can be seen from TABLE 8 that the objective evaluation values of different image fusion results of different algorithms are fluctuating. In the fusion results of FIGURES 15.2, 15.3 and 15.5, the proposed algorithm is lower than the optimal result in the objective evaluation value of Q_{CB}. It is shown that in the processing of these three images, our algorithm does not deal well with the features of image level and has poor visual effects. However, in terms of integrating multiple evaluation indicators, our algorithm has the most optimal value among the objective evaluation result values used. Thus, our algorithm has good results in terms of human visual perception, source image information retention, structural information similarity or frequency domain information similarity. Therefore, the overall fusion result of the proposed algorithm is satisfactory.

D. COMPARISON OF DIFFERENT ALGORITHM DECISION MAPS

In order to further show the comparison results of different fusion methods, we give the decision maps of five different images of different fusion methods in FIGURE 18. It can be clearly seen that the boundary division and region selection of the focusing and defocusing regions in FIGURE 18. In the fusion decision map, from top to bottom are the fusion decision maps generated by different algorithms, and the last line is the fusion image generated by our algorithm. It is found from the decision maps that SR, GFF and DCT have the wrongly selected areas, which will result in the defocused areas of the final fused images. The boundary classification of NSCT-PCNN and MWG is fuzzy, and the boundary area of the image cannot be correctly distinguished. P-CNN is well divided in the boundary region, but there is a lack of an image features segmentation in the upper right corner of 7(b) of FIGURE 18. In FIGURE 18, the outline portion of 8(c)

