ISSN: 2664-9640

Multi-focus image fusion network model combining multi-scale residual characteristics and spatial frequency

Yangchuan Tian [1], Mingju Chen[2], Huixian Xu [2] and Hong Wang [2, *]

[1] Sichuan University of Science & Engineering Zigong 643000, Sichuan, China;

[2] Sichuan Key Laboratory of artificial intelligence, Zigong 643000, Sichuan China.

Abstract

In order to improve the effect of multi focus image fusion, an unsupervised learning method is proposed to train and introduce multi-scale residual encoder decoder network model. The model obtains the deep features from the input source image, calculates the activity level of the deep features by using the spatial frequency, and obtains the initial decision graph. Then the consistency verification method is applied to optimize the initial decision graph and the source image, and the final fusion result graph is calculated. The experimental results show that the proposed model achieves better fusion performance in subjective and objective evaluation.

Keywords

Keywords unsupervised learning multiscale residuals spatial frequency consistency verification method.

1. Introduction

An important issue in image processing is multi-focus image fusion. Due to the limitations of the optical lens, only the objects in the depth of field (DOF) in the photo have a focused and clear appearance, and most of the objects outside this range are blurred. Therefore, objects at different distances will cause the object to be photographed not in full focus. So many algorithms are designed to have different focal points in different scenes and to build fully focused images from multiple source images. The fusion image can be used to visualize the direction. Therefore, in the field of medical image fusion, remote sensing images and infrared and visible light fields, there are applications of multi-focus image fusion technology. Deep learning has developed rapidly in recent years, and its application fields are also very wide. The application of multi-focus image fusion also has good results. A method based on the use of unsupervised learning under convolutional neural network (CNN) on multi-focus image fusion is proposed, which is mainly used to train classification multi-level image network. Through a supervised learning strategy, the main purpose of the network is to distinguish whether the pixel block is in the focus area, so as to obtain the fusion result map. Since the advantage of the convolutional neural network is that a large number of data labels are needed to label the training data set, the data labeling workload of this algorithm is relatively large and it is not easy to implement. And the generated training data set is slightly different from the actual situation, and it is difficult to get the actual fusion effect. This paper proposes a multi-focus image fusion algorithm that uses an unsupervised learning method to train and introduces a multi-scale residual encoder-decoder network model. Training the network through unsupervised learning is mainly to deal with the problem of multi-focus data labels, and then use the spatial frequency to get the initial decision diagram. The consistency verification method is applied to optimize the initial decision diagram, and the final fusion result diagram is calculated with the source image. The experimental results show that compared with the seven fusion methods, this method achieves good fusion performance in subjective and objective evaluation.

2. Related work

2.1. Traditional algorithm

Traditional fusion algorithms are divided into two categories, one is based on transform domain methods. Another is based on the spatial domain method. The decomposition coefficient of the transform domain is mainly used to calculate the activity level of the source image. This is the fusion method of the transform domain. The more classics are based on the theory of multi-scale transformation, mainly non-subsampled contourlet transform (NSCT) [1], Laplace pyramid (LP) [2] dual complex wavelet transform (DTCWT) [3], discrete Wavelet transform (DWT) [4], curvelet transform (CVT), sparse representation (SR) [5]. The transform domain method is relatively easy to implement. Therefore, the transform domain algorithm based on the decomposition coefficients is usually difficult to obtain a good fusion effect, but the generated fused image has low contrast and darker brightness, and the result of the fused image is not clear. The spatial domain fusion method is based on gradient information to measure the activity level. Early methods used manual fixed pixel block strategy to calculate activity level, such as spatial frequency, this algorithm is prone to some artifacts. In order to solve this problem, some researchers have proposed an adaptive size based on the differential evolution algorithm to obtain the optimal block of the decision graph [6]. As more and more people study the spatial domain, some people propose a spatial domain algorithm based on pixel gradient [10], which can capture the details of the source image and have a good effect on the edge contour. Such as guided filtering (GF) [7], multi-scale weighted gradient (MWG), DSIFT. However, in the spatial domain method, Gaussian noise often affects the fusion image. The resulting image of the fusion result will have errors in the edge detection information..

