2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-2923-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISBI52829.2022.9761531

A NOVEL MULTI-FOCUS FUSION NETWORK FOR RETINAL MICROSURGERY

Xinyi Zhou[1], *Louying Hao*[1], *Qiushi Nie*[1], *Yingquan Zhou*[1], *Lihui Wang*[2], *Yan Hu*[1,⋆], *Jiang Liu*[1]

[1] Department of Computer Science and Engineering,
Southern University of Science and Technology, Shenzhen 518055, China
[2]Institute of Semiconductors,
Guangdong Academy of Sciences,Guangzhou, Guangdong 510650, China

ABSTRACT

Retinal microsurgery requires high precision. Due to the limited depth of field (DOF) of the ophthalmic microscope and eyeball's spherical construction, doctors observe the retina with partially in focus and partly out of focus. To solve this problem, we propose a deep-learning-based multi-focus fusion model to reconstruct an all-in-focus image. A focus measure block (FMB) is proposed to obtain the focus area in an image, and a fusion network (FN) is adopted to fuse the selected focus areas to produce the all-in-focus image. Considering the characteristics of retinal images, we propose to adopt two new losses to constrain our network. Based on our in-house dataset, extensive experiments prove the effectiveness of our algorithm.

Index Terms— Retinal microsurgery, depth of field, fusion, deep-learning, focus measure

1. INTRODUCTION

The retina is a layer of tissue in the back of the eyeball with only about 300-micrometer thickness. Retinal microsurgery is widely adopted for retinal diseases, such as macular holes, retinal detachment, branch\central retinal vein occlusion, and so on. Such operation requires extremely high precision. For example, in the treatment for central retinal vein occlusion named endovascular microsurgery, surgeons perform cannulation of retinal vessels with an injection of tissue plasminogen activator [1] as shown in Fig.1. Clear images captured by ophthalmic microscopes are fundamental for high-precision surgery. However, due to the limited DOF of the high magnification microscope camera and the eye's spherical construction, images are captured partially out of focus. It is difficult to locate the exact vessel position to inject and supervise the blood flow with a clear vision, but the defocused areas make it even harder. In this paper, we propose to adopt multi-focus image fusion [2] to solve this problem. It aims to combine the focus parts from multiple regional out-of-focus image stacks into one all-in-focus image. With surgical videos where the multi-focus image stack can be obtained, a sharp and clear image can be reconstructed to assist ophthalmic surgeons.

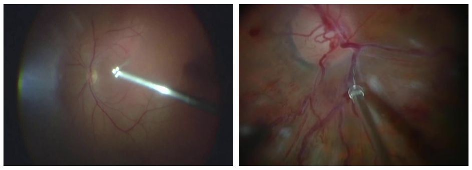


Fig. 1. Ophthalmic surgical microscope images in vitrectomy surgery and retinal endovascular surgery.

Multiple approaches have emerged to solve the multi-focus image fusion problem [2, 3]. Transform domain-based solutions as [4, 5] first convert the source images into the transform domain, in which operations based on predefined fusion rules can be applied. The results are then inversely converted back to the original domain. The conversion between the two domains makes this kind of solution time-consuming. Pixel-based methods [6, 7, 8], as the most representative methods of spatial domain-based multi-focus image fusion solutions, utilize spatial features of the source image to generate a weight map and calculate each pixel value of the fused image as the weighted average of all source images. But the above traditional approaches fail to address the problem of the boundary region between the in-focus region and the out-of-focus region, restricting their performance.

Deep learning-based solutions [9, 10, 11] also emerged, which are mainly designed for public datasets capturing natural scenes, with limited works aiming at medical image fusion [12, 13]. U-Net is a frequently applied deep-learning-based neural network model for its high performance on biomedical image segmentation. Nevertheless, U-Net-based models still suffer from their limited contextual information extraction ability and generate more blurred details than traditional methods as discussed in [12]. Therefore, we propose a novel fusion model for ophthalmic surgical microscope multi-focus image fusion.

