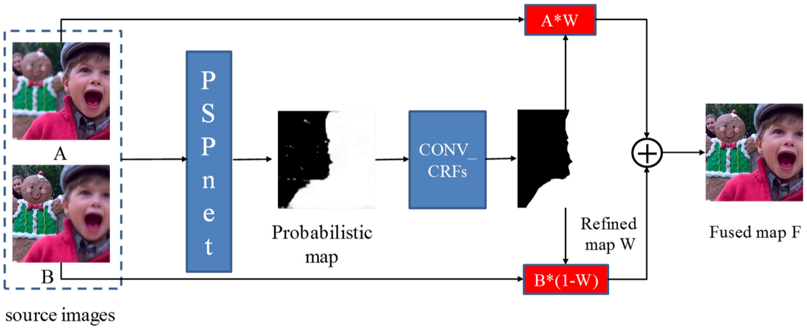
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Fusion PSPnet Image Segmentation Based Method for Multi-Focus Image Fusion

Volume 11, Number 6, December 2019

Jingchun Zhou
Mingliang Hao
Dehuan Zhang
Peiyu Zou
Weishi Zhang



DOI: 10.1109/JPHOT.2019.2950949

Fusion PSPnet Image Segmentation Based Method for Multi-Focus Image Fusion

Jingchun Zhou, Mingliang Hao, Dehuan Zhang, Peiyu Zou, and Weishi Zhang

Information Science and Technology, Dalian Maritime University, Dalian 116026, China

DOI:10.1109/JPHOT.2019.2950949

Manuscript received August 14, 2019; revised October 23, 2019; accepted October 29, 2019. Date of publication November 1, 2019; date of current version December 17, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61702074, in part by the Liaoning Provincial Natural Science Foundation of China under Grant 20170520196, in part by the Fundamental Research Funds for the Central Universities under Grant 3132019205, and in part by the Fundamental Research Funds for the Central Universities under Grant 3132019354. Corresponding author: Weishi Zhang (e-mail: teesiv@dlmu.edu.cn).

Abstract: To address the problem that the traditional multi-focus images fusion methods cannot fully use of spatial context information. A novel image segmentation method for multi-focus image fusion is proposed. This method consists of two main steps, using PSP-net combined, and the region optimization using ConvCRFs for multi-focus image fusion. PSPnet is utilized to extract the focused region of the source image, and ConvCRFs is used to optimize the segmentation map to obtain a refined map. Finally, 20 couples of color multi-focus images are employed as experimental datasets, and the contrast results show that the proposed method has a better fusion visual effect than the other state-of-the-art in subjective and objective point of view.

Index Terms: Multi-focus image fusion, image segmentation, PSPnet, ConvCRFs.

1. Introduction

Image fusion is not a simple superposition of multiple images and it can generate a new image with more valuable information. Image fusion technology belongs to the fusion of visible information in multi-sensor information fusion. It uses computer technology to process these multi-source images after using the multi-source sensor to framing the same scene, so as to maximize the effective information obtained from different channels. Image fusion can combine effective information into a high-quality image to improve the utilization of image information, spatial resolution and spectral resolution for monitoring. The emergence of this technology has overcome the shortcomings of the limited depth of field of traditional optical lenses and other imaging equipment, and fully utilizes the information complementarity and redundancy carried by the multi-source image itself, as shown in Fig. 1, Lytro-03-A is one of the source images with foreground is focused, Lytro-03-B is another source image with the background is focused, respectively, and Lytro-03-F is the final fused image with the whole scene is focused, so it is clear to see that all the targets are clear, which indicates the fused image contains much more image information than the source images. Therefore, image fusion technology has been successfully applied to infrared target recognition [1], remote sensing image detection [2], underwater image processing [3] and digital products *etc*. [4], as a very effective image processing method.

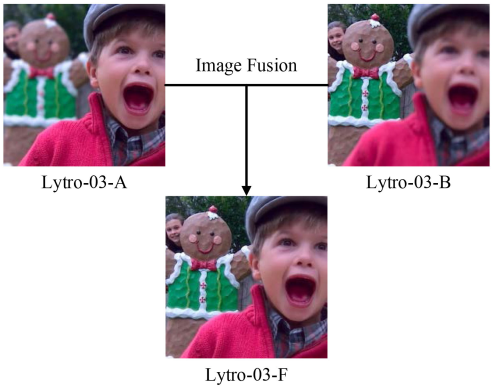


Fig. 1. Multi-focus color image fusion.

Spatial domain methods and transform domain methods are the two general methods used in image fusion. For spatial domain algorithms, the images are fused by a linear combination which can be divided into three categories in general: pixel based, block based, and region based methods [5]. The advantages of the pixel-based method are that it preserves more detail information and higher accuracy. However, its shortcomings are also obvious, which result in the loss of information about the source image.

Multi-scale transform (MST) domain-based algorithms, such as LP-based [6], WT-based [7], NSCT-based *et al*. [8], the fusion process of MST-based methods can be briefly summarized in three steps: a. The source images are decomposed into a high-frequency component and low-frequency component based on the image multi-scale characteristics; b. Different kinds of fusion rules are selected to get the high-fused and low-fused maps; c. The final fused map is obtained through inverse MST. In recent years, some artificial neural network methods like PCNN-based [9] and CNN-based [10], which have the ability to accurately extract the effective information of image under complex backgrounds. However, the traditional methods require manual setting of fusion criteria and features. The CNN-based method uses image block to calculation, while the FCN-based [12] method does not consider the context information, which ignore the incorporate suitable global features.