2.2. Fusion algorithm based on deep learningage

The emergence of Convolutional Neural Network (CNN) has brought new vitality to the field of image fusion. It serves as a driving model for data sets. Compared with the traditional algorithm, the algorithm is more robust and the result of image fusion is clearer. Ram.Prabhakar [8] and others proposed an unsupervised method based on CNN for the fusion problem, which is mainly based on the measurement of the activity level based on the deep fusion strategy, and proposed a pixel block neural network to solve the multi-focus image. The fusion problem is of great significance to the help of multi-focus fusion images. A manual design template was used to create the data set. The algorithm includes post-processing of the decision graph. A fusion image with both focused and defocused parts. Due to the defects of artificial design, the training image is very different from the real multi-focus image and the resulting fusion effect map is average. Zhang Yong et al. proposed a CNN-based algorithm to handle different types of image fusion tasks, generating large-scale multi-focus image data sets that are more similar to real multi-focus image data in RGB images and depth images. The algorithm is designed in an end-to-end manner, and it is proposed to train on a data set. Because it is designed as a general image fusion framework and has a good fusion effect for specific types of images, this method has great limitations. The fusion algorithm of traditional methods and deep learning is at a disadvantage in extracting the deep features of the image [9]. The clear image is randomly blurred and used as a training set for network training. The convolutional network is used to continuously extract the deep features of the image. Finally, The multi-layer network will obtain the integrated feature map to fuse to obtain the final fusion image. The entire fusion process is completed in the network, which means that the training of the network is very difficult, the parameters need to be adjusted and the training time is long, and the fusion effect cannot reach the ideal value. This paper proposes a fusion method based on unsupervised deep

ISSN: 2664-9640

convolutional networks, which mainly combines deep learning with traditional methods. It consists of a multi-scale residual module and spatial frequency. The multi-scale residual feature extraction module is introduced in the encoder- In the decoder network model, unlabeled training is performed on the data set in an unsupervised training method. The training cost is relatively low and the sum of the pixel loss and the structural similarity loss is used as the loss function to realize unsupervised learning. Obtain the deep features from the input source image, use the spatial frequency to calculate the activity level of the deep features, and obtain the initial decision diagram. The consistency verification method is applied to optimize the calculation of the initial decision diagram and the source image. Get the final fusion result map. The method in this paper has the advantages of both traditional methods and deep learning. Deep learning is easy to implement and train in fusion images. Compared with traditional algorithms, it is easier to process details, and the definition of fusion images is higher.

3. Method implementation

3.1. Algorithm strategy

The method proposed in this paper is shown in Figure 1. In the training stage, the deep features of the image are extracted through the multi-scale residual feature extraction module through two source images with different focus. In the test phase, the extracted image features are input to the spatial frequency to calculate and measure the activity level and get the initial decision diagram, and apply consistency verification methods and morphological filtering to eliminate edge errors, and finally generate and optimize the decision diagram. The optimized decision diagram is combined with the source image to obtain the final fused image.

