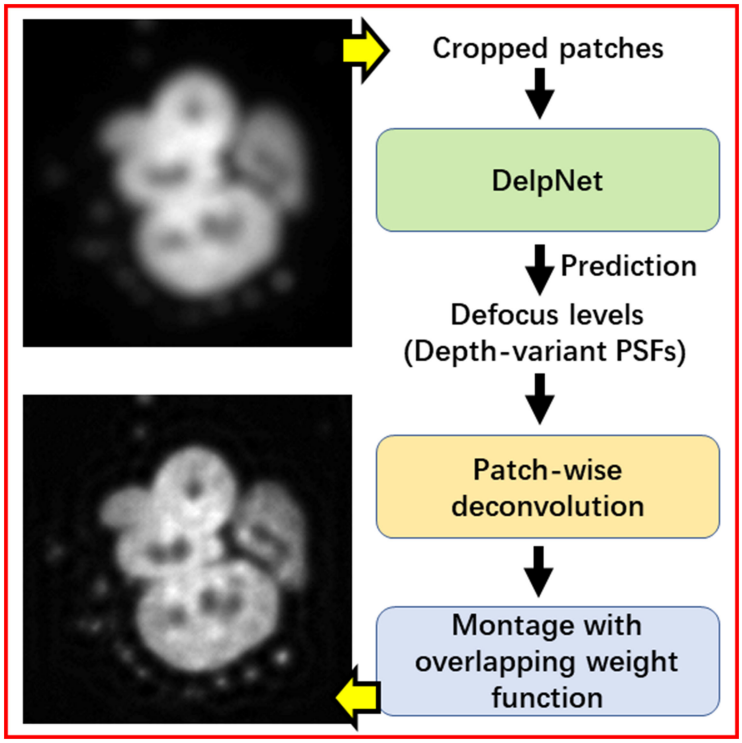
IEEE Photonics Journal
An IEEE Photonics Society Publication

Restoration of Out-of-Focus Fluorescence Microscopy Images Using Learning-Based Depth-Variant Deconvolution

Volume 12, Number 2, April 2020

Da He
De Cai
Jiasheng Zhou
Jiajia Luo
Sung-Liang Chen



DOI: 10.1109/JPHOT.2020.2974766

IEEE photonics SOCIETY

IEEE

Restoration of Out-of-Focus Fluorescence Microscopy Images Using Learning-Based Depth-Variant Deconvolution

Da He,[1] **De Cai**,[1] **Jiasheng Zhou**,[1] **Jiajia Luo**,[2] **and Sung-Liang Chen** [1]

[1]University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Biomedical Engineering Department, Peking University, Beijing 100191, China

DOI:10.1109/JPHOT.2020.2974766

Manuscript received February 11, 2020; accepted February 13, 2020. Date of publication February 17, 2020; date of current version March 9, 2020. This work was supported by the National Natural Science Foundation of China under Grants 31870942 and 61775134. *(Da He and De Cai contributed equally to this work.)* Corresponding authors: Jiajia Luo; Sung-Liang Chen (e-mail: jiajia.luo@pku.edu.cn; sungliang.chen@sjtu.edu.cn).

Abstract: Image quality is degraded in the out-of-focus region because of the depth-variant (DV) point spread function (DV-PSF) of a fluorescence microscope. Either non-blind or blind deconvolution for restoration results in limited improvement. In this work, we propose a two-step learning-based DV deconvolution (LB-DVD) to restore the out-of-focus image. In the first step, DV-PSF is predicted by a defocus level prediction convolutional neural network (DelpNet). In the second step, the extracted DV-PSF is used for DV deconvolution. To our knowledge, LB-DVD is proposed and demonstrated for the first time. DelpNet achieves an accuracy of 98.2% for predicting defocus levels of image patches (84×84 pixels). The subsequent DV deconvolution gives rise to good performance in peak signal-to-noise ratios and structural similarity index, which are improved by up to 6.6 dB and 11%, respectively, before and after the deconvolution. As for a wide-field image, there exist different DV-PSFs within the two-dimensional fluorescence image due to the surface undulation. An overlapping weighting patch-wise LB-DVD is used in image montage to eliminate patch boundary artifacts. As a result, our LB-DVD shows the feasibility and promise to be applied to typical fluorescence microscopy in practical applications.

Index Terms: Fluorescence microscopy, convolutional neural network, deconvolution.