The major contributions of our proposed algorithm are listed as follows:

∗ Correspondence: huy3@sustech.edu.cn

Authorized licensed use limited to: University of Southern California. Downloaded on August 13,2023 at 05:11:39 UTC from IEEE Xplore. Restrictions apply.

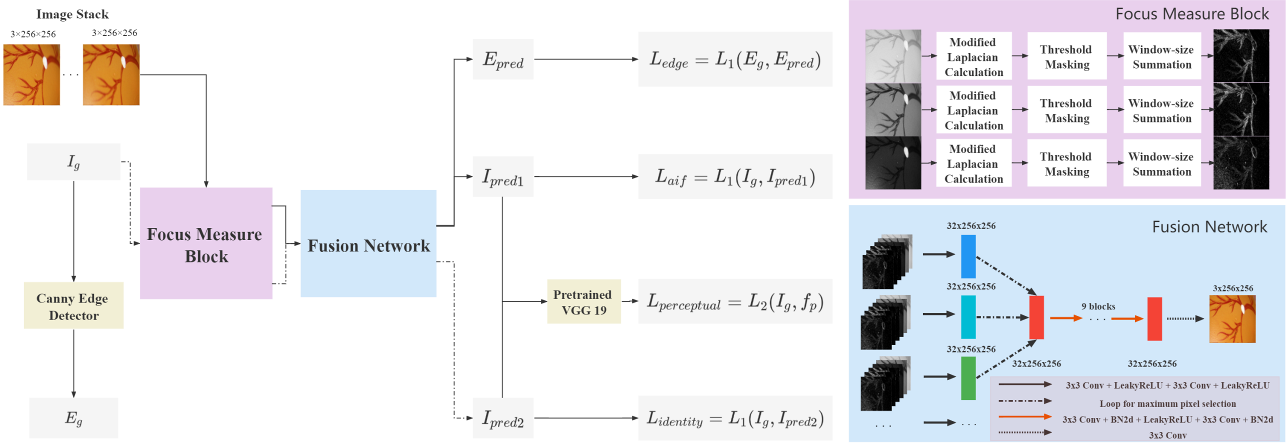


Image Stack
3×256×256
3×256×256
I_g
Canny Edge Detector
E_g
Focus Measure Block
Fusion Network
E_{pred}
I_{pred1}
I_{pred2}
Pretrained VGG 19
$L_{edge} = L_1(E_g, E_{pred})$
$L_{aif} = L_1(I_g, I_{pred1})$
$L_{perceptual} = L_2(I_g, f_p)$
$L_{identity} = L_1(I_g, I_{pred2})$
Focus Measure Block
Modified Laplacian Calculation
Threshold Masking
Window-size Summation
Fusion Network
32x256x256
9 blocks
3x256x256
3x3 Conv + LeakyReLU + 3x3 Conv + LeakyReLU
Loop for maximum pixel selection
3x3 Conv + BN2d + LeakyReLU + 3x3 Conv + BN2d
3x3 Conv

Fig. 2. The general framework of the proposed model. The Focus Measure Block and Fusion Network are depicted in detail.

1) We develop a novel deep-learning-based multi-focus image fusion model for ophthalmic microscope images to obtain all-in-focus images, which provides ophthalmic surgeons with a clearer and sharper view when performing retinal microsurgery in order to lower the operation risk.

2) We propose a Focus Measure Block for extracting the focus measure of the input images to provide the deep neural network with more information. Then we innovatively regularize the performance of the fusion network by two new losses.

3) A dataset is collected with an ophthalmic surgical microscope. With these images, various ablation studies are conducted to prove the significance of our proposed block and loss function. Meanwhile, comprehensive comparison experiments are done, and the results have illustrated the effectiveness of our proposed model.