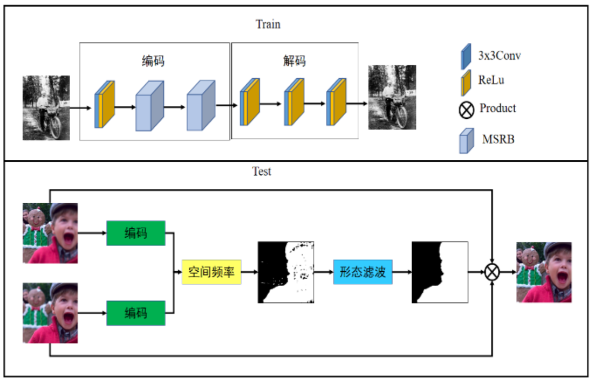

Figure 1: Encoder-decoder network frame structure diagram

3.2. Multi-scale residual feature extraction module

The multi-scale residual feature extraction module is mainly composed of a codec network, and its structure is shown in Figure 1. The encoder is composed of two parts, the upper part is composed of a 3x3 convolution kernel, and the lower part is composed of a 3x3 convolution kernel. The main function of the convolution kernel here is filtering. After passing, the two parts are output in series (concat), so the number of channels of the feature map is twice as much as before. And here as the input of the back part, connected with the front part, the number of channels obtained is 4 times that of the beginning. In this way, to realize the residual operation, the number of channels must be the same as the input at the beginning. Therefore, the following 1x1 convolution kernel realizes the compression of the number of channels of the feature map, and finally adds the ReLu function as the activation function. The filtering in the image can realize the extraction of image features when doing the convolution operation. The image features that can be extracted by different sizes of convolution kernels are different, so if two convolution kernels are used in the same network, then different features can be obtained in one network, thereby achieving a higher image resolution Requirements. The decoder network is composed of these 4 C2 convolutional layers, and its main function is to reconstruct the characteristics of the input image. In the entire training phase of the network, we only consider the encoder-decoder network, we mainly use the encoder-decoder network to reconstruct the input image in the training phase. In this experiment, the MS-COCO data set [11] was used to train in the encoder-decoder network. There are more than 120,000 pictures in total. This article divides the two data sets into training set and validation set for use in this experiment. The training set has 82,783 images, and the remaining 40,504 images are used as the verification set, which is mainly used to verify the convergence of the network loss during the training process. The image size adjustment of the image is set to 256×256 and needs to be converted to a grayscale image. The learning rate is 1×10-4, and the batch size and number of training rounds are 16 and 30, respectively. Since the training parameters are fixed, this article will test the fusion image on the test set. The realization of this article comes from the public Pytorch framework.

As shown in Figure 2, the multi-scale residual network is composed of a feature extraction part and a reconstruction part [12], and the feature extraction part is divided into: multi-scale residual block and hierarchical feature fusion structure. This article mainly uses multi-scale residual blocks. The L1 loss function is used. The output of each layer is connected to the other layers. This enhances feature propagation and reduces the number of parameters. This is to reconstruct the image more accurately and improve the fusion efficiency.

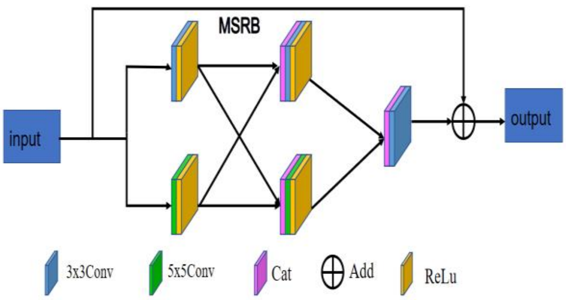

Figure 2 Multi-scale residual module

Local residual learning: In order to improve the efficiency of the network, we use residual learning for each MSRB. Experimentally, the formula for describing the multi-scale residual block in this article is:

$$M_n = S' + M_{n-1} \quad (1)$$

Among them, Mn and Mn-1 represent the input and output operations of MSRB, respectively. S'+Mn-1 is executed by shortcut connection and addition of elements in order. It is worth mentioning that the use of local residual learning greatly reduces the computational complexity and at the same time improves the performance of the network.

Multi-scale feature fusion: Different from the previous work, we constructed a double bypass network, and different bypasses use different convolution kernels. In this way, the information between these bypasses can be shared with each other, so that image features of different scales can be detected. The formula is defined as:

$$\begin{aligned} S_1 &= \sigma(W^1_{3\times3}*\mathrm{M}_{n-1}+b^1) \\ P_1 &= \sigma(W^1_{5\times5}*\mathrm{M}_{n-1}+b^1) \\ S_2 &= \sigma(W^2_{3\times3}*[\mathrm{S}_1, \ \mathrm{P}_1]+b^2) \\ P_2 &= \sigma(W^2_{5\times5}*[\mathrm{P}_1, \ \mathrm{S}_1]+b^2) \\ S^{'} &= (W^3_{1\times1}*[\mathrm{S}_2, \ \mathrm{P}_2]+b^3) \end{aligned} . \tag{2}$$

w and b represent weights and bias values, the superscript represents the number of layers in which they are located, and the subscript represents the size of the convolution kernel used in the layer. Represents the ReLu function and represents the connection operation.

