



Correction of out-of-focus microscopic images by deep learning

Chi Zhang^{a,b}, Hao Jiang^a, Weihuang Liu^{a,c}, Junyi Li^{d,*}, Shiming Tang^e, Mario Juhas^f, Yang Zhang^{a,*}



^a College of Science, Harbin Institute of Technology, Shenzhen, China

^b School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

^c Department of Computer and Information Science, University of Macau, Macau, China

^d School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong 518055, China

^e School of Computing and Engineering, University of Missouri-Kansas City, MO, United States

^f Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland

ARTICLE INFO

Article history:

Received 18 November 2021

Received in revised form 1 April 2022

Accepted 2 April 2022

Available online 20 April 2022

Keywords:

Microscopic image
Out-of-focus correction
Confocal fluorescence microscope
Bright-field microscope
Mammalian cell
Leishmania parasite
Deep learning
CycleGAN

ABSTRACT

Motivation: Microscopic images are widely used in basic biomedical research, disease diagnosis and medical discovery. Obtaining high-quality in-focus microscopy images has been a cornerstone of the microscopy. However, images obtained by microscopes are often out-of-focus, resulting in poor performance in research and diagnosis.

Results: To solve the out-of-focus issue in microscopy, we developed a Cycle Generative Adversarial Network (CycleGAN) based model and a multi-component weighted loss function. We train and test our network in two self-collected datasets, namely Leishmania parasite dataset captured by a bright-field microscope, and bovine pulmonary artery endothelial cells (BPAEC) captured by a confocal fluorescence microscope. In comparison to other GAN-based deblurring methods, the proposed model reached state-of-the-art performance in correction. Another publicly available dataset, human cells dataset from the Broad Bioimage Benchmark Collection is used for evaluating the generalization abilities of the model. Our model showed excellent generalization capability, which could transfer to different types of microscopic image datasets.

Availability and Implementation: Code and dataset are publicly available at: <https://github.com/jiangdat/COMI>.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The microscopic imaging of cellular and subcellular structures belongs among the most widely used techniques in biomedical research and clinical diagnosis. With the improvements of imagery hardware, the advent of super-resolution microscopy imaging provides more details and information for related biomedical research. This led to a surge of interest in super-resolution microscopy techniques, such as Structured Illumination Microscopy (SIM) [5], Stochastic Optical Reconstruction Microscopy (STORM) [26], Photo Activated Localization Microscopy (PALM) [1] and Stimulated Emission Depletion (STED) Microscopy [4]. However, obtaining high-quality in-focus microscopy images has been a cornerstone of the microscopy [36]. Images obtained by microscopes are often out-of-focus, resulting in poor performance in research and diag-

nosis. These low-quality images are usually degraded due to the diffraction barrier, astigmatism, optical system or camera defects, and the human error in sample preparation and image acquisition [7]. Human operator mistakes and autofocus system errors lead to low-quality out-of-focus, blurry microscopic images with worse performance in scientific research and clinical diagnosis. Employing deblurring algorithms in out-of-focus microscopic images will assist pathologists in making a better diagnostic decision, and extend the biological phenomena range observed by microscopy [24].

Various methods have been proposed for microscopic image deblurring, including probabilistic model [28], L0-regularization [30] and dark channel prior [22]. Furthermore, the computer-assisted image processing approaches, such as the nearest neighbor technique, inverse filtering and constrained iterative algorithm [17,31,33], provide alternative to improve the quality of blurry microscopic images. With the successful application of deep learning in biological analysis [3,10,16,35], convolutional neural networks (CNN) provide solutions for deblurring microscopic

* Corresponding authors.

E-mail addresses: lijunyi@hit.edu.cn (J. Li), zhangyang07@hit.edu.cn (Yang Zhang).

images, which circumvent the complicated pre-processing and achieve better results. CNN can extract prior knowledge from numerous high-resolution images and produces superior images from a low-resolution counterpart based on the extracted high-resolution prior knowledge. Specifically, Schuler *et al.* developed a deep learning based image deblurring approach with stacked multiple CNN [27]. However, the performance of this approach drops in the case of large blur kernels. To avoid problems related to blur kernel estimation, Nah *et al.* designed a multi-scale CNN with a coarse-to-fine multi-scale loss function [20]. Furthermore, the Recurrent Neural Network (RNN) is also implemented in the image reconstruction. A scale RNN with a simple network structure and a small number of parameters was proposed to deblur the image by utilizing multi-scale spatial information [29]. Zhang *et al.* proposed a spatial variant RNN with weights generated by a CNN to dynamically deblur image [34]. Persch *et al.* utilized an image enhancement method which combined deconvolution, denoising and inpainting for 3-D confocal based STED microscopy imager [23]. Jin *et al.* applied U-Net to reconstruct the super-resolution images from limited number of extreme low-light SIM raw data, but this research lacks the exclusive design of the network structure [11]. Huang *et al.* proposed a pure-hardware solution, whole-cell 4Pi single-molecule switching nanoscopy (W-4PiSMSN), which can output ultra-high-resolution 3D image of subcellular structure at 10 to 20 nm [8]. Rivenson *et al.* tried to reconstruct a high magnification objective through a low magnification objective to deblur low-resolution images [25]. This method provides a new framework in holographic image reconstruction. However, the limitation is that it was developed for phase recovery of bright-field microscopy images. The microscopic features in bright-field and fluorescence microscopes are dissimilar. Therefore, employing trained neural network for phase recovery does not align well with the use of fluorescence images.

Advances in the field of generative adversarial networks (GAN), aiming to translate images from the source domain to the target domain, have led to great success in image-to-image translation [9,32,38,47]. Although GAN were originally not designed for microscopic images, image deblurring can be considered a special image-to-image translation, GAN can therefore be introduced into image deblurring task [14,15]. Ouyang *et al.* implemented a GAN method based on Pix2Pix to reconstruct high-quality super-resolution images from sparse and rapidly acquired single-molecule localization data and wide-field images [21]. GAN with U-Net architecture for generator and Markovian for discriminator achieves perceptually superior results on image deblurring, suggesting that GAN-based deblurring models need to focus on global style transfer and local texture synthesis simultaneously [18]. Combalia *et al.* used CycleGAN to combine the confocal fluorescence microscopy and reflectance confocal microscopy modes into a digitally stained hematoxylin and eosin (H&E) slide [2]. Lim *et al.* proposed a deep learning model for microscopic image blind deconvolution based on cycle consistency and point spread function modeling layers [19]. Different from the above two work, we tried to correct the out-of-focus microscopic images by establishing a CycleGAN-based deblurring framework with multi-component weighted loss function. The proposed model shows superior performance and yield sharper and more plausible textures compared to other GAN-based classic methods. Our main contributions are summarized as follows:

- We apply Cycle Generative Adversarial Network with a new multi-component weighted loss function to solve the out-of-focus issue in microscopy.
- Two self-collected datasets are used for model training and testing, namely Leishmania parasite dataset captured by a bright-field microscope and BPAEC dataset captured by a confocal fluorescence microscope.

- We evaluate our trained model in another publicly available out-of-focus microscopy image dataset and the result shows that our model has excellent generalization capability.