1. Introduction

Fluorescence microscopy is widely used in biomedical applications such as visualizing structures of cells and tissues [1]. However, the depth-variant (DV) point spread function (PSF) (DV-PSF) of the microscope and noise degrade the image quality of the two-dimensional (2D) microscopy images. The former, DV-PSF, indicates different 2D PSFs when samples are placed at various depths, and is a result of the limited depth of field in a fluorescence microscope, causing low-quality images in the out-of-focus region [2]. For example, Fig. 1 shows representative examples of in-focus and out-of-focus (or defocused) images, the latter obviously showing much-blurred patterns. Although some samples can be re-aligned at the focus of the microscope to restore the blurred image, some

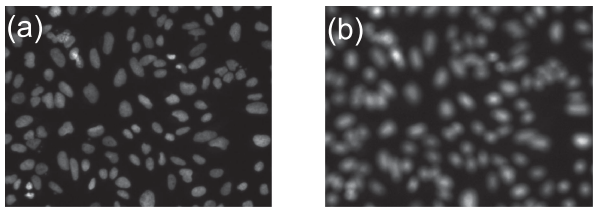


Fig. 1. Representative examples of fluorescence microscopy images. (a) An in-focus image. (b) An out-of-focus image.

particular samples such as those with non-flat surface suffer from regional blurs within a wide-field image and thus cannot be retrieved by focus re-alignment.

Since the defocused image can be modeled as the convolution between the latent clear image and a PSF, clear images may be obtained if the convolution process can be reversed (i.e., deconvolution). Therefore, DV deconvolution has been intensively studied to restore the out-of-focus images through non-blind and blind ways, depending on how much the system's PSF is known. Non-blind deconvolution requires the known PSF typically from experimental measurement [3], [4]. However, experimental measurement of PSF is usually inconvenient and time consuming, which may result in restrictions in biomedical applications. On the other hand, blind deconvolution sidesteps the acquirement of PSF by using only the blurred image [5], yet the performance is usually impeded due to a lack of system's prior knowledge.

Recently convolutional neural network (CNN) has shown its great potential for image processing, which implements various computer vision tasks such as image classification by ResNet [6], object detection by YOLO [7], and instance segmentation by Mask R-CNN [8]. CNN is a learning-based method and has the power to work for a variety of input images including medical and biological images. It has been shown that CNN can be used to predict defocus level and thus DV-PSF [9]. CNN-based methods have also been employed to deblur images and remove artifacts. For example, some large convolution kernels as well as traditional deconvolution schemes can be combined for efficient image restoration [10]. Another example is that direct mapping from blurred images to clear images is able to be learned by CNN to handle adaptive optics retinal images [11].

As a matter of fact, semi-blind deconvolution is a promising method to restore image quality. Unlike blind deconvolution without knowing PSF, the semi-blind method estimates some parameters of a given PSF model, followed by the deconvolution using this estimated PSF. The estimation is learned from the relation between the PSF parameters and blurred images. It has been shown that semi-blind deconvolution can increase the overall robustness and generality in image restoration [12], [13]. For example, a learning-based spatially-variant (SV) deconvolution was demonstrated to deal with the regional blurs within a wide-field image [12], i.e., SV image blurs in 2D. However, although promising, the semi-blind deconvolution has not been investigated to recover DV image blurs.

In this paper, aiming at restoring images blurred by DV-PSF, we therefore propose learning-based DV deconvolution (LB-DVD) to restore the out-of-focus images in fluorescence microscopy. To our knowledge, this is the first study to utilize learning-based DV-PSF to handle out-of-focus microscopy images. As mentioned previously, CNN can be used to estimate DV-PSF [9], and thus, semi-blind deconvolution to deal with DV image blurs is feasible. Compared with existing DV deconvolution using non-blind and blind approaches, as illustrated above, the semi-blind LB-DVD circumvents the challenges of acquiring DV-PSF in the experiment encountered by the non-blind approach and also enjoys more accurate DV-PSF than the blind approach. Specifically, DV-PSF is predicted by defocus level prediction CNN (DelpNet) and is then used for DV deconvolution. The DelpNet achieves an extremely high accuracy of 98.2% for classifying the defocus levels of image patches (84×84 pixels). On the other hand, in part by virtue of highly accurate DV-PSF, the subsequent DV deconvolution gives rise to good performance in peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [14] of the DV-deconvolved images, which are improved by up to 6.6 dB and