In order to train the encoder-decoder network better, its main function is to extract the deep features of the image, so this paper adopts the weighted combination of structural similarity function loss (Lsimm) and pixel function loss (LP) and weight value (λ) , As shown in formula (1), structural similarity function loss (Lsimm) and pixel function loss (LP) are shown in formula (2) and formula (3) respectively, and the weight λ is set in the interval [1,9], do experiments In the process of, it is found that when λ is set to [2-5], the training network has the best effect. This article sets λ to 3. In addition, all weight adjustments of the network layer in this paper use the Adam optimizer [10] to optimize the objective function.

$$L = \lambda L_{\mathrm{ssim}} + \mathrm{L}_{\mathrm{p}} . \tag{3}$$

The pixel loss is determined by the Euclidean distance between the output O of the (LP) network and the input I:

$$L_{\mathrm{p}} = \| O - I \|_2 . \tag{4}$$

Structural similarity function loss (Lsimm) input and output structure difference:

$$L_{\mathrm{ssim}} = 1 - SSIM(O, I) . \tag{5}$$

Among them, SSIM stands for structural similarity operation. The expressions for brightness, contrast, and structural similarity are on the right side of the formula, from left to right; is the average value; is the standard deviation. C1, C2 and C3 are constants [13]:

$$\begin{aligned} SSIM(O,I) = &\frac{2\mu_O\mu_I + C_1}{\mu_O^2 + \mu_I^2 + C_1} \\ &\times \frac{2\sigma_O\sigma_I + C_2}{\sigma_O^2 + \sigma_I^2 + C_2} \\ &\times \frac{\sigma_{O,I} + C_3}{\sigma_O^2\sigma_I^2 + C_3} \end{aligned} . \tag{6}$$

3.3. Experimental results

In order to verify and evaluate the effectiveness of the algorithm in this paper, the experimental work was verified and compared with seven representative image fusion algorithms, discrete wavelet transform (DWT), dual-tree complex wavelet transform (DTCWT), curvelet transform (CVT)) [14], Multi-scale Weighted Gradient Fusion (MWG) [15], Convolutional Neural Network (CNN) [16], SESF-Fuse [17] and CRF [18] and other algorithms. This article uses 38 pairs of published multi-focus images as the test set for evaluation. A group of color images, grayscale images, and images with complex boundaries are selected for experimental introduction, and the algorithm is evaluated qualitatively and quantitatively. Through multi-focus image fusion, as much transparency and clarity attributes of the source image as possible are merged into the source image. The proposed fusion method is compared with other 7 representative image fusion methods: CVT, DWT, DTCWT, MWG, CNN from Liu Yu's website [19]. SESF-Fuse comes from Zhu Yu's website. This article compares the advantages and disadvantages of different fusion methods from subjective visual perception. The experimental part of this article is to

ISSN: 2664-9640

show the subjective visual difference of different fusion result images by means of fusion result graph and difference graph.