Therefore, in order to improve these defects, and considering that multi-focus image fusion is a dichotomous problem actually, a multi-focus image fusion framework based on image segmentation using PSPnet is proposed in this paper. PSPnet utilizes the context information of multi-focus image to effectively distinguish between focused and de-focused areas, at the same time, the convolution condition random field (ConvCRFs) is added to the processing procedure to effectively optimizes the network prediction probability graph, which helps to improve the segmentation result and enhance the fusion effect, and a large number of experiments are carried out to verify the effectiveness and superiority of our algorithm.

The rest of this paper is summarized as follow: Section 2 introduces some related works about PSPnet and ConvCRFs; Section 3 puts up the fusion framework and explains the training process of the net; Section 4 demonstrates several experimental results, and Section 5 is the conclusion.

2. Related Work

Scene parsing here is exactly the semantic segmentation based on the pixel level of the image. Objects of different sizes need to be separated from the background relatively easily within the range of different sensory fields. When it involves target detection, the mainstream algorithms based on deep learning after the year of 2015 are greatly improved from the SPPnet (spatial pyramid pooling net) [11], which fuse the features of various sizes to overcome the problem of scale variation.

CNN based methods have a better performance than traditional methods on image classification which has been proved in the past few years, but how to identify specific parts of an image remains a global problem. Jonathan Long *et al.*, researchers of neural networks, put a new module named fully convolutional networks (FCN) [12] makes a big step for semantic segmentation, in which the full connection layer has been replaced by a full convolution layer in CNN module, this change helps to remove the limitations of the full connection layer on the network structure, and at the same time, reduce the number of parameters and computational complexity, this two contributions enhanced the performance of networks. In 2016, a DeepLab [13] module was proposed based on FCN, this structure employs a process called atrous convolution operation, at the same time adds a rate to skip several adjacent convolution kernels, which replaces the max-pooling and convolution layer to avoid resolution loss problem. Besides, considering that some traditional methods of converting the image to the same size can easily cause some features to be distorted or disappeared, the DeepLab using an atrous SPP to conquer this shortage.

Du *et al.* [14] adopted a convolutional neural network to segmentation, then produced a fused feature map, the fused map is post-processed using initial segmentation, morphological operation, and watershed to obtain the map. Zhang *et al.* [15] proposed a multi-scale morphological focus-measure to detect the boundaries between the focused and defocused regions. Du *et al.* [16] used Orientation Information Motivated Pulse Coupled Neural Networks to fuse the multi-focus images, and get the decision map, which modified by a mathematical morphology post-processing technique. Farid *et al.* [17] detected the focused regions by using a Content Adaptive Blurring (CAB) algorithm, which adopted morphological operators and graph-cut techniques.

Image segmentation networks, such as SPPnet, FCN, DeepLab, and some other popular deep learning based frameworks, almost have the same problems: the mismatch of global context information and the misclassification of similar targets. Zhao *et al.* [18] put up a scene parsing module named PSPnet to improve the scene resolution effect, and improve the recognition efficiency of the target object in the complex scene at the same time. Inspired by their conclusion, due to the special imaging mechanisms of multi-focus images, it is always a big challenge to discern the boundary of the focus region. The PSPnet has a superior performance to identify the target accurately in a complex scene, so attempted to utilize this module for image segmentation in multi-focus image fusion field.

2.1 Image Segmentation via PSPnet

PSPnet, which main architecture is based on FCN module, combined with a pyramid structure and then the feature of global pyramid pooling is proposed. This feature extraction method integrates features of different scales, to obtain more global information and provide more effective features for pixel classification. As for multi-focus image fusion, under the precondition that the source image has only one focused region and the other part of the image is defocused, then how to accurately identify whether the pixel is located in the focused area or the non-focused area has a great influence on the final fusion result. Considering that this is a two-classification task, an image segmentation method based on scene parsing using PSPnet is shown in Fig. 2, through this structure, the focus area of source image can be accurately identified by taken full advantage of the hierarchical global prior of PSPnet, which help to avoid the loss of context information between different sub-regions.

As displayed in Fig. 2, different sub-region representations are obtained by four pooling scales through the pyramid pooling module the main structure of PSPnet, and then followed by an up-sampling operation before concat in the four pooling layers to rebuild the final feature representation, which contains both local and global context information, Fig. 4 shows the segmentation results of source image A, and this segmentation strategy can be organized as follow:

I. The feature map with a certain size of source image is obtained via a pretrained CNN model named ResNet101 [19] with the dilated network strategy, as shown in Fig. 2. At the final layer, we adopt a binary cross entropy (BCE) loss instead of using softmax loss to train the last classifier for this binary classification problem, which be showed in Fig. 3. The binary cross

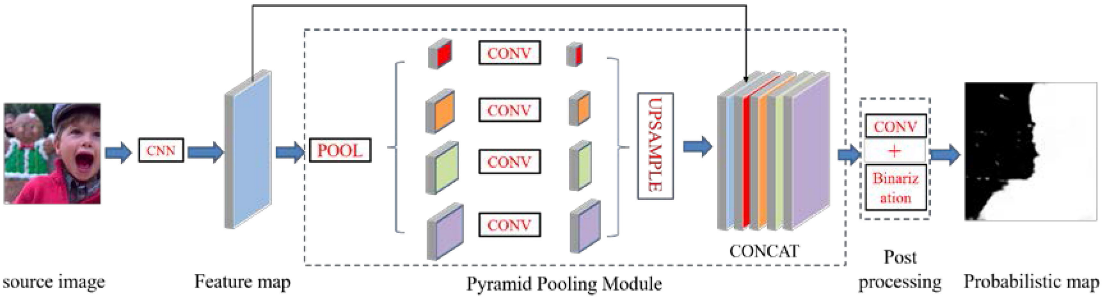


Fig. 2. Multi-focus image segmentation framework using PSPnet.

