BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 70(1), 2022, Article number: e140352
DOI: 10.24425/bpasts.2022.140352

The automatic focus segmentation of multi-focus image fusion

K. HAWARI[1,2] and ISMAIL[1*]

[1] Universiti Malaysia Pahang, Faculty of Electrical and Electronics Engineering,26300 Kuantan, Malaysia
[2] Politeknik Negeri Padang, Electrical Engineering Department, 25162, Padang, Indonesia

Abstract. Multi-focus image fusion is a method of increasing the image quality and preventing image redundancy. It is utilized in many fields such as medical diagnostic, surveillance, and remote sensing. There are various algorithms available nowadays. However, a common problem is still there, i.e. the method is not sufficient to handle the ghost effect and unpredicted noises. Computational intelligence has developed quickly over recent decades, followed by the rapid development of multi-focus image fusion. The proposed method is multi-focus image fusion based on an automatic encoder-decoder algorithm. It uses deeplabV3+ architecture. During the training process, it uses a multi-focus dataset and ground truth. Then, the model of the network is constructed through the training process. This model was adopted in the testing process of sets to predict the focus map. The testing process is semantic focus processing. Lastly, the fusion process involves a focus map and multi-focus images to configure the fused image. The results show that the fused images do not contain any ghost effects or any unpredicted tiny objects. The assessment metric of the proposed method uses two aspects. The first is the accuracy of predicting a focus map, the second is an objective assessment of the fused image such as mutual information, SSIM, and PSNR indexes. They show a high score of precision and recall. In addition, the indexes of SSIM, PSNR, and mutual information are high. The proposed method also has more stable performance compared with other methods. Finally, the Resnet50 model algorithm in multi-focus image fusion can handle the ghost effect problem well.

Key words: deep learning; ResNet50; multi-focus image fusion.

1. INTRODUCTION

The multi-focus image fusion is a sort of image processing field. It is a process that patchworks sharper regions of the same scene camera images [1]. The process increases image quality, reduces randomness and redundancy to intensify the usability of implementation. All the focus information is placed in one image, termed a fused image. The focus information is the scene information that occupies the zone of a depth of field (DOF) camera lens. Otherwise, the scene information beyond the DOF zone is known as blurred information. The multi-focus image fusion process is shown in Fig. 1 below. It contains two images. They have different focus regions from each other. Image 1 has a focus region at the bottom part, whereas image 2 has a focus region at the top part. The fused image has focus regions for all image surfaces.

The clear information image, free of blurred parts and noises is very important in many fields such as medical diagnostics [2], robotics, and remote sensing [3, 4]. As multi-focus image fusion is important, there are a lot of techniques in multi-focus image fusion available nowadays. However, scientific articles distinguish mostly the methods between spatial and spectral domains [5]. The spatial domain method directly converts the focus regions of camera images into a focus map. This method

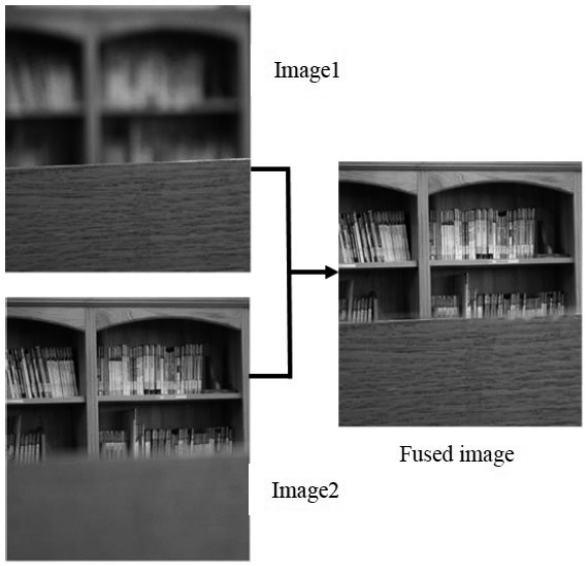

Fig. 1. Multi-focus image fusion process

can be found in generative gradient maps to generate focus maps [6], the region-based kernel to generate focus maps [7] and the use of normalized generative gradient can generate focus maps [8].

In other words, the spectral domain method transforms the multi-focus images into the frequency domain, fuses the images, and inverses the image back to the spatial domain again.

*e-mail: ismail@pnp.ac.id

Manuscript 2021-04-28, revised 2021-09-05, initially accepted for publication 2021-09-10, published in February 2022.