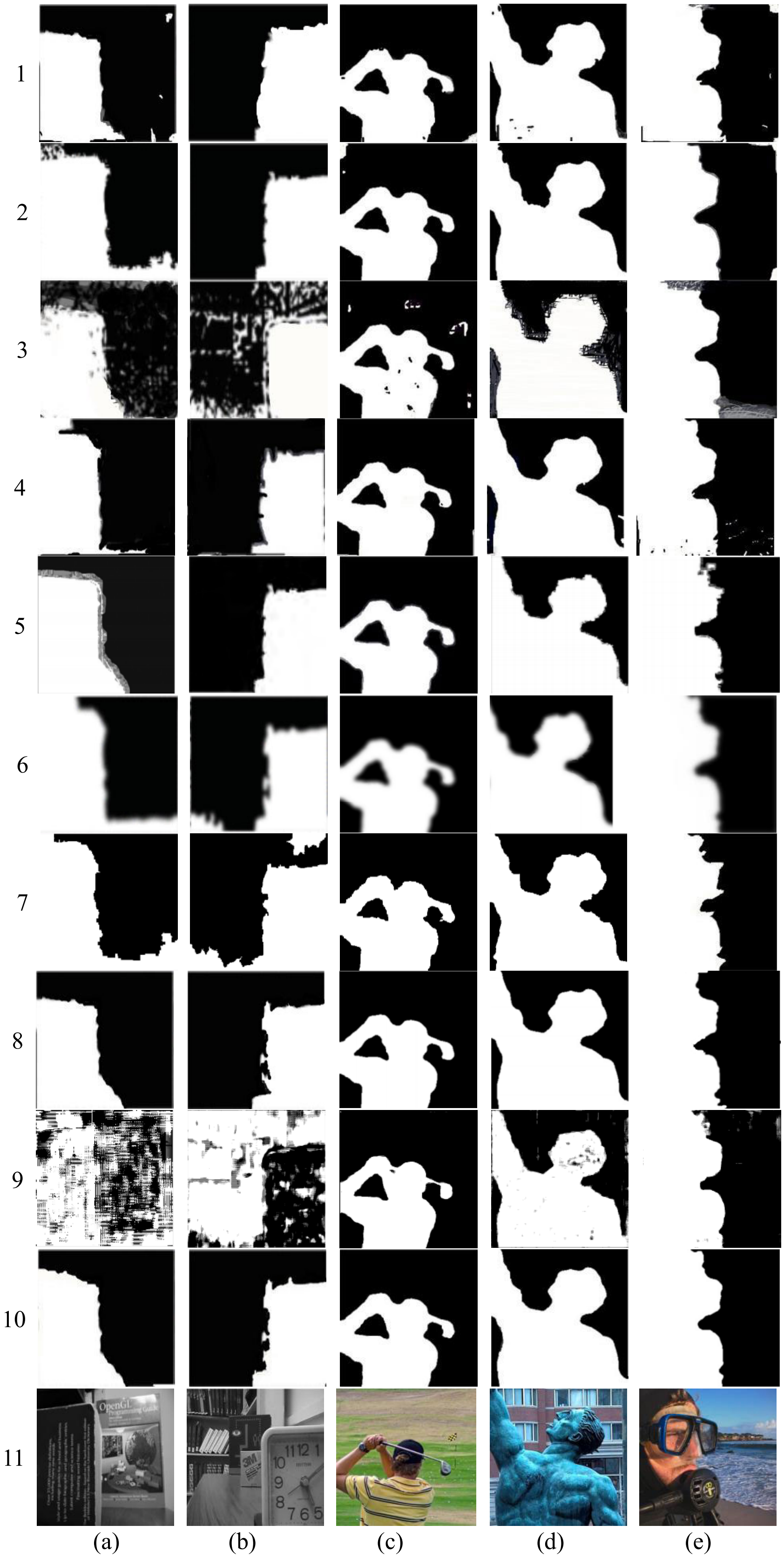


FIGURE 18. The decision map obtained by SR, GFF, DCT, NSCT-PCNN, MWG, CNN, P-CNN, NSST-SCM, BADNN,NSST-ResNet.

is shaded. In FIGURE 18.9, it can be seen that in addition to FIGURE 18.9 (c), it is possible to clearly distinguish the different focal regions of the image. The focus and defocus areas are not well differentiated in the remaining decision diagrams. The decision map of the proposed algorithm gives a clearer sense of sight compared with CNN's. As can be seen from the multiple decision maps given in the figure below, our algorithm can effectively select the focus areas of the image and form a fused image with more clear structure.

In the above experimental results, our algorithm gives an outstanding performance in both subjective visual effects and objective evaluation indicators. The residual network selected in this paper has a deep convolutional level, which can better extract the hierarchical features of the image. Compared with the other nine representative fusion algorithms, our algorithm can make full use of all the information of the image and fused them in the fused image. However, it is time-consuming and computationally inefficient in the image processing process, which is a disadvantage of the algorithm.

VI. CONCLUSION

Based on the theory of deep learning, we propose a multi-focus image algorithm by combining with NSST and ResNet. The proposed algorithm uses ResNet fusion rules based on the NSST transform domain to extract deep information of the images. The proposed algorithm fully considers the time-frequency excellent characteristics of NSST. At the same time, ResNet is used to extract and retain the information of the source images for the low frequency images containing the global information, and IGSML is used to process the high frequency image containing the detailed information. Furthermore, the spatial continuity of the image is better improved, and the fusion result is more in line with the visual nervous system. Combining the above experimental results, the fused image produced by our method contains more clear overall information features and details. However, the network model in the running process is relatively time consuming. So, saving time is one of the directions for future research.

REFERENCES

[1] C.-P. Tu, J.-S. Du, K.-H. Du, and B.-S. Yi, "Multi-focus image fusion algorithm based on the anisotropic thermal diffusion equation," *Acta Electronica Sinica*, vol. 43, no. 6, pp. 1192–1199, Jun. 2015.

[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] J. H. Tan, H. Fujita, S. Sivaprasad, S. Sivaprasad, S. V. Bhandary, A. K. Rao, K. C. Chua, and U. R. Acharya, "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Inf. Sci.*, vol. 420, pp. 66–76, Dec. 2017.

[4] Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus images," *Inf. Fusion*, vol. 20, pp. 60–72, Nov. 2014.

[5] S. Liu, M. Shi, Z. Zhu, and J. Zhao, "Image fusion based on complex-shearlet domain with guided filtering," *Multidimensional Syst. Signal Process.*, vol. 28, no. 1, pp. 207–224, Jan. 2017.

[6] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.

[7] Y. Yang, S. Huang, and J. Gao, "Multi-focus image fusion using an effective discrete wavelet transform based algorithm," *Meas. Sci. Rev.*, vol. 14, no. 2, pp. 102–108, 2014.

[8] Y. Tian, J. Luo, W. Zhang, T. Jia, A. Wang, and L. Li, "Multifocus image fusion in Q-Shift DTCWT domain using various fusion rules," *Math. Problems Eng.*, vol. 2016, Sep. 2016, Art. no. 5637306.