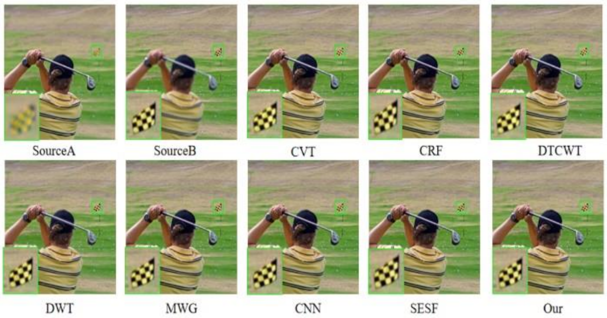

Figure 3 "Men in yellow" fusion result map

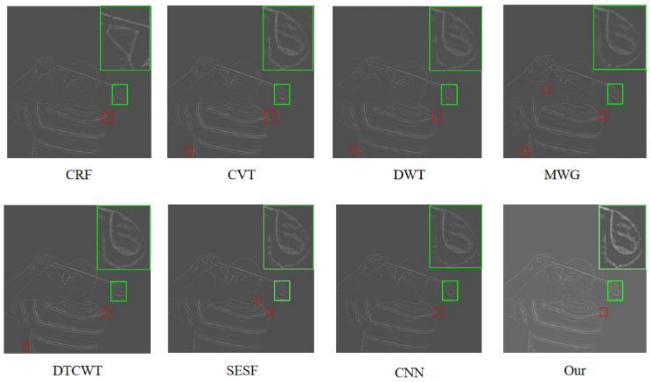

Figure 4 The difference of the fusion results of the "Men in Yellow"

Figure 3 shows the fusion results of each fusion method on the color image set "Men in Yellow". In addition, the boundary area of the fusion image is partially enlarged and placed in the upper right corner of each fusion image for better comparison. Observed from the perspective of the image outline, the fusion effect of each method is not bad. The outline of the "little yellow flag" in the partially enlarged image is relatively smooth. From the partially enlarged image, the upper part of the CVT and the background part appear. Mixed, this small part is severely blurred. From the perspective of clarity, the partially enlarged parts of CRF, DTCWT, and DWT are blurred compared to CNN and SESF-Fuse. In addition, the brightness of the image of MWG is obviously abnormally increased, and the overall image is blurred. The image of the algorithm in this paper is relatively clear, and the contour of the partially enlarged part of the image is relatively smooth, and no artifacts appear. However, only observation from the fusion image cannot effectively confirm the effectiveness of the proposed method. In order to compare the image fusion effect of each method more intuitively from subjective vision, Figure 4 visualizes each fusion image in Figure 3 minus the first source image. The difference image is obtained from one of the images, and the value of each difference image is standardized to the range of 0 to 1. After the subtraction is completed, the cleaner the subtracted part in the difference image, the lighter the boundary contour indicates the fusion The effect is better. On the contrary, if the subtracted part of the difference map still has image traces, the boundary contour is inaccurate and the boundary is obvious, then the fusion effect is not ideal. Zoom in and observe the part of the boundary in the difference map. It can be seen from Figure 4 that the CRF effect is the worst, the boundary trace is heavier, and the contour shape of the subtracted part can be clearly seen in the right half of the image, Figure 4 The SESF-Fuse in Figure 4 has heavy boundary traces and small recognition errors in the boundary contour. The subtracted parts in the CVT, DWT, and MWG difference map in Figure 4 are very clean, and the boundary contour is accurate but obvious. The effect of this algorithm is Preferably, the boundary contour is accurate and no obvious artifacts are visible. From the two sets of experiments in Fig. 3 and Fig. 5, it is shown that the algorithm proposed in this paper provides the best fusion results on the color image set of "Golf Man", and the center or boundary area near the focus area performs well.