© 2022 The Author(s). This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

This method is applied in spectrum comparison [9], steerable local frequency [10], and wavelet and guided image filter [11]. Furthermore, a few other articles mention some additional categories, such as deep learning and hybrid [12]. The multi-focus image fusion based on the deep learning method can be found in using deep CNN to obtain focus maps [5], the use of pixel-wise to distinguish focus and blur [13], using ensemble learning to CNN [14] and U net [15]. The last is the hybrid method. The hybrid method is a combination of spatial domain and spectral domain. The hybrid method is implemented in multiscale Top Hat and dual sliding windows to generate accurate focus maps [16], using multiscale and a multi-direction neighborhood filter to obtain focus maps [17] and using multiscale analysis filter to obtain initial focus maps and block consistency evaluation to generate focus maps [18].

The multi-focus image fusion process enhances the camera image perfectly and the final image has no blurred and redundant pixels problems. There are some common problems in multi-focus image fusion, such as rice field effects, ghost effects, etc. This effect is caused by an inaccurate prediction of the focus map. The very common features to detect focus maps are variance, pixels activities, pixel gradient, and saliency. The variance of the focus region is higher than the blurred region of the image [19]. The focus region also has higher pixel activities rather than the blurred region [7]. The saliency of the focus region is used to predict focus maps [20] and the last is the using pixel gradient to detect the focus map [21]. The pixel gradient of the focus region can detect the focus pixel accurately, but under certain conditions, it failed. The result still left the ghost effect. This problem makes the fused image become low quality and lose some important information. To eliminate all unwanted effects, the proposed method uses a deep learning algorithm. The method uses a dataset and ground truth to learn the model. The rest of the paper consists of Section 2 *The deep learning principle*; Section 3 *The proposed method*; Section 4 *The experiment and result*; Section 5 ends with *Conclusions* (Fig. 2). The architecture of deep learning [22]

2. DEEP LEARNING

2.1. Principle

Deep learning is a sophisticated computational model that consists of many processing layers to learn the feature of data through a lot of layers abstraction [22]. Deep learning is a part of machine learning with a bit different operation. Deep learning is an end-to-end learning process. This algorithm can solve the problems such as identification of the object of images, converting audio to text, texture recognition [23], and failure detection in robotic applications [24]. The convolution neural network is common in deep learning architecture. It has a series of layers. The beginning layers are convolutional and max-pooling layers. The architecture is shown in Fig. 2.

The unit in the convolution layer regulates the feature map. Each unit connects local patches in the feature map to the previous layer through a set of weights called filter banks. A result of this local weight sum is then passed through a non-linear function, such as ReLu. The same filter bank is utilized by all units in a feature map. A different filter bank is used by a different feature map in a layer. This architecture is termed twofold. The convolutional layer has the task to detect local conjunction of features from the previous layer. Otherwise, the max-pooling layer has a role as an operator to merge semantically similar features into one.

There are many types of applications of deep learning nowadays. It successfully detects, recognizes, and predicts tasks of objects and surrounding in an image. Consequently, deep learning is applied in many fields of technology such as autonomous mobile robots and self-driving cars, in which they need accurate data, especially in an image form.

2.2. Residual network (Resnet50)

A deep convolution neural network has become a powerful method to generate image classification [25]. A deeper network does not guarantee increasing learning accuracy. A deeper network makes saturation of accuracy happen and then degradation increase rapidly.