[9] Y. Dongsheng, H. Shaohai, L. Shuaiqi, M. Xiaole, and S. Yuchao, "Multi-focus image fusion based on block matching in 3D transform domain," *J. Syst. Eng. Electron.*, vol. 29, no. 2, pp. 415–428, Apr. 2018.

[10] L. Cao, L. Jin, H. Tao, G. Li, Z. Zhuang, and Y. Zhang, "Multi-focus image fusion based on spatial frequency in discrete cosine transform domain," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 220–224, Feb. 2015.

[11] X. Luo, Z. Zhang, C. Zhang, and X. Wu, "Multi-focus image fusion using HOSVD and edge intensity," *J. Vis. Commun. Image Represent.*, vol. 45, pp. 46–61, May 2017.

[12] 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, Jun. 2017.

[13] Y. Yang, M. Ding, S. Huang, Y. Que, W. Wan, M. Yang, and J. Sun, "Multi-focus image fusion via clustering PCA based joint dictionary learning," *IEEE Access*, vol. 5, pp. 16985–16997, 2017.

[14] S. Liu, J. Zhao, and M. Shi, "Medical image fusion based on improved sum-modified-laplacian," *Int. J. Imag. Syst. Technol.*, vol. 25, no. 3, pp. 206–212, Sep. 2015.

[15] S. Liu, J. Wang, Y. Lu, H. Li, J. Zhao, and Z. Zhu, "Multi-focus image fusion based on adaptive dual-channel spiking cortical model in non-subsampled shearlet domain," *IEEE ACCESS*, vol. 7, pp. 56367–56388, 2019.

[16] X. Bai, M. Liu, Z. Chen, P. Wang, and Y. Zhang, "Multi-focus image fusion through gradient-based decision map construction and mathematical morphology," *IEEE Access*, vol. 4, pp. 4749–4760, 2016.

[17] M. Nejati, S. Samavi, N. Karimi, S. R. Soroushmehr, S. Shirani, I. Roosta, and K. Najarian, "Surface area-based focus criterion for multi-focus image fusion," *Inf. Fusion*, vol. 36, pp. 284–295, Jul. 2017.

[18] H. Li, H. Qiu, Z. Yu, and B. Li, "Multifocus image fusion via fixed window technique of multiscale images and non-local means filtering," *Signal Process.*, vol. 138, pp. 71–85, Sep. 2017.

[19] B. Zhang, X. Lu, H. Pei, H. Liu, Y. Zhao, and W. Zhou, "Multi-focus image fusion algorithm based on focused region extraction," *Neurocomputing*, vol. 174, pp. 733–748, Jan. 2016.

[20] S. Liu, Y. Lu, J. Wang, S. Hu, J. Zhao, and Z. Zhu, "A new focus evaluation operator based on max-min filter and its application in high quality multi-focus image fusion," in *Multidimensional Systems and Signal Processing*. Dordrecht, The Netherlands, Springer, 2019, pp. 1–22.

[21] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multiple visual features measurement with gradient domain guided filtering for multisensor image fusion," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 4, pp. 691–703, Apr. 2017.

[22] J. Markoff, "Scientists See Promise in Deep-Learning Program," *The New York Time*, pp. 11–23, 2012.

[23] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[24] C. Du and S. Gao, "Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network," *IEEE Access*, vol. 5, pp. 15750–15761, 2017.

[25] H. Tang, B. Xiao, W. Li, and G. Wang, "Pixel convolutional neural network for Multi-Focus image fusion," *Inf. Sci.*, vols. 433–434, pp. 125–141, Apr. 2017.

[26] Y. Yang, Z. Nie, S. Huang, P. Lin, and J. Wu, "Multilevel features convolutional neural network for multifocus image fusion," *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 262–273, Jun. 2019.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, Jun. 2016, pp. 770–778.

[28] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM J. Math. Anal.*, vol. 39, no. 1, pp. 298–318, 2008.

[29] X. Ma, S. Liu, S. Hu, P. Geng, M. Liu, and J. Zhao, "SAR image edge detection via sparse representation," *Soft Comput.*, vol. 22, no. 8, pp. 2507–2515, 2018.

[30] H. Hermessi, O. Mourali, and E. Zagrouba, "Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain," *Neural Comput. Appl.*, vol. 30, no. 7, pp. 2029–2045, Oct. 2018.

[31] W. Q. Lim, "The discrete shearlet transform: A new directional transform and compactly supported shearlet frames," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1166–1180, May 2010.