2. Data collection and evaluation

2.1. Data collection

We collect two datasets for network training and testing, and make them publicly available in Mendeley Data: <https://data.mendeley.com/datasets/m3ixgb54c9/4>. The first dataset is a protozoan parasite microscopy image dataset of Leishmania, obtained from the preserved slides stained with Giemsa. The paired blur-sharp images are acquired by employing a bright-field microscope (Olympus IX53) with 100× magnification oil immersion objectives (Fig. 1A). We first capture the sharp images as ground truth, then acquire its corresponding out-of-focus images. The extent and nature of defocusing are random along the optical axis, where the degree of out-of-focus is inconsistent from image-to-image. This dataset includes 764 in-focus and 764 corresponding out-of-focus images, where each image is composed of 2304 × 1728 pixels in 24-bit JPG format.

The second dataset named bovine pulmonary artery endothelial cells is collected by a confocal fluorescence microscope (high-resolution confocal scanning microscope ZEISS-LSM880 incorporated with a Zyla sCMOS). ZEISS-LSM880 offers high sensitivity and enhanced resolution in x-, y- and z-axis. The fluorescent signals emitted by samples with different depths on the z-axis are captured. The slides of mitochondria, actin and nuclei are stained by MitoTracker® Red CMXRos, Alexa Fluor® 488 phalloidin, and blue-fluorescent DNA stain (DAPI) individually. For each excitation line (DAPI, Alexa Fluor® 488, and TxRed), a set of in-focus and out-of-focus images are captured by using z-stack modular with a 100×/1.4-NA objective. The confocal z-stacking typically refers to a series of images that were acquired at different locations in the sample along the optical axis. Images acquired at the optimal focal plane are seldom blurry and are usually considered in-focus images with sharp edges and clear textures. Images acquired below and above the optimal focal plane are blurry and out-of-focus images. In detail, scans were acquired in z-stack of 15 layers spanning the depth (8.4 μm) with 0.6 μm between each slice, z = 7 is the optimal focal plane, z = 1–6 are below the focal plane, z = 8–15 are above the focal plane. We make layers from z = 4 to z = 10 publicly available. Their visual variations in 3-dimensional structure can be negligible. In the end, datasets of actin and nucleus contain 100 in-focus images and 600 out-of-focus images respectively, dataset of mitochondria contain 97 in-focus images and 582 out-of-focus images. Each image is composed of 1024 × 1024 pixels in 8-bit JPG format. The details of two dataset parameters are summarized in Table 1. The parameters and examples of each categories are shown in Fig. 1B.

2.2. Evaluation metrics

Several common image quality evaluation metrics are used to evaluate the quality of restored images. Specifically, Peak signal-to-noise ratio (PSNR), Structural similarity (SSIM) and Pearson correlation coefficient (PCC) are calculated as the restored image quality evaluation metrics. Details of these evaluation metrics are as follows. The PSNR is defined as:

$$PSNR(X, Y) = 10 \log_{10} \frac{MAX^2 \times H \times W}{\sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2} \quad (1)$$

Where H , W are the height and width of the image X and Y , while MAX is the maximum value of image pixel. A higher PSNR value provides a higher image quality and lower numerical differences

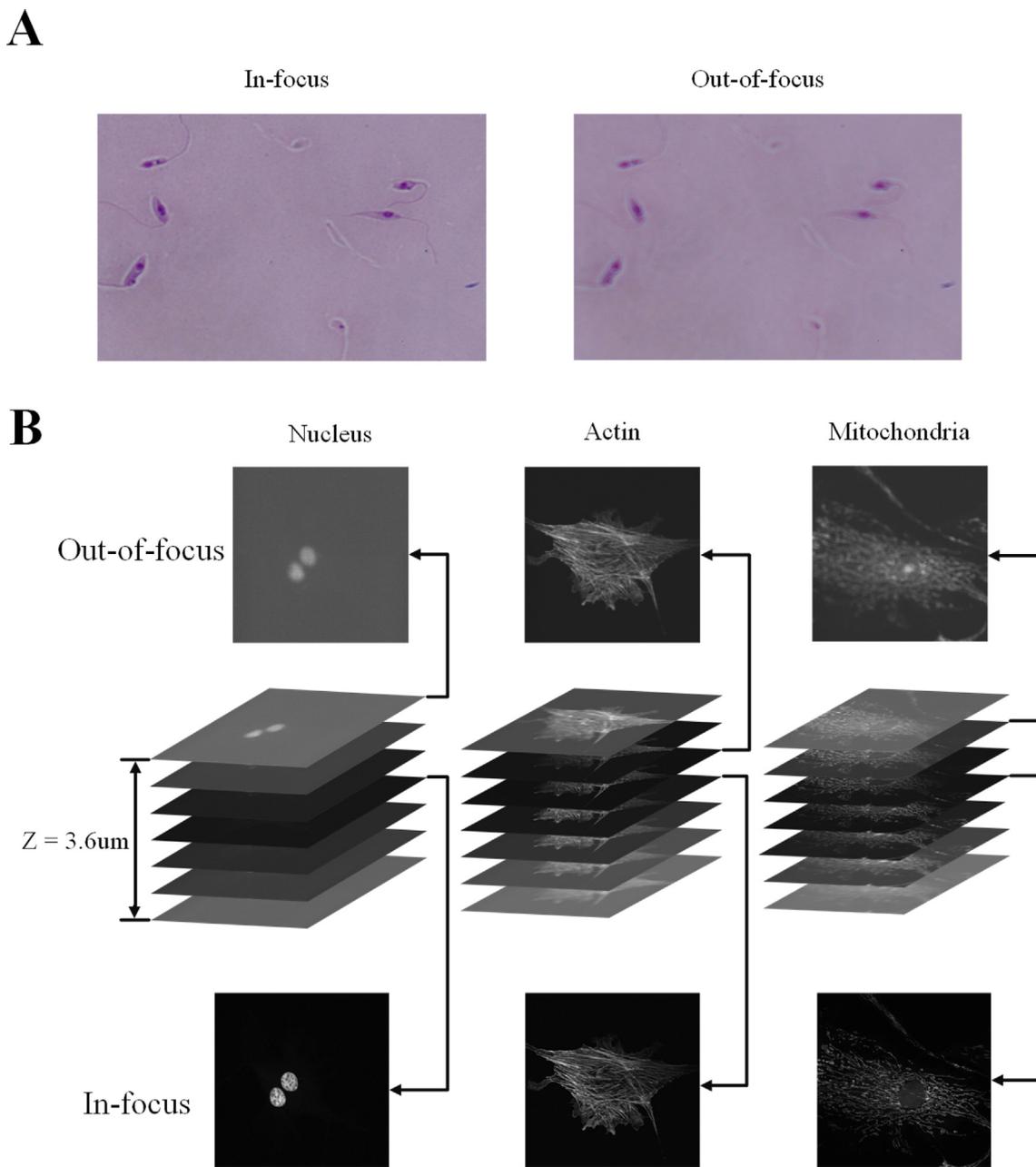


Fig. 1. Sharp in-focus images from the optimal focal plane and blurred out-of-focus images from outside the focal plane. A. The paired in-focus and out-of-focus images of Leishmania. B. Bovine pulmonary artery endothelial cells collected by a confocal microscopy. In-focus (layer $z = 7$) and out-of-focus (layer $z = 4, 5, 6, 8, 9, 10$) images of nucleus, actin and mitochondria collected by a confocal fluorescence microscope in z-stack with $0.6 \mu\text{m}$ distance between each slice. The layer $z = 7$ is ground truth while other layers have different degree of out-of-focus.