2. METHOD

To address the deficiencies of current models in contextual information extraction and boundary region fusion, we propose a novel multi-focus image fusion algorithm. The general framework of the proposed method is shown in Fig.2. A stack of images with different defocused areas is input into the proposed Focus Measure Block (shown in Fig.2 light purple block) to measure their degree of focus based on the gradient of each image channel. Then images and their focus measures are concatenated and sent into our Fusion Network (shown in Fig.2 light blue block), where all in-focus pixels are extracted and fused into one all-in-focus image. In order to upgrade the network's learning ability, we innovatively propose to make use of the boundary and hues information to regularize its parameters. Detailed explanations about each part of the solution are presented in this section, including three parts: the Focus Measure Block (FMB), the Fusion Network (FN), and the loss function.

2.1. Focus Measure Block (FMB)

As can be seen from Fig.2, the focus measure is calculated by every channel of the input image. Inspired by [14], Focus Measure Block (FMB) adopts the three steps of sum-modified-laplacian in extracting focus measure, which is one of the frequently used spatial domain focus measure operators [15]. The discrete approximation to the modified Laplacian $\nabla^2_{ML}I(x,y)$ of each pixel at location (x,y) in the one channel matrix I is calculated by Eqs.(1).

$$\begin{aligned}\nabla^2_{ML}I(x,y) =& |2I(x,y) - I(x-step,y) - I(x+step,y)| + \\ & |2I(x,y) - I(x,y-step) - I(x,y+step)|\end{aligned} \tag{1}$$

In order to accommodate for possible variations in the size of texture elements, a variable spacing is applied, denoted by *step* between the pixels to computing the partial derivatives. Therefore, the focus measure at point (x,y) takes the form of Eqs.(2) as the sum of modified Laplacian greater than a threshold value from a window area around the point:

$$FMB = \sum_{i=x-N}^{i=x+N} \sum_{j=y-N}^{j=y+N} \nabla^2_{ML}I(i,j)\ , for\ \nabla^2_{ML}I(x,y) \geq T \tag{2}$$

where, T is the discrimination threshold value and N is the pixel window size. T and N are used in the steps of threshold masking and window-size summation illustrated in Fig.2.

2.2. Fusion Network (FN) Architecture

Concatenating each RGB image with its channel-by-channel focus measure, FN takes in a six-dimensional input. As shown at the bottom right of Fig.2, four kinds of network modules are included in FN based on CNN. The first module is used to do sampling on the input matrices multiple times and store the information in feature maps with more channels but the same resolution. Then, all the feature maps from different inputs are combined during which the second kind of module selects the maximum values. The motivation to

Authorized licensed use limited to: University of Southern California. Downloaded on August 13,2023 at 05:11:39 UTC from IEEE Xplore. Restrictions apply.

directly select the maximums of the feature maps is inspired by traditional pixel-based fusion methods. After selection and fusion, the feature maps are sent into nine blocks of the third kind for optimization and refinement. At last, a layer of CNN is implemented to generate the final RGB all-in-focus image result based on previously extracted features.

2.3. Loss Function

Generally, to optimize the multi-focus fusion network, the loss function adopts L_1 loss (L_{aif}) and perceptual loss ($L_{perceptual}$) instead of the common MSE loss so as to avoid the boundary smoothing effect of L_2-norm [10], as shown in Eqs.3 and Eqs.4:

$$L_{aif} = \frac{1}{3H_gW_g}||I_g - I_{pred1}||_1 \tag{3}$$

$$L_{perceptual} = \frac{1}{C_pH_pW_p}||I_g - f_p||_2^2 \tag{4}$$

where I_g, I_{pred1} and f_p represent the ground-truth image, the predicted image and the feature map produced by pretrained CNN model acting as the feature extractor in calculating perceptual loss. H_g and W_g denote the height and width of the ground truth image. C_p, H_p and W_p denote the channel number, height and width of feature map f_p.