[32] S. Liu, S. Hu, and Y. Xiao, "Image separation using wavelets-complex shearlets dictionary," *J. Syst. Eng. Electron.*, vol. 25, no. 2, pp. 314–321, 2014.

[33] P. Shivakumara, D. Tang, M. Asadzadehkaljahi, T. Lu, U. Pal, and M. H. Anisi, "CNN-RNN based method for license plate recognition," *CAAI Trans. Intell. Technol.*, vol. 3, no. 3, pp. 169–175, 2018.

[34] S. Liu, T. Liu, L. Gao, H. Li, Q. Hu, J. Zhao, and C. Wang, "Convolutional neural network and guided filtering for SAR image denoising," *Remote Sens.*, vol. 11, no. 6, pp. 702–720, Mar. 2019.

[35] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[36] E. A. Smirnov, D. M. Timoshenko, and S. N. Andrianov, "Comparison of regularization methods for ImageNet classification with deep convolutional neural networks," *AASRI Procedia*, vol. 6, pp. 89–94, Jan. 2014

[37] H. Li, X. J. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infr. Phys. Technol.*, vol. 102, Nov. 2019, Art. no. 103039.

[38] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2014, *arXiv:1312.6229*. [Online]. Available: https://arxiv.org/abs/1312.6229

[39] S. Farfade, M. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, Jun. 2015, pp. 643–650.

[40] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[41] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2705–2710.

[42] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.

[43] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[44] M. Amin-Naji and A. Aghagolzadeh, "Multi-focus image fusion in DCT domain using variance and energy of Laplacian and correlation coefficient for visual sensor networks," *J. AI Data Mining*, vol. 6, no. 2, pp. 233–250, 2018.

[45] X.-B. Qu, J.-W. Yan, H.-Z. Xiao, and Z.-Q. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Autom. Sin.*, vol. 34, no. 12, pp. 1508–1514, 2008.

[46] H. Ma, J. Zhang, S. J. Liu, and Q. Liao, "Boundary aware multi-focus image fusion using deep neural network," 2019, *arXiv:1904.00198*. [Online]. Available: https://arxiv.org/abs/1904.00198

[47] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'Information measure for performance of image fusion,'" *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.

[48] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

[49] J. Zhao, R. Laganiere, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *Int. J. Innov. Comput. Inf. Control*, vol. 3, no. 6, pp. 1433–1447, 2007.

[50] X. B. Qu, J. W. Yan, and G. D. Yang, "Multifocus image fusion method of sharp frequency localized Contourlet transform domain based on sum-modified-laplacian," *Opt. Precis. Eng.*, vol. 17, no. 5, pp. 1203–1212, May 2009.

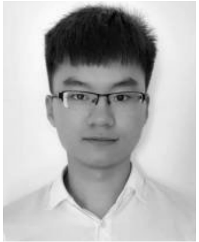
SHUAIQI LIU received the B.S. degree from the Department of Information and Computer Science, Shandong University of Science and Technology, in 2009, and the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University, in 2014. He was a Visiting Scholar with Ottawa University, from August 2016 to January 2017. He is currently an Associate Professor with the College of Electronic and Information Engineering, Hebei University. His current research interests include image processing and signal processing.

JIE WANG received the B.S. degree from the School of Industry and Commerce, Hebei University, in 2017, where she is currently pursuing the B.Eng. degree with the College of Electronic and Information Engineering. Her current research interest includes image processing.

YUCONG LU received the B.S. degree from the College of Electronic and Information Engineering, Hebei University, in 2016, where she is currently pursuing the B.Eng. degree with the College of Electronic and Information Engineering. Her current research interests include the image denoising, software development, and image fusion.

SHAOHAI HU was born in 1964. He received the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University, in 1991, where he has been a Professor, since 2008. His current research interests include signal processing and information fusion, especially image fusion, image denosing, and sparse representation.

XIAOLE MA was born in 1991. She received the B.S. degree in communication engineering from the Department of Electronic Information Engineering, Hebei University. She is currently pursuing the Ph.D. degree in signal and information processing with the Institute of Information Science, Beijing Jiaotong University. Her current research interests include the image de-noising, image fusion, and signal processing.



YIFEI WU received the B.S. degree from Xidian University, Xi'an, China, in 2018. He is currently pursuing the M.S. degree with the University of California at San Diego, San Diego, USA. His current research interests include signal and image processing.