IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 24, 2020, accepted April 30, 2020, date of publication May 8, 2020, date of current version May 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993404

Multi-Focus Color Image Fusion Algorithm Based on Super-Resolution Reconstruction and Focused Area Detection

SHUAIQI LIU[1,2], JIAN MA[1,2], LU YIN[1,2], HAILIANG LI[3], SHUAI CONG[4], XIAOLE MA[5], AND SHAOHAI HU[5]

[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]Hong Kong Applied Science and Technology Research Institute Company Limited (ASTRI), Hong Kong
[4]Industrial and Commercial College, Hebei University, Baoding 071000, China
[5]Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

Corresponding authors: Shuai Cong (congshuai@hbu.edu.cn) and Xiaole Ma (maxiaole@bjtu.edu.cn)

This work was supported in part by Natural Science Foundation of China under grant 61401308 and 61572063, Natural Science Foundation of Hebei Province under grant F2016201142, F2019201151, F2020201025 and F2018210148, Science research project of Hebei Province under grant BJ2020004,QN2016085 and QN2017306, Opening Foundation of Machine vision Engineering Research Center of Hebei Province under Grant 2018HBMV01 and 2018HBMV02, Natural Science Foundation of Hebei University under Grant 2014-303 and 8012605, Social science project of Hebei University under Grant 202002. This work was also supported by the High-Performance Computing Center of Hebei University.

ABSTRACT Multi-focus image fusion is an image processing that generates an integrated image by merging multiple images from different focus area in the same scene. For most fusion methods, the detection of the focus area is a critical step. In this paper, we propose a multi-focus image fusion algorithm based on a dual convolutional neural network (DualCNN), in which the focus area is detected from super-resolved images. Firstly, the source image is input into a DualCNN to restore the details and structure from its super-resolved image, as well as to improve the contrast of the source image. Secondly, the bilateral filter is used to reduce noise on the fused image, and the guided filter is used to detect the focus area of the image and refine the decision map. Finally, the fused image is obtained by weighting the source image according to the decision map. Experimental results show that our algorithm can well retain image details and maintain spatial consistency. Compared with existing methods in multiple groups of experiments, our algorithm can achieve better visual perception according to subjective evaluation and objective indexes.

INDEX TERMS Multi-focus image fusion, super-resolution, dual convolutional neural network, focus area detection.

I. INTRODUCTION

Due to the limitation of the depth of field in the optical lens, objects at different distances in the same scene cannot be fully focused by cameras. The area within the depth of field is usually a sharp focus area, while the area outside the depth of field is usually a blurry defocus area [1]. Multi-focus image fusion technology is used to extract different focus areas from multiple images in the same scene to synthesize a clear image. The fusion process can also be regarded to improve the image quality on image processing [2], since clear all-focus images are more suitable for human visual perception systems and post-processing.

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

Generally, the existing image fusion algorithms can be divided into pixel-level fusion, feature-level fusion and decision-level fusion [3]. Pixel-level image fusion directly processes the single pixel of the images to generate the fusion result. Feature-level image fusion works in a way by extracting the edge texture and structure information in an image. The region-based image fusion algorithm is an important method of feature-level fusion [4]. Decision-level image fusion makes use of the feature information from the feature-level image according to the specific requirements and then makes the optimal decision directly according to certain criteria and the credibility of each decision [5]. Pixel-level image fusion is the fundamental fusion of the three levels, on which our works focus on this paper.

Pixel-level image fusion algorithms are generally divided into three categories: methods based on transform domain,

based on spatial domain and based on sparse representation [6]. The methods based on the transform domain are generally implemented by three steps: image multi-scale decomposition, fusion of the coefficients generated by the transformation, and multi-scale reconstruction based on the fused coefficients. Among the representative algorithms is image fusion based on Laplace pyramid transform [7], discrete wavelet transform [8], curvelet transform [9], contourlet transform [10], and shearlet transform [11] and so on. Spatial domain-based methods include image fusion based on: max-min filtering [12], image block matching [13], guided filter [14], dense-scale feature-invariant methods [15] and multi-focus image fusion algorithms based on boundary discovery [16]. Methods based on sparse representation include multi-focus image fusion based on sparse representation [17], convolution sparse representation [18], dictionary sparse representation [19], multi-scale transform and sparse representation [20], and so on. Although these methods can improve the fusion quality of the image to a certain extent, they still are not able to obtain the optimal fusion result due to the artifacts caused by low contrast and block effect.

In recent years, with the significant improvement in computing power, convolutional neural networks (CNN) have been widely used in image fusion, image segmentation, image classification, image recognition, image denoising and other fields. CNNs demonstrate their powerful automatic feature learning capability by designing a multi-layer network structure [21]–[25]. As an end-to-end model, CNN can deeply learn its features of multiple levels by setting with different levels. A shallow network has a convolutional layer with a small perceptual domain and can only learn features of some local areas, such as the structure of an image. A deeper network has a larger convolution perception layer and can learn more abstract features, such as image detail textures. Liu *et al.* [21] are the pioneers to bring convolutional neural networks into image fusion. They designed a dual network with two-channel weights to classify the focused and defocused pixels of the source image, then obtain the initial decision map, later segment the image and optimize it with post-processing algorithms, such as consistency verification. Tang *et al.* [26] proposed a pixel-level convolutional neural network to identify focused and unfocused pixels. They also created a multi-focus image training set to simulate different focus situations through masks with 12 geometric models. Mustafa *et al.* [27] proposed a multi-focus image fusion algorithm with a multi-scale convolutional neural network framework. Except for the basic convolutional layer, they introduced a set of 3 convolutional layers with different sizes of convolution kernels and added jump connections between convolutional layers to extract more features in multi-focus images. Ma *et al.* [28] proposed a boundary sensing method based on the residual network, which extracts features directly from the source image through a dual network and refined the decision map by post-processing.

Because multi-focus image fusion can effectively synthesize focused areas in multiple images in the same scene,

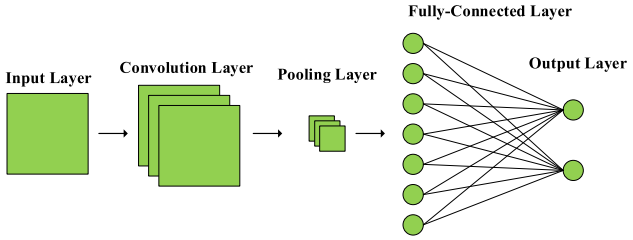

FIGURE 1. Block diagram of a convolutional neural network.

the quality of the fused image depends largely on the successful detection of the focus area of the image. Authors in [29] proposed a multi-scale weighted gradient-based multi-focus image fusion algorithm. This algorithm used a large scale to solve the fusion problem caused by image anisotropy blur and registration error, and small scale to determine the focus area boundary. In [30], Ma *et al.* proposed a two-scale multi-focus image fusion algorithm based on an enhanced random walk algorithm. This method uses a random walk algorithm on two-scale to detect focus areas. Although this method can improve the accuracy of detection on focused and defocused areas, when processing multi-focus images with complex backgrounds, a small number of areas with incorrect recognition still appear. In [15], the authors proposed a multi-focus image fusion algorithm based on dense scale-invariant feature transformation. This algorithm uses a sliding window to extract the local feature of an image and detect focused areas of the image. Although this method can improve the fusion quality, artificial artifacts may still generate. In order to effectively eliminate noise and reduce artificial textures, in [31], the authors bring guided filter for the focus area detection. Because the guided filter can maintain the structure and edges of an image, this fusion algorithm proposed in [31] can effectively enhance image details. So we also employ a guided filter for focus area detection in this work.

In the detection of the focus area, the above methods usually use the original images to detect the focus area, which often leads to a large boundary error of the focus area. It is more serious for color image fusion. In order to better detect the focus area, many scholars try to carry out super-resolution for each source image to recover the lost image details during image acquisition, and then detect the focus area, so as to obtain better image fusion effect. In [32], Aymaz *et al.* firstly introduced the image super-resolution algorithm into multi-focus image fusion. They used the bi-cubic interpolation based super-resolution algorithm to retain more details of the original image in image fusion. In [33], Yang *et al.* proposed a multi-resolution image fusion algorithm based on super-resolution via convolutional neural network, which uses an end-to-end CNN mapping to improve the quality of the source image. In general, image super-resolution algorithms can be implemented by two types of methods, that is, simple interpolation methods and neural network based methods. These simple interpolation methods can be found in the industry: nearest neighbor

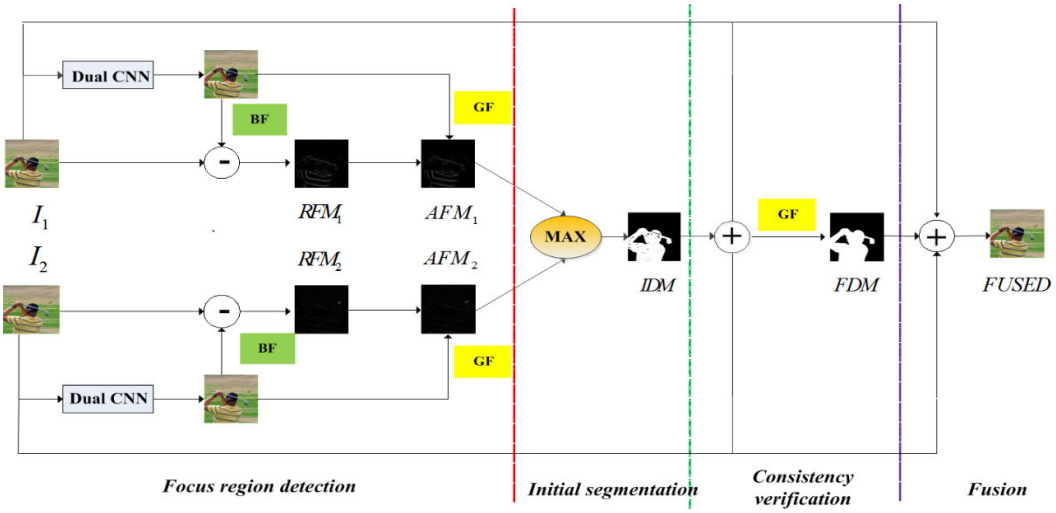

FIGURE 2. Flow chart of the proposed fusion method.

interpolation, bi-cubic interpolation and bilinear interpolation [32]. Bi-cubic interpolation, for example, uses the relationship between adjacent pixels to do interpolation. This method is particularly prominent in retaining edge details, but this linear model is still insufficient to restore high-frequency details. The method based on neural network is to learn non-linear feature mapping of the low-resolution input image to the high-resolution target image, which can recover more details at high frequencies. With the development of the convolutional neural network, CNN is widely used in image super-resolution analysis. In [34], the authors proposed a deep learning method for single image super-resolution (named as SRCNN). SRCNN has a lightweight structure, which can achieve fast speed for practical on-line use. However, the structure of SRCNN is relatively simple, only three convolutional layers are used for feature extraction, which will lose some image details. In [35], in order to extract more image detailed information, a 20-layer deep neural residual network was introduced into super-resolution image reconstruction. The introduction of residual structure not only improves the convergence of model training speed, but also brings a deep network structure with a wider receptive field. Although residual learning can better retain the high-frequency details, it still brothers the low-frequency errors in the structure. In order to better restore the image, we use the DualCNN in [36] for image super-resolution reconstruction. The DualCNN network is divided into two channels, the deep-level network is used to estimate the details, and the shallow-level network is used to estimate the structure. By learning the non-linear feature mapping between input and output, a high-quality image is finally generated.

In multi-focus image fusion, the resolution of the source images has a great influence on the quality of the fused image. Low-resolution images lead to poor fusion effect in the image fusion processing. The super-resolution reconstruction can not only make the image contain more image information, but also improve the quality of the image. Simultaneously, the performance of multi-focus image fusion is affected by the accuracy of image focus area detection. The higher the sharpness of the image is, the more accurately the focus area can be identified, which can improve the fusion effect. Therefore, we use a super-resolution image to improve the quality of source images. Color images contain rich colors and texture information. However, the defocus areas in color images have poor edge and texture information. Through super-resolution reconstruction, we can restore the edge and texture information in defocus area of color images, and can obtain better fusion performance.

