A Self-Supervised Deep Model for Focal Stacking

Weizhi Du[1,2*], Qichen Fu[2*], Zhengyu Huang[1,2]
1.Center for Ultrafast Optical Science, University of Michigan, Ann Arbor, Michigan, 48109, USA
2.University of Michigan, Ann Arbor, Michigan, 48109, USA
**These authors contributed equally.*
wzd@umich.edu

Abstract: We propose to train a self-supervised autoencoder to extract image features and fuse focal stack images. Numerical experiments show the proposed method achieves better fusion performance, compared to traditional fusion method using Laplacian operator.

1. Introduction

Due to the very shallow depth of field (DOF) of the high magnification object lens, only the portion of the sample within the DOF are sharply imaged. Focal stacking has been widely used in the field of microscopic measurement to obtain an all-in-focus image by fusing multiple images focused on different sample depths [1,2]. Traditional focal stacking method uses Laplacian operator for fusing images. On the other hand, deep learning based methods have improved the performance of various optical image processing tasks including image denoising, image segmentation, image reconstruction [3-5]. In this report, we propose a self-supervised focal stacking method based on deep learning, which shows improved performance over the traditional Laplacian operator based method [6].

2. Methods and Algorithms

The proposed fusion method pipeline shown in Fig.1 illustrates the autoencoder training process and fusing process. In the training process, the image is fed into an Encoder-Decoder Network. A five-layer Dense-Net encoder is used to extract deep features. A four-layer decoder then reconstructs the input image from the extracted features.

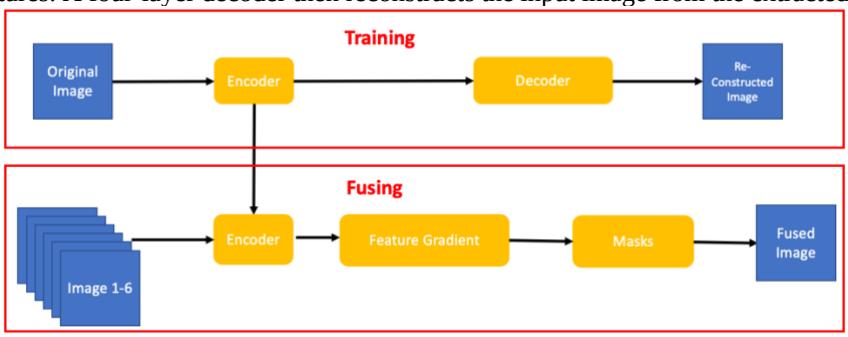

Fig.1. Schematic of image fusion pipeline

The loss function includes two terms L_p and L_s. L_p describes the pixel difference between the original image and the reconstructed image, and is given by:

$$L_p = ||R - O||_{2,}^2$$

where O is the original image, R is the reconstructed image. L_s describes the structural difference between the original image and the reconstructed image considering their brightness, contrast, and structural similarity. The L_s is given by:

$$L_s = \frac{2\mu_R\mu_O+C_1}{\mu_R^2+\mu_O^2+C_1}\frac{2\delta_R\delta_O+C_2}{\delta_R^2+\delta_O^2+C_2}\frac{\delta_{R,O}+C_3}{\delta_{R,O}^2+C_3},$$

where μ_R, δ_R are the mean and standard deviation of the reconstructed image, μ_O, δ_O are the mean and standard deviation of the original image, $\delta_{R,O}$ is correlation of reconstructed image and original image, and C_1, C_2, and C_3 are predefined constants [7]. The total training loss is the combination of the pixel loss and the structure loss, which is given by:

$$L_{tot} = L_p + \lambda(1 - L_s),$$

where λ is a hyper-parameter. During the fusion process, deep features of the blurred images are generated by the encoder network. We calculate feature gradients to determine the image fusing mask (decision map), which is then used to fuse images. The fusing algorithm is shown as follows:

$$M_{x,y} = argmax_i(FG_{x,y}^{(i)}), \quad i = 1,2...6,$$
$$FI_{x,y} = (Img_{x,y}^{M_{x,y}}),$$

where M is the mask, FG is the feature gradient and FI is the fused image. The pixels in the fused image are selected from the focal stack images according to the decision mask.

3. Results

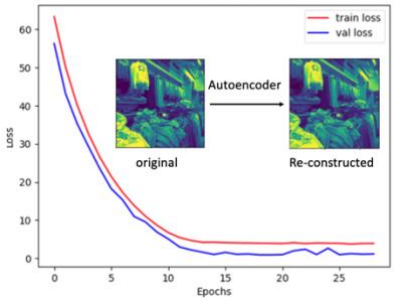

Fig.2. Training process of Autoencoder and reconstruction example

We choose NYU-v2 dataset [8] to test the effectiveness of our proposed method. The loss curve is shown in Fig.2. The image reconstruction example is embedded in Fig.2. The difference between the input image and the output reconstructed image shows little difference. In the fusion process, six focal stacked images are generated from NYU-v2 datasets, each focusing at a different depth. The testing MAE (mean-absolute-error) between the fused image and the original image is 0.0051, which is much lower than the traditional method (MAE = 0.0111). We visualize the fusing result in Fig.3. Although we tested our method using NYU-v2 dataset, we expect it to be also applicable to the microscopic focal stacking applications.

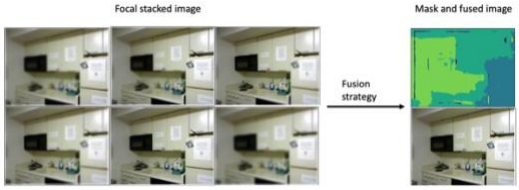

Fig.3. visualization of fused result. The left part is focal stacked images, the upper right corner is the mask (decision map), lower right corner is fused image

4. Conclusions

In this work, we proposed a self-supervised autoencoder to fuse focal stack images, which outperforms the traditional Laplacian operator fusion method on the NYU-v2 dataset. The proposed method can be applied to microscopic focal stack images in the future.

4. References

[1] Clark, Douglas, and Brian Brown. "A rapid image acquisition method for focus stacking in microscopy." Microscopy Today 23.4 (2015): 18-25.
[2] Wu, Ang, et al. "Sequence Image Registration for Large Depth of Microscopic Focus Stacking." IEEE Access 8 (2020): 6533-6542.
[3] Yedder, Hanene Ben, Ben Cardoen, and Ghassan Hamarneh. "Deep learning for biomedical image reconstruction: A survey." Artificial Intelligence Review 54.1 (2021): 215-251.
[4] Lahmiri, Salim, and Mounir Boukadoum. "Biomedical image denoising using variational mode decomposition." 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings. IEEE, 2014.
[5] Li, Yifei, et al. "FrequentNet: A Novel Interpretable Deep Learning Model for Image Classification." Available at SSRN 3895462 (2021).
[6] Nayar, Shree K., and Yasuo Nakagawa. "Shape from focus." IEEE Transactions on Pattern analysis and machine intelligence 16.8 (1994): 824-831.
[7] Ma, Boyuan, et al. "SESF-fuse: An unsupervised deep model for multi-focus image fusion." Neural Computing and Applications 33.11 (2021): 5793-5804.
[8] Silberman, Nathan, et al. "Indoor segmentation and support inference from rgbd images." European conference on computer vision. Springer, Berlin, Heidelberg, 2012.