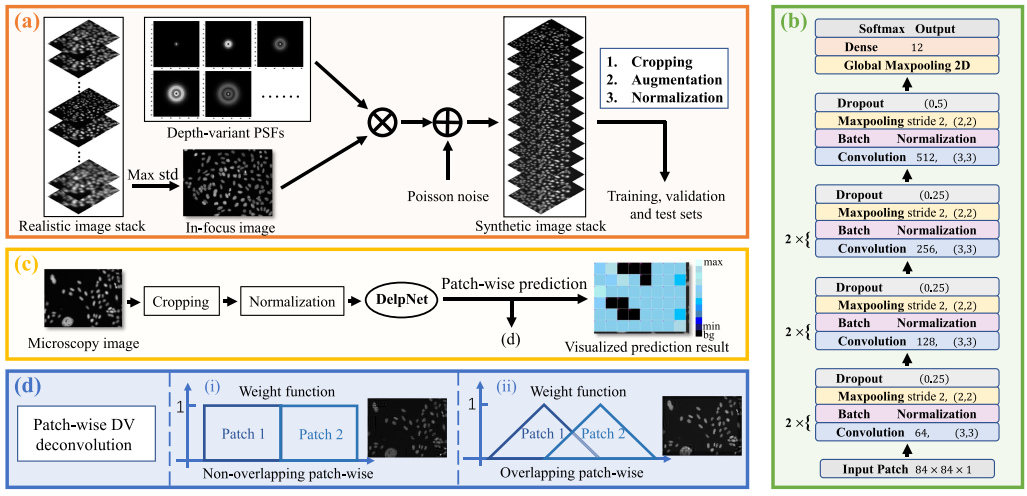


Fig. 2. The overview of the method. (a) The process of generating synthetic datasets. (b) The architecture of DelpNet. (c) An example of the patch-wise defocus level prediction process. The prediction results are expressed in different colors, and the "bg" here indicates a background patch. (d) The patch-wise DV deconvolution for two cases: (i) the stride is the same as the patch size and it is just the non-overlapping patch-wise deconvolution with rectangular weight function and (ii) the stride is smaller than the patch size and triangular function is used as the weights.

11%, respectively, better than those of the depth-invariant (DI) deconvolved counterparts. Further, "patch-wise" LB-DVD is explored to restore the image quality of a wide-field image, which better recovers local image characteristics. A weighting technique is adopted in image montage to remove patch boundary artifacts. The results suggest that LB-DVD is an effective way to restore the out-of-focus microscopy images and is applicable to wide-field images in practical applications.

2. Methods

Firstly, we classified the defocus levels of patches in fluorescence microscopy images by DelpNet. The DV-PSF can then be recognized and used in deconvolution for image restoration. For wide-field image processing, patch-wise deconvolution with an overlapping weighting technique was employed. The overview of our pipeline is shown in Fig. 2.

2.1 Fluorescence Microscopy Images

As mentioned previously, out-of-focus blurred fluorescence microscopy images of cells are shown in Fig. 1(b). This imaging process can be modeled as follows. The fluorescence microscopy image $i(x, y; z)$ can be represented as:

$$i(x, y; z) = \wp(o \otimes h)(x, y; z) \tag{1}$$

where $(x, y; z)$ represents a 2D function over x, y with a variable z showing the depth of objects along the optical axis, $\wp$ is the Poisson noise model, $\otimes$ stands for the convolution operator, $o(x, y; z)$ is the imaged object and $h(x, y; z)$ is DV-PSF.

Note that in Eq. (1), the 2D image (or image patch) $i(x, y; z)$ lying on the same 2D plane is assumed, and the 2D DV-PSF $h(x, y; z)$ is used, which is valid for a small image patch. For a large image or a three-dimensional (3D) object that can lie in multiple 2D planes, a more complicated integral equation using 3D DV-PSFs to model the DV imaging problem should be used [15], [16].

If DV-PSF $h(x, y; z)$ can be estimated and used for deconvolution, one can recover clear 2D images $o(x, y; z)$ at different depth z. For an optical or fluorescence microscope, the information of the depth z of the PSF $h(x, y; z)$ varies depending on the depth where the object is placed.

That is, the depth z is a parameter that cannot be determined from the specification of the microscope. Fortunately, the depth z can be inferred from the captured image $i(x, y; z)$, as indicated in Eq. (1). This lays a foundation to estimate depth information of an image, i.e., DV-PSF, by DelpNet (details described later). Note that instead of a regression problem, we handle DV-PSF as a classification problem, in part because accurately measuring and labeling DV-PSF is not easy. Moreover, the approach to classify DV-PSF into several depth ranges can be more robust and practical.

We generated synthetic images with 11 defocus levels from realistic (i.e., experimental) focal images for training (Fig. 2(a)), and applied the trained model to realistic images with unknown defocus level for application (Fig. 2(c)). For the training dataset, instead of using realistic defocused images, synthetic ones were used because of more accurate defocus level labels in the synthetic defocused images.

We used fluorescence images of U2OS cells with Hoechst stain in the BBBC006 dataset [17] in this study. The dataset acquired by an ImageXpress Micro automated cellular imaging system (Molecular Devices, CA) consists of image stacks, each stack having images of a specimen at different depths. The 381 in-focus images of size 696×520 were first selected based on the maximum standard deviation of image intensity from 381 image stacks. They are regarded as the reference in-focus images. Synthetic defocused image dataset was then obtained by convolving the selