Convolutional Neural Network Based Multi-Focus Image Fusion

Huaguang Li
Information College
Yunnan University
Kunming 650504, China
huaguangli@mail.ynu.edu.cn

Rencan Nie
Information College
Yunnan University
Kunming 650504, China
rcnie@ynu.edu.cn

Dongming Zhou
Information College
Yunnan University
Kunming 650504, China
zhoudm@ynu.edu.cn

Xiaopeng Gou
Information College
Yunnan University
Kunming 650504, China
657717191@qq.com

ABSTRACT

In this study, this paper mainly focuses on the use of convolutional neural network (CNN) to improve the clarity of multi-focus image and the fusion effect. First, set up an image picture data set and we convert labels tags to the dataset into binary images. After that , then use the CNN network to train the established datasets. In the end, we need to learn a direct mapping between source and focus map.

CCS Concepts

Applied computing~Cartography

Keywords

Multi-focus image fusion；Image fusion；Convolutional neural network；Deep learning；Fusion rule

1. INTRODUCTION

More focus on image fusion will focus on the target of different source image fusion processing [1], from the source image clear part eventually in the synthesis of different goals are in the same scene, clear images, thus more comprehensive and fairly reflect the scene information, conducive to the human eye observation to make accurate analysis and understanding of the images [2], according to the characteristics of the more focused image, more focus on image fusion method based on CNN, compared with some classical algorithms improved to focus on efficiency and performance of image fusion, can extract the image feature of a more effective, more suitable for image fusion [3,24]. With the rise of deep learning, the deep learning algorithm has shown great potential in the field of computer vision.

In the image fusion algorithm can be divided into two kinds, respectively is spatial domain algorithm and transform domain algorithm These algorithms can be mainly classified as the energy of Laplacian (EOG) [4],such as the spatial frequency (SF) [11],Some representative examples include the morphological pyramid (MP)-based method [5], the discrete wavelet transform (DWT)-base method [6], the dual-tree complex wavelet transform (DTCWT)-based method [7,25]. the source image in the spatial domain algorithm through the linear combination of the direct integration, these algorithms can be roughly divided into: based on region, based on the block, based on pixels, and so on.

However, using these algorithms usually leads to the appearance of the fused images some adverse side effects such as fuzzy phenomenon, at the integration of boundaries and lose some of the original image information. Transform domain algorithm in more focus is very important in image fusion algorithm [3], this method is characteristic of the original image into another domain, we put forward in this paper is now very popular image fusion algorithm based on CNN[3,6,11].This method mainly includes: making data sets, establishing CNN model, training data set and image fusion. Finally, we can get a better fusion image.

ICACS '18, July 27–29, 2018, Beijing, China

ACM ISBN 978-1-4503-6509-3/18/07…$15.00
https://doi.org/10.1145/3242840.3242863

2. CNN MODEL FOR MULTI-FOCUS IMAGE FUSION

2.1 CNN model

The deep of the convolutional neural network (CNN) is a classic learning network model in the image semantic segmentation field by pixel - level classification[5-8]. In addition, a large number of field use of CNNs [3] algorithm is effectively achieve the desired purpose such as speech recognition, text processing, and other fields. As is known to all, CNN is a multi-stage feed forward neural network that can be trained, with a certain number of feature mappings at each level [5]. Each unit in the feature graph is called a neuron, and each neuron is connected to a different feature graph by means of linear convolution, nonlinear activation, and spatial pooling.The basic framework of CNN includes three parts: local receiving domain, weight sharing and sub-sampling. The first is that a neuron is connected to a neighboring spatial domain, which is very similar to the mammalian visual cortex. The second part is that the space of the kernel will not change during the feature mapping and the number of weights will be greatly reduced. The third part mainly focuses on sub-sampling, and pooling can reduce the dimension of data. Generally, our maximum pooling and average pooling are used by us [6]. The sub-sampling is also known as polling, which can reduce data dimension, Maximum pool and average pool is often used method in the CNN model, because the average pooling can decrease limited because field size estimation variance caused by the increase, and the ability to retain more of the background of the image information, so we choose the average pooling is applied in our CNN model, the algorithm is as follows