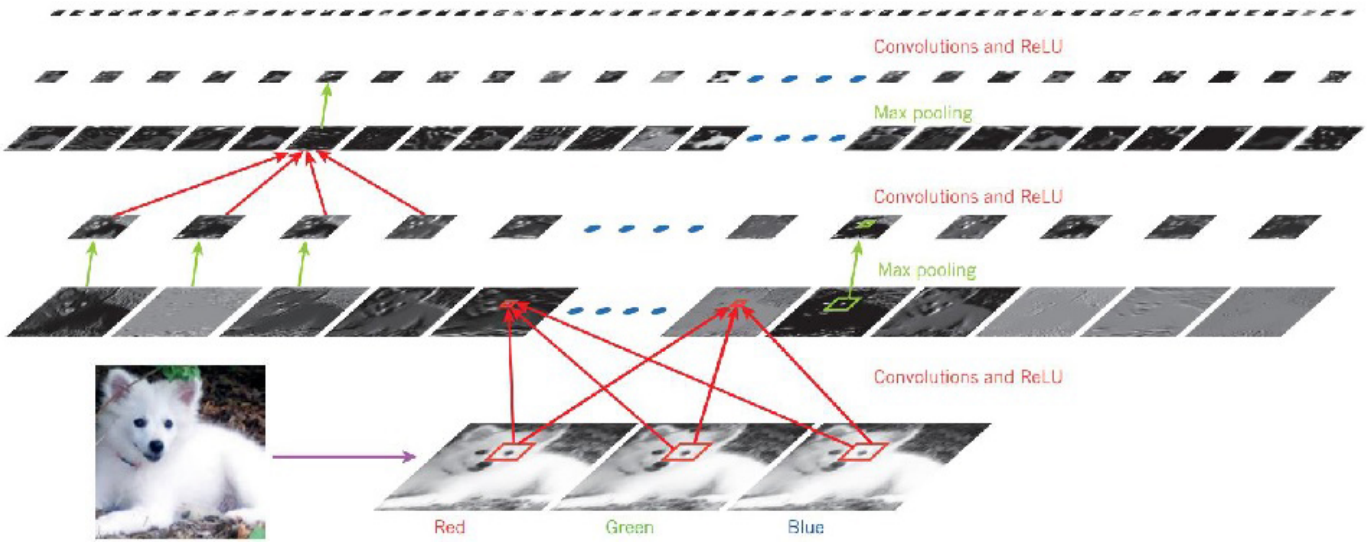

Fig. 2. The architecture of deep learning [22]

The most important thing is that the problem does not connect to overfitting. Commonly, the increase in deeper layers makes a new training error grow even bigger. To handle the problem, the network applies an additional identity mapping. The other layers are copied from the learned shallowed model. It is a mapping of the input node into the output node. The details of the residual network (ResNet50) are depicted in Fig. 3. Figure 3 shows that x is an input vector, y is an output vector. The function $F(x\{W\_i\})$ represents the residual mapping to be learned. Mathematically, equation (1) for the block is

$$y = F(x, \{W_i\}) + x. \tag{1}$$

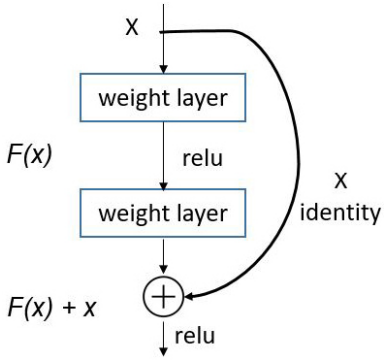

Fig. 3. The residual network architecture block with identity mapping

Based on Fig. 2, detail equation (2) can be drawn

$$F = W_2\sigma(W_1 x), \tag{2}$$

where σ is a ReLu function.

The operation $F + x$ is performed by short cut connection and pixel-wise addition. The non-linearity is performed after the addition or ReLu operation. For instance, there is an equity between input and output channel (e.g. F and x are equal), then the shortcut connection can be formulated as in equation (3) below

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

where W_s is a linear projection matrix and used if both dimension input and output are the same. F is a residual function, and it is flexible, it can have two, three layers, and more. Nevertheless, if F has only a single layer, equation (1) becomes linear and simplified as

$$y = W_1 x + x. \tag{4}$$

The element-wise addition can be carried out through channel by channel.

3. THE PROPOSED METHOD

The proposed method uses a residual network with 50 blocks (Resnet50). The Resnet50 has 177 layers and 192 connections. The proposed method uses a grayscale dataset and labels images. Figure 4 shows the details of the proposed method architecture.

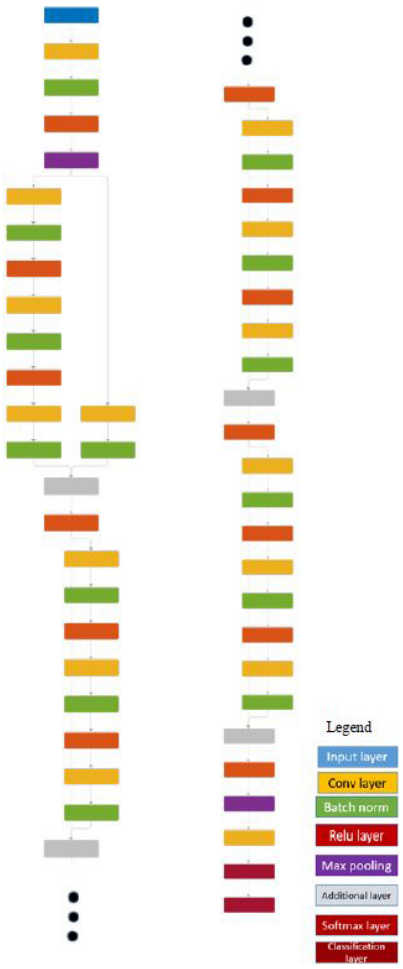

Fig. 4. The Resnet50 architecture

The proposed method uses a common dataset [26] and a homemade dataset. The original network of Resnet50 is a kind of classification model, in which the output of the network is the name of an object, name of region, text, etc. It is a vector. To obtain the spatial information of the input image, the output image must have a similar size to the input image. The proposed method creates a deep lab, it is a *deeplabv3 plus*. This layer is generated by involving a deep neural network and this function is a Matlab function. The Deeplabv3 plus consists of a convolutional neural network, delated convolution, and skip connection for semantic image segmentation [27]. As shown in Fig. 5, it is similar to encoder-decoder architecture, Atrous convolution, and fully connected CRFs [28].