To get better color fusion image, we propose a color image fusion algorithm based on image super-resolution reconstruction. Firstly, we use DualCNN to reconstruct the super-resolution of color images. Secondly, in order to suppress the noise generated by super-resolution, we introduce bilateral filtering to denoise the reconstructed super-resolution images and optimize the image edges. Then, the difference images are generated by the super-resolution reconstruction images and the source images. From the initial focus area detected by combining with the above difference images and guide filter, we can obtain the initial decision map. Later, the final decision map is obtained by small area removal strategy and refined by the guide filter, and the fusion image is further obtained. The main contributions of the proposed method can be summarized as follows.

(1) Super-resolution reconstruction based on DualCNN is applied to enhance the detailed information and improves the image resolution of the color images, which can improve the clarity of the color images and can increase the accuracy in the recognition of the focus area to generate better fused images.

(2) The bilateral filter is used for suppressing the noise in super-resolution image and guided filtering is used for enhancing the image details in the focus area detection, which can maintain the edge structure and the spatial consistency

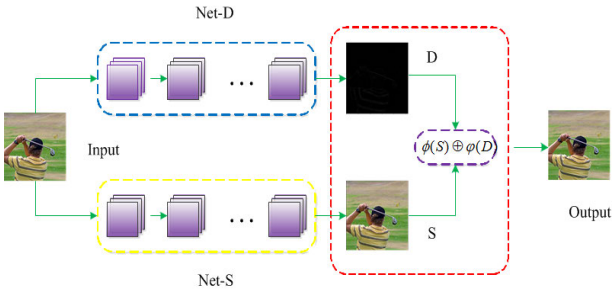

FIGURE 3. The dual convolutional neural network model.

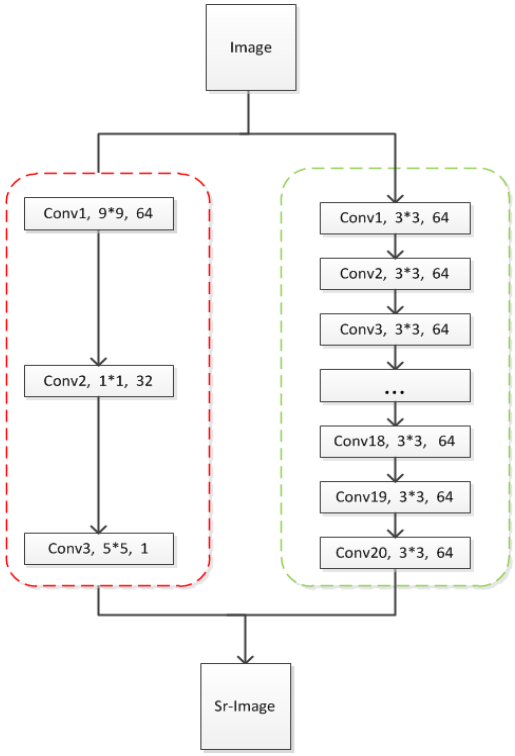

FIGURE 4. Network architecture and parameters of each layer of the DualCNN.

of the images. Compared with other algorithms, the artificial artifacts generated after fusion can be effectively avoided by the proposed method. Our algorithm is very efficient and suitable for big data applications.

The rest of this article is arranged as follows: In section 2, we mainly introduce some related work used in our algorithm. In section 3, we illustrate the framework of our algorithm. In section 4, we compare the proposed algorithm and other algorithms in color multi-focus images and gray-scale multi-focus images separately. Finally, we give some conclusions in section 5.

II. RELATE WORK

A. CONVOLUTIONAL NEURAL NETWORK

Generally, the framework of CNN is divided into 3 parts, namely the convolutional layer, the pooling layer and the fully connected layer [37]. A classic convolutional neural network framework is shown in FIGURE 1. Among all the layers, the convolution layer is the core layer for designing a neural network, which uses the convolution kernel to extract different levels of features from an image.

CNN convert the input image into a feature map by using a set of filters in the convolution layer and transform this set of feature maps into another set of feature maps by all convolution layers. This process can be described as follows.

$$y^j = \max(0, \sum\nolimits_i k^{ij} * x^i + b^j), \tag{1}$$

where x^i is the $i-th$ feature map of the input, y^j is the $j-th$ feature map of the output, k^{ij} is the convolution kernel of x^i and x^j, b^j is the bias, and $*$ is the convolution operator.

Conventionally, super-resolution can be processed by using simple nearest-neighbor interpolation or bi-cubic interpolation, but artificial noise often appears in the recovered images. Due to the advantages of CNN in feature extraction, we choose it for super-resolution reconstruction.

B. BILATERAL FILTER AND GUIDED FILTER

The bilateral filter is a kind of widely used non-linear filter, which can effectively reduce the superimposed noise while maintaining the edge details of the image [38]. The output pixel value is obtained by the weighted combination of neighboring pixels, as shown in formula (2):

$$BF(I)_p = \frac{1}{w_p} \sum\nolimits_{q \in N(p)} G_{\sigma_s}(\|p-q\|) G_{\sigma_r}(|I_p - I_q|) I_q, \tag{2}$$

where I_p is the input image, $BF(I)_p$ is the filtered image, p is the pixel points in the image, and $N(p)$ is the neighborhood pixels near the pixel p whose size is $(2N+1) \times (2N+1)$. G_{σ_s} and G_{σ_r} represent the weights of the spatial domain according to the geometric space distance and the intensity domain according to the pixel difference, respectively. σ_s is the spatial proximity factor, σ_r is the intensity similarity factor, G_{σ_s} and G_{σ_r} are two factors that affect the filtering effect, G_{σ_s} and G_{σ_r} are calculated as shown in formula (3-4).

$$G_{\sigma_s}(\|p-q\|) = \exp(-\frac{(p_x - q_x)^2 + (p_y - q_y)^2}{2\sigma_s^2}), \tag{3}$$

$$G_{\sigma_r}(\|p-q\|) = \exp(-\frac{(I_p - I_q)^2}{2\sigma_r^2}), \tag{4}$$

where W_p is the product of the weight coefficients in the spatial domain and the intensity domain, which can be calculated as formula (5).

$$W_p = \sum\nolimits_{q \in N(p)} G_{\sigma_s}(\|p-q\|) G_{\sigma_r}(|I_p - I_q|), \tag{5}$$

In general, the larger the value of σ_s, the more blur the flat area of the image, and the larger the value of σ_r, the more blur the edge area of the image. So, the effect of bilateral filter is controlled by σ_s and σ_r. Because the fusion effect of multi-focus images is usually disturbed by noise, we use bilateral filter to denoise the multi-focus images after super-resolution, as well as maintain good edge details. In order

FIGURE 5. Lytro color multi-focus image fusion test data. (a1)-(a8) are the left-focused images, and (b1)-(b8) are the right-focused image.

to obtain higher quality images, σ_s and σ_r are set to 3 and 0.6 respectively. At the same time, we set N in the neighborhood size of the pixel to 3.

The guided filter is a linear filter, which can better retain the details of the image and avoid the computation depending on the size of the filter [14], [39]–[41]. Assuming that the output image is q and the guide image is I, the local linear relationship between the output image q and the guide image I can be described as in formula (6).

$$q_i = \alpha_k I_i + b_k, \quad \forall i \in \omega_k, \tag{6}$$

where i and k are the indices of the pixels, and a, b are the coefficients of the linear function when the neighborhood center is at k. ω_k is a local window of size $(2r+1) \times (2r+1)$ with the pixel k as the center.

The next step is to solve the coefficients of the linear function by minimizing the error between the output value q and the filtering image p, that is:

$$E(\alpha_k, b_k) = \sum\nolimits_{i \in \omega_k} ((\alpha_k I_i + b_k - p_i)^2 + \varepsilon \alpha_k^2), \tag{7}$$

where ε is a regularization parameter, α_k and b_k can be calculated by linear regression, as:

$$\alpha_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \varepsilon}, \tag{8}$$

$$b_k = \bar{p}_k - \alpha_k \mu_k, \tag{9}$$

where μ_k and σ_k^2 are the mean and variance of the input image I in the local window ω_k, and $|\omega|$ is the number of pixels in the local window ω_k. $\bar{p}_k$ represents the mean in the local window ω_k.

In our experiments, the guided filter process can be represented by $G_{r,\varepsilon}(p, I)$, which is determined by the parameter r and the regularization parameter ε, where we set r and ε to 3 and 0.3, respectively.

III. PROPOSED ALGORITHM

In order to overcome the problem of inaccurate detection of focus area caused by low resolution in color image fusion, we propose a color image fusion algorithm based on super-resolution reconstruction.

The steps of the algorithm proposed in this paper are shown in FIGURE 2. As for two source images, firstly, DualCNN is used to recover the super-resolution source images from the two components of structure and detail. Since the super-resolution images usually contain slight noise, the bilateral filter is employed for denoising. After filtering, the difference between the super-resolution image and the source image is used to obtain an initial difference image, and the initial difference image is refined by the guided filter to obtain a refined difference image. Secondly, we use pixel-max rules to obtain the initial decision map (IDM) from the refined difference image. Then, the small region removal strategy and the guided filter is used to optimize the IDM to get the final decision map (FDM). Finally, the final fusion image is obtained as a summed image from the source images weighted by the FDM.

Ideally, the focus area has more high-frequencies than the defocus area, so focus area detection is an essential process in image fusion. Our algorithm mainly includes the following steps: image super-resolution, focus area detection and image fusion.

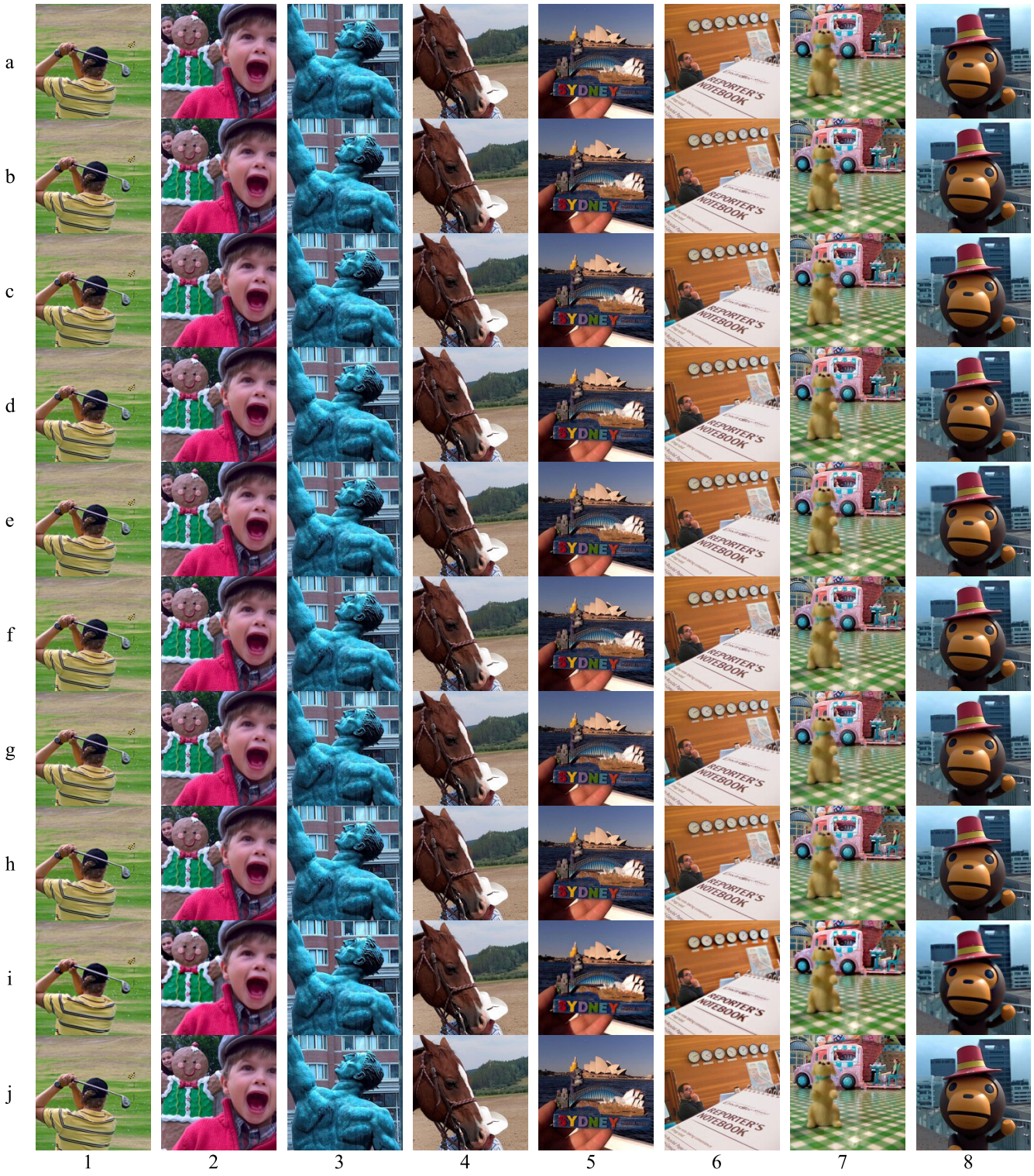

FIGURE 6. Fusion results of different algorithms, where from (a) to (j) are: CNN, DSIFT, BRW, GFF, MWGF, IM, SR, PCNN, MGF, and Ours.

A. IMAGE SUPER-RESOLUTION

In order to better recover the structure and details of the image, we use DualCNN for super-resolution reconstruction, whose network model is shown in FIGURE 3.

In FIGURE 3, with a low-resolution image as the input, the detail component and structural component of the input low-resolution image are estimated through the upper channel: Net-D and the lower channel: Net-S, respectively, and the

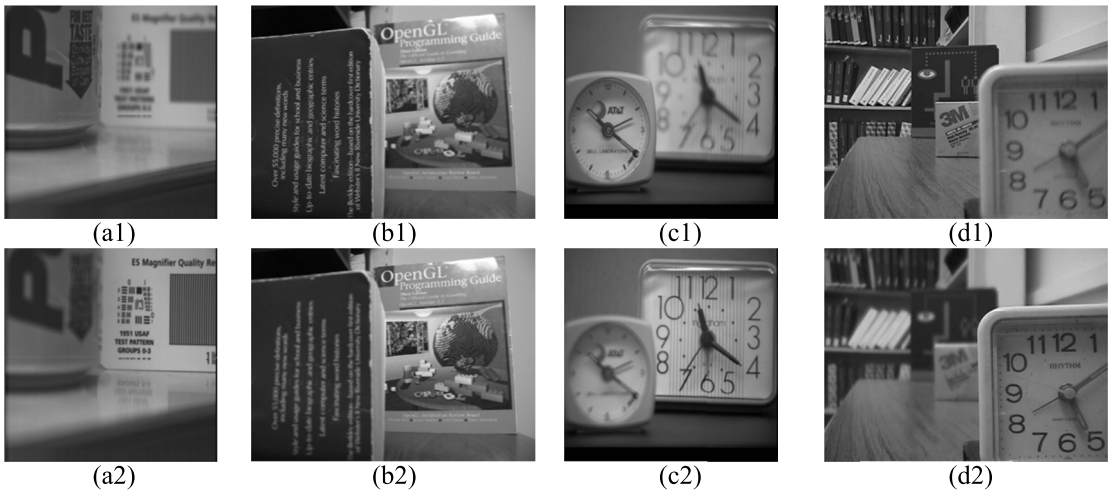

FIGURE 7. Four pairs of gray-scale multi-focus images.

structure and details of the image are restored in an end-to-end model. As shown in FIGURE 3, image D is the detail component generated by the source image through the Net-D network, and image S is a structural component of the source image through the Net-S. Among them, the network structure and parameters of each layer of Net-D and Net-S are shown in FIGURE 4.

The structure and details of DualCNN used in this paper are like those in [36]. In FIGURE 4, the red dotted block on the left is the Net-S network structure. The network structure is with 3 convolution layers, and the filter size of each layer is 9×9, 1×1, 5×5, the depth of the feature map of each layer is 64, 32 and 1 respectively. In FIGURE 4, the green dotted block on the right side is the network structure of Net-D. The network structure contains 20 convolutional layers, the size of each layer of filters is 3×3, and the depth of the feature map of each layer is 64.

In FIGURE 3, let X, S, and D be the truth image (high-resolution image), the output of Net-S, and the output of Net-D, respectively. When training the network, the error between the final output of the network and the truth image should be minimized, which can be defined as the loss function of the network, such as

$$L_x(S, D) = \|\phi(S) + \varphi(D) - X\|_2^2, \tag{10}$$

where $\phi(S)+\varphi(D)$ is the output of DualCNN, $\phi()$ and $\varphi()$ are the same as the functions used in [36].

Since DualCNN has two channels, if we only use the loss function of formula (10), it will be unstable in model training. Then, we use two separate loss functions to normalize the upper and lower channels. The loss functions of Net-S and Net-D are shown in formulas (11) and (12).

$$L_s(S) = \left\|S - S_{gt}\right\|_2^2, \tag{11}$$

$$L_d(D) = \left\|D - D_{gt}\right\|_2^2, \tag{12}$$

where S_{gt} and D_{gt} are the corresponding ground-truth values of Net-S and Net-D output. Combining the equations (10)-(12), the overall loss function of the DualCNN can be defined as:

$$L = \alpha L_x + \lambda L_s + \gamma L_d, \tag{13}$$

where α, λ and γ are regularization parameters, which is set to 1, 0.001 and 0.01.

The training data set we used is from ImageNet. During the training, the selected 300 high-quality images (the ground-truth image X) are first iterated 5 times by Gaussian blurring, and a total of 1500 low-resolution images are generated namely S_{gt}. The low-resolution image S_{gt} is up-sampled by the bi-cubic up-sampling method as the network input. The ground truth value D_{gt} is obtained from the difference between the truth label X and the structure S_{gt}. During the training process, we set the learning rate to 0.0001. We used the random gradient descent optimizer to minimize the loss function, set a reasonable regularization term of the loss function to prevent the model from over-fitting, and compared the PSNR between super-resolution images generated by the training and the original image to evaluate the performance of the training model.

B. FOCUS AREA DETECTION AND IMAGE FUSION

Let I_1 and I_2 represent as two color source images. Firstly, the structure and details of the source images I_1 and I_2 are restored by using DualCNN. Then, a bilateral filter is used for denoising, we can get image Sr_1 and Sr_2. The initial difference image IDI_1 and IDI_2 can be calculated by

$$\begin{cases} IDI_1(x, y) = |Sr_1(x, y) - I_1(x, y)| \\ IDI_2(x, y) = |Sr_2(x, y) - I_2(x, y)| \end{cases} \tag{14}$$

where Sr_1 and Sr_2 are super-resolution images after bilateral filter, and $|\cdot|$ is the absolute value operation.

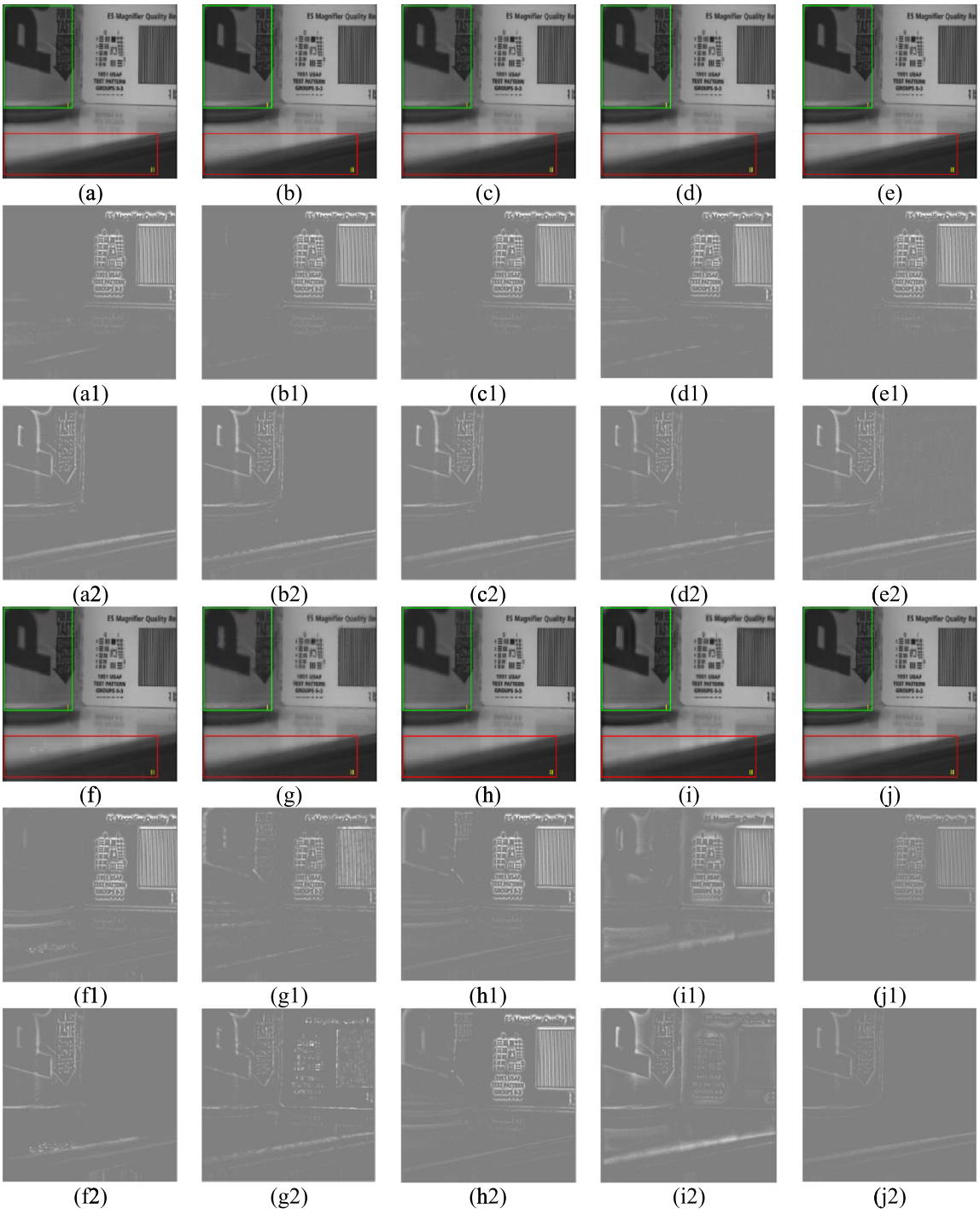

FIGURE 8. The fusion images and difference images tested on the images of Pepsi. (a-j) are fusion results based on CNN, DSIFT, BRW, GFF, MWGF, IM, SR, PCNN, MGF and Ours. (a1-j1), (a2-j2) are difference images with the left focused image Figure 7(a1) and the right focused image Figure 7(a2), respectively.