$$j = \arg\min \| x_i - d_x \|_2^2, \qquad (1)$$

$$h_m = \frac{1}{|N_m|}\sum_{i\in N_m}\alpha_i, \qquad (2)$$

$$\alpha_i \in \{0,1\}^k, \ \alpha_{ij} = 1,$$

Let x_i denote the i-th input feature map of a convolutional layer, where k is the convolutional kernel ,and $\alpha_{i,j}$ is the neuron , and effect, because the CNNS has powerful GPU computing power and effective training conditions and can carry on the effective processing of large data [9].

2.2 CNNs for multi-focus image fusion

According to the above introduction, we know that the activity level measurement is known as feature extraction, while the role of fusion rule is similar to that of a classifier used in general classification tasks. Thus, it is theoretically feasible to employ CNNs for image fusion. The CNN architecture for visual classification is an end-to-end framework [9], and CNNS has a good effect on the fusion of focusing images. Therefore, we use d_x is the bias.when there are M input maps and N output maps,this layer will contain N 3D kernels of size $d \times d \times M$ and each kernel owns a bias. In the past few years, CNNS applied in various areas especially in the field of speech processing and computer vision, such as: speech recognition, face recognition, image fusion, etc., which based on the research of CNNS are more efficient than the traditional method with better

CNNs method for multi-focus image fusion. First of all, we use the method of semantic segmentation of all pixels in the image tag, using digital target different Numbers represent different marks [10-13], we will work this part is called image segmentation, we put the zero mark is a focus area, with the number 1 is marked as focus area, we can understand it into binary classification problems. Through the database we create a lot more focused image, in the depth of the powerful learning framework PYTORCH it is easy to build a complex network training we want CNNs model, so compared with the traditional algorithm we use CNNs algorithm can get better more focused image fusion effect [2].

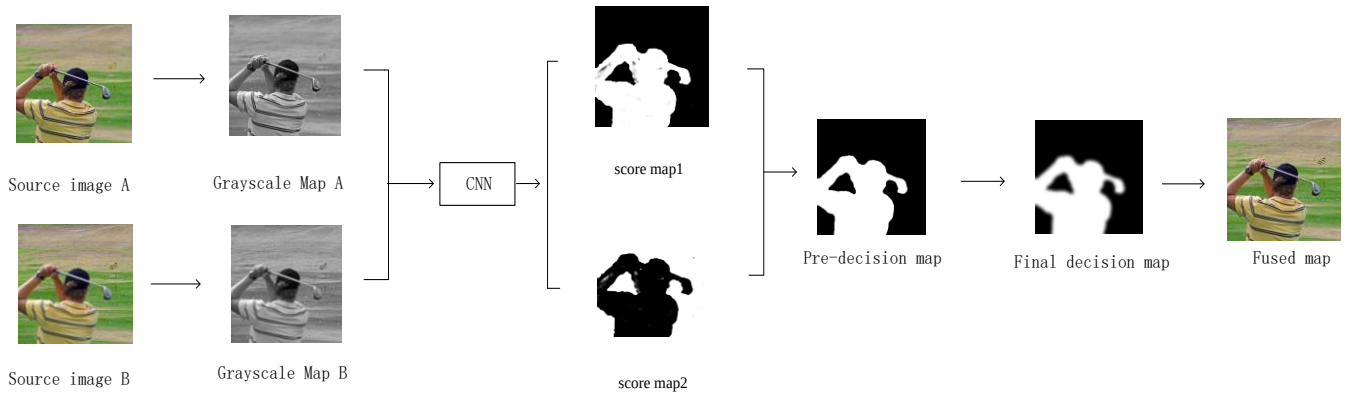


Fig. 1. Our schematic diagram of the proposed CNN-based multi-focus image fusion algorithm

Fig. 2. Some pairs of artificial multi-focus images and theirs corresponding labels

3. THE PROPOSE METHOD

A schematic diagram of the CNN-based multi-focus image fusion algorithm is illustrated in Fig.1.Obviously, there are only two source images in Fig1.However, we can apply this algorithm to more than two source images. It can be seen from Fig. 1 that our algorithm includes five steps. First of all, We first gray scale the source image for the next operation,shown in Grayscale Map A

and Grayscale Map B . CNN model is trained via our artificial multi-focus image dataset. Second,Then a pair of multi-focus images is fed to the trained model and output one score maps.We add corresponding pixels of score map 1 and score map 2 and output focus map . Accordingly, we compare the score map 1 and score map 2 pixel-wisely and get the binary focus map shown in pre_decsion map . Finally, we reverse the pre-decision map and take away the max areas, other areas will be cancelled. Using the reversal operation again, the final decision map will be generated shown on final decision map . With this final decision map, the fused image will be produced. The detailed procedures are illustrated as follows.

3.1 Network Design

In this section, we have the image patch with the same scene, S1, S2, and a CNN with the output range of 0 to 1.That is to say, when the S1 is focus S2 the output value should be close to 1, when the focus S1 to S2 is focused on the output value is close to 0. Some pairs of artificial multi-focus images and theirs labels are showed in Fig.2.In short, the output value shows that the attribute of this pair of image patch. Finally, we used a lot of image patches to train the data. We define a positive example when S1 is clearer than S2, its label is labeled 1.On the contrary, we define a negative example, when clear than S1 S2 we marked as 0. Multi-focus images with arbitrary space size in practical application, and a kind of method is sliding window method is applied to separate the image overlapping patches.as was applied in [6,9,13],We get a score for each pair of image patches on the network. This path-based approach is very consuming, considering that there is a large number of overlapped image patches. The other approach is to create a dense prediction map of the original image of the input network as a whole without separating them into patches [8,15].Therefore, both the input and output data of the full connection layer have a fixed size, which is converted into a volume set layer by reconstructing parameters in the full connection layer. After the transformation, the network only has the convolution layer and the maximum pooling layer, so the network can process the source image of any size to produce intensive prediction [12,16,17].Since Siamese is more natural in image fusion, the method of feature extraction and activity level testing in two equally weighted branches is identical to the source diagram. In addition, Siamese can effectively reduce the problem of small space. So we chose Siamese type network to apply to the method based on CNN image fusion.

In [10],Another important issue is the size of the patches required for web design. When we set up 32 * 32 size due to the use of more of the image content classification accuracy is relatively high as its defects can't be ignored, because patch size setting of 32 * 32 biggest pooling layer is not easy to be defined, when there are two or more of the biggest in branch pooling layer patch the pace of at least four pixels, the fusion results will block the effects of artifacts. When there is only one maximum pool layer in a branch, the size of the CNN model is usually large, because the value of the weight of the full connection layer increases significantly. Training CNN's patch is too small to ensure the accuracy of the classification when we select a patch size of 8*8.So we selected 16*16 size patches for research. In this training network has five convolutional layers and three max-pooling layer. The kernel size and stride of each convolutional layer are set to 3 × 3 and 1, respectively. The kernel size and stride of the max-pooling layer are set to 2 × 2 and 2, respectively. The 256 feature maps and then have two fully-connected layers with a 256-dimensional feature vector and other is a 128-dimensional feature vector. The output of the network is a 2-dimensional vector that is fully-connected with the 128-dimensional vector. Actually, the 2-dimensional vector is fed to a 2-way softmax layer which produces a probability distribution over two classes.

3.2 Training

We do with the VCO2012 datasets with 10000 pieces of high quality natural images [17], for each image must first be converted into gray space, and then, we use the standard deviation of 2 of gaussian filter to process data set for the first of two gaussian filter for the first time after processing by fuzzy images and then repeated processing, get the data set we need. At last, we randomly divided each original picture and fuzzy picture into 16*16 size patches, and we got 20,000 patches from the dataset.

we define our loss function as soft max loss. The optimization method we use in this work is stochastic gradient descent (SGD). The base learning rate is set to 1e-6. The momentum and weight decay are set to 0.9 and 0.0005, respectively. The weights are updated with the following rule

$$v_{i+1} = 0.9 \times v_i - 0.0005 \times \alpha \times \frac{\partial L}{\partial w_i}, \qquad w_{i+1=} w_i + v_{i+1}, \quad (4)$$

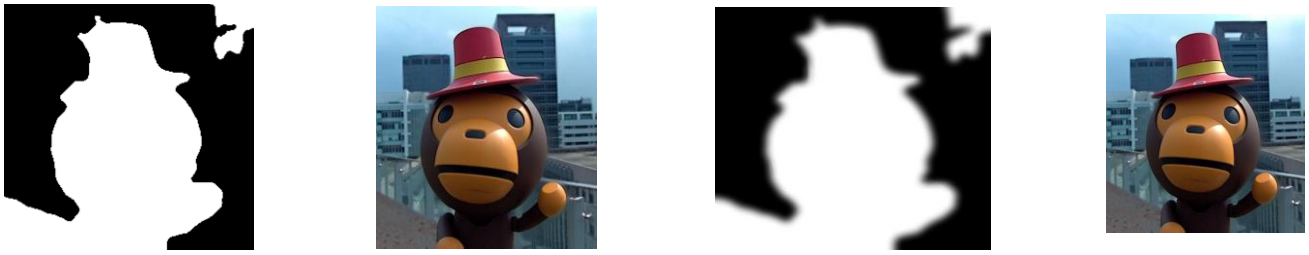
Where v is the momentum variable, i is the iteration index,α is the learning rate,L is the loss function,and $\frac{\partial L}{\partial w_i}$ is the derivative of the loss with respect to the weights at w_i . We use the “step” learning policy to train the model and the parameter gamma is set to 0.5. We train our CNN model using the popular deep learning framework Caffe [18]. Our CNN model contains convolution layers, pooling layers, relu layers and the SUM layers. At every convolution layers, the learning rate and the decay rate are both set 1.The bias’s learning rate and the decay rate are set to 2 and 0, respectively. In pooling layers, we use the max pooling method to subsample the feature maps.

3.3 Fusion Criteria

In this part , we discuss the detail of the fusion criteria. As shown in Fig. 1, when a pair of natural multi-focus image is fed to the trained CNN model, we can observed the score maps will be produced， at the same time we let s1, s2 ,s3 denote the score map 1, score map2,score map3.We can see that s1 and s2 focus the man and the ball park roughly. Though in the s1 the boy is also partly focused and the ball park is partly unfocused, if we add s1 and s2 pixel-wisely, the score map3 will be produced. It can be seen that in s3 the man and the ball park is totally focused. However, the man is still focused partly. Observation finds that s1 and s2 are approximately complementary focused images from the same scene. we compare s1and s2 pixel-wisely, We define s1(x,y) as 1 when s1 is higher than s2 and We define s1(x,y) as 0 when s1(x,y) is low than s2(x,y). so, we can get the pre-decision map. However, In reality pre-decision map will still appear the label error phenomenon. In order to make more clear image edge, we also implement the morphological operations. dilation to produce the final decision map. Using the final decision map and the source map, the fused map can be generated from the formula as follows:

$$F(x,y)=D(x,y)S(x,y)+(1-D(x,y))S_1(x,y) \quad (3)$$

Our example in Fig. 3 shows the concrete operation.



Initial decision map | Initial fused image | Final decision map | Fused image

Fig. 3.Our formula diagram of the fused image

4. EXPERIMENTS

We evaluate the proposed CNN algorithm on two image datasets. where the pairs of color multi-focus images of size 520*520 pixels .Some examples of the first one is shown in Fig. 4. In this work, we compare our proposed CNN algorithm with five representative multi-focus image fusion algorithms, which are the multi-focus image fusion based on discrete wavelet transform(DWT)-based one[20],The Laplace pyramid transformation based fusion method(LPT)-based one[21], the image fusion based on Non-lower sampling contour wave transform(NSCT)-based one[19], the Unsampled shear wave transform based fusion for multi-focus images(NSST)-based one[22],the image fusion based Pulse Coupled Neural Network (PCNN)-based one[23].To compare with these methods, firstly, we obtain the fused image via using the original code provided by theirs authors. Then using the objective quality metrics illustraed as follows, we can measure the quanlity of different method quantitativly.

5. OBJECTIVE QUALITY METRICS

In order to verify the rationality of the parameters and the fusion strategies in our proposed method, several experiments were conducted on the image sets, At the beginning, four groups of different fusion strategies are compared, Group1:information entropy(IE)[19],It is used to measure the information richness in the fusion image. This index does not care about the relationship between the fusion image and the source image, but only focuses on the gray distribution of the fusion image itself.Group2:average gradient(AG) [6],It's similar to SF.Group3:standard deviation(SD)[22],application in image quality assessment can measure the richness of image information.Group4:spatial frequency (SF)[19],It is applied to reflect the rate of change of image gray scale and reflect image clarity, generally, the clearer the image, the higher the spatial frequency.

5.1 Fusion Results and Discussions

In the first experment, we evaluate our proposed algorithm in the gray dataset. As shown in Fig. 5, In order to verify the algorithm has better convergence in multi-source image fusion effect, this experiment adopts has good registration “clock” more focused image as the input source images, in order to make the comparison results more clearly, this article will compare significantly enlarge the details of the operation so that we can better observe shown in Fig.6.It can be seen from the figure that the image fusion using DWT algorithm is not clear, especially the edge part. Virtual shadow also exists in image edge using LPT algorithm. NSCT and NSST are adopted to realize information fusion basically, but virtual shadows also exist. The edge effect of the graph implemented by the algorithm in this paper is better, and the brightness and saturation are also higher.

On the above experiments, this article also from the IE, AG, SD, SF, four objective index to compare the proposed algorithm, the experiment of the experimental data are shown in Table.1 below, to make it more intuitive, we added the bar graph in Fig.7, from the table we can see IE algorithm in this paper, such as SD index is significantly higher than other algorithms. Therefore, the fusion of this algorithm is better.

Fig. 4 some example from the multi-foucs image dataset

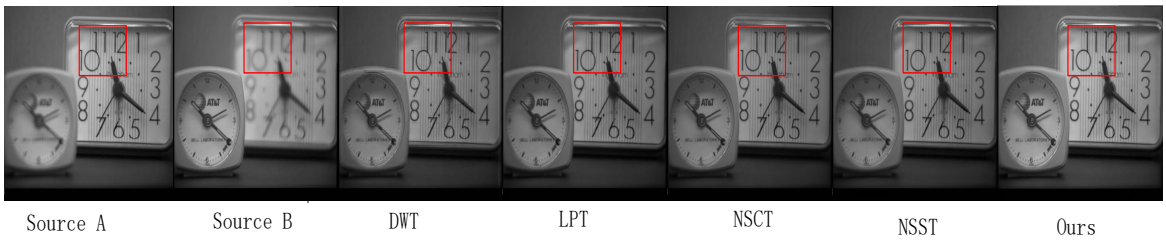


Fig. 5 The "clock" source image pair and the fused image obtained via various fused algorithms