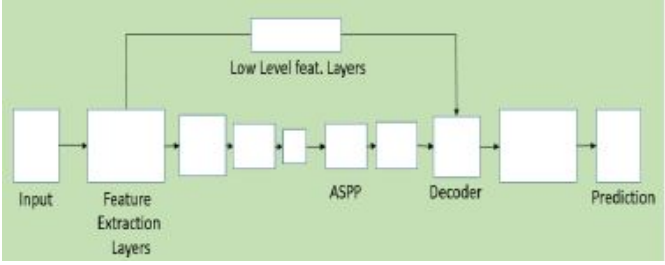

Fig. 5. The general architecture of DeepLabV3+ [27]

The deeplabv3plusLayers needs the input size, number of classes, and the applied network model. The proposed method uses a pre-trained model of residual network with 50 blocks (Resnet-50).

4. EXPERIMENT AND RESULT

The proposed method has two main steps. The first is generating the deeplabv3+ architecture with Resnet-50 pre-trained model inside. This architecture is a model for the semantic focus segmentation process.

The Generating of deeplabv3plus network with residual network architecture needs the image size, number of classes, and the selected pre-trained model. The proposed method uses Resnet 50 model in features extraction of deeplabv3plus architecture. The time needed to process the forming of *deeplabv3 plus* network is a few seconds.

The second stage is training to generate a deeplabv3plus network with the dataset. This procedure involves parameters such as dataset, *deeplabv3 plus* architecture, and training variables. The training variables are as follows: the optimization method is Adam, mini-batch size is 8 and maximum epoch is 3. This training process requires a bit more time. For the above parameters, it requires about a few minutes. The model is a generalization of the weight of the focus map with input multi-focus images. Then this model is applied to do focus segmentation of the test set.

Since the deeplabv3plus uses a pre-trained model, this transfer-learning model does not need a huge number of a dataset to generate a semantic focus segmentation model. The dataset consists of 16 images or 8 pairs of multi-focus images. They are divided into the training set and test set with 8 images, respectively. The training set and test set have 50% composition. The training dataset is composed of multi-focus images and ground truth.

The detailed dataset structure composition is a very common organization. The multi-focus images are placed in the form of collection images, called image datastore. Otherwise, the image labels are placed in another datastore, called label datastore. It is in the array data object formation. The multi-focus image and ground truth are processed to develop an image label datastore. This data store is used to support semantic focus segmentation using deep learning [29].

The process of building a semantic focus segmentation model is very flexible. Many properties should be clear to use. Nevertheless, by a few arrangements of the properties training process, it also produces an accurate model. It has some properties, but in this operation, the two selected properties are input size and color operation type. This datastore is in form of a categorical label. The training set is as shown in Fig. 6. The label pixels are the ground truth of images. The black label shows the focus region, otherwise, the white label denotes the non-focus region of multi-focus images.

The generating of deeplabv3+ involves input size, several classes, and a selected pre-trained network. The proposed method selects Resnet-50. The task uses parameters such as max epoch, mini-batch size, and optimization method. They

Fig. 6. Training set

are 20, 5, and 'sdgm', respectively. It does not use a validation set and augmented data. This process produces deeplabv3+ architecture. It is useful for semantic segmentation using deep learning. Matlab facilitates all the processes needed to build deeplabv3+, such as removing and updating the pre-trained model layers. Matlab also makes the output size equal to the input size.

To generate deeplabv3+ architecture, Matlab provides the function *deeplabv3+Layers* [27]. This function requires three main variables. They are input size, number of the class label, and the selected pre-trained network. The proposed method determining the input size is 512×512, the number of classes is two and Resnet-50 is the selected pre-trained network. The deeplabv3+ architecture is shown in Fig. 7.