Secondly, we use Sr_1 and Sr_2 as guide images to enhance the high-frequency information in the initial difference image IDI_1 and IDI_2 to obtain the refined difference image RDI_1 and RDI_2

$$\begin{cases} RDI_1 = G_{r,\varepsilon}(IDI_1, Sr_1) \\ RDI_2 = G_{r,\varepsilon}(IDI_2, Sr_2) \end{cases} \tag{15}$$

Then, in order to retain more information about the image, the pixel-value-max rule is used to get the initial decision map (IDM) shown in formula (16):

$$IDM(x, y) = \begin{cases} 1, & if\ RDI_1(x, y) > RDI_2(x, y) \\ 0, & otherwise. \end{cases} \tag{16}$$

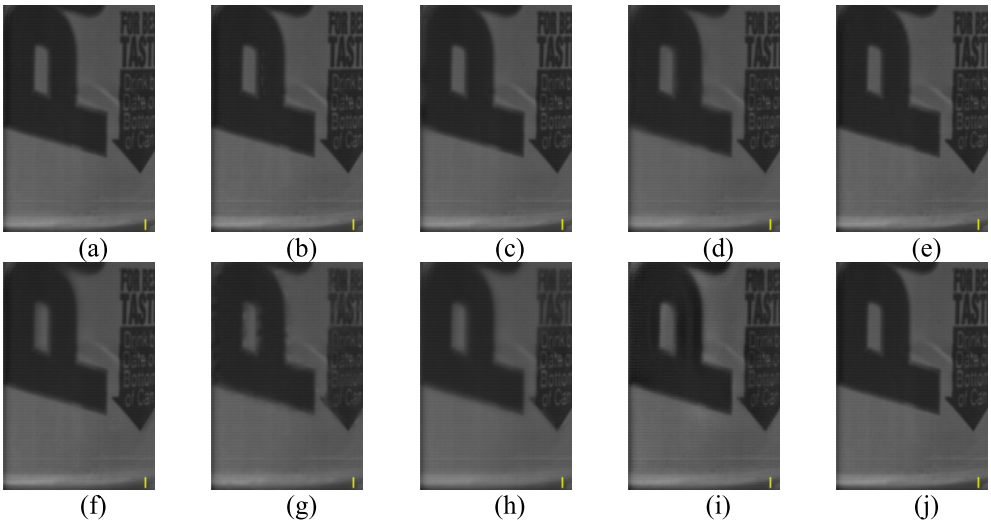

FIGURE 9. The enlarged area of the green rectangular frame.

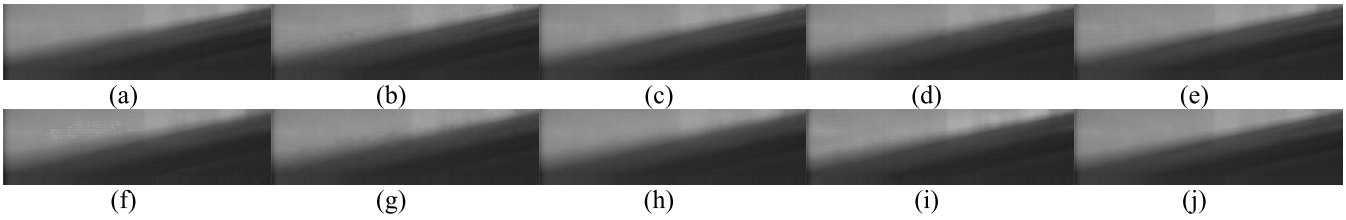

FIGURE 10. The enlarged area of the red rectangular.

Although previous focus area detection methods have obvious effects, some outliers and isolated regions often appeared in the initial decision map (IDM). In our method, we use small area removal strategy in [21] to optimize the initial decision map. In this paper, the region threshold is set to $0.01 \times H \times W$, where H and W are height and width of the source image. After the small area removing, we can get the refined decision map (IDMR). The initial fusion image I_{CF} is obtained by using IDMR as shown in formula (17).

$$I_{CF}(x, y) = IDMR(x, y)I_1(x, y) + (1 - IDMR(x, y))I_2(x, y), \tag{17}$$

Finally, the final decision map can be got by $FDM = G_{r,\varepsilon}(IDMR, I_{CF})$. And the final decision map (FDM) and the source image are subjected to pixel weighted fusion rules to obtain the final fused image I_F, as shown in formula (18):

$$I_F(x, y) = FDM(x, y)I_1(x, y) + (1 - FDM(x, y))I_2(x, y), \tag{18}$$

IV. EXPERIMENTAL RESULTS

In order to effectively evaluate the fusion performance of the proposed algorithm in different degrees of focus images, we test our method in color multi-focus image fusion test set named Lytro database (as shown in FIGURE 5) and commonly used gray multi-focus images (as shown in FIGURE 7). The performance indicators from both subjective and objective aspects are compared with the other nine representative multi-focus fusion algorithms. The compared algorithms are listed in the following. (1) A multi-focus image fusion algorithm based on a convolutional neural network (CNN) proposed in [21]. (2) A multi-focus image fusion algorithm based on dense scale-invariant feature transformation (DSIFT) proposed in [15]. (3) A multi-focus image fusion algorithm by using boosted random walk based algorithm with two-scale focus maps (BRW) proposed in [30]. (4) Image fusion processing based on guided filter (GFF) proposed in [14]. (5) Multi-focus image fusion algorithm based on multi-scale weighted gradient fusion (MWGF) proposed in [29]. (6) Multi-focus image fusion algorithm based on image matching (IM) proposed in [13]. (7) Image fusion algorithm based on sparse representation (SR) proposed in [20]. (8) Multi-focus image fusion algorithm based on pixel-level convolutional neural network (PCNN) proposed in [26]. (9) Image fusion algorithm based on multi-scale guided (MGF) proposed in [42]. All comparison fusion algorithms use the default settings from the listed papers. In addition, all the comparison algorithms are implemented in MATLAB 2014B and the evaluation of the fusion quality is achieved in a machine with an Intel Core i5-4210, 2.6Ghz and 12GB RAM.

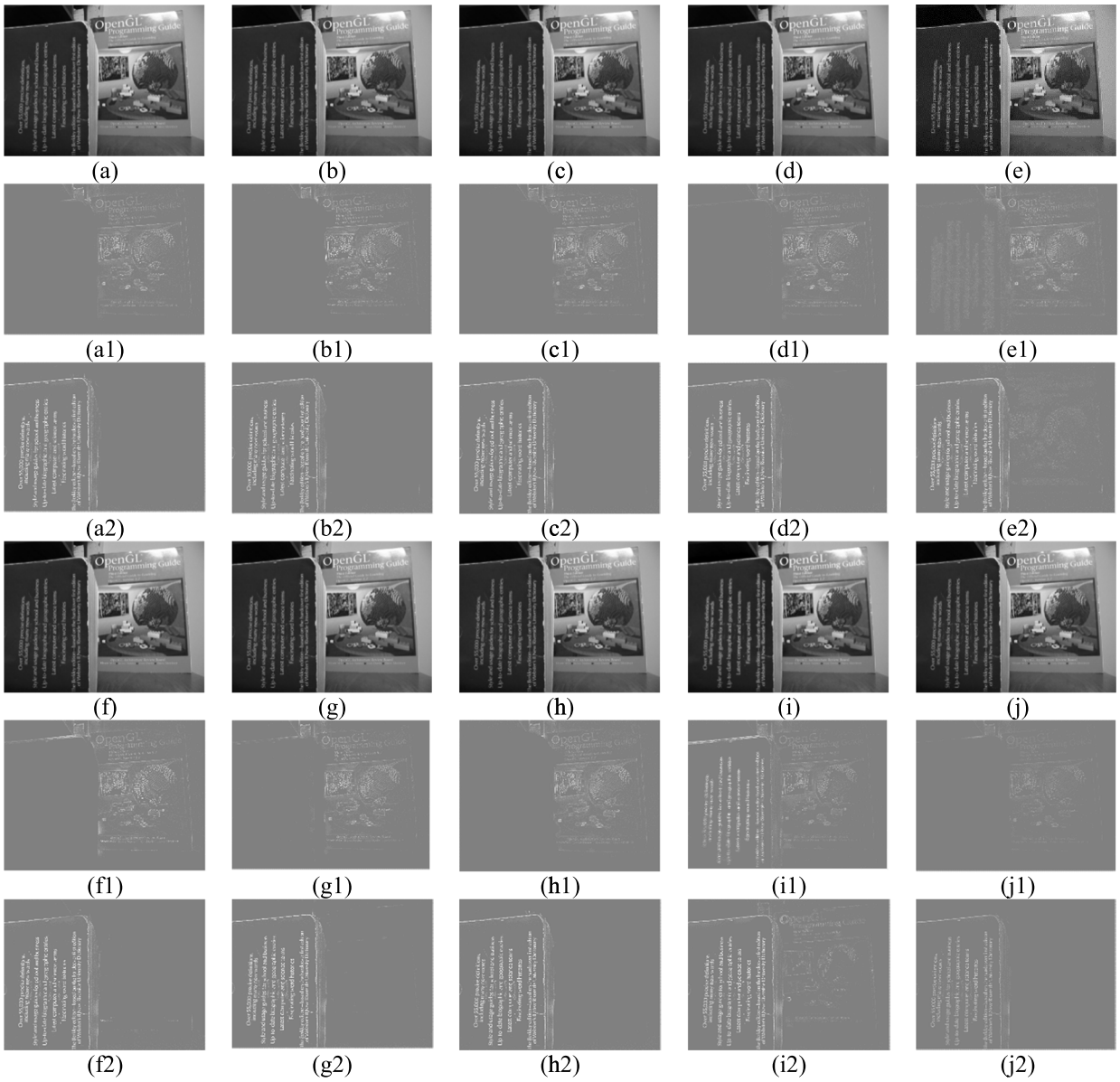

FIGURE 11. The fusion images and difference images tested on the images of Book. (a-j) are fusion results based on CNN, DSIFT, BRW, GFF, MWGF, IM, SR, PCNN, MGF and Ours. (a1-j1), (a2-j2) are difference images with the left focused image Figure 7(b1) and the right focused image Figure 7(b2), respectively.

We also make subjective and objective evaluations of the fused images generated by different algorithms. In the objective evaluation of images, we use eight kinds of objective evaluation indicators including indexes based on information theory (included the measures of feature mutual information Q_{FMI} [31], the measures of normalized mutual information MI [43], Tsallis entropy Q_{TE} [43], and nonlinear correlation information entropy Q_{NCIE} [43]), index based on structural similarity Q_Y [43], index based on human perception Q_{CB} [43] and index based on image features (included metric based on edge information Q_G [43] and metric based on phase consistency Q_P [43]). The value of the index is larger, the effect of the fused image is better. At the same time, we also compared the computational efficiency of algorithms with the criteria of runtime.

A. COLOR MULTI-FOCUSED IMAGE

In order to test the proposed algorithm, the Lytro color multi-focus image fusion test data set [19] is fused with the above image fusion algorithm. FIGURE 5 shows eight pairs of color source images with test data set.

FIGURE 6 is the listed fused images obtained from the eight pairs of multi-focus images in FIGURE 5. Each row in FIGURE 6 represents all the fused image obtained by one fusion algorithm, and the last row is the fusion result obtained by our algorithm. FIGURE 6.1 (e) produces a slight

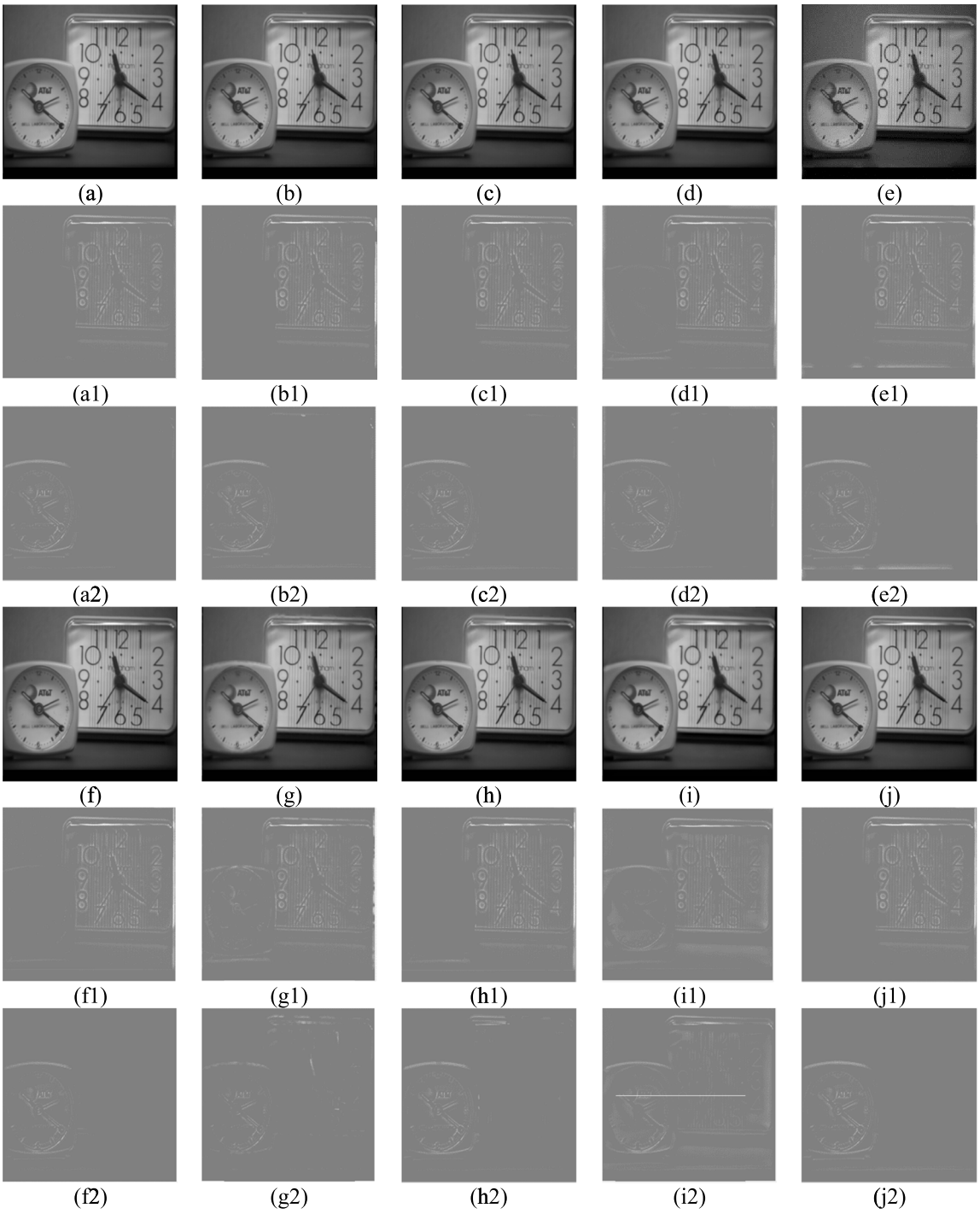

FIGURE 12. The fusion images and difference images tested on the images of Clock. (a-j) are fusion results based on CNN, DSIFT, BRW, GFF, MWGF, IM, SR, PCNN, MGF and Ours. (a1-j1), (a2-j2) are difference images with the left focused image Figure 7(c1) and the right focused image Figure 7(c2), respectively.

blur under the person's arm and a defocused pixel block at the edge of the ''person'', which affects the visual effect. In FIGURE 6.1 (f) and 6.1 (i), some of the edge texture information is lost at the ''hat''. FIGURES 6.2 (a) and (d) have slight artifacts at the edge of the ''clothes'' on the left part near the ''red child''. FIGURE 6.2 (i) is too smooth on the ''clothes'' of the ''little green snowman'' and some details are missing. On FIGURE 6.4 (e) and 6.4 (g), the images have poor spatial continuity. FIGURE 6.5 (a), 6.5 (d), and 6.5 (e) are too blurry on the ''border'' between the front ''hand'' and the back background. FIGURE 6.5 (c-d) and 6.5 (f) have a defocused area in the center of the ''hand''.

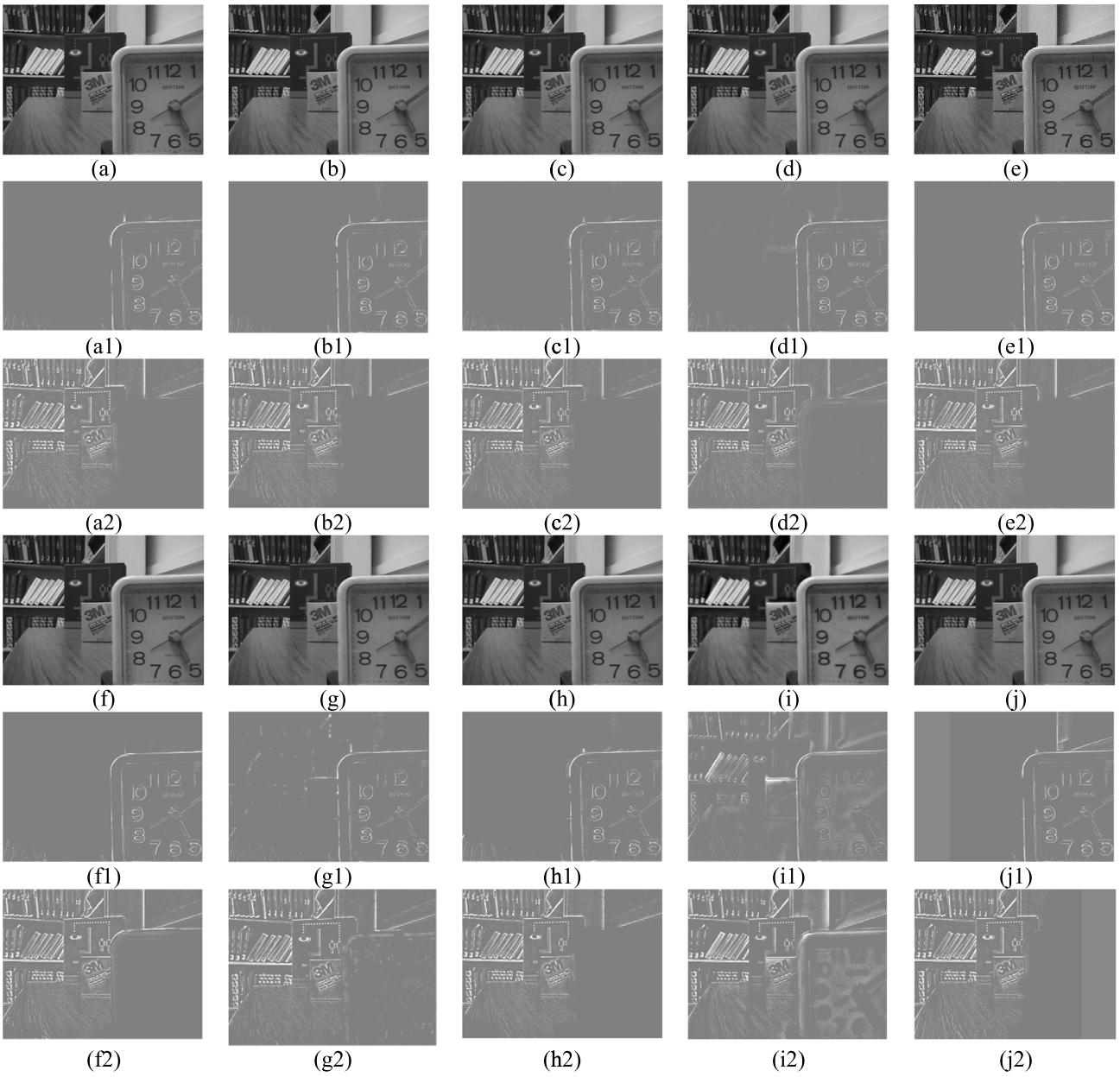

FIGURE 13. The fusion images and difference images tested on the images of Desk. (a-j) are fusion results based on CNN, DSIFT, BRW, GFF, MWGF, IM, SR, PCNN, MGF and Ours. (a1-j1), (a2-j2) are difference images with the left focused image Figure 7(d1) and the right focused image Figure 7(d2), respectively.

In FIGURE 6.6 (d), the upper boundary part of the image has obvious artificial texture. The fused image in FIGURE 6.7 can fully fuse the focus area of the two source images, retain the detail information of the image. FIGURES 6.8 (d) and 6.8 (g) do not show the small ''monkey'' outline well, the spatial consistency is poor and some spatial information is lost.

In general, MGF smooths the high-frequency details of the image. Although SR, IM and GFF can improve the quality of fused image, they cannot avoid the block effect produced on the edge areas. PCNN and CNN can effectively classify the pixels in the focus area and the pixels in non-focus area through an end-to-end non-linear mapping. However, in smooth area where the focus area and the defocus area intersected, there are often classification errors by PCNN and CNN. DSIFT, BRW and MWGF are three representative methods for measuring the focus area, which can get relatively accurate in focus areas of image but are insufficient to handle the edge parts. Our method can classify these smooth areas well and retain more detailed texture.

To show the fusion effect of different fusion algorithms, TABLE 1 and TABLE 2 give all objective evaluation indexes of each fusion algorithm. In the given evaluation indexes, larger value means the algorithm has better performance.

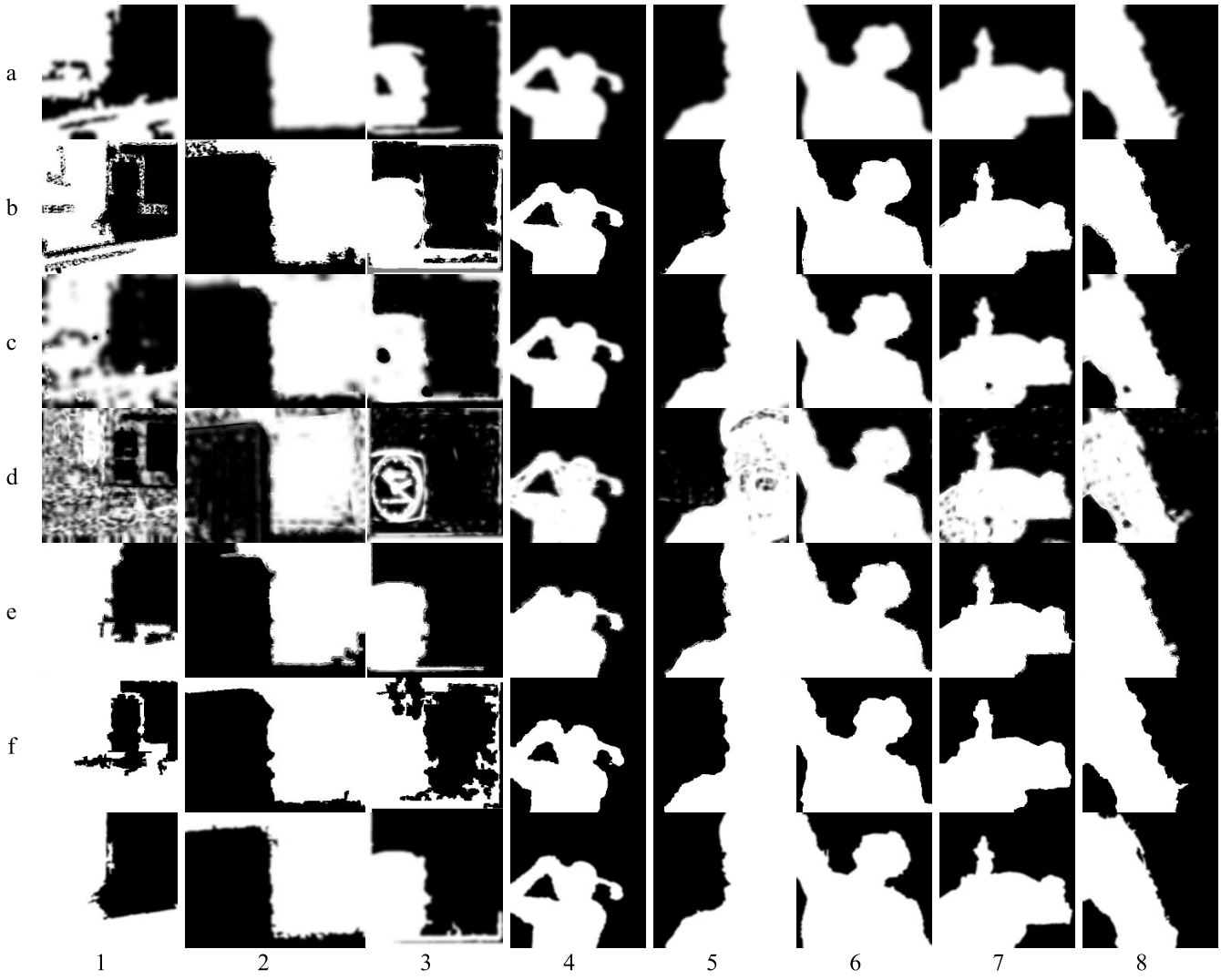

FIGURE 14. Decision maps of CNN, DSIFT, BRW, GFF, MWGF, PCNN and Ours.