Table 1

The parameters of two self-collected datasets. The first dataset includes sharp-blur pairs of Leishmania image. The second dataset contains the confocal fluorescence microscopy images of nucleus, actin and mitochondria of BPAEC, where each clear image corresponds to 6 out-of-focus images with different degree of blurring. In training and testing, each in-focus image from the layer $z = 7$ in BPAEC forms a clear-blur image pair with the out-of-focus image from each other layers ($z = 4, 5, 6, 8, 9, 10$). The in-focus image of the testing dataset is only used to calculate PSNR, SSIM and PCC. The training and testing dataset proportion is 8:2.

Data types	First dataset (Parasite)	Second dataset (BPAEC)		
	Leishmania	Nucleus	Actin	Mitochondria
Out-of-focus (single)	764	600	600	582
In-focus (single)	764	100	100	97
Train dataset (pairs)	611	480	480	466
Test dataset (pairs)	153	120	120	116
Resolution (pixel)	2304×1728	1024×1024		
Bit depth (bit)	24	8		

between ground truth [39]. SSIM can measure the image similarity in brightness, contrast and structure, which is considered to correlate with the quality perception of human visual system [39,45]. SSIM can be calculated as:

$$SSIM(X, Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (2)$$

The parameters μ_X and μ_Y respectively represent the average of the image X and Y , σ_X and σ_Y represent the variance of image X and Y , σ_{XY} represents the covariance of image X and Y . Both c_1 and c_2 are small positive constants which can keep the denominator non-zero when either $(\mu_X^2 + \mu_Y^2)$ or $(\sigma_X^2 + \sigma_Y^2)$ converges to 0. SSIM value is greater than 0 and less than 1. The higher SSIM indicates the smaller difference between restored images and ground truth. PCC reflect the linear correlation degree between restored images and ground truth. Given image X and Y with $N \times N$ pixels, the PCC between X and Y can be calculated using the following formula:

$$Pearson(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X\sigma_Y} \quad (3)$$

Where the $E[.]$ represents the mathematical expectation. The PCC value ranges from -1 to 1 , where 0 means the absence of relationship, negative value is negative correlation and positive value is positive correlation.

3. Methods

3.1. Network architecture

This work applies the GAN structure as backbone network. GAN try to replicate the probability distribution by employing two essential components: the discriminator and the generator. The generator G generates new samples according to the potential data distribution of target image, and the discriminator D is a binary classifier used to determine whether the input samples are ground truth or generated samples [38]. After the alternate optimization and multiple training, the capabilities of both generator and discriminator are enhanced. The final objective is finding the optimal generator and discriminator, meaning that generator G can minimize the difference between generated samples and target image as much as possible, and discriminator D can distinguish the generated samples from G and target image as much as possible. The process can be summarized as follows:

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (4)$$

Where the \log is the abbreviation of logarithm function. The x represents target image which follows the distribution P_{data} , and z is the input random noise vector which is sampled from a prior distribution P_z . Maximize D means that when we input generated image from G and the target image, the discriminator can identify the target image as accurately as possible, so that the probability value of $D(x)$ to be as close to 1 as possible. Minimize G means that the generated samples $G(z)$ from generator G are supposed to be as close as possible to the target image, so that the value of $1 - D(G(z))$ is as close to 0 as possible. The discriminator tries to maximize this loss while generator tries to minimize this loss. Both of them are iteratively optimized in adversarial mode.

Specially, our method is inspired by the CycleGAN, a derivate GAN structure that has two mirror-symmetric GAN structures and each GAN has a generator and a discriminator individually. The proposed GAN structure is shown in Fig. 2. In this structure, two mapping functions $G_s : source \rightarrow target$ and $G_t : target \rightarrow source$

are considered simultaneously, and two adversarial discriminators D_s and D_t are created relatively. The generator architecture used the style-transfer type [12]. The encoder-decoder structure contains two stride convolution blocks, 9 residual blocks (ResBlocks), and two transposed convolution blocks. In ResBlocks, the global skip connections pass the image details to upper layers and back-propagation gradients to lower layers, thus successfully solving the degradation problem [6]. The discriminator employed the Markovian discriminator in our network, which is purely composed of convolutional layers and its final output is an $n \times n$ matrix. The discriminate probability is the average value in $n \times n$ matrix.

3.2. Multi-component weighted loss

In order to improve performance of CycleGAN and generate visually comfortable image, three component weighted losses are applied to guide training stage, including adversarial loss l_{GAN} , content loss l_{cont} , cycle consistency loss l_{cycle} . The adversarial loss l_{GAN} is applied to generators G_s and G_t , and discriminators D_s and D_t , which design to capture and minimize the distribution distance between generated samples and target images. For example, generator G_s tries to generate in-focus image $G_s(I^s)$ that look similar to the image I^t from domain t , and D_t tries to discriminate generated samples $G_s(I^s)$ and target image I^t . The adversarial loss in CycleGAN can be expressed as:

$$l_{GAN} = E_{t \sim P_{data}(t)} [\log D_t(t)] + E_{s \sim P_{data}(s)} [\log(1 - D_t(G_s(s)))] + E_{s \sim P_{data}(s)} [\log D_s(s)] + E_{t \sim P_{data}(t)} [\log(1 - D_s(G_t(t)))] \quad (5)$$

We apply the content loss l_{cont} to generators G_s and G_t . The content loss ensures that the reconstructed image has similar features to the target image, which can be regarded as the Euclidean distance loss of the feature map between the reconstructed deblurring image and the target image. Classical loss functions like L_1 loss and L_2 loss generally lead to blurred images due to the pixel-wise average of possible solutions in the pixel space. Perceptual loss is a simple L_2 loss which can measure perceptual differences in content and style between two images. For example, generator G_s calculates its perceptual loss l_{cont-G_s} between the in-focus image I^t and generated in-focus image $G_s(I^s)$, which can be expressed as:

$$l_{cont-G_s} = \frac{1}{W_{ij}H_{ij}} \sum_{s=1}^{W_{ij}} \sum_{t=1}^{H_{ij}} (\varphi_{ij}(I^t)_{s,t} - \varphi_{ij}(G_s(I^s))_{s,t})^2 \quad (6)$$

Here (W, H) is the image size, φ_{ij} is the feature map obtained by the j -th convolution (after activation) before the i -th maxpooling layer. In this work, we use activations from VGG-19 convolutional layer. Similarly, the generator G_t calculates its perceptual loss l_{cont-G_t} between the out-of-focus image I^s and generated out-of-focus image $G_t(I^t)$, which can be expressed as:

$$l_{cont-G_t} = \frac{1}{W_{ij}H_{ij}} \sum_{s=1}^{W_{ij}} \sum_{t=1}^{H_{ij}} (\varphi_{ij}(I^s)_{s,t} - \varphi_{ij}(G_t(I^t))_{s,t})^2 \quad (7)$$

W_{ij} and H_{ij} are the dimensions of the respective feature maps. The total content loss includes losses in both directions and can be expressed as:

$$l_{cont} = l_{cont-G_s} + l_{cont-G_t} \quad (8)$$

Cycle consistency loss allows images to be translated back again after being translated from one domain to the other, preventing the learned mapping $G_s : source \rightarrow target$ and $G_t : target \rightarrow source$ from contradicting each other [47]. The cycle consistency loss can be totally expressed as:

$$l_{cycle} = E_{s \sim P_{data}(s)} [\|G_t(G_s(s)) - s\|_1] + E_{t \sim P_{data}(t)} [\|G_s(G_t(t)) - t\|_1] \quad (9)$$

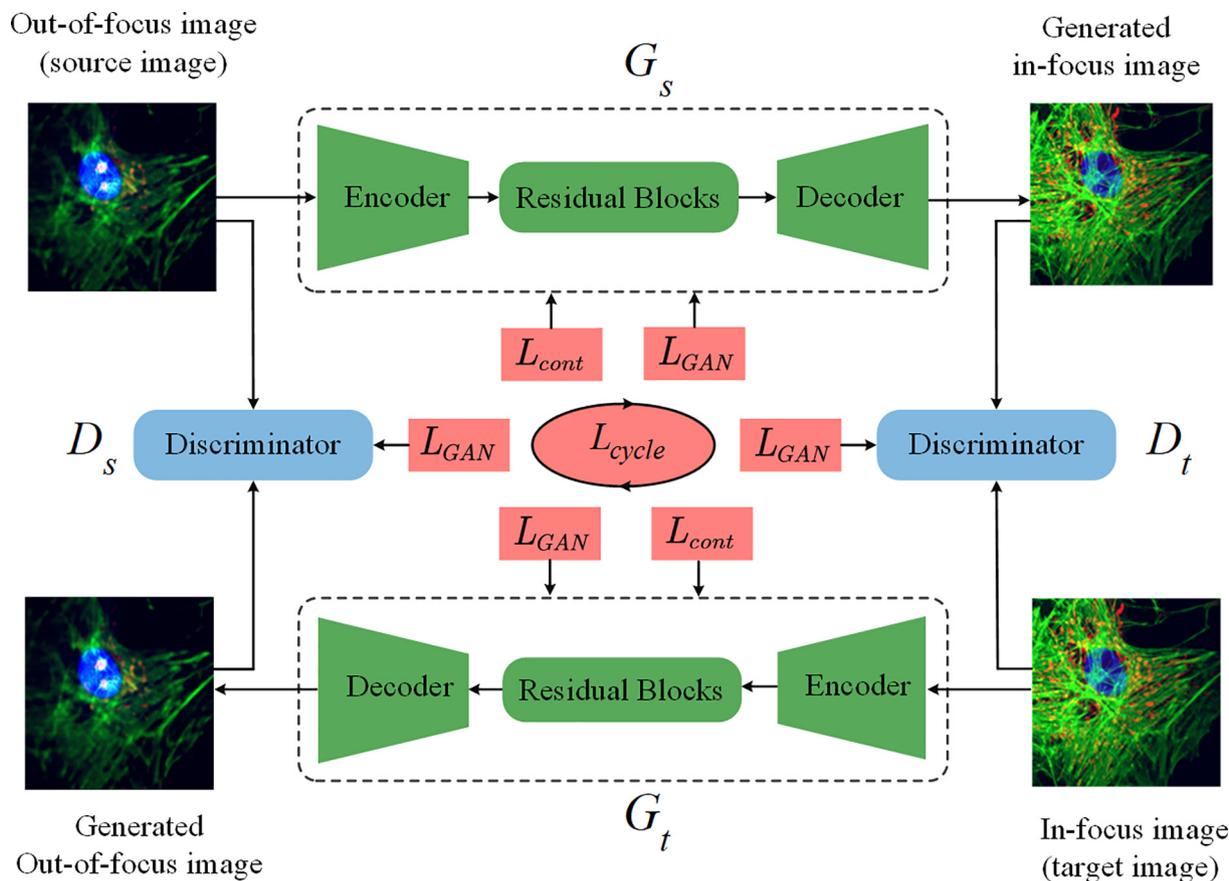


Fig. 2. Overview of CycleGAN-based deep learning for the correction of out-of-focus microscopy images. The proposed method contains two generators (G_s and G_t) and two discriminators (D_s and D_t). Generator G_s translates out-of-focus image to in-focus image, and discriminator D_t tries to distinguish real in-focus image and generated in-focus image. Generator G_t translates in-focus image to out-of-focus image, and discriminator D_s tries to distinguish real out-of-focus image and generated out-of-focus image.

Finally, multi-component weighted loss function is designed. The loss consists of three components and can be expressed as follows:

$$l = \lambda_1 l_{GAN} + \lambda_2 l_{cont} + \lambda_3 l_{cycle} \tag{10}$$

Where l_{GAN} is the adversarial loss, l_{cont} is the content loss and l_{cycle} is the cycle consistency loss. λ_1, λ_2 and λ_3 are the non-negative hyper-parameters which constant and adjust different influence on overall deblurring effects. Same with other methods [42,46,47], these weights in our multi-components weighted loss function are set according to the data characteristics for different cases and we weight each loss item empirically to balance the importance of each component. In this experiment, we set λ_1, λ_2 and λ_3 to 1, 1, and 0.001 respectively.

3.3. Training

The proposed model is built in Ubuntu 16.04 system configured with Tesla K40C GPU, and implemented in Keras of Tensorflow. Rotation, flip, translation, and scale are used for data augmentation. 5-fold cross validation is used to evaluate the performance of the models. We use the Adam optimizer [13] with parameters $\beta_1 = 0.5$ and $\beta_2 = 0.99$. The total training iteration times is set as 100,000 times. The learning rate is set initially to 10^{-4} for both of generator and discriminator. Then, the learning rate is linearly decayed to 0 after 50,000 training iterations. The batch size is set to 1 to train our model. Both Leishmania and BPAEC dataset are randomly divided into two parts in a proportion of 8:2, which mean that the testing dataset is unseen to the trained model. We

use the testing set to evaluate the trained model, and repeat three times to obtain the standard deviation in Table 2 and Table 3. The codes, implementation details and datasets are available online: <https://github.com/jiangdat/COMI>.

4. Results

4.1. Comparison of different models on the parasite dataset

We compare the performance of the proposed method with GAN-based classic deblurring models in the first self-collected dataset. We train and test our proposed method along with other GAN-based deblurring methods, including DeblurGAN, DeblurGAN-V2, Pix2Pix, CycleGAN. Pix2Pix is a GAN-based model for pixel-level image translation, which is widely used to reconstruct high-quality super-resolution images. CycleGAN is the basis

Table 2 Comparison of different methods on Leishmania parasite dataset. The results of different methods were generated by testing them in the same Leishmania parasite dataset. PSNR, SSIM and PCC were calculated. The higher the score the better the result. The highest values are highlighted in bold.