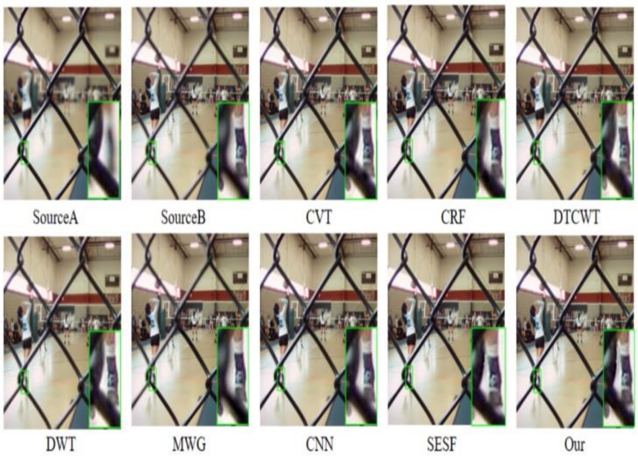

Figure 5 "Volleyball female athletes" fusion results

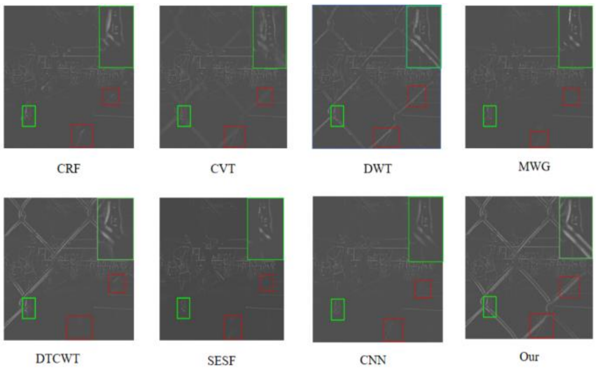

Figure 6 The difference of the fusion results of the "volleyball female athletes"

In Figure 5, this article visualizes the "volleyball female athlete" image and its fusion example. In this group of color fusion images, the CRF and SESF-Fuse in Figure 5 are relatively fuzzy, and some irregular fuzzy spots appear around the partially enlarged image. The CVT in Figure 5 is relatively clear as a whole, but there are unclear dots on the edge of the sock. DWT and DTCWT also have unclear dots on the edge of the sock, and the brightness of the sock has an abnormal increase. While there is a grayish-white blur in the partially enlarged area in the CNN [20], the overall effect of the algorithm in this paper is better, the floor in the partially enlarged image is clear, the edge of the barbed wire is smooth, and there is no artifact. In addition, observe the difference map of this group of gray-scale images. The contours of CVT and DTCWT are clear in Figure 6, but there are a lot of jagged artifacts around the contour of the "barbed wire" in the partially enlarged area, the contour of "barbed wire" The loss trace is too heavy and distorted. The image in MWG is blurry in the whole world, the fusion effect is the worst, and the local enlarged area is blurry, and there are many points with unclear boundaries. The CNN as a whole is seriously blurred and artifacts are blurred. The SESF-Fuse and CRF subtraction parts are relatively clear, but the edges are clearly identified by errors and incomplete images, and the contours of some edges have been weakened. DWT is better overall [21], but there are incomplete border images and artifacts in the local enlarged area. The boundary definition of the algorithm in this paper is visible, the outline texture is clearly visible and the effect is relatively excellent. Observation of the two sets of images in Fig. 5 and Fig. 6 shows that the fusion result on the color image set "Barbed Wire" is not bad.

Figure 7 Visualization of the fusion result.

As shown in Figure 7, the first line is a near-focused image, the second line is a far-focused image, the third line is a decision map, and the last line is a fusion result map.

This article uses 5 indicators: AVG (average gradient), STD (standard deviation), VIF (visual information fidelity), IFC and QABF, etc. As a fusion quality index to objectively evaluate performance. As shown in see Table 1., compared with the other 7 algorithms, the method proposed in this paper achieves the maximum value in most of the five index evaluations. The method in this paper only lags behind other algorithms in a few indicators, especially in IFC and QABF lag behind the CNN and SESF-Fuse methods. Then it ranks first in all other indicators. In general, in the objective level of evaluating algorithm performance, the method in this paper has a good advantage over the other seven algorithms [22]. The method in this paper has stronger generalization ability and can retain relatively high fidelity of visual information. Can retain more detailed and structural information and generate information-rich fusion images[23]

Table 1 Evaluation indicators of each method

Approachs	AVG	STD	VIF	IFC	QABF
CRF	7.7603	62.1742	0.69790	6.9580	0.7297
CVT	8.2652	62.1742	0.66717	5.5527	0.7113
S R	8.2142	62.2010	0.68320	6.0646	0.7298
IMF	6.3718	60.8008	0.59700	8.1700	0.6189
DWT	8.6385	61.6848	0.55608	4.7748	0.6839
CNN	8.2536	62.3058	0.70872	9.5648	0.7394