The bold value indicates the best objective evaluation index. From TABLE 1 and TABLE 2, we can find that our method has the best objective evaluation indexes except just little weak on Q_{CB} in FIGURE 6.1 and Q_Y in FIGURE 6.6. So, our algorithm is superior to other algorithms for both the visual effects of the image and the objective evaluation indexes.

B. GRAY MULTI-FOCUSED IMAGE

Our algorithm not only has a good fusion effect on color multi-focus images, but also can be applied to gray-scale multi-focus images. Here we use 4 pairs of widely used multi-focus images as the source images to compare with all the algorithms. The gray-scale multi-focus images are shown in FIGURE 7.

FIGURE 7 (a1) and (a2) are left-focus image and right-focus image of multi-focus images of *Pepsi*, respectively. FIGURE 7 (b1) and (b2) are left-focus image and right-focus image of multi-focus images of *Books*, respectively. FIGURE 7 (c1) and (c2) are left-focus image and right-focus image of multi-focus images of the *Clock*, respectively. FIGURE 7 (d1) and (d2) are left-focus image and right-focus image of multi-focus images of the *Desk*, respectively.

In order to verify the effect of the fusion algorithm on gray-scale images, we still use the same algorithm for gray-scale multi-focus images as for color multi-focus images. The comparison of the four groups of images is shown in FIGURE 8-13.

FIGURE 8 gives all of the fusion images and difference images of the *Pepsi* with all image fusion algorithms. FIGURE 8(g), FIGURE 8(h), and FIGURE 8(i) show that SR, PCNN and MGF cannot accurately detect the focus and defocus areas. FIGURE 8(a), FIGURE 8(b) and FIGURE 8(d) show that CNN, DSIFT and GFF have detection errors at the smooth area. FIGURE 8 (c) shows that BRW brings a small amount of artificial texture in the edge area of the image. FIGURE 8 (e) and FIGURE 8 (f) show that MGWF and IM cannot distinguish the boundary between the focus region and the defocused region well. FIGURE 8 (j)

TABLE 1. The evaluation indexes of different fusion algorithms for figure 6.1-6.4.

	Methods	MI	Q_G	Q_{PC}	Q_Y	Q_{CB}	Q_{FMI}	Q_{TE}	Q_{NCIE}	$Time(s)$
Figure 6.1	CNN	1.0817	0.7163	0.8233	0.9840	0.8091	0.5790	0.3729	0.8315	172.9012
	DSIFT	1.1198	0.7217	0.8245	0.9866	**0.8152**	0.5869	0.3750	0.8335	7.4648
	BRW	1.1022	0.7197	0.8251	0.9864	0.8126	0.5842	0.3743	0.8325	1.0926
	GFF	1.0469	0.7165	0.8241	0.9845	0.8067	0.5766	0.3712	0.8300	**0.6160**
	MWGF	1.0624	0.7019	0.8112	0.9874	0.8036	0.5805	0.3688	0.8301	8.2414
	IM	1.0500	0.7052	0.7800	0.9737	0.7961	0.5649	0.3650	0.8301	4.3402
	SR	0.9632	0.6762	0.7733	0.9675	0.7696	0.5111	0.3647	0.8254	217.9408
	PCNN	1.1110	0.7160	0.8144	0.9858	0.8119	0.5847	0.3734	0.8328	12.5100
	MGF	0.7705	0.5731	0.6498	0.9006	0.6914	0.4136	0.3484	0.8181	1.9434
	Ours	**1.1259**	**0.7244**	**0.8290**	**0.9882**	0.8144	**0.5902**	**0.3762**	**0.8338**	1.2235
Figure 6.2	CNN	1.0654	0.7078	0.8606	0.9866	0.7827	0.5881	0.3693	0.8350	175.3970
	DSIFT	1.0808	0.7084	0.8587	0.9865	0.7855	0.5900	0.3698	0.8359	6.4850
	BRW	1.0726	0.7090	0.8609	0.9869	0.7844	0.5896	0.3696	0.8354	0.8897
	GFF	1.0082	0.7027	0.8582	0.9804	0.7744	0.5750	0.3650	0.8317	**0.5467**
	MWGF	1.0241	0.7052	0.8588	0.9864	0.7800	0.5832	0.3645	0.8326	7.1341
	IM	1.0597	0.7055	0.8483	0.9851	0.7817	0.5851	0.3678	0.8346	4.1553
	SR	0.9621	0.6781	0.8270	0.9657	0.7386	0.5223	0.3629	0.8291	216.2767
	PCNN	1.0751	0.7060	0.8517	0.9859	0.7818	0.5851	0.3694	0.8355	10.6367
	MGF	0.7045	0.5617	0.6907	0.8867	0.6321	0.4233	0.3341	0.8181	1.8390
	Ours	**1.0912**	**0.7121**	**0.8635**	**0.9887**	**0.7874**	**0.5972**	**0.3709**	**0.8365**	1.2671
Figure 6.3	CNN	1.0819	0.7504	0.9083	0.9767	0.8479	0.6693	0.3969	0.8385	182.9058
	DSIFT	1.0985	0.7510	0.9084	0.9765	0.8479	0.6720	0.3977	0.8396	6.7120
	BRW	1.0912	0.7510	0.9087	0.9767	0.8480	0.6709	0.3977	0.8391	0.9447
	GFF	1.0557	0.7494	0.9079	0.9765	0.8422	0.6675	0.3953	0.8372	**0.4700**
	MWGF	1.0566	0.7405	0.9039	0.9754	0.8393	0.6648	0.3918	0.8372	6.8508
	IM	1.0521	0.7402	0.8867	0.9717	0.8312	0.6583	0.3915	0.8367	3.9165
	SR	1.0707	0.7473	0.9072	0.9755	0.8449	0.6638	0.3960	0.8379	232.3637
	PCNN	1.0959	0.7502	0.9085	0.9760	0.8475	0.6710	0.3972	0.8394	10.2064
	MGF	0.7877	0.6499	0.8132	0.9367	0.7701	0.5629	0.3670	0.8232	1.8576
	Ours	**1.1416**	**0.7571**	**0.9118**	**0.9781**	**0.8530**	**0.6825**	**0.4012**	**0.8425**	1.2165
Figure 6.4	CNN	1.3220	0.7495	0.9008	0.9870	0.8500	0.6615	0.4463	0.8583	179.2466
	DSIFT	1.3278	0.7498	0.9006	0.9868	0.8501	0.6618	0.4466	0.8588	6.1081
	BRW	1.3263	0.7511	0.9007	0.9866	0.8499	0.6615	0.4469	0.8587	0.9010
	GFF	1.3194	0.7510	0.9002	0.9838	0.8487	0.6599	0.4468	0.8582	**0.6168**
	MWGF	1.2844	0.7140	0.8689	0.9864	0.8342	0.6570	0.4409	0.8551	7.0334
	IM	1.2891	0.7385	0.8876	0.9783	0.8304	0.6495	0.4402	0.8560	4.1448
	SR	1.2761	0.7399	0.8934	0.9780	0.8349	0.6448	0.4434	0.8553	212.8595
	PCNN	1.3266	0.7488	0.9010	0.9853	0.8493	0.6615	0.4464	0.8587	10.7365
	MGF	1.0695	0.6410	0.8026	0.9515	0.7577	0.5561	0.4302	0.8413	1.8244
	Ours	**1.3434**	**0.7535**	**0.9016**	**0.9884**	**0.8531**	**0.6651**	**0.4488**	**0.8601**	1.3713

shows that our algorithm can retain more image texture, and the fused image has a better visual effect.

In order to better show the performance of the proposed algorithm, we zoom in on some areas. FIGURE 9 and FIGURE 10 are enlarged areas of the green frame part and the red frame part of FIGURE 8. FIGURE 9(a) and 9(c) have artificial artifacts on the edge part of the *Pepsi* cup, and the fusion image is blurred. FIGURE 10 (f) and 10 (g) have breakpoints at the edges of *Desk*. Figure 9 (j) and FIGURE 10 (j) show that the fusion image obtained by our method has a better visual effect.

TABLE 3 gives objective evaluation indicators for all algorithms in FIGURE 8. TABLE 3 shows that our algorithm has the best values in the evaluation indexes of Q_{PC}, Q_Y, Q_{CB}, Q_{FMI} and Q_{NCIE}, which means that our algorithm can effectively maintain edge details. Although the proposed algorithm does not have the best results in every evaluation indicator, it has six good objective values and has satisfactory results on processing the details.

FIGURE 11 gives all of the fusion images and difference images of the *Book* with all image fusion algorithms. FIGURE 11(e) and FIGURE 11(i) show that MWGF and MGF bring block effect on the right side of the fused image, which leads to poor image fusion quality. There are some spots in FIGURE 11 (f1) and FIGURE 11 (g1), which are pixels caused by IM and SR misclassification. FIGURE 11(b) and FIGURE 11(c) show that DSIFT and BRW has a good fusion effect, but some details of texture are missing. FIGURE 11 (h) shows that PCNN brings a small amount of artificial texture in the corner of the *Book*. FIGURE 11 (d) shows that GFF smooths the details of the image too much, which makes the fused image have poor visual effects. Compared with other algorithms, the CNN-based algorithm and our algorithm can well retain the image details and obtain better visual effects.

TABLE 4 gives objective evaluation indicators for all algorithms in FIGURE 11. TABLE 4 shows that our algorithm has the best values in the evaluation indexes of Q_G, Q_{PC}, Q_Y

TABLE 2. The evaluation indexes of different fusion algorithms for figure 6.5-6.8.