Model	PSNR	SSIM	PCC
L0-regularized	34.97 ± 0.00	0.8034 ± 0.0000	0.9349 ± 0.0000
Pix2Pix	36.18 ± 0.06	0.8868 ± 0.0010	0.9532 ± 0.0004
CycleGAN	32.35 ± 0.10	0.8306 ± 0.0008	0.9098 ± 0.0012
DeblurGAN	36.15 ± 0.15	0.8789 ± 0.0009	0.9513 ± 0.0006
DeblurGAN-V2	36.86 ± 0.08	0.8724 ± 0.0012	0.9597 ± 0.0022
Ours	38.62 ± 0.09	0.8951 ± 0.0006	0.9726 ± 0.0009

Table 3

The quantitative evaluation results of our model on the BPAEC dataset. The images coming from layer $z = 7$ are the in-focus images. PSNR, SSIM and PCC were calculated for each cellular structure (nucleus, actin and mitochondria). The higher the score the better the result. The highest values are highlighted in bold.

Layer	Nucleus			Actin			Mitochondria		
	PSNR	SSIM	PCC	PSNR	SSIM	PCC	PSNR	SSIM	PCC
$z = 4$	33.56 ± 0.19	0.8293 ± 0.0174	0.9690 ± 0.0006	28.11 ± 0.30	0.4711 ± 0.0027	0.9141 ± 0.0017	30.44 ± 0.43	0.5519 ± 0.0076	0.9138 ± 0.0063
$z = 5$	31.89 ± 0.09	0.6350 ± 0.0047	0.9361 ± 0.0070	30.82 ± 0.16	0.4925 ± 0.0030	0.9577 ± 0.0057	31.68 ± 0.39	0.3923 ± 0.0132	0.9629 ± 0.0045
$z = 6$	33.65 ± 0.33	0.6880 ± 0.0242	0.9851 ± 0.0020	30.61 ± 0.16	0.5630 ± 0.0034	0.9596 ± 0.0037	34.45 ± 0.28	0.8310 ± 0.0165	0.9709 ± 0.0071
$z = 8$	35.44 ± 0.33	0.9432 ± 0.0035	0.9779 ± 0.0031	30.16 ± 0.33	0.7047 ± 0.0101	0.9508 ± 0.0048	32.74 ± 0.55	0.8769 ± 0.0129	0.9538 ± 0.0114
$z = 9$	32.78 ± 0.13	0.6990 ± 0.0099	0.9673 ± 0.0040	28.05 ± 0.27	0.5745 ± 0.0056	0.9174 ± 0.0102	29.87 ± 0.44	0.4490 ± 0.0061	0.8893 ± 0.0041
$z = 10$	33.29 ± 0.17	0.8775 ± 0.0053	0.9581 ± 0.0026	26.53 ± 0.42	0.3547 ± 0.0240	0.8718 ± 0.0122	28.09 ± 0.48	0.4362 ± 0.0057	0.8232 ± 0.0092
Average	33.44 ± 0.21	0.7787 ± 0.0108	0.9656 ± 0.0032	29.05 ± 0.27	0.5268 ± 0.0081	0.9286 ± 0.0064	31.21 ± 0.43	0.5896 ± 0.0103	0.9190 ± 0.0071

of our proposed model. DeblurGAN is a deep learning method based on conditional GAN and content loss for macroscopic image deblurring. DeblurGAN-V2 is improved version of DeblurGAN, which not only improves the deblurred image quality, but can also design a model with low computational cost. Besides, we also reproduce L0-regularized, which is a representative of the traditional algorithm. The sparse L0 approximation scheme consisting of a family of loss functions to approximate the L0 cost into the objective to remove blur. The qualitative evaluation results in first self-collected dataset are shown in Fig. 3. The traditional method generates image that has some abnormal high-frequency signals, thus resulting in bad quality image (Fig. 3 B). Deep learning-based methods achieve better results and the generated image is smoother, especially when using DeblurGAN and DeblurGANv2 (Fig. 3 E and F). However, Pix2Pix and CycleGAN (Fig. 3, C and D) were not designed specifically for microscopy images and there are many blocking artifacts on the edge of Leishmania with inter-

mittent flagella, as shown in the zoomed-in patches. The image deblurred by our method has the highest similarity with the in-focus image with more details in textures (Fig. 3 G). In contrast with the result of CycleGAN and ours, the effectiveness of our multi-component weighed loss are verified. Compared with the result of Pix2Pix, our method can clearly restore sharp and continuous boundaries in flagella of Leishmania. The quantitative evaluation results with standard metrics (PSNR, SSIM and PCC) are shown in Table 2, which is consistent with the result of qualitative evaluation result. Our method outperforms other GAN-based classic models and traditional method in all three evaluation metrics.

4.2. Comparison of different blur-levels on the BPAEC dataset

BPAEC experiment is designed for evaluating the deblurring capability of our model in confocal fluorescence microscopy image dataset. The proposed model is trained and tested in the BPAEC