But for retinal images, the vascular details are vital for surgery. A clear display of vascular details on the fundus has a significant impact on the visual effect of the generated all-in-focus image. Inspired by [9] and [16], to improve our model's capability to maintain texture details such as the blood vessels and recover a sharp edge between in-focus pixels and out-of-focus pixels, we introduce an edge loss into the loss function. The edge information of the real image setting is extracted from ground-truth images by Canny edge detectors for its superior performance [17]. The edge loss is defined as Eqs.5:

$$L_{edge} = \frac{1}{H_{edge}W_{edge}}||E_g - E_{pred}||_1 \tag{5}$$

where E_g and E_{pred} represent the edge map of ground-truth image and predicted image, H_{edge} and W_{edge} represent the height and width of the edge map.

Meanwhile, inspired by [18], the identity loss encourages the network to learn identical color mapping by forcing the model to generate the same image of the target domain. Thus, it is adopted to solve the problem that the output images have a slightly different color tone compared to the ground-truth images and the original image stacks. The identity loss is defined as:

$$L_{identity} = \frac{1}{H_gW_g}||I_g - I_{pred2}||_1 \tag{6}$$

where I_{pred2} indicates the predicted image used to compute $L_{identity}$.

Finally, the overall loss function of our network is defined as follows:

$$L = L_{aif} + L_{perceptual} + L_{edge} + L_{identity} \tag{7}$$

3. EXPERIMENTS AND RESULTS

3.1. Dataset

To solve the in-focus and defocus problem, we collected a dataset containing 3718 microscope fundus images with a fake eye model. The microscope is an EDER Surgical Microscope SM2000J with a resolution of 1920x1080. Six image stacks are collected under commonly used magnifications of 10, 16, and 20 respectively in different numbers of images according to different DOF. The DOF under different magnification is measured using a DOF 5-15 Depth of Field Target from Edmund Optics. The DOF for magnification 10, 16, 20 are 1.5mm, 0.87mm, and 0.56mm, respectively.

Five image stacks of each magnification class are for training and another stack for testing. The images are rescaled to 256x256. Due to the specialty of microscopic images, there are neither instruments to capture real all-in-focus ground-truth images nor methods to make a synthetic dataset. Inspired by [12], a well-performed traditional fusion algorithm [7] is adopted to generate ground-truth all-in-focus images.

3.2. Implementation Details

In the FMB, the focus measure of each channel is calculated with a 3x3 sliding window size and a threshold value of 7. We implemented the algorithm by PyTorch with Adam optimization algorithm and a learning rate of 0.0001. The model is trained with 1000 epochs. The Canny edge detector is applied with zero low threshold and high threshold. The Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR) are adopted as our evaluation metrics.

3.3. Ablation Study

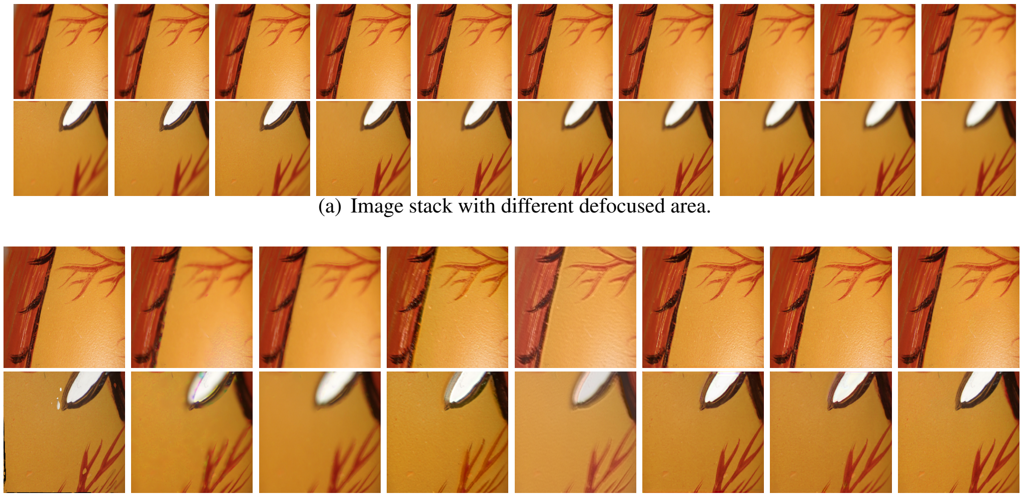
In this paper, we proposed the focus measure block (FMB) and loss function based on the fusion network (FN). Thus, FN is considered the baseline. To prove the effectivity of our FMB, which calculates focus measure for each channel (named as FMB3), we also gave out the experiment results of focus measure block for the gray channel of images, short as FMB1. As shown in Table 1, the contrast of FMB1 and FMB3 experiments shows that utilizing all three channel to calculate focus measure map has significant improvements on performance. Adding edge loss and identity loss also improves the model's performance. Our proposed algorithm, combining FN, FMB3 and two losses, gives the best SSIM and PSNR.

3.4. Comparison Experiments

We compared our method with Wang [8], ASR [4], CBF [6], IFCNN [10], U2Fusion [11] and AiFNet [16] on the dataset. The evaluation metrics are shown in Table.2 indicating that our proposed method outperforms other networks.

Qualitative results are shown in Fig.3. There are two sets

Authorized licensed use limited to: University of Southern California. Downloaded on August 13,2023 at 05:11:39 UTC from IEEE Xplore. Restrictions apply.



(a) Image stack with different defocused area.

(b) Wang (c) ASR (d) CBF (e) IFCNN (f) U2Fusion (g) AiFNet (h) Ours (i) GT

Fig. 3. Qualitative comparison of our method with 7 methods and ground truths (GT) on the fake fungus image dataset.

Table 1. Ablation Study

Models	SSIM	PSNR
FN	0.8942	30.64
FN+FMB1	0.8860	29.86
FN+FMB3	0.9071	30.80
FN+FMB3+Edge loss	0.9079	31.30
FN+FMB3+Identity loss	0.9193	30.82
Ours	**0.9230**	**32.10**

of image stack in Fig.3(a), capturing retinal blood vessels, optic nerves and the edge of fundus. Images generated by ASR[4], CBF[6], IFCNN[10] and U2Fusion[11] suffer from blurring effect to some extent while Wang [8] has a problem of pixel mismatching. Therefore, both the qualitative and quantitative results prove the effectiveness of our proposed method.

Table 2. Comparison experiments metrics results.

Models	Wang	ASR	CBF	IFCNN	U2Fusion	AiFNet	Ours
PSNR	25.04	29.26	26.43	29.35	25.64	31.91	**32.10**
SSIM	0.6782	0.8054	0.7180	0.8583	0.7620	0.9054	**0.9230**

3.5. Fusion Application

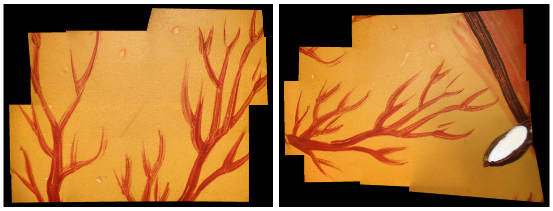


Fig. 4. Panorama image results stitched from generated all-in-focus images.

It is difficult to get a panoramic view of the fundus through microscope imaging due to the curvature of the eyeball. With partially blurred images, image stitching algorithms also have limited power to generate satisfactory results. However, the stitching outputs such as Fig.4 are more satisfying when we can get panoramas with image stitching code[1] and the all-in-focus images generated by our model.

4. CONCLUSIONS

In this paper, we proposed a multi-focus image fusion model to generate all-in-focus images for ophthalmic microsurgery, which provides the surgeons with a clear view of target tissues. A focus measure block was proposed to obtain the focus areas, and then a fusion network fused them to produce all-in-focus images. Evaluated on our collected database, the ablation and comparison experiments proved the effectiveness of the proposed modules and loss function. In the future, we would like to extend our algorithm to achieve advanced performance on real optic microsurgery images and videos.