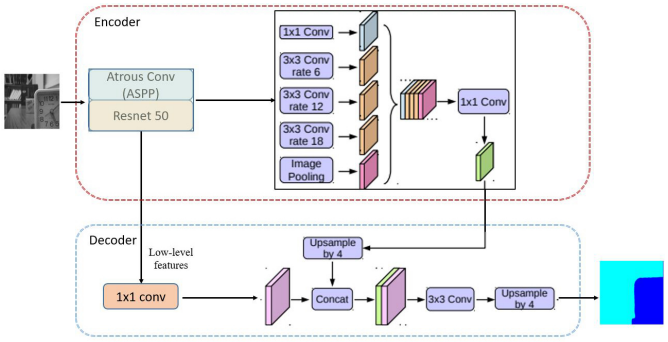

Fig. 7. The *deeplabv*3+ architecture of the proposed method

The second step is training the network to obtain the model. This operation needs some variables. They are pixel-labeled datastore, layers, and training properties. The pixel label data-

store is a categorical label assigned to a pixel at a certain location. The layer is a convolutional neural network for semantic segmentation. It is deeplabv3+ with Resnet-50. The last is learning properties. They include the optimization learning mode, mini-batch size, and max epoch. The proposed method applies those three properties. The others are in default of Matlab. Through this setting, the training process is done. Finally, the expected model of the network is obtained. This model is implemented for semantic segmentation.

The semantic segmentation process is performed by using two variables. The first is the test set. The test set is a multi-focus image fusion excluding the training set. Here, the test set is 8 images [26]. The test set is shown in Fig. 8. The test set is grayscale multi-focus images. This dataset is very commonly used in the multi-focus image fusion process. The other dataset is the hand-made dataset of the author.

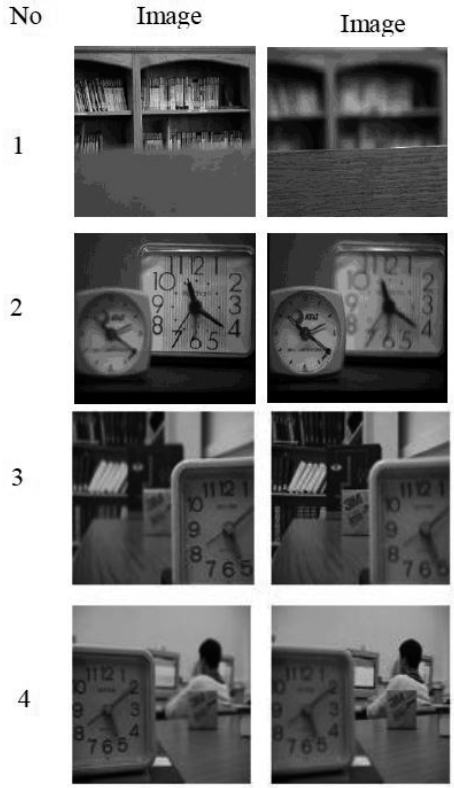

Fig. 8. The test set [26]

The second requirement is the model network. This model is produced from the above training process. Then the semantic segmentation task can be carried out. The result of this operation is a focus map. The focus map shows the categorical label of regions, as focus or blurred regions. The focus maps are shown in Fig. 9. The blue map represents the focus region, and the beach blue indicates the blurred region.

The training of the network is processed to obtain the network model. This model is a weight network with huge weight variables inside.

While performing the training process, there are many training parameters, such as epoch, iteration, image size, etc. The

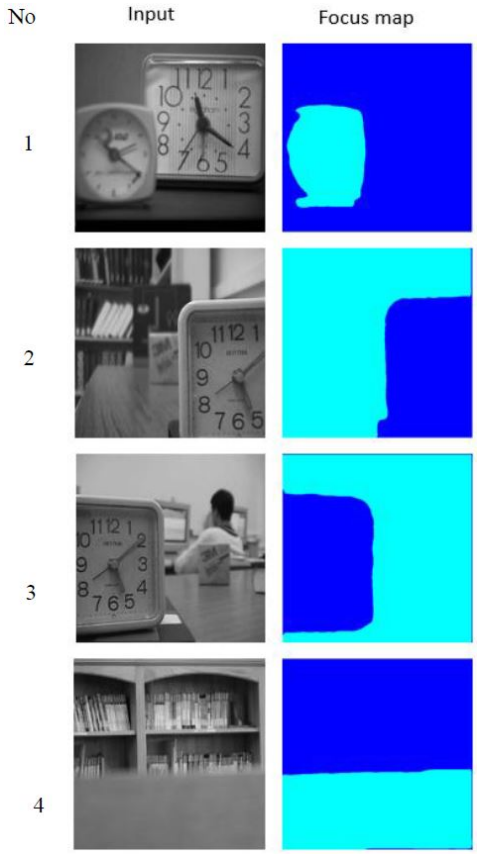

Fig. 9. The focus map of the test set

training parameter makes the training process accurate faster, and free of overfitting problems. The training parameters of the training process are shown in Table 1 below.