	Methods	MI	Q_G	Q_{PC}	Q_Y	Q_{CB}	Q_{FMI}	Q_{TE}	Q_{NCIE}	$Time(s)$
Figure 6.5	CNN	1.0978	0.7344	0.8377	0.9876	0.7814	0.5709	0.3854	0.8381	177.2950
	DSIFT	1.1283	0.7394	0.8383	0.9872	0.7880	0.5756	0.3902	0.8400	6.7759
	BRW	1.1119	0.7376	0.8396	0.9875	0.7846	0.5734	0.3878	0.8389	0.9969
	GFF	1.0454	0.7323	0.8368	0.9858	0.7708	0.5652	0.3821	0.8350	**0.7583**
	MWGF	1.0723	0.7361	0.8377	0.9782	0.7771	0.5704	0.3820	0.8365	7.5854
	IM	1.0830	0.7258	0.8105	0.9776	0.7696	0.5627	0.3811	0.8372	4.6770
	SR	1.0863	0.7259	0.8316	0.9757	0.7650	0.5527	0.3859	0.8374	207.5283
	PCNN	1.1264	0.7374	0.8359	0.9869	0.7871	0.5749	0.3883	0.8398	13.1469
	MGF	0.8056	0.5928	0.7140	0.9453	0.6218	0.4549	0.3539	0.8233	1.9165
	Ours	**1.1541**	**0.7436**	**0.8442**	**0.9898**	**0.7943**	**0.5830**	**0.3918**	**0.8416**	1.3303
Figure 6.6	CNN	1.1992	0.7300	0.8058	0.9876	0.7693	0.5574	0.4406	0.8409	170.7704
	DSIFT	1.2155	0.7316	0.8048	**0.9888**	0.7504	0.5794	0.4340	0.8420	5.4914
	BRW	1.2056	0.7315	0.8052	0.9886	0.7699	0.5781	0.4417	0.8413	0.9322
	GFF	1.1490	0.7265	0.8039	0.9829	0.7599	0.5730	**0.4468**	0.8381	**0.5995**
	MWGF	1.1678	0.7262	0.8026	0.9863	0.7545	0.5738	0.4241	0.8389	6.9925
	IM	1.1821	0.7188	0.7717	0.9742	0.7383	0.5689	0.4263	0.8402	4.0028
	SR	1.1284	0.7178	0.7974	0.9751	0.7557	0.5587	0.4257	0.8362	198.0687
	PCNN	1.2183	0.7293	0.8038	0.9887	0.7564	0.5792	0.4337	0.8423	10.4901
	MGF	0.7415	0.5833	0.7108	0.8787	0.6269	0.4607	0.3728	0.8189	1.8581
	Ours	**1.2244**	**0.7338**	**0.8094**	0.9853	**0.7937**	**0.5892**	0.4353	**0.8424**	1.2468
Figure 6.7	CNN	0.9281	0.7032	0.8627	0.9817	0.8298	0.5375	0.3551	0.8282	175.8174
	DSIFT	0.9489	0.7019	0.8667	0.9829	0.8404	0.5414	0.3566	0.8289	6.5495
	BRW	0.9429	0.7026	0.8682	0.9828	0.8390	0.5409	0.3565	0.8287	0.8701
	GFF	0.8757	0.6977	0.8549	0.9742	0.8113	0.5257	0.3513	0.8257	**0.5784**
	MWGF	0.9093	0.6980	0.8667	0.9819	0.8350	0.5352	0.3526	0.8267	6.8781
	IM	0.9048	0.6855	0.7990	0.9707	0.8231	0.5220	0.3499	0.8267	4.1678
	SR	0.8467	0.6823	0.8292	0.9617	0.7809	0.4961	0.3516	0.8248	214.5657
	PCNN	0.8994	0.6902	0.8065	0.9694	0.7942	0.5173	0.3506	0.8278	10.5416
	MGF	0.6255	0.5950	0.7157	0.9047	0.6816	0.4240	0.3264	0.8141	1.8084
	Ours	**1.0177**	**0.7136**	**0.8788**	**0.9889**	**0.8534**	**0.5932**	**0.3628**	**0.8324**	1.4149
Figure 6.8	CNN	1.1760	0.6625	0.7821	0.9806	0.7847	0.5786	0.4022	0.8457	173.9832
	DSIFT	1.2052	0.6658	0.7767	0.9806	0.7873	0.5803	0.4065	0.8478	8.9407
	BRW	1.1902	0.6648	0.7805	0.9827	0.7887	0.5818	0.4048	0.8467	0.9146
	GFF	1.0925	0.6502	0.7798	0.9682	0.7522	0.5554	0.3969	0.8402	**0.6088**
	MWGF	1.1848	0.6269	0.7320	0.9858	0.7413	0.5854	0.4007	0.8486	7.0960
	IM	1.1535	0.6454	0.7358	0.9594	0.7646	0.5622	0.3964	0.8442	4.5350
	SR	1.0713	0.6243	0.7406	0.9438	0.6989	0.5080	0.3953	0.8389	208.7715
	PCNN	1.2015	0.6567	0.7554	0.9772	0.7824	0.5756	0.4053	0.8475	10.3382
	MGF	0.8363	0.5215	0.6670	0.8428	0.6069	0.4198	0.3718	0.8269	1.8441
	Ours	**1.2180**	**0.6725**	**0.7899**	**0.9876**	**0.7971**	**0.5919**	**0.4074**	**0.8487**	1.3060

TABLE 3. Objective evaluation indexes of different fusion algorithms for Pepsi images.

Methods	MI	Q_G	Q_{PC}	Q_Y	Q_{CB}	Q_{FMI}	Q_{TE}	Q_{NCIE}	$Time(s)$
CNN	1.1871	0.7649	0.8696	0.9648	0.7189	0.6378	0.4464	0.8394	170.5958
DSIFT	1.2078	0.7483	0.8571	0.9525	0.7206	0.6241	0.4439	0.8407	12.5331
BRW	1.1622	**0.7676**	0.8655	0.9638	0.7177	0.6317	0.4439	0.8379	1.1015
GFF	1.0393	0.7637	0.8378	0.9553	0.6296	0.5930	0.4306	0.8316	**0.4077**
MWGF	0.8946	0.5391	0.6272	0.7915	0.6589	0.2408	0.4885	0.8250	2.6519
IM	1.1056	0.7416	0.8175	0.9335	0.6478	0.5926	0.4944	0.8360	5.3877
SR	0.9696	0.7265	0.6761	0.9309	0.5620	0.5617	0.4241	0.8282	71.4543
PCNN	1.0717	0.7646	0.8123	0.9406	0.6065	0.5996	**0.5006**	0.8341	10.5567
MGF	0.8319	0.6954	0.7265	0.9158	0.6022	0.5406	0.4822	0.8227	0.6198
Ours	**1.2465**	0.7625	**0.8738**	**0.9643**	**0.7555**	**0.6492**	0.4493	**0.8432**	0.9380

and Q_{FMI}, which means that our algorithm can effectively maintain spatial consistency of image. Compared with DSIFT and CNN, the evaluation indexes of our method are slightly worse in MI, Q_{CB}, Q_{TE} and Q_{NCIE} but the overall evaluation index shows that our algorithm produces a relatively satisfactory fusion result. Although the proposed algorithm has no advantage in terms of computation compared with GFF and MGF, it has better image clarity compared with other algorithms.

FIGURE 12 gives all of the fusion images and difference images of the *Clock* with all image fusion algorithms. There are spots in FIGURE 12 (b) and FIGURE 12 (e),

TABLE 4. Objective evaluation indexes of different fusion algorithms for Book images.

Methods	MI	Q_G	Q_{PC}	Q_Y	Q_{CB}	Q_{FMI}	Q_{TE}	Q_{NCIE}	$Time(s)$
CNN	1.2865	0.6741	0.9315	0.9837	**0.8373**	0.6292	0.4283	0.8489	141.4901
DSIFT	**1.2988**	0.6736	0.9300	0.9794	0.8361	0.6265	**0.4295**	**0.8497**	5.3582
BRW	1.2847	0.6724	0.9314	0.9811	0.8364	0.6271	0.4282	0.8488	1.4269
GFF	1.2421	0.6586	0.9310	0.9673	0.8262	0.6052	0.4249	0.8459	**0.4291**
MWGF	0.8776	0.3988	0.7892	0.6965	0.7697	0.2990	0.3740	0.8273	2.0132
IM	1.2738	0.6704	0.9254	0.9795	0.8292	0.6220	0.4256	0.8480	3.6841
SR	1.2153	0.6271	0.9284	0.9431	0.8096	0.5683	0.4225	0.8442	53.5003
PCNN	1.2940	0.6708	0.9304	0.9786	0.8353	0.6266	0.4291	0.8494	8.1121
MGF	0.9803	0.5798	0.8581	0.8852	0.7629	0.5178	0.3992	0.8318	0.4334
Ours	1.2905	**0.6741**	**0.9317**	**0.9834**	0.8367	**0.6302**	0.4286	0.8491	0.8830

TABLE 5. Objective evaluation indexes of different fusion algorithms for Clock images.

Methods	MI	Q_G	Q_{PC}	Q_Y	Q_{CB}	Q_{FMI}	Q_{TE}	Q_{NCIE}	$Time(s)$
CNN	1.1888	0.6734	0.8216	0.9785	0.7745	0.5967	0.4302	0.8409	177.5960
DSIFT	1.2088	0.6713	0.8041	0.9690	0.7584	0.6022	0.4296	0.8404	11.8258
BRW	1.1948	0.6711	**0.8313**	0.9790	**0.7766**	0.6022	0.4290	0.8403	0.9719
GFF	1.1170	0.6595	0.7759	0.9515	0.7289	0.5706	0.4297	0.8386	**0.4062**
MWGF	0.9522	0.4880	0.5735	0.7687	0.6974	0.3222	0.4220	0.8284	2.7010
IM	1.1818	0.6703	0.7685	0.9646	0.7367	0.5763	**0.4370**	**0.8422**	4.9719
SR	1.0771	0.6348	0.6590	0.9386	0.6730	0.5138	0.4338	0.8354	66.2356
PCNN	1.1718	0.6418	0.7343	0.9425	0.7155	0.5691	0.4294	0.8393	11.6596
MGF	0.8491	0.5514	0.6086	0.8323	0.6371	0.4645	0.4244	0.8238	0.5625
Ours	**1.2088**	**0.6749**	0.8119	**0.9817**	0.7641	**0.6122**	0.4314	0.8414	0.9457

and the image information is not sufficiently extracted by DSIFT and MWGF. FIGURE 12(d), FIGURE 12 (i) and FIGURE 12 (g) are affected by the block effect, which deteriorates the fused image obtained by GFF, MGF and SR lead to poor fusion quality. The edges of the large *Clock* in FIGURE 12 (h) obtained by PCNN look unnatural. FIGURE 12(f) shows that IM cannot detect focus area accurately, which leads to some details of texture lost in the left of the small *Clock* FIGURE 12(a) and FIGURE 12(c) show that CNN and BRW bring slight artificial texture on the fused image. Our algorithm can fully extract the focus area between small *Clock* and large *Clock* without information lost and achieve satisfactory fusion results.

TABLE 5 gives objective evaluation indicators for all algorithms in FIGURE 12. TABLE 5 shows that our algorithm has the best values in the evaluation indexes of MI, Q_G, Q_Y and Q_{FMI}, which means that our algorithm can accurately identify the focus area of the image. Compared with BRW and IM, the evaluation indexes of our method are slightly worse in Q_{PC}, Q_{CB}, Q_{TE} and Q_{NCIE}, but the overall evaluation index shows that our algorithm produces a relatively satisfactory fusion result.

FIGURE 13 gives all of the fusion images and difference images of the *Desk* with all image fusion algorithms. Compared with other algorithms, our algorithm eliminates the texture artifacts in the lower-left corner of the image, and correctly detects the focus and defocus areas.

Table 6 gives objective evaluation indicators for all algorithms in FIGURE 13. Table 6 shows that our algorithm has the best values in the evaluation indexes of Q_{TE}, Q_{PC}, Q_Y and Q_{FMI}. Although Q_G and Q_{CB} are not optimal in all comparison indexes, they are comparable to the best indexes.

C. DECISION DIAGRAM

In order to evaluate the performance of the proposed algorithm, we compare the decision map of different algorithms. In this section, we select the five groups of color multi-focus images in FIGURE 5 (a1-a5) and FIGURE 5 (b1-b5) in Section 4.1, and the three groups of gray multi-focus images in FIGURE 7 (a1-c1) and FIGURE 7 (b1-c1) in Section 4.2. The algorithms for comparison are CNN, DSIFT, TS, GFF, MWGF, PCNN and our algorithm. All the decision maps are shown in FIGURE 14 (a-g).

FIGURE 14 lists the decision maps of the seven algorithms. From top to bottom, they are the decision maps of CNN, DSIFT, BRW, GFF, MWGF, PCNN and the proposed method. FIGURE 14 (d) shows that the GFF loses more detailed information compared with other algorithms in the defocus and focus areas. FIGURE 14 (e) shows that MWGF can obtain good performance in most cases, but often damage the boundaries of some targets during fusion. As shown in FIGURE 14(a) and FIGURE 14(f), CNN and PCNN can effectively segment the focus area of a color multi-focus image, but it still has some incorrectly segmented pixels in some gray-scale multi-focus images. As shown in FIGURE 14(b), DSIFT incorrectly detected the focus area or defocus area that appears at the edges of the image. In FIGURE 14(c), although the BRW can well identify the focus area and defocus area, in the center of the color multi-focus image there is still some defocused area that is

TABLE 6. Objective evaluation indexes of different fusion algorithms for Desk images.