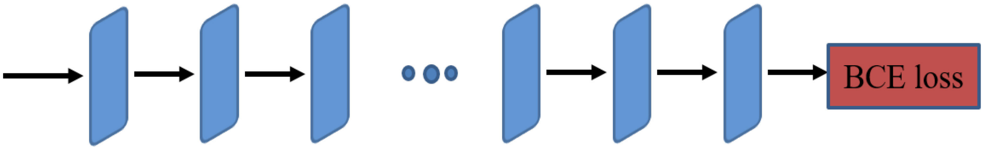


Fig. 3. Illustration of our ResNet101. (Each blue block means a residue block).

entropy (BCE) loss expands to:

$$L(\hat{y}, y) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}) \tag{1}$$

$$\hat{y} = \frac{1}{1 + \exp\left(-\left(\sum_j w_j x_j + b\right)\right)} \tag{2}$$

where $y \in \{0, 1\}$, $x = (x_1, \ldots, x_j, \ldots, x_n)$ represent the n features, $w = (w_1, \ldots, w_j, \ldots, w_n)$ corresponds to x. The binary cross entropy (BCE) loss can effectively avoid the gradient descent.

II. The extracted feature map is processed by the pyramid pooling module with four level pyramid to obtain the context information. As we can see, the output feature maps of different pyramid layers have varied feature sizes. Then, the features of four levels are upsampled to the same scale as the original features, and then the original features are combined with multi-scale information to obtain the CONCAT feature map through a 1 × 1 convolution dimensionality reduction.

III. The convolution and binarization operations are combined as the post-processing to work on the CONCAT map to obtain a probabilistic map.

2.2 Convolutional Conditional Random Fields

Although the PSPnet has taken full advantage of the global context information of the source image, it is still clear to see that there are some misclassified pixels in the probabilistic map in Fig. 4. To achieve a better and more accurate segmentation performance, a convolutional conditional random field (ConvCRFs) [20] is applied to optimize the probabilistic map.

ConvCRFs is proposed based on the fully connected CRFs (FullCRFs) [21], in which the input image I can be segmented by solving $\arg\max_X P(X|I)$:

$$P(X = \hat{x} | \hat{I} = I) = \frac{1}{Z(I)} \exp(-E(\hat{x}|I)) \tag{3}$$

where $X = \{X_1, \ldots, X_n\}$ is called as a random field and the energy function $E(\hat{x}|I)$ is denoted as:

$$E(\hat{x}|I) = \sum_{i \le N} \psi_u(\hat{x}_i|I) + \sum_{i \ne j \le N} \psi_p(\hat{x}_i, \hat{x}_j|I) \tag{4}$$

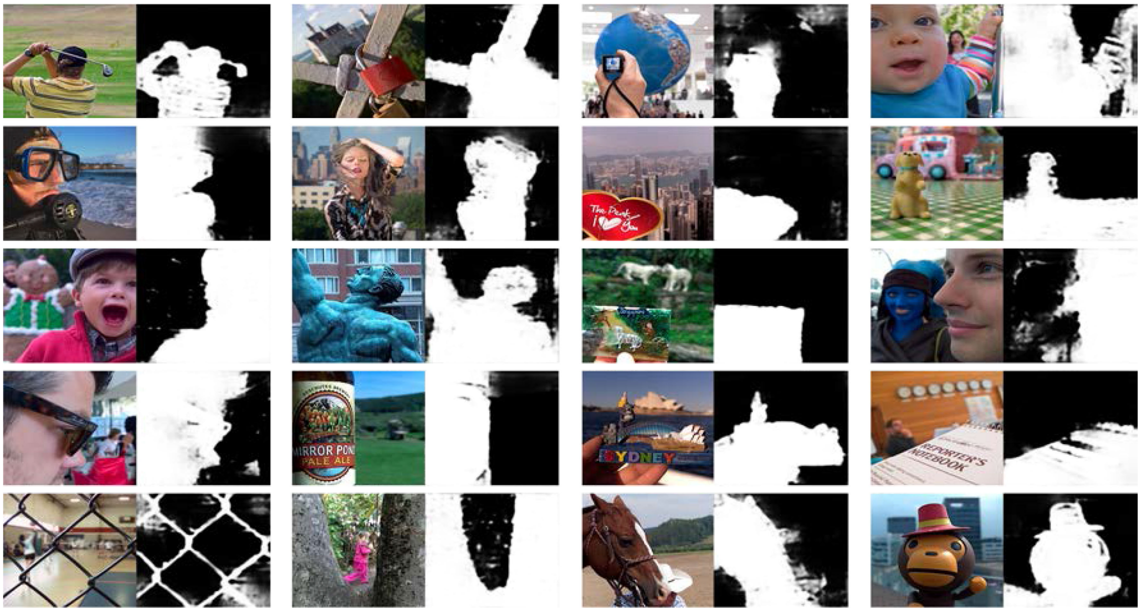


Fig. 4. The segmentation results via PSPnet.

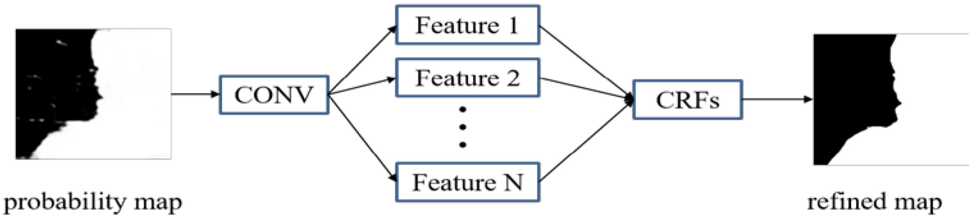


Fig. 5. Optimization process through ConvCRFs.

where the function $\psi_u(\hat{x}_i|I)$ represents unary potential, which is always computed through CNNs in current approaches [13], [22], [23]. And $\psi_p(\hat{x}_i, \hat{x}_j|I)$, the pairwise potential, indicates the joint distribution of pixels x_i and x_j.

Adding conditional independence assumptions to the framework of FullCRFs, this allows us to infer efficiently on the GPU using convolution operations, which called ConvCRFs, a model combines the feature extraction mechanism of convolutional neural network with the modeling capability of a random field, thus, ConvCRFs has the ability to pass messages efficiently.

In order to realize the accurate saliency detection of the target region, ConvCRFs is used to integrate boundary information, local information and global information of the probability map, which helps to obtain a refined map—the optimization of the probability map, and Fig. 5 is the optimization process, the multi-features of probability map are calculated through a convolution operation and then CRFs is utilized to integrate these features into a refined map, the optimization results of the datasets are displayed in Fig. 6.

3. The Proposed Architecture

3.1 The Fusion Framework

As shown is Fig. 7, the proposed method can be summarized as follow: Putting forward a multi-focus image segmentation method using PSPnet for image fusion, and different from the origin

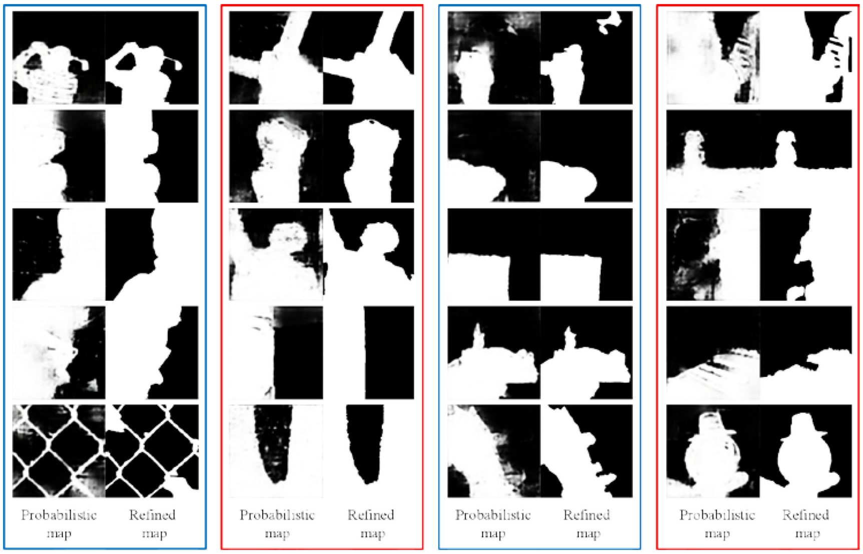


Fig. 6. 20 pairs of probabilistic map and the refined map (In each block, Left: the probabilistic map; Right: the refined map after using ConvCRFs to optimization).

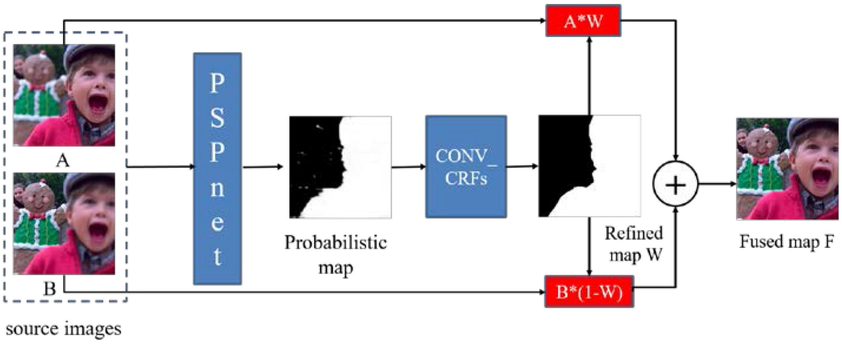


Fig. 7. Framework of the proposed method for multi-focus image fusion.

PSPnet [18], the auxiliary loss is wiped off from the net and a BCE loss was utilized to take the place of the softmax loss to train the last classifier of the pretrained CNN model Resnet101. Besides, for the very first layer of the network, it was set as a 6-channel input layer, which allows a pair of images to be input to the network at the same time; and then ConvCRFs is adopted for subsequent treatment. Therefore, we suppose that there are two source multi-focus images of the same scene, and then the fusion process [24], [25] is concluded as follow:

I. A pair of source images A and B are input to the PSPnet, which is utilized to extract the probabilistic map of image A, and the process details can be seen in Section 2.1;
II. ConvCRFs is set as the optimization strategy to get a refined map W of probabilistic map A, seeing in Section 2.2;
III. The final fused map F is obtained through the following calculation scheme:

$$F = A * W + B * (1 - W) \tag{5}$$

3.2 Training Data Synthesize

In the training phase, using the VOC2012 dataset, which contains a total of 17125 images, a total of 20 categories, in which the original images and their corresponding segmentation images

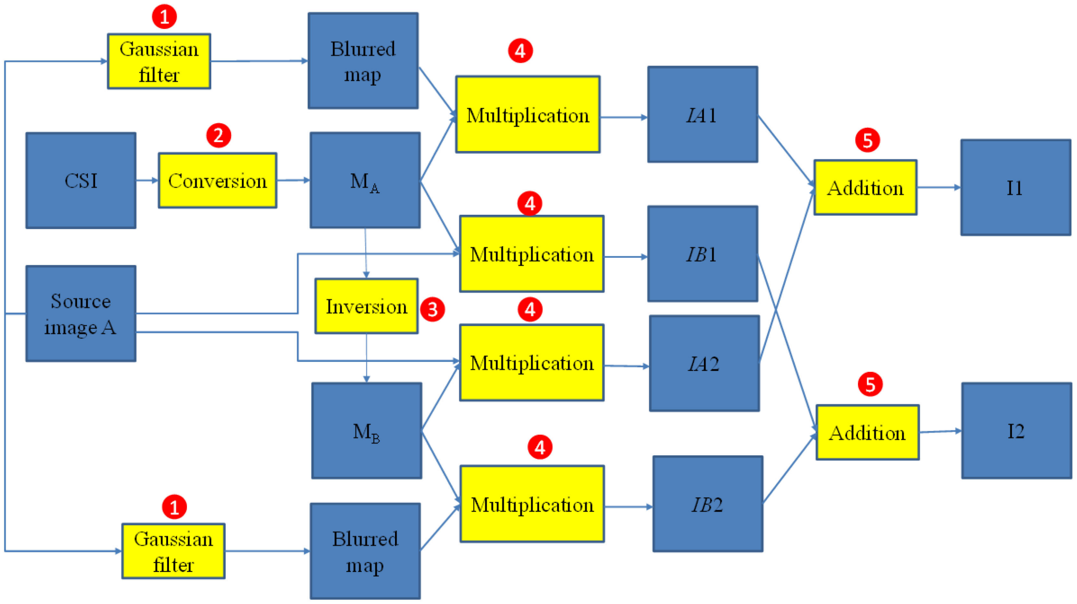


Fig. 8. Process of training data synthesize.

are available to assist realize this synthesis task. Here, we selected 2250 images to synthesize multi-focus images and create the data sets we need. We perform analog multi-focus on the 20 types of regions according to the corresponding segmented image of the source image, that is, the background focus, the target object is not focused, and vice versa. Synthesize-multi-focus is divided into the following five steps [26]:

Step 1: Gaussian blur: Gaussian filtering is used to blur the source image A to simulate the unfocused situation, the standard deviation of Gaussian filtering is 2, the window is 7×7.

Step 2: Image conversion: A binary operation is adopted to the corresponding segmentation images (CSI) of source image A to get a mask map M_A.

Step 3: Image in version: Another mask map M_B is obtained by inverting the image M_A in **Step 2**.

Step 4: Pixel-by-pixel multiplication: To obtain the focused and unfocused images, we multiply the mask maps in **Step 2** and **Step 3** with the blurred map in **Step 1** and source image, respectively, and sequentially obtain differently focused images $IA1$, $IA2$, $IB1$, $IB2$.

Step 5: Pixel-by-pixel addition: To synthesize multi-focus images, we add the differently focused images simulated in **Step 4** to obtain multi-focus images $I1$, $I2$ with foreground and background is focused, respectively.

$$I1 = IA1 + IA2 \tag{6}$$

$$I2 = IB1 + IB2 \tag{7}$$

After the above five steps, we synthesized 2250 pairs of multi-focus images, that is, a total of 4500 sheets.

3.3 Training Details

This paper trains a pair of $480 \times 320 \times 3$ RGB images, which are combined to form a $480 \times 320 \times 6$ 6-channel image, which ensures that the network feeds a one-to-many focused image each time. This operation is also used for the corresponding segmentation image.

The experimental platform of this paper is Pytorch, which is trained on the graphics card 1080. The loss function of the network is binary cross entropy (BCE loss). Using Adam gradient optimization strategy, the learning rate is 0.001. The whole training process has passed 20 epochs,

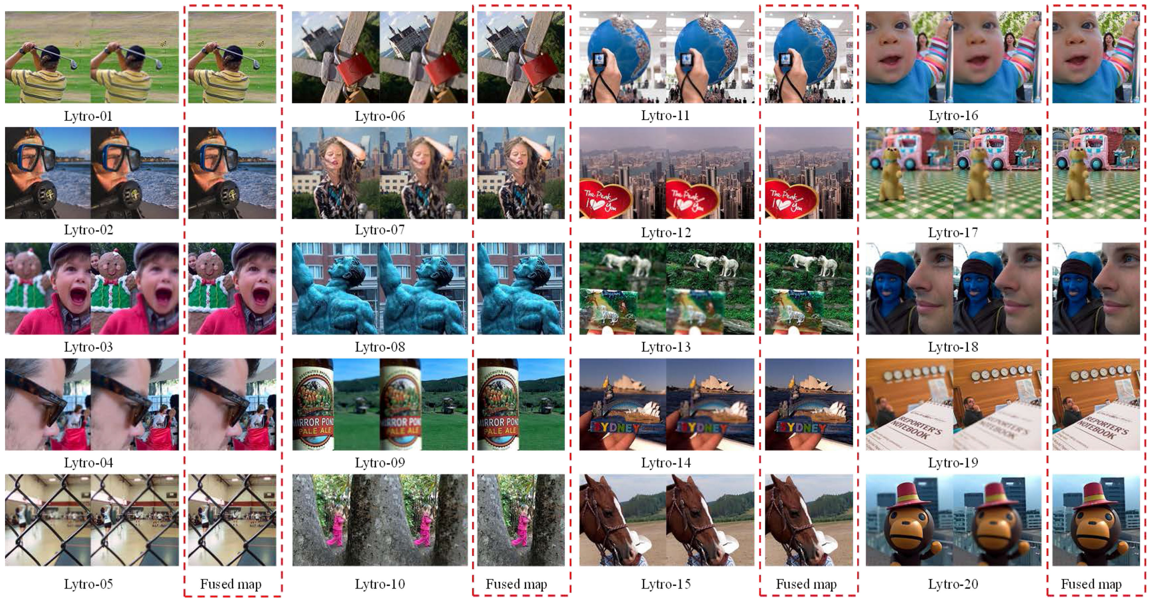


Fig. 9. 20 pairs of color multi-focus image datasets and fusion results.

and the batch size of each iteration is 16. Other parameters of the network are consistent with the literature PSPnet [18].