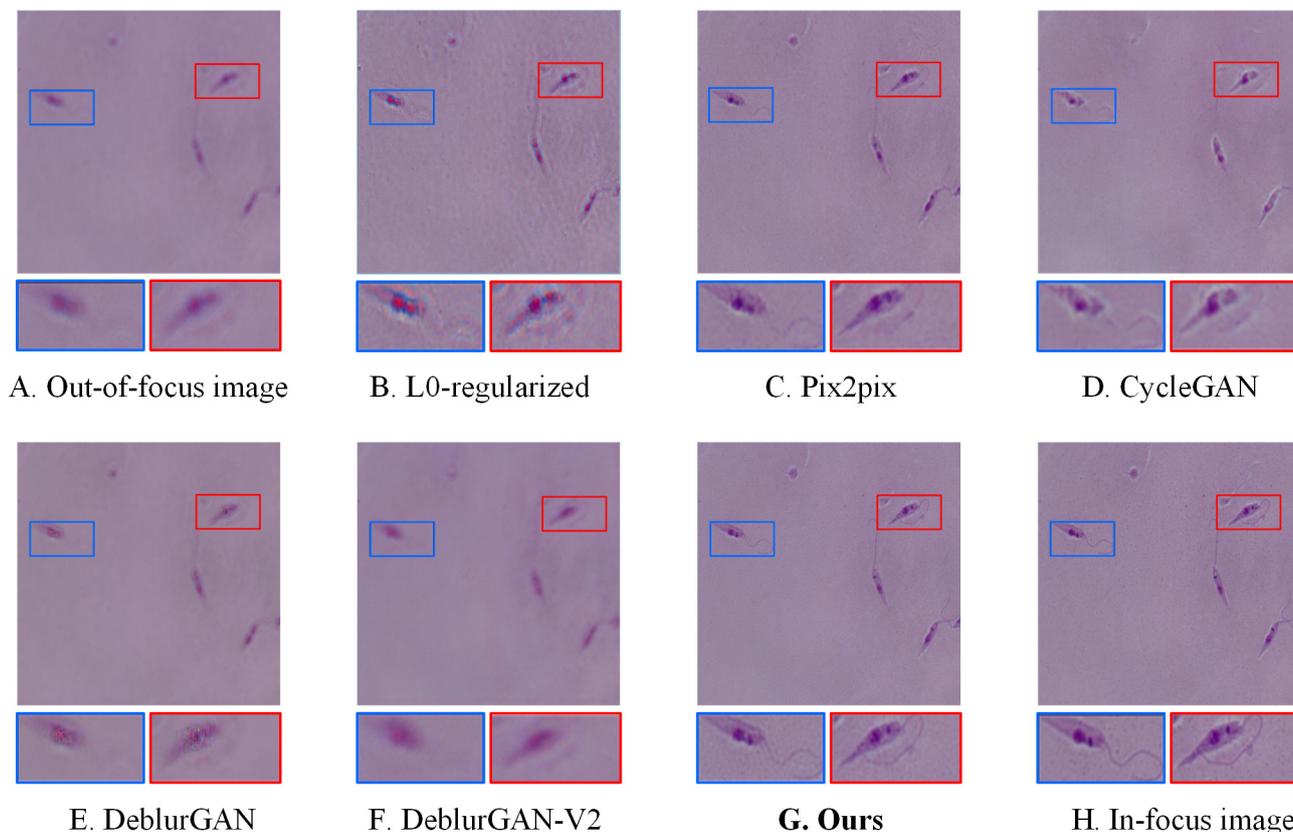


Fig. 3. The qualitative evaluation results in the Leishmania parasite dataset. Red and blue zoomed-in regions correspond to two different field of views. The results of different methods were generated by testing them in the same Leishmania parasite dataset. Zoomed-in regions of Leishmania parasites were shown at the bottom of the image. Our method significantly improves the quality of out-of-focus microscopic images and generates sharp boundaries and clear textures compared to other GAN-based deblurring methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dataset. Before feeding them into network, all raw images with the resolution of 1024×1024 pixels are resized to 512×512 pixels. Multiple trainings are performed for each layer of mitochondria, actin, and nuclei, and the unique mapping from the single blurred layer to the sharp layer is learned by the model. The restored result is shown in Fig. 4. The first row represents raw data inputted into the trained model, the second row represents the restored results by our trained model, and the third row is the ground truth. Observed from these figures, the fine features of microtubules, mitochondria and nuclei, which are in agreement with the ground truth images are enhanced. As shown in Table 3, image quality metrics (PSNR, SSIM, PCC) are computed to quantitatively evaluate the model performance. The deviation values of all metrics fluctuate in a small range, which reflects the stability of our method. In general, the evaluation metrics size restored from the layers $z = 6$

and $z = 8$ are higher than in any other layers. This result denotes that our model is convergent and the quality of the restored images is proportional to the quality of the input raw images. The evaluation metrics size of nucleus is significantly better than the metrics size of actin and mitochondria. Under the premise that the data instruments, environment and other variables in capturing images are the same, the only variable is the shape of the target. Compared with actin and mitochondria, the fixed morphology features of the nucleus, such as the fixed shape, clear boundaries and weak background interference reduced difficulty for model to recover images.

4.3. Generalizability of our model on the BBBC006 dataset

We evaluate model generalizability in another publicly available confocal fluorescence microscopy image dataset. We employ

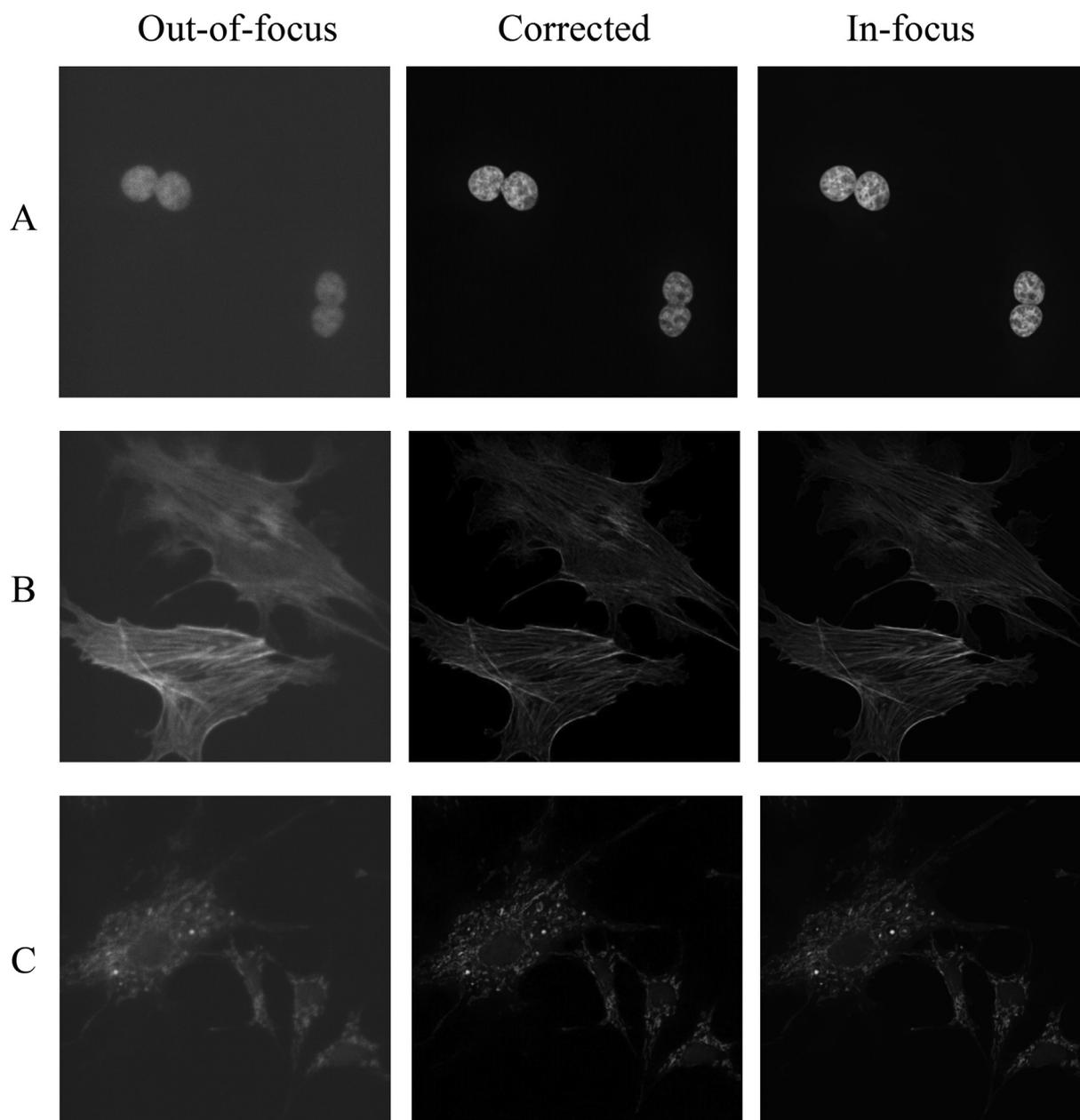


Fig. 4. The reconstruction results of our method in the BPAEC dataset. The first row represents the out-of-focus images inputted into the trained model. The out-of-focus images come from layer $z = 4$. The second row represents the restored results by our trained model, and the third row is the ground truth in-focus image. The in-focus images come from layer $z = 7$. Line A is nucleus, B is actin and C is mitochondria. The in-focus image has sharp edges while the out-of-focus images are blurred and noisy. The images corrected by our method are very similar to the in-focus images in both boundaries and textures.