Table 1 shows that the accuracy increases when the iteration increases, too. Otherwise, the mini-batch loss decreases. The highest accuracy is obtained at 68.06%. Yet, it can deliver the high accuracy of the focus map.

Table 1
The output of the training process

Epoch (No)	Iteration (No)	Time-elapsed (hh:mm:ss)	Mini-batch accuracy (%)	Mini-batch loss (%)	Base learning rate parameter
1	1	00:02:42	50.27	44.44	0.0100
25	50	00:02:42	65.66	0.28	0.0100
50	100	00:02:42	67.18	0.16	0.0100
75	150	00:02:42	67.83	0.12	0.0100
100	200	00:02:42	68.06	0.09	0.0100

The focus map shows the indicator of focus and non-focus region. The blue color denotes the focus label, and the blue beach color is the blurred label. The size of the focus region and focus label almost are the same. To assess the accuracy of the predicted focus map, a confusion matrix is adopted. The confusion matrix of the proposed method is shown in Fig. 10.

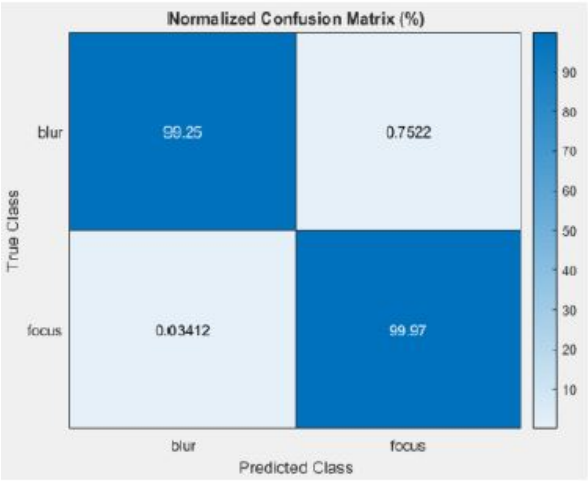

Fig. 10. The confusion matrix of the proposed method

Based on the accuracy of the predicted focus map, the fused image is generated through the following equation (5)

$$\text{Fused} = \alpha \times \text{Image } 1 + (1-\alpha) \times \text{Image } 2, \tag{5}$$

where the α is the focus map. Image 1 and Image 2 are a pair of multi-focus images. The block diagram of equation (1) is shown in Fig. 11.

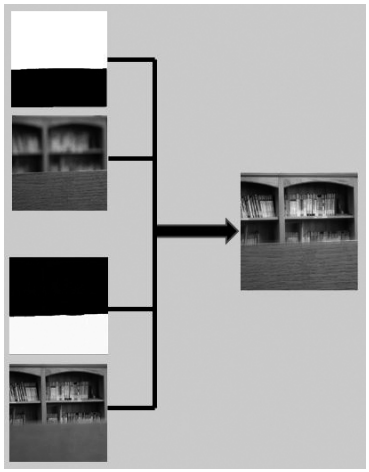

Fig. 11. The block diagram of the fusion process

The proposed method generates fused images as shown in Fig. 12.

The obtained fused images are very clear. The quality of the fused images is very high. There is no ghost effect or ring effect. To measure the accuracy of the proposed method, a comparison with the other two popular methods is performed. The first is multi-focus image fusion based on gradient domain [30] and the second is boundary finding through the morphological filter [31]. The comparison is shown in Fig. 13.

To measure the result objectively, the fused images are measured by SSIM [33], PSNR, and a normalized weighted performance metric, $Q_p^{AB/F}$ [34] indexes. The results are shown in Table 2 below.