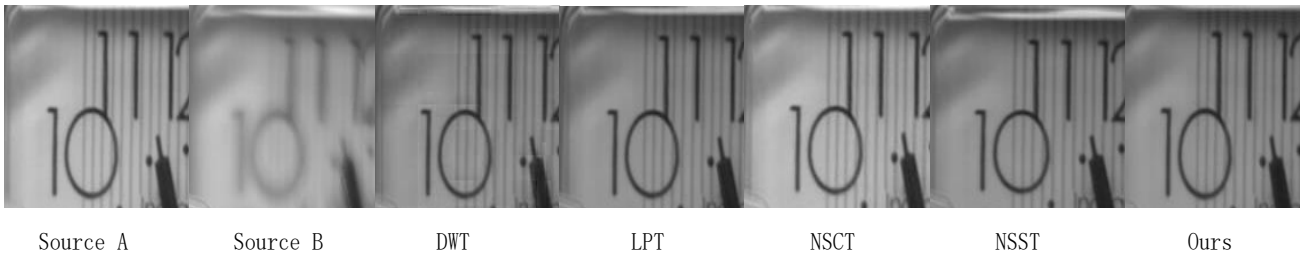


Fig.6 The magnifed regions in the red rectangles in the Fig.5

Table.1 Objective assessment of various fusion algorithms for grayscale image dataset.（Best results are in bold.）

	IE	AG	SD	SF
DWT	7.0576	3.4230	41.0685	8.2859
LPT	7.0481	3.2057	41.1402	7.9901
NSCT	7.0197	3.1882	41.1203	8.0377
NSST	7.0197	3.2853	40.8655	8.1314
Ours	7.0616	3.3327	41.6960	8.3218

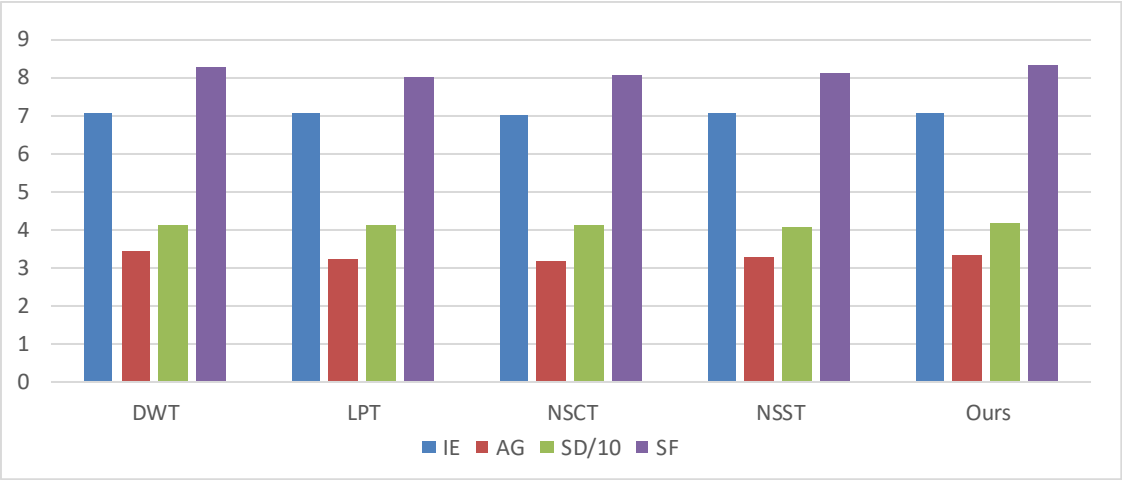


Fig.7 The average score for four metrics obtained by using various fused algorithms

Fig.8 more results using our proposed algorithm

6. CONSLUSIONS

In this paper,we presented a novely multi-focus image fusion algorithm based on the CNN.The proposed algorithm learned a direct mapping between source images and the focus map.The post-treatment of our proposed algorithm is so little.Compared with other methods, the experimental results show that the CNN structure is simple which has fewer parameters to set, low computational costs, and objective is outstanding in the fused image, the outline is clear, rich background details in the fused image, the fusion performance is better than the other state-of-the-art methods both the subjective and objective evaluation.