SESF	8.3444	62.4296	0.70894	9.5141	0.7374
Our	8.3472	62.4351	0.70896	9.5082	0.7374

In addition, an average time for image synthesis was done with other algorithms. The average time of image synthesis means the time for image fusion of the two source images at the time when the model of each algorithm is called. It can be seen from Figure 8 that the call time of the synthetic model in the CNN algorithm is the longest. Other algorithms do not have obvious advantages in the time of calling the model. At present, only SESF and the algorithm in this paper have relatively obvious advantages in this issue. Among the 8 algorithms, the advantage of this paper is the image synthesis that calls the model the most and takes the shortest time.

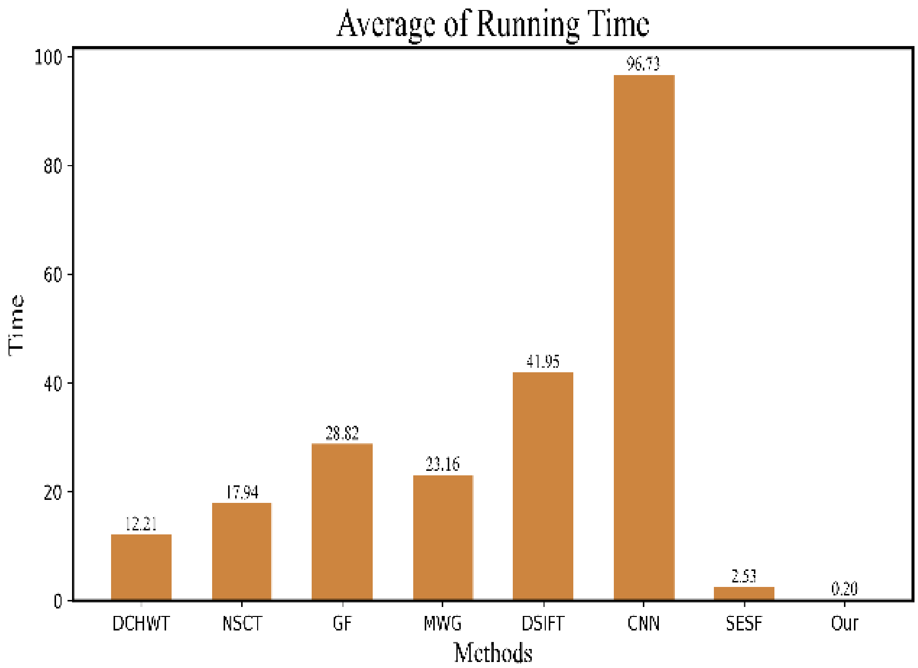

Figure 8 Average time chart of image synthesis

4. Conclusion

Aiming at the problem of multi-focus image fusion, an unsupervised deep learning model is proposed. The key point of the model in this paper is to train the encoder-decode network in an unsupervised way to obtain the deep features of the input image. Then the spatial frequency is used to calculate the activity level from these features, and the result of image fusion is obtained according to the decision diagram. Experimental results show that compared with the existing fusion methods, this method has better fusion performance. This article also demonstrates the feasibility of combining unsupervised learning with traditional image processing algorithms.

Acknowledgements

Fund Project: Sichuan Science and Technology Department Project(2020JDJQ0075, 2020YFSY0027);

Artificial intelligence Sichuan Key Laboratory Project(2020RZY02).

ISSN: 2664-9640

References

[1] Zhao Xin, Liu Huaixia. Multi-focus image fusion based on NSST-CS [J]. Journal of Xinyu University, 2020, 25(02): 37-43.

[2] Burt P, Adelson E (1983) The laplacian pyramid as a compact image code. IEEE Trans Commun 31(4):532–540.

[3] https://doi.org/10.1109/TCOM.1983.1095851.