Fig. 12. The fused images

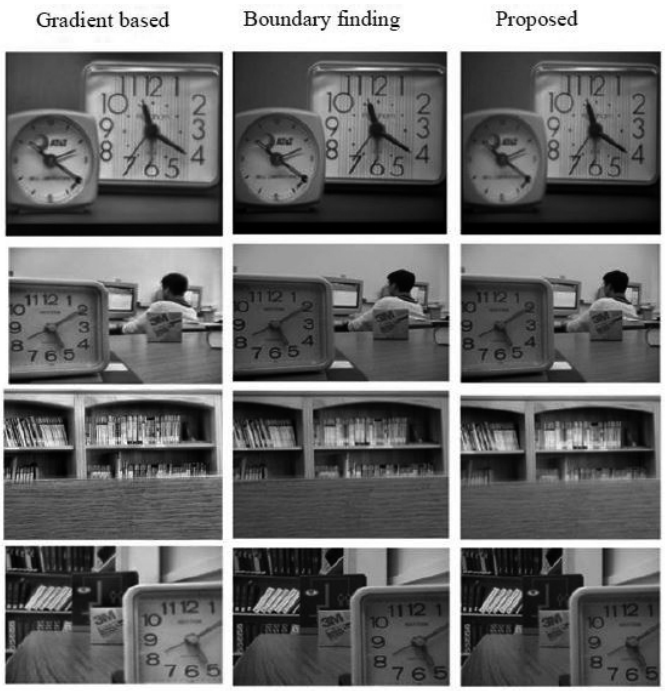

Fig. 13. The comparison proposed method with gradient-based [32] and boundary finding [31]

Table 2 shows the SSIM, PSNR and $Q_p^{AB/F}$ indexes of gradient-based, boundary finding and proposed from left to right side. The SSIM indexes are the metrics that measure the structural alteration of the image. The highest value is 1. It means perfect and there are no structural changes. These indexes present the highest values delivered by the proposed method. It gets almost 1 value to all datasets. The other two have the index lower. Furthermore, the PSNR indexes are a kind

Table 2
The Quality Assessment indexes of gradient-based, boundary finding, and the proposed method

	SSIM index			PSNR (dB)			$Q_p^{AB/F}$ index		
Dataset	Image fusion in gradient domain	Boundary find based	Proposed	Image fusion in gradient domain	Boundary find based	Proposed	Image fusion in gradient domain	Boundary find based	Proposed
1	0.8185	0.8382	**1.0000**	18.2909	26.0218	**83.8452**	0.6965	**0.7172**	0.6513
2	0.8347	0.8768	**0.9999**	17.7320	26.2011	**74.8848**	0.6735	**0.7464**	0.6766
3	0.6147	0.6457	**0.9999**	20.9542	20.7809	**76.4259**	**0.9753**	0.8123	0.7425
4	0.7582	0.7652	**0.9999**	17.2469	23.5760	**71.7448**	0.6794	**0.7284**	0.6836

of image metrics that measures signal-to-noise ratio. The best value is between 60 and 80. These indexes show the proposed method has the highest indexes for all datasets, too. Otherwise, the other two methods also have lower indexes. The last $Q_p^{AB/F}$ indexes measure the loss information as the algorithm runs. The best value of this index is 1, which means no information loss. The minimum acceptable standard quality of this index is 0.6. Otherwise, 0 means a complete loss of information. Based on these indexes of some datasets, the boundary finding [31] has the highest score for datasets no. 1, 2, and 4. It gets more than a 0.7 score. The gradient-based method [30] has more than a 0.9 score for just dataset no. 3. The other two methods show they get higher results in $Q_p^{AB/F}$ indexes. The proposed method has indexes more than 0.6 and close to 0.7 $Q_p^{AB/F}$ score for all datasets. According to these three metric indexes, the proposed method has highest score on SSIM and PSNR for all dataset and has acceptable result indexes for $Q_p^{AB/F}$ metric. Otherwise, the other two methods have a higher score on $Q_p^{AB/F}$ indexes. In general, the proposed method can handle the ghost effect and noise. Furthermore, it has more stable performance than other two methods. It means the proposed method can handle the limitation of sample dataset but not the other two methods. The proposed method preserves more adaptable generalization model to wide variety of dataset. This proposed method with different dataset for training set and test set shows the predicted robust focus map and generated fused images. This proposed method can handle a wider variety and larger dataset. It can be potentially used in many industrial applications.

5. CONCLUSIONS

The proposed method is an implementation of deep learning with semantic focus segmentation to compose a higher quality image, a fused image. The proposed method demonstrates the high stability of the detected focus map. This focus map can build fused images accurately. The comparison of the proposed method with other methods shows the stability performance of the proposed method to all datasets and robust to noise. Possibly, this implemented method is easier and simpler in many fields nowadays. This simple algorithm makes it easier to develop multi-focus image fusion in many new fields.

REFERENCES

[1] X. Zhang, "Multi-focus Image Fusion: A Benchmark", arXiv preprint arXiv:2005. 01116.

[2] E. Kot, K.S.Z. Krawczyk, L. Królicki, and P. Czwarnowski, "Deep learning based framework for tumour detection and semantic segmentation", *Bull. Pol. Acad. Sci. Tech. Sci.,* vol. 69, no. 3, p. e136750, 2021.

[3] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A Review of Multimodal Medical Image Fusion Techniques", *Comput. Math. Methods Med.,* vol. 2020, pp. 1–16, 2020.