7. REFERENCES

[1] Aslantas V, Kurban R. Fusion of multi-focus images using differential algorithm[J].Expert Systems with APPlications,2010,37(12):886-8870.

[2] GanasalaP, Kumar V.CT and MR image fusion scheme in nonsubsampled contourlet transform domain[J].Joumal of Diging,2014.27(3):407-418

[3] Yu Liua , Xun Chena,* , Hu Penga , Zengfu Wang b Multi-focus image fusion with a deep convolutional neural network [J] Information Fusion 2016.12(12):341-8960

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

[5] https://en.wikipedia.org/wiki/Deep_learning

[6] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891 – 1898.

[7] A. Toet, A morphological pyramidal image decomposition, Pattern Recognit. Lett. 9 (4) (1989) 255 – 261.

[8] H. Li, B. Manjunath, S. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models Image Process. 57 (3) (1995) 235 – 245. [8] J. Lewis, R. O ī Callaghan, S. Nikolov, D. Bull, N. Canagarajah, Pixel- and regionbased image fusion with complex wavelets, Inf. Fusion 8 (2) (2007) 119 – 130.

[9] X. Qu, J. Yan, H. Xiao, Z. Zhu, Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain, Acta Autom. Sin. 34 (12) (2008) 1508 – 1514.

[10] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: International Conference on Artificial Intelligence and Statistics, 2010.

[11] V. Petrovic, V. Dimitrijevic, Focused pooling for image fusion evalution, Inf. Fusion 22 (1) (2015) 119 – 126.

[12] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, W. Wu, Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study, IEEE Trans. Pattern Anal. Mach. Intell. 34 (1) (2012) 94 – 109.

[13] M. Hossny, S. Nahavandi, D. Creighton, Comments on information measure for performance of image fusion, Electron. Lett. 44 (18) (2008) 1066 – 1067.

[14] C.S. Xydeas, V.S. Petrovic, Objective image fusion performance measure, Electron. Lett. 36 (4) (2000) 308 – 309.

[15] C. Yang, J. Zhang, X. Wang, X. Liu, A novel similarity based quality metric for image fusion, Inf. Fusion 9 (2) (2008) 156 – 160.

[16] Y. Chen, R. Blum, A new automated quality assessment algorithm for image fusion, Image Vis. Comput. 27 (10) (2009) 1421 – 1432.

[17] http://pjreddie.com/projects/pascal-voc-dataset-mirror

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the ACM International Conference on Multimedia, 2014, pp. 675 – 678.

[19] Hou, R., Nie, R., Zhou, D., Cao, J., & Liu, D. (2018). Infrared and visible images fusion using visual saliency and optimized spiking cortical model in non-subsampled shearlet transform domain. Multimedia Tools & Applications, 1-24.

[20] Li H, Manjunath BS, Mitra SK (1995) Multisensor image fusion using the wavelet transform. Graph Model Image Process 57(3):235 – 245

[21] [21]Toet A (1989) Image fusion by a ratio of low-pass pyramid. Pattern Recogn Lett 9(4):245 – 253

[22] Kong W, Wang B, Lei Y (2015) Technique for infrared and visible image fusion based on non-subsampled shearlet transform and spiking cortical model. Infrared Phys Technol 71:87 – 98

[23] Yang, Y., Que, Y., Huang, S., Lin, P.: Multimodal sensor medical image fusion based on type-2 fuzzy logic in NSCT domain. IEEE Sens. J. 16, 3735 – 3745 (2016)

[24] Hongxia Wang , Bangxu Yin. Watermarking-Based Blind QoS Assessment for Wireless Image Communication, Journal of communications, 8(3),207-215(2013)

[25] Moad Y. Mowafi, Fahed H. Awad, Eyad S. Taqieddin, Omar Q. Banimelhe, A Practical Study of Jointly Exploiting Multiple Image Compression Techniques for Wireless Multimedia Sensor Networks,in: Journal of communications , 7(4), 309-320(2012)