4. Experiments

In this section, in order to prove the effectiveness of the proposed algorithm, multi-focus color image datasets named Lytro [27], which contains 20 pairs of color multi-focus images, are used for experiments. At the same time, some popular image fusion methods such as multi-focus image fusion based on image matting (IFM) [28], the guided filtering-based fusion method (GFF) [29], discrete cosine harmonic wavelet transform-based method (DCHWT) [30], multi-scale weighted gradient-based fusion method (MWGF) [31], cross-bilateral filter-based method (CBF) [32], and CNN-based (CNN) [10] are set as the contrast methods. And the experiment results told that the proposed method has a superior performance and can achieve a better visual effect for multi-focus image fusion from both intuitive visual perception and objective perception.

4.1 Subjective Visual Effect

Fig. 9 demonstrates 20 pairs of color multi-focus image sets Lytro and their fusion results of our proposed method, respectively, Fig. 10 is the fusion results of ten couples of the datasets, in order to see the fusion effect between the algorithms more intuitively, one pair of the source images are adopted to show the difference fusion effect of these seven methods. Fig. 11 displays the Partial enlargements of fusion results of Lytro-03, in which, it is clear to see that “the ear edge of the boy" is blurred in the contrast methods, and the proposed fusion result has a relatively clear marginal structure obtained through the proposed method’s better edge information extraction capability. Fig. 12 shows the difference map between fusion result and source image Lytro-03-A, and it is evident that the less trace of the focused area left in different map, means the more image information of the focused part is extracted in to the fused map, which indicates a better fusion effect, as we can see, the difference maps of CBF, and DCHWT have distinct trace of the focused area in Lytro-03-A, and when paying attention to the border of MWGF, GFF, IFM, and CNN, it is easy to find out they are not as clear as the border in the proposed method. Thus, we can claim

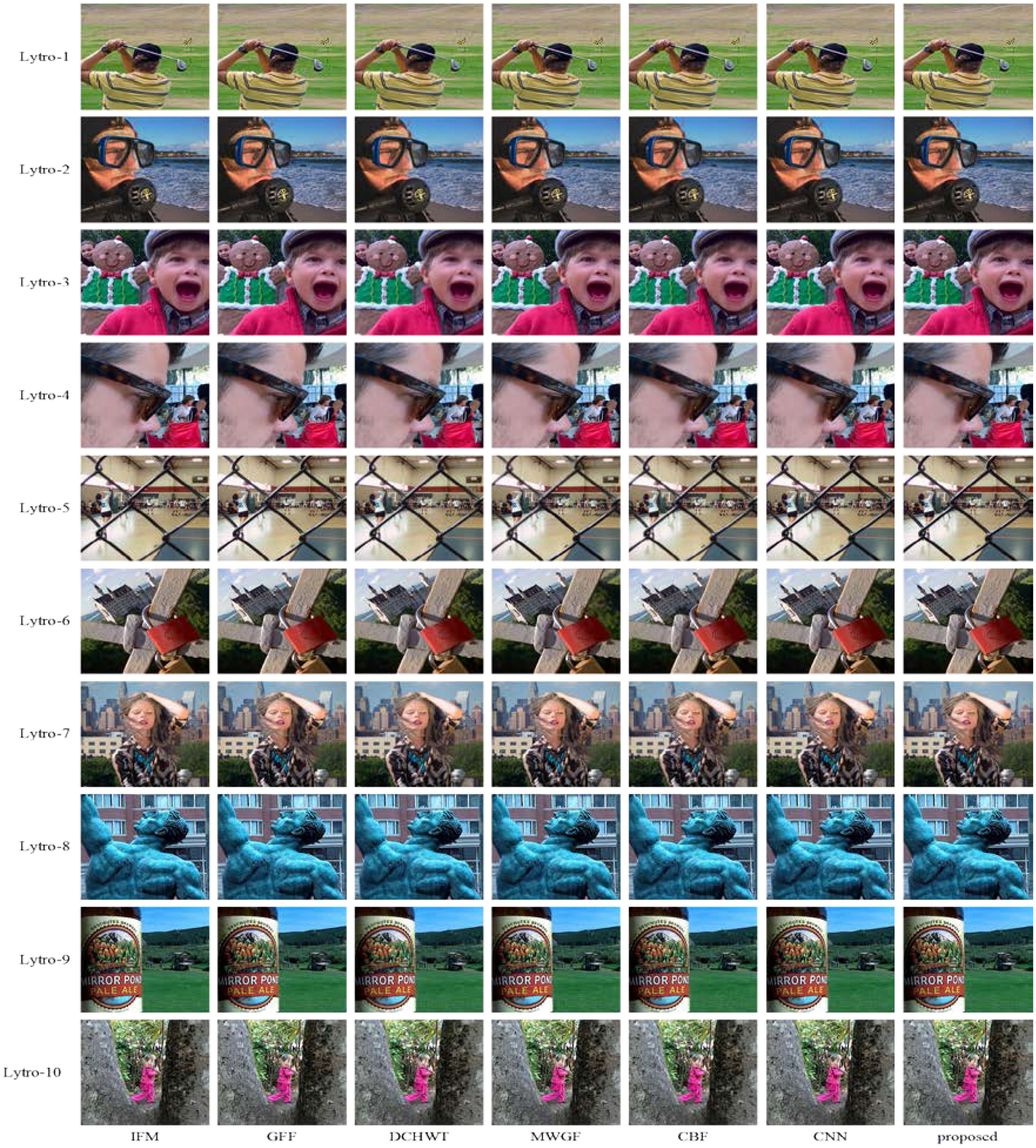


Fig. 10. Exhibition of a part of the fusion results.

that our proposed method achieves a better fusion performance than the other six comparison algorithms.