[4] K. Kulpa, M. Malanowski, J. Misiurewicz, and P. Samczynski, "Radar and optical images fusion using stripmap SAR data with multilook processing", *Int. J. Electron. Telecommun.*, vol. 57, no. 1, pp. 37–42, 2011.

[5] 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, 2017.

[6] Ismail and K.H. Bin Ghazali, "The Multifocus Images Fusion Based on a Generative Gradient Map", *Lect. Notes Electr. Eng.*, vol. 632, pp. 401–413, 2020.

[7] Ismail and K. Hawari, "Multi Focus Image Fusion with Region-Center Based Kernel", *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 11, no. 1, pp. 57–63, 2021.

[8] K. Hawari, "The Normalized R and om Map of Gradient for Generating Multifocus Image Fusion", *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 1063–1069, 2020.

[9] X. Zhang, X. Li, and Y. Feng, "A new multifocus image fusion based on spectrum comparison", *Signal Process.*, vol. 123, pp. 127–142, 2016.

[10] V.N. Gangapure, S. Banerjee, and A.S. Chowdhury, "Steerable local frequency based multispectral multifocus image fusion", *Inf. Fusion*, vol. 23, pp. 99–115, 2015.

[11] A. Jameel, A. Ghafoor, and M.M. Riaz, "Wavelet and guided filter based multifocus fusion for noisy images", *Optik (Stuttg).*, vol. 126, no. 23, pp. 3920–3923, 2015.

[12] S. Bhat and D. Koundal, "Multi-focus image fusion techniques: a survey", *Artif. Intell. Rev.*, vol. 54, no. 6, pp. 1–53, 2021.

[13] H. Tang, B. Xiao, W. Li, and G. Wang, "Pixel convolutional neural network for multi-focus image fusion", *Inf. Sci.*, vol. 433–434, pp. 125–141, 2018.

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

[15] Z. Krawczyk and J. Starzyński, "Segmentation of bone structures with the use of deep learning techniques", *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 69, no. 3, p. e136751, 2021.

[16] H. Li, Y. Chai, and Z. Li, "A new fusion scheme for multifocus images based on focused pixels detection", *Mach. Vis. Appl.*, vol. 24, no. 6, pp. 1167–1181, 2013.

[17] H. Li, X. Liu, Z. Yu, and Y. Zhang, "Performance improvement scheme of multifocus image fusion derived by difference images", *Signal Process.*, vol. 128, pp. 474–493, 2016.

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

[19] Y. Yang, W. Zheng, and S. Huang, "Effective multifocus image fusion based on HVS and BP neural network", *Sci. World J.*, vol. 2014, p. 281073, 2014.

[20] R. Hong, C. Wang, Y. Ge, M. Wang, X. Wu, and R. Zhang, "Salience preserving multi-focus image fusion", *Proc. 2007 IEEE Int. Conf. Multimed. Expo*, *ICME 2007*, 2007, pp. 1663–1666.

[21] X. Bai, M. Liu, Z. Chen, P. Wang, and Y. Zhang, "Multi-Focus Image Fusion Through Gradient-Based Decision Map Construction and Mathematical Morphology", *IEEE Access*, vol. 4, no. 1, pp. 4749–4760, 2016.

[22] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[23] R. Kapela, "Texture recognition system based on the Deep Neural Network", *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 68, no. 6, pp. 1503–1511, 2020.

[24] A. Świetlicka and K. Kolanowski, "Robot sensor failure detection system based on convolutional neural networks for calculation of Euler angles", *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 68, no. 6, pp. 1525–1533, 2020.

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

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

[27] "Create DeepLab v3+ convolutional neural network for semantic image segmentation – MATLAB deeplabv3plusLayers". [Online]. Available: https://www.mathworks.com/help/vision/ref/deeplabv3pluslayers.html. [Accessed: 15-Jul-2021].

[28] L.C. Chen, Y. Zhu, G. Pap, R.F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11211 LNCS, pp. 833–851, 2018.

[29] "Semantic Segmentation – MATLAB & Simulink". [Online]. Available: https://www.mathworks.com/solutions/image-video-processing/semantic-segmentation.html. [Accessed: 16-Jul-2021].

[30] S. Paul, I.S. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain", *J. Circuits, Syst. Comput.*, vol. 25, no. 10, pp. 1–18, 2016.

[31] 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.

[32] S. Paul, I.S. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain", *J. Circuits, Syst. Comput.*, vol. 25, no. 10, 2016.

[33] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[34] S. Xydeas, C. and P. Petrović, "Objective Image Fusion Performance Measure", *Electron. Lett.* vol. 36, no. 4, pp. 308–309, Feb. 2000.