Methods	MI	Q_G	Q_{PC}	Q_Y	Q_{CB}	Q_{FMI}	Q_{TE}	Q_{NCIE}	$Time(s)$
CNN	1.1104	**0.6882**	0.8040	0.9888	**0.7838**	0.6443	0.3821	0.8369	211.0208
DSIFT	**1.1334**	0.6848	0.7967	0.9808	0.7779	0.6357	0.3846	**0.8381**	8.2041
BRW	1.1072	0.6859	0.8030	0.9853	0.7792	0.6393	0.3819	0.8365	1.0579
GFF	0.9731	0.6599	0.7921	0.9411	0.7246	0.5677	0.3721	0.8291	**0.4736**
MWGF	1.0019	0.6609	0.7830	0.9734	0.7586	0.5785	0.3738	0.8305	3.0249
IM	1.0943	0.6776	0.7736	0.9739	0.7653	0.6207	0.3840	0.8362	4.0460
SR	0.9666	0.6296	0.7472	0.9185	0.6927	0.5191	0.3700	0.8288	76.0440
PCNN	1.1178	0.6819	0.7896	0.9814	0.7743	0.6393	0.3827	0.8369	13.7173
MGF	0.7089	0.5204	0.6222	0.8219	0.6270	0.4378	0.3470	0.8178	0.7065
Ours	1.1282	0.6829	**0.8047**	**0.9916**	0.7703	**0.6492**	**0.3847**	0.8376	1.0608

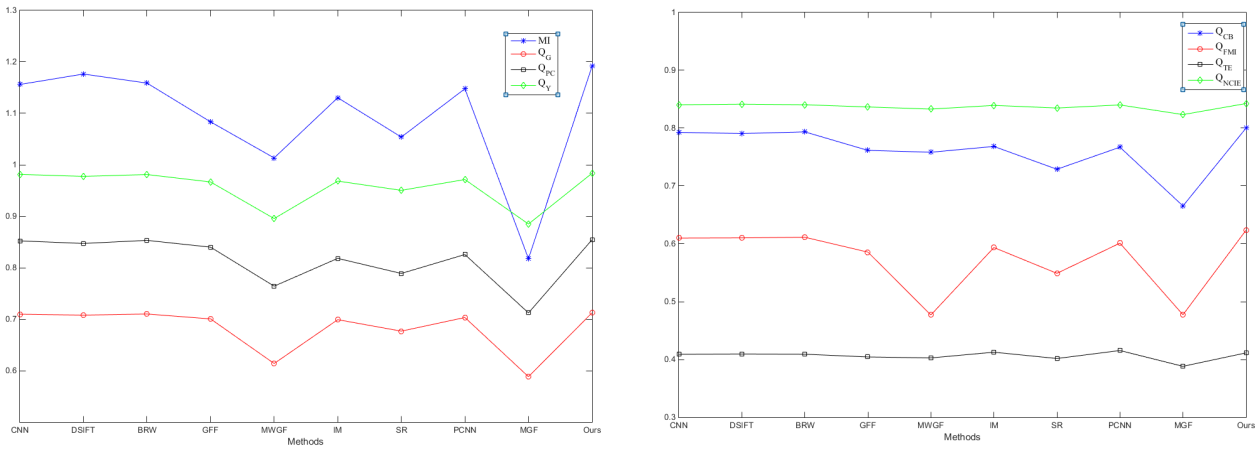

FIGURE 15. The mean of all the objective evaluation indexes of all test fusion methods in all the test data.

incorrectly identified. It can be seen from FIGURE 14 (g) that our algorithm can effectively detect the focused and defocused areas in the fused image whether it is a color multi-focus image or a gray-scale multi-focus image, which proves the performance of the proposed algorithm.

Because the color multi-focus images contain more image information and higher contrast than the gray-scale multi-focus images, the color multi-focus images can recover more high-frequency details after super-resolution reconstruction. Super-resolution reconstruction plays a very important role in focus area detection and generating high quality fused images. Therefore, our algorithm is more effective when applied to color multi-focus images.

D. COMPARED ON LARGER DATA

In order to fully demonstrate the effectiveness of the algorithm, we compare the proposed method with the other nine fusion methods in a lager data. The data contains Lytro color multi-focus image data in [19], multi-focus image data in [16] and other multi-focus image data in [15] and [37]. We calculate the mean value of all the evaluation indexes of all test fusion methods in all the test data, as shown in FIGURE 15.

FIGURE 15 (a) gives the graph of the mean objective evaluation indexes of MI, Q_G, Q_{PC} and Q_Y of all the test images of the proposed algorithm and nine representative multi-focus fusion algorithms. It can be seen from FIGURE 15 (a) that MGF and MWGF have no advantage with other algorithms, the proposed algorithm has obtained the best indexes in MI, Q_G, Q_{PC} and Q_Y indexes. FIGURE 15 (b) gives the graph of the mean the objective evaluation indexes of Q_{CB}, Q_{FMI}, Q_{TE} and Q_{NCIE} of all the test images of the proposed algorithm and nine representative multi-focus fusion algorithms. It is not difficult to see that there is little difference between all the comparison algorithms under the indexes of Q_{NCIE} and Q_{TE}. The other indexes Q_{CB} and Q_{FMI} in FIGURE 15 (b) show that the proposed algorithm has the best value among all the fusion methods. FIGURE 15 shows that the proposed algorithm has better performance than other algorithms.

V. CONCLUSION

This paper proposes a novel image fusion algorithm based on focus area detection via image super-resolution reconstruction with DualCNN. This method uses the DualCNN to restore the source image details, combined with bilateral filter and guided filter, which can well maintain the spatial consistency of the fused image. The experimental results show that the proposed algorithm can detect the focus area better than other algorithms and produce a clearer fusion image. The fusion image is not only with better visual perception

than other algorithms, but also with an advantage in terms of computational efficiency.

REFERENCES

[1] S. Liu, J. Wang, Y. Lu, S. Hu, X. Ma, and Y. Wu, "Multi-focus image fusion based on residual network in non-subsampled shearlet domain," *IEEE Access*, vol. 7, pp. 152043–152063, 2019.

[2] X. Ma, S. Hu, S. Liu, J. Fang, and S. Xu, "Multi-focus image fusion based on joint sparse representation and optimum theory," *Signal Process., Image Commun.*, vol. 78, pp. 125–134, Oct. 2019.

[3] 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.

[4] H. Li, L. Li, and J. Zhang, "Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering," *Opt. Commun.*, vol. 342, pp. 1–11, May 2015.

[5] 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.

[6] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.

[7] N. Radha and T. Babu, "Multifocus color image fusion based on walsh Hadamard transform and sum-modified-Laplacian focus measure," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 142–150, 2019.

[8] J. Dou, Q. Qin, and Z. Tu, "Image fusion based on wavelet transform with genetic algorithms and human visual system," *Multimedia Tools Appl.*, vol. 78, no. 9, pp. 12491–12517, May 2019.

[9] Y. Yang, S. Tong, S. Huang, P. Lin, and Y. Fang, "A hybrid method for multi-focus image fusion based on fast discrete curvelet transform," *IEEE Access*, vol. 5, pp. 14898–14913, 2017.

[10] D. Yang, S. Hu, and S. Liu, "Multi-focus image fusion based on block matching in 3D transform domain," *J. Syst. Eng. Electron.*, vol.29, no. 2, pp. 415–428, 2018.

[11] 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.

[12] 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," *Multidimensional Syst. Signal Process.*, vol. 31, no. 2, pp. 569–590, Apr. 2020.

[13] S. Li, X. Kang, J. Hu, and B. Yang, "Image matting for fusion of multi-focus images in dynamic scenes," *Inf. Fusion*, vol. 14, no. 2, pp. 147–162, Apr. 2013.

[14] 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.

[15] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Inf. Fusion*, vol. 23, pp. 139–155, May 2015.

[16] Y. Zhang, X. Bai, and T. Wang, "Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure," *Inf. Fusion*, vol. 35, pp. 81–101, May 2017.

[17] 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.

[18] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[19] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.

[20] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol.24, pp. 147–164, Jul. 2015.

[21] 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.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[23] Z. Gong, P. Zhong, Y. Yu, W. Hu, and S. Li, "A CNN with multiscale convolution and diversified metric for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3599–3618, Jun. 2019.

[24] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5297–5307.

[25] 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 2019.

[26] 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. 2018.

[27] H. T. Mustafa, J. Yang, and M. Zareapoor, "Multi-scale convolutional neural network for multi-focus image fusion," *Image Vis. Comput.*, vol. 85, pp. 26–35, May 2019.

[28] H. Ma, J. Zhang, S. Liu, and Q. Liao, "Boundary aware multi-focus image fusion using deep neural network," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 1150–1155.

[29] 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.

[30] J. Ma, Z. Zhou, B. Wang, L. Miao, and H. Zong, "Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps," *Neurocomputing*, vol. 335, pp. 9–20, Mar. 2019.

[31] X. Qiu, M. Li, L. Zhang, and X. Yuan, "Guided filter-based multi-focus image fusion through focus region detection," *Signal Process., Image Commun.*, vol. 72, pp. 35–46, Mar. 2019.

[32] S. Aymaz and C. Köse, "A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion," *Inf. Fusion*, vol. 45, pp. 113–127, Jan. 2019.

[33] B. Yang, J. Zhong, Y. Li, and Z. Chen, "Multi-focus image fusion and super-resolution with convolutional neural network," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 15, no. 4, Jul. 2017, Art. no. 1750037.

[34] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 184–199.

[35] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[36] J. Pan, S. Liu, D. Sun, J. Zhang, Y. Liu, J. Ren, Z. Li, J. Tang, H. Lu, Y.-W. Tai, and M.-H. Yang, "Learning dual convolutional neural networks for low-level vision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3070–3079.

[37] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "Ensemble of CNN for multi-focus image fusion," *Inf. Fusion*, vol. 51, pp. 201–214, Nov. 2019.

[38] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 568–580.

[39] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2012.

[40] L. Shuaiqi, Z. Jie, and S. Mingzhu, "Medical image fusion based on rolling guidance filter and spiking cortical model," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–9, Jun. 2015.

[41] X. Ma, S. Hu, S. Liu, J. Fang, and S. Xu, "Remote sensing image fusion based on sparse representation and guided filtering," *Electronics*, vol. 8, no. 3, pp. 303–319, 2019.

[42] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, "Multi-scale guided image and video fusion: A fast and efficient approach," *Circuits, Syst., Signal Process.*, vol. 38, no. 12, pp. 5576–5605, Dec. 2019.

[43] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.

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 research interests include image processing and signal processing.

JIAN MA received the B.S. degree from the School of Industry and Commerce, Hebei University, in 2018, where he is currently pursuing the B.Eng. degree with the College of Electronic and Information Engineering. His research interest is image processing.

LU YIN received the B.S. degree from the College of Electronic and Information Engineering, Lanzhou University of Finance and Economics, in 2017. She is currently pursuing the B.Eng. degree with the College of Electronic and Information Engineering, Hebei University. Her research interest is image processing.

HAILIANG LI received the Ph.D. degree from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, in 2018, and the M.Sc. degree from the Department of Automation, Xiamen University, China, in 2004. He is currently working as an Algorithm Software Engineer with the Hong Kong Applied Science and Technology Research Institute Company Limited (ASTRI). His work is related to image processing, computer vision, natural language processing, and machine learning.

SHUAI CONG received the bachelor's degree in information management and information technology from Hebei University, in 2009, and the master's degree in computer engineering from the School of Mathematics and Computer Science, Hebei University, in 2013. He is currently a Lecturer with the Industrial and Commercial College, Hebei University. His research interests include image processing and big data technology.

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 research interests include image de-noising, image fusion, and signal processing.

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