4.2 Objective Evaluation Indexes

To give an effective validation of the proposed algorithm from objective perspective, three objective evaluation indexes named mutual information (*MI*) [33], nonlinear correlation information entropy (*NCIE*) [34] and image edge information preservation estimates (Q^{AB}) [35] are applied to evaluate the fusion results objectively, the greater of the values, suggests a better fusion visual effect of the

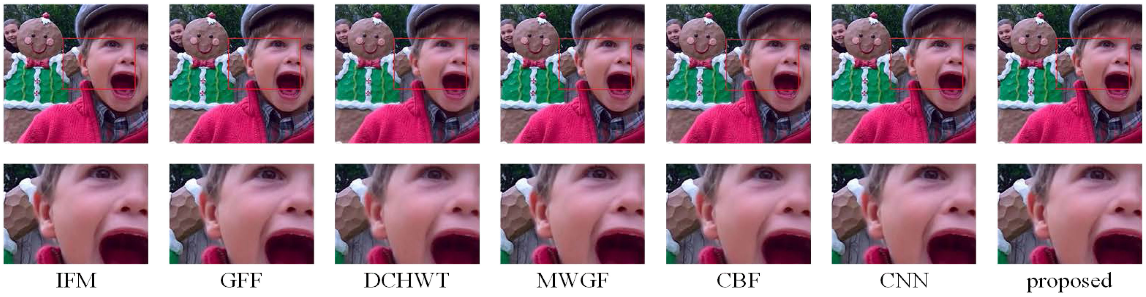


Fig. 11. Partial enlargements of fusion results of Lytro-03(The top row are 7 fusion results and the bottom row are 7 enlargements in the red block of each algorithm, respectively).

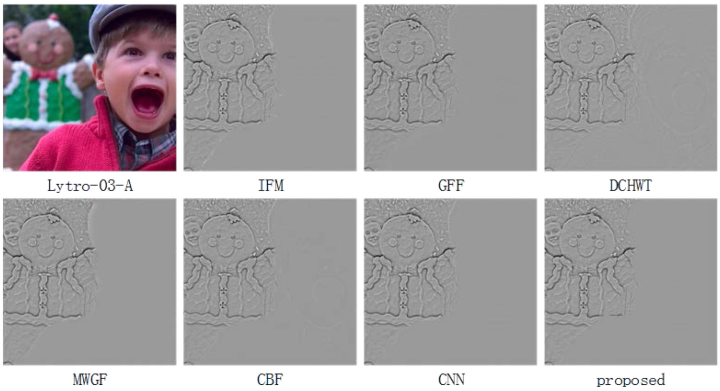


Fig. 12. The difference map between Lytro-03-A and the fused map.

algorithms, as shown in Table 1, where the best result for each assessment is highlighted in bold, and the first ten couples of the dataset's objective indexes are demonstrated in the table.

Mutual information (*MI*). *MI* (an important concept in information theory) is always used to denote the correlation between two information sets, or the measure of the information contained in one information set about another information set. As one of the most commonly used fusion effect evaluation measures among many objective evaluation indexes of image fusion quality, the larger the value of mutual information is, the more information in the source image contained in the fusion result, the better the fusion effect will be.

Nonlinear Correlation Information Entropy (*NCIE*). Information entropy is a common method to quantify the information contained in images, for multiple variables, *NCIE* can effectively measure the degree of correlation between any two variables through nonlinear correlation coefficients, that is to say a larger value of *NCIE* means there is a stronger correlation between the information contained in the fused image and the information contained in the source image, and obviously, it also indicates a better fusion performance.

Image edge information preservation estimates ($Q^{AB/F}$). $Q^{AB/F}$ is founded on the association of important visual information to the "edge" information that is present in each pixel of an image, it is a wildly used image fusion evaluation index to calculate the edge information transferred from source images to the final fused image, therefore, the larger value of $Q^{AB/F}$ means a better fusion effect too

As we can see from Table 1, it is obvious that almost all the metrics of the proposed method achieve the greatest over the other state-of-the-art algorithms. Although some metrics in the experimental data are less than the maximum value, the differences are tiny. In general, this also suggests that the proposed method can acquire a relatively better result in objective evaluation indexes for multi-focus image fusion.