the out-of-focus microscopy image set BBBC006 from the Broad Bioimage Benchmark Collection, which contains a z-stack of human cell image sets stained by Hoechst and phalloidin [41]. The layer $z = 16$ is at the optimal focal plane, which was used as the ground truth. The layers $z = 0$ to $z = 15$ are above the optimal focal plane, and $z = 17$ to $z = 32$ are below the optimal focal plane. We randomly choose images from 6 representative out-of-focus layers, namely $z = 5, 10, 15, 20, 25$ and 30 . Each layer has resolution with 696×520 pixels. The out-of-focus images of BBBC006 dataset were directly fed into our model pre-trained on BPAEC without further training. The correction results of different field of views (FoVs) from different layers are displayed in Fig. 5. Compared with out-of-focus reconstruction in nuclei that using the same dataset [43], our results are comparable to the autoencoder-based method. This experiment show that our method are of excellent generalization capabilities, which can be transferable to various scenarios of microscopy images. However, our model restores nuclei better than actin. Some regions in actin that has the irregular shape were not recovered well. This conclusion is consistent with the result in BPAEC, as shown in Fig. 4 B and Table 3. In addition, some layers were not recovered well which are very far away from the optimal

focal plane (Fig. 5, B and L), suggests this model may need to be retrained to achieve a better correction result.

5. Discussion and conclusion

Microscopic imaging is widely used in biomedical research and medical discovery. We design a deep learning based deblurring method to learn mapping from out-of-focus microscopy images to in-focus images. Our model performs well in a wide range of microscopy images ranging from protozoan parasite, to nucleus, actin and mitochondria of mammalian cells, thus showing its great potential in the correction of out-of-focus microscopy images. Our model employs CycleGAN with multi-component weighted loss, significantly improving image quality in the post-processing of out-of-focus microscopy images. Our method utilizes multi-layer CNNs to correct out-of-focus images by learning regression mapping functions between in-focus and the corresponding out-of-focus images. The proposed algorithm is trained on two self-collected datasets. The testing results on two self-collected datasets and one publicly available dataset demonstrate that CycleGAN combined with multi-component weighted loss can restore images

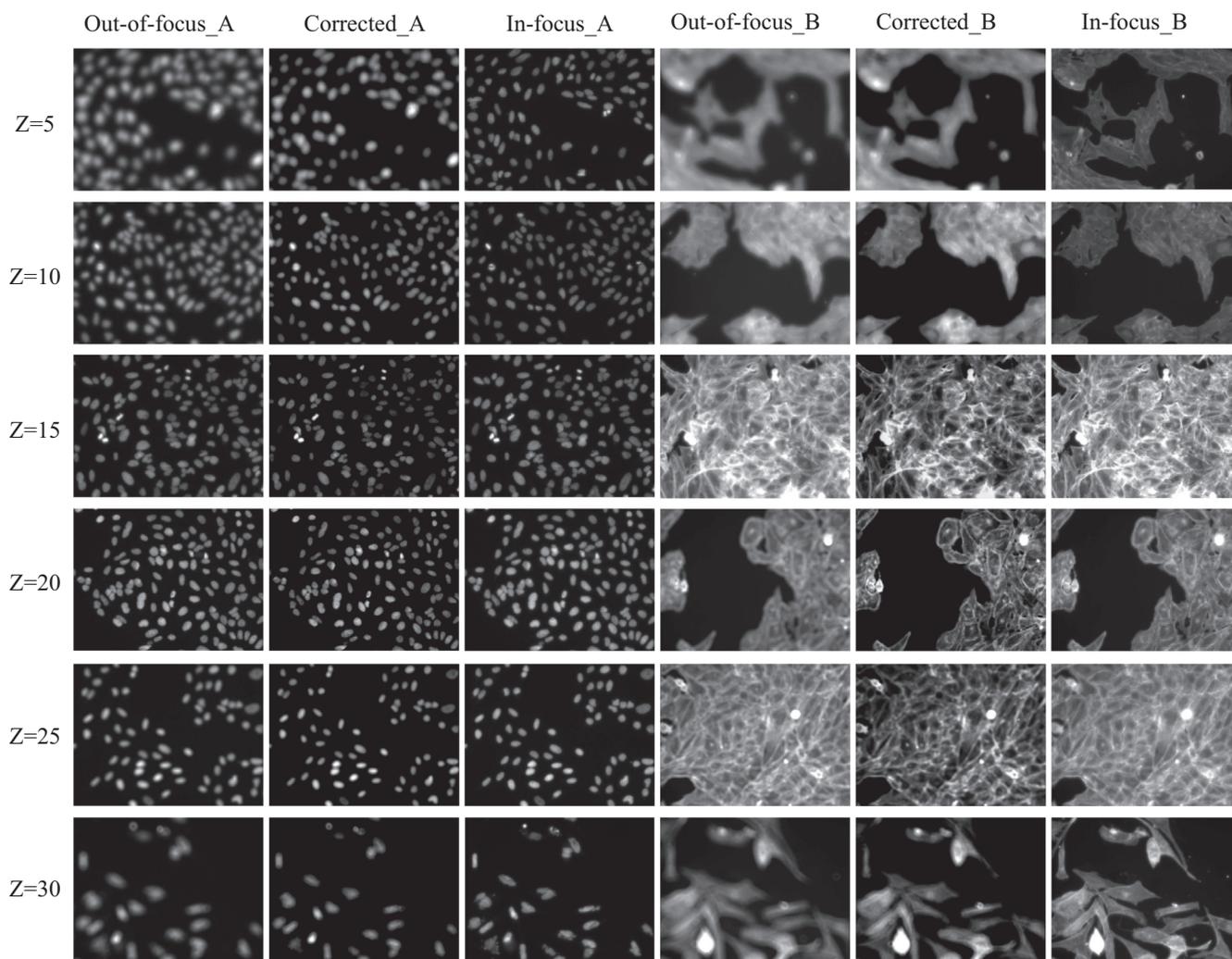


Fig. 5. The generalizability of our trained model was evaluated on the BBBC006 dataset. The BBBC006 dataset contains 32 layers of microscopy images. The focal plane is $Z = 16$, which was used as the ground truth in-focus layer. The results on six representative out-of-focus layers ($Z = 5, 10, 15, 20, 25$ and 30) from different field of views were shown. The out-of-focus images were directly inputted into the trained model without further training, and the corrected images were generated by our trained model. Comparing to the corresponding in-focus images, our model showed generalization capability in the correction of those blurring microscopy images. The character 'A' denotes Hoechst stained image for labeling nuclei, while 'B' denotes phalloidin stained image for labeling actin.

well from both bright-field microscope and confocal fluorescence microscope. Experiment of generalizability demonstrates that our proposed method can be applied to various scenarios and fields where microscope is needed, e.g. diagnosis, where optical microscopy is among the most widely used and deployed techniques.