5. ACKNOWLEDGMENTS

This work was supported by the National College students' Innovative Entrepreneurial Training Plan Program (Grant No. 202114325010), The National Natural Science Foundation of China (8210072776),The Science and Technology Innovation Committee of Shenzhen City (JCYJ20200109140820699 and 20200925174052004).

[1] https://github.com/yrlu/image_mosaic_stitching

Authorized licensed use limited to: University of Southern California. Downloaded on August 13,2023 at 05:11:39 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] Leon A Bynoe, Robert K Hutchins, Howard S Lazarus, and Mark A Friedberg, "Retinal endovascular surgery for central retinal vein occlusion: initial experience of four surgeons," *Retina*, vol. 25, no. 5, pp. 625–632, 2005.

[2] Yu Liu, Lei Wang, Juan Cheng, Chang Li, and Xun Chen, "Multi-focus image fusion: A survey of the state of the art," *Information Fusion*, vol. 64, pp. 71–91, 12 2020.

[3] Xingchen Zhang, "Deep learning-based multi-focus image fusion: A survey and a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, ¡br/¿.

[4] Yu Liu and Zengfu Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Processing*, vol. 9, pp. 347–357, 5 2015.

[5] Yu Liu, Xun Chen, Rabab K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.

[6] BK Shreyamsha Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, image and video processing*, vol. 9, no. 5, pp. 1193–1204, 2015.

[7] Yu Zhang, Xiangzhi Bai, and Tao Wang, "Boundary finding based multi-focus image fusion through multiscale morphological focus-measure," *Information Fusion*, vol. 35, pp. 81–101, 5 2017.

[8] Lihui Wang, Jianjiang Cui, Satoshi Tabata, and Masatoshi Ishikawa, "Low-cost, readily available 3d microscopy imaging system with variable focus spinner," *Optics Express*, vol. 26, pp. 30576, 11 2018.

[9] Jinxing Li, Xiaobao Guo, Guangming Lu, Bob Zhang, Yong Xu, Feng Wu, and David Zhang, "Drpl: Deep regression pair learning for multi-focus image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4816–4831, 2020.

[10] Yu Zhang, Yu Liu, Peng Sun, Han Yan, Xiaolin Zhao, and Li Zhang, "Ifcnn: A general image fusion framework based on convolutional neural network," *Information Fusion*, vol. 54, pp. 99–118, 2 2020.

[11] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling, "U2fusion: A unified unsupervised image fusion network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 7 2020.

[12] Vidas Raudonis, Agne Paulauskaite-Taraseviciene, and Kristina Sutiene, "Fast multi-focus fusion based on deep learning for early-stage embryo image enhancement," *Sensors*, vol. 21, no. 3, pp. 863, 2021.

[13] Lingbo Jin, Yubo Tang, Yicheng Wu, Jackson B Coole, Melody T Tan, Xuan Zhao, Hawraa Badaoui, Jacob T Robinson, Michelle D Williams, Ann M Gillenwater, Rebecca R Richards-Kortum, and Ashok Veeraraghavan, "Deep learning extended depth-of-field microscope for fast and slide-free histology," .

[14] Shree K Nayar and Yasuo Nakagawa, "Shape from focus," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[15] Wei Huang and Zhongliang Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 493–500, 2007.

[16] Maxim Maximov, Kevin Galim, and Laura Leal-Taixé, "Focus on defocus: bridging the synthetic to real domain gap for depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1071–1080.

[17] Pinaki Pratim Acharjya, Ritaban Das, and Dibyendu Ghoshal, "Study and comparison of different edge detectors for image segmentation," *Global Journal of Computer Science and Technology*, 2012.

[18] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

Authorized licensed use limited to: University of Southern California. Downloaded on August 13,2023 at 05:11:39 UTC from IEEE Xplore. Restrictions apply.