TABLE 1
Objective Indexes of the First Ten Datasets

dataset	index	IFM	GFF	DCHWT	MWGF	CBF	CNN	Proposed
Lytro-01	*MI*	3.741	3.659	2.954	3.718	3.333	3.793	**3.990**
	NCIE	0.831	0.830	0.821	0.830	0.825	0.832	**0.835**
	$Q^{AB/F}$	0.741	0.753	0.694	0.741	0.742	0.754	**0.756**
Lytro-02	*MI*	4.417	4.394	3.627	4.337	3.966	4.464	**4.521**
	NCIE	0.843	0.843	0.831	0.842	0.836	0.844	**0.845**
	$Q^{AB/F}$	0.745	0.753	0.712	0.749	0.744	**0.754**	0.752
Lytro-03	*MI*	4.043	3.813	3.124	3.877	3.483	4.050	**4.116**
	NCIE	0.836	0.832	0.823	0.833	0.827	0.836	**0.837**
	$Q^{AB/F}$	0.743	0.744	0.696	0.743	0.731	**0.747**	0.742
Lytro-04	*MI*	4.633	4.460	3.796	4.566	4.182	4.666	**4.729**
	NCIE	0.848	0.845	0.834	0.847	0.840	0.849	**0.850**
	$Q^{AB/F}$	0.740	0.744	0.695	0.741	0.734	**0.746**	0.742
Lytro-05	*MI*	4.354	4.115	3.327	4.521	3.945	4.276	**4.715**
	NCIE	0.849	0.841	0.827	**0.855**	0.838	0.844	0.852
	$Q^{AB/F}$	0.708	0.722	0.668	0.693	**0.738**	0.720	0.713
Lytro-06	*MI*	4.498	4.218	3.518	4.569	4.105	4.532	**4.685**
	NCIE	0.845	0.840	0.830	**0.849**	0.838	0.845	0.848
	$Q^{AB/F}$	0.734	0.739	0.679	0.669	0.734	**0.742**	0.741
Lytro-07	*MI*	4.191	4.019	3.387	4.124	3.668	4.257	**4.406**
	NCIE	0.839	0.836	0.827	0.838	0.831	0.840	**0.843**
	$Q^{AB/F}$	0.710	0.715	0.673	0.716	0.706	**0.719**	0.715
Lytro-08	*MI*	4.220	4.139	3.462	4.176	3.692	4.282	**4.436**
	NCIE	0.840	0.839	0.828	0.839	0.831	0.841	**0.844**
	$Q^{AB/F}$	0.769	0.783	0.760	0.771	**0.789**	0.785	0.775
Lytro-09	*MI*	4.081	3.852	2.912	3.924	3.638	4.071	**4.116**
	NCIE	0.836	0.832	0.820	0.833	0.829	0.836	**0.836**
	$Q^{AB/F}$	0.799	0.800	0.745	0.798	0.792	0.801	**0.801**
Lytro-10	*MI*	3.781	3.586	2.581	3.632	3.101	3.763	**3.942**
	NCIE	0.834	0.832	0.816	0.831	0.823	0.834	**0.838**
	$Q^{AB/F}$	0.746	0.749	0.725	0.746	0.749	**0.749**	0.746

5. Conclusions

In this paper, we put up a multi-focus image fusion diagram named “FusionPSPnet: Image segmentation based method for multi-focus image fusion". Multi-focus image fusion task can be treat as a dichotomous problem as a matter of fact, because the main assignment is to define the edges between the focused region and de-focused area, thus the image semantic segmentation model PSPnet is utilized to implement the segmentation task, and a convolutional conditional random fields are used to refine border, and a large number of experiments are carried out to show the superiority of the proposed method over the other popular contrast methods, the experiment results verify the proposed method can achieve a better visual effect for multi-focus image fusion.

References

[1] J. Ma *et al.*, “Infrared and visible image fusion via gradient transfer and total variation minimization,” *Inf. Fusion*, vol. 31, pp. 100–109, 2016.

[2] L. Teng, F. Xue, and Q. Bai, “Remote sensing image enhancement via edge-preserving multiscale retinex,” *IEEE Photon. J.*, vol. 11, no. 2, Apr. 2019, Art. no. 7000310.

[3] Y. Tian, B. Liu, X. Su, L. Wang, and K. Li, "Underwater imaging based on LF and polarization," *IEEE Photon. J.*, vol. 11, no. 1, Feb. 2019, Art. no. 6500309.

[4] H. Li *et al.*, "Multifocus image fusion by combining with mixed-order structure tensors and multiscale neighborhood," *Inf. Sci.*, vol. 349, pp. 25–49, 2016.

[5] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Inf. Fusion*, vol. 8, no. 2, pp. 131–142, 2007.

[6] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[7] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.

[8] Q. Zhang and B. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, 2009.

[9] Z. Wang, Y. Ma, and J. Gu, "Multi-focus image fusion using PCNN," *Pattern Recognit.*, vol. 43, no. 6, pp. 2003–2016, 2010.

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

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

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

[13] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[14] 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.

[15] 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, 2017.

[16] C. Du and S. Gao, "Multi-focus image fusion algorithm based on pulse coupled neural networks and modified decision map," *Optik*, vol. 157, pp. 1003–1015, 2018.

[17] S. Farid M, A. Mahmood, and A. Al-Maadeed S, "Multi-focus image fusion using content adaptive blurring," *Inf. Fusion*, vol. 45, pp. 96–112, 2019.

[18] H. Zhao *et al.*, "Pyramid scene parsing network," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

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

[20] M. T. Teichmann and R. Cipolla, "Convolutional CRFs for semantic segmentation," 2018, *arXiv:1805.04777*.

[21] P. Krahenbuhl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.

[22] A. G. Schwing and R. Urtasun, "Fully connected deep structured networks," 2015, *arXiv:1503.02351*.

[23] S. Zheng *et al.*, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1529–1537.

[24] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3253–3261.

[25] W. Ren *et al.*, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *Int. J. Comput. Vis.*, pp. 1–20, 2019.

[26] X. Guo *et al.*, "Fully convolutional network-based multifocus image fusion," *Neural Comput.*, vol. 30, no. 7, pp. 1775–1800, 2018.

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

[28] S. Li *et al.*, "Image matting for fusion of multi-focus images in dynamic scenes," *Inf. Fusion*, vol. 14, no. 2, pp. 147–162, 2013.

[29] 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.

[30] B. K. S. Kumar, "Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform," *Signal, Image Video Process.*, vol. 7, no. 6, pp. 1125–1143, 2013.

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

[32] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image Video Process.*, vol. 9, no. 5, pp. 1193–1204, 2015.

[33] K. He *et al.*, "Multi-focus image fusion combining focus-region-level partition and pulse-coupled neural network," *Soft Comput.*, vol. 23, no. 13, pp. 4685–4699, 2019.

[34] Q. Wang and Y. Shen, "Performances evaluation of image fusion techniques based on nonlinear correlation measurement," in *Proc. 21st IEEE Instrum. Meas. Technol. Conf.*, 2004, vol. 1, pp. 472–475.

[35] C. S. Xydeas and V. S. Petrovic, "Objective pixel-level image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.