[4] Lewis JJ, O'Callaghan RJ, Nikolov SG, Bull DR, Canagarajah N,(2007) Pixel- and region-based image fusion with complex wavelets. Inf Fusion 8(2):119–130. https://doi.org/ 10.1016/j.inffus. 2005.09.006 Special Issue on Image Fusion: Advances in the State of the Art.

[5] Li H, Manjunath B, Mitra S (1995) Multisensor image fusion using the wavelet transform. Graph Models Image Process57(3):235–245.

[6] https://doi.org/10.1006/gmip.1995.1022.

[7] Yang B, Li S (2010) Multifocus image fusion and restoration with sparse representation. IEEE Trans Instrum Meas 59(4):884–892.

[8] https://doi.org/10.1109/TIM.2009.2026612

[9] Aslantas V, Kurban R (2010) Fusion of multi-focus images using differential evolution algorithm. Expert Syst Appl 37(12):8861–8870.

[10] https://doi.org/10.1016/j.eswa.2010.06.011.

[11] Li S, Kang X, Hu J (2013) Image fusion with guided filtering.IEEE Trans Image Process 22(7):2864–2875.

[12] https://doi.org/10.1109/TIP.2013.2244222.

[13] Prabhakar R (2017) Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In: The IEEE international conference on computer vision (ICCV).

[14] Chen Qingjiang, Li Yi, Chai Yuzhou. A multi-focus image fusion algorithm based on deep learning [J]. Progress in Laser and Optoelectronics, 2018, 55(07): 246-254.Xu H, Fan F, Zhang H, Le Z, Huang J (2020) A deep model for multi-focus image fusion based on gradients and connected regions. IEEE Access 8:26316–26327.

[15] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D,Dolla ́r P, Zitnick CL (2014) Microsoft coco: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision—ECCV 2014. Springer, Cham, pp 740–755

[16] Juncheng Li, Faming Fang, Kangfu Mei, Guixu Zhang; Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 517-532.

[17] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity.IEEE Trans Image Process 13(4):600–612.

[18] Nencini F, Garzelli A, Baronti S, Alparone L (2007) Remote sensing image fusion using the curvelet transform. Inf Fusion 8(2):143–156. https://doi.org/10.1016/j.inffus.2006.02.001 (Spe-cial Issue on Image Fusion: Advances in the State of the Art)

[19] Zhou Z, Li S, Wang B (2014) Multi-scale weighted gradient-based fusion for multi-focus images. Inf Fusion 20:60–72.

[20] https://doi.org/10.1016/j.inffus.2013.11.005.

[21] Liu Y, Chen X, Peng H, Wang Z (2017) Multi-focus image fusion with a deep convolutional neural network. Inf Fusion36:191–207. https://doi.org/10.1016/j.inffus.2016.12.001.

[22] Ma, B., Zhu, Y., Yin, X. et al. SESF-Fuse: an unsupervised deep model for multi-focus image fusion. Neural Comput & Applic (2020).

[23] https://doi.org/10.1007/s00521-020-05358-9.

[24] Yan Yulu, Wu Jin, Deng Huiping, Song Can. Multi-focus image fusion based on CRF pre-segmentation and NSCT [J]. Television Technology, 2018, 42(01): 7-12.

[25] Liu Y (2019) Image fusion. http://www.escience.cn/people/liuyu1/Codes.html.

ISSN: 2664-9640

[26] Riani M, Simonotto E (1994) Stochastic resonance in the per-ceptual interpretation of ambiguous figures: a neural network model. Phys Rev Lett 72(19):3120

[27] Liang Cong, Tang Zhenhua. A multi-focus image fusion algorithm based on local energy in DCT domain[J]. Computer Applications and Software, 2016, 33(05):235-238.King ma DP, Ba J (2015) Adam: a method for stochastic optimization. In: International conference on learning representations.

[28] Wang Changcheng, Zhou Dongming, Liu Yanyu, Xie Shidong. Multi-focus image fusion algorithm for unsupervised deep learning model [J/OL]. Computer Engineering and Applications: 1-12 [2021-02-10].