Although the proposed method is robust for different types of microscopy image datasets, some limitations remain. As shown in the zoomed-in patches in Fig. 3, our restored images generate slight checkerboard artifacts. This is caused by CNN in both of the two processes: forward-propagation of upsampling layers and deconvolution layers with non-unit strides [44]. Some methods successfully alleviate or avoid generating checkerboard artifacts and the effectiveness is verified in natural image and medical image [37,44]. The idea extending the conventional condition for linear systems to non-linearity CNNs is worth checking out for future improvement [44]. Secondly, our model speeds up training by importing ImageNet pre-trained parameters, but ImageNet is a natural image dataset with feature different from microscopy images. Importing pre-trained parameters from a more correlated dataset may further improve the model performance in restoring microscopy image. In addition, simultaneously acquiring the paired blurred-sharp images may not be easy to obtain in some cases and the quality of the sharp images in training set directly affect the model performance. Out-of-focus correction based on unsupervised learning can be considered in the future work. Moreover, empirically setting the hyperparameters λ_1 , λ_2 and λ_3 in multi-components weighted loss function is a difficult and expensive process. Automatic loss weighting strategies, such as using homoscedastic uncertainty as a basis for weighting losses in a multi-task learning may be useful for further improving model performance [40]. At last, the 3-dimensional structures of nuclei, actin filaments, and mitochondria are not considered in this study. Nuclei, for example, are often roughly spheroid/ellipsoid in shape, suggesting that different cross-sections along the optical axis may have different areas, depending on the spacing between the z-stacked images. However, the qualities of the reconstructions are computed versus the central (“in-focus”) image in the z-stack without regarding to spatial variation. The spacing of the z-stacked images should be considered more thoroughly for future study. The general applicability of our model was tested under different experimental circumstances. But considering instrument-to-instrument fluctuation, this model may need to be retrained to be transferable to different types of microscopy images.

Funding

The work was supported by the Shenzhen Science and Technology Program the university stable support program (20200821222112001), and the Guangdong Basic and Applied Basic Research Foundation (2021A1515220115).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Betzig E et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 2006;313:1642–5.
- Combalia M et al. Digitally stained confocal microscopy through deep learning. In: Proceedings of 2nd Intl Conf. medical imaging with deep learning, p. 121–9.
- Do DT et al. Using deep neural networks and biological subwords to detect protein S-sulfenylation sites. *Brief Bioinform* 2020;bbaa128.
- Dyba M et al. Immunofluorescence stimulated emission depletion microscopy. *Nat Biotechnol* 2003;21:1303–4.
- Gustafsson MGL. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J Microsc* 2000;198:82–7.
- He K et al. Deep residual learning for image recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 770–8.
- Huang B et al. Breaking the diffraction barrier: Super-resolution imaging of cells. *Cell* 2010;143:17565–9.
- Huang F et al. Ultra-high resolution 3D imaging of whole cells. *Cell* 2016;166:1028–40.
- Isola P et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 5967–76.
- Jiang H et al. Geometry-aware cell detection with deep Learning. *mSystems* 2020;5:e00840–19.
- Jin L et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nat Commun* 2020;11:1934.
- Johnson J et al. Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision, p. 694–711.
- Kingma, D.P. and Ba, J.L. (2015) Adam: A method for stochastic optimization. *preprint arXiv: 1412.6980*.
- Kupyn O et al. DeblurGAN: blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 8183–92.
- Kupyn O et al. DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE international conference on computer vision, p. 8878–87.
- Le NQK. Fertility-GRU: identifying fertility-related proteins by incorporating deep-gated recurrent units and original position-specific scoring matrix profiles. *J Proteome Res* 2019;18:3503–11.
- Leonard AP et al. Quantitative analysis of mitochondrial morphology and membrane potential in living cells using high-content imaging, machine learning, and morphological binning. *Biochim Biophys Acta - Mol Cell Res* 2015;1853:348–60.
- Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: European conference on computer vision, p. 702–16.
- Lim S, Ye J.C. Blind deconvolution microscopy using cycle consistent CNN with explicit PSF layer. In: Proceedings of the international workshop on machine learning for medical image reconstruction, p. 173–80.
- Nah S et al. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 257–65.
- Ouyang W et al. Deep learning massively accelerates super-resolution localization microscopy. *Nat Biotechnol* 2018;36:460–8.
- Pan J et al. Blind image deblurring using dark channel prior. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 1628–36.
- Persch N et al. Enhancing 3-D cell structures in confocal and STED microscopy: A joint model for interpolation, deblurring and anisotropic smoothing. *Meas Sci Technol* 2013;24:125703.
- Weigert M et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods* 2018;15:1090–7.
- Rivenson Y et al. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci Appl* 2018;7:17141.
- Rust MJ et al. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 2006;3:793–5.
- Schuler CJ et al. Learning to Deblur. *IEEE Trans Pattern Anal Mach Intell* 2016;38:3503–11.
- Shan Q et al. High-quality motion deblurring from a single image. *ACM Trans Graph* 2008;27:73.
- Tao X et al. Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 8174–82.
- Xu L et al. Unnatural L0 sparse representation for natural image deblurring. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 1107–14.
- Yang Z et al. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* 2012;20:237–47.
- Yi Z et al. DualGAN: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE international conference on computer vision, p. 2868–76.
- Young SS et al. Adaptive deblurring of noisy images. *Appl Opt* 2007;46:744–52.
- Zhang J et al. Dynamic scene deblurring using spatially variant recurrent neural networks. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, p. 2521–9.
- Zhang Y et al. Deep Learning for Imaging and Detection of Microorganisms. *Trends Microbiol.* 2021. S0966-842X.
- Bian Z et al. Autofocusing technologies for whole slide imaging and automated microscopy. *J Biophotonics* 2020;13(12):e202000227.
- Cheng K et al. Depth Estimation for Colonoscopy Images with Self-supervised Learning from Videos. *International conference on medical image computing and computer-assisted intervention*, 2021.
- Goodfellow, I., et al. (2014). Generative adversarial nets. 27.
- Hore, A., et al. (2010). Image quality metrics: PSNR vs. SSIM. 2010 20th international conference on pattern recognition, IEEE.

- [40] Kendall A et al. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.
- [41] Ljosa V et al. Annotated high-throughput microscopy image sets for validation. *Nature Methods* 2012;9(7):637.
- [42] Nah S et al. Deep multi-scale convolutional neural network for dynamic scene deblurring. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [43] Raureau A. Transfer learning for cell image reconstruction. University of Twente; 2021.
- [44] Sugawara, Y., et al. (2019). Checkerboard artifacts free convolutional neural networks. 8.
- [45] Wang, Z., et al. (2004). Image quality assessment: from error visibility to structural similarity. 13(4): 600-612.
- [46] Wen Y et al. Structure-aware motion deblurring using multi-adversarial optimized cyclegan. *IEEE Trans Image Process* 2021;30:6142–55.
- [47] Zhu J-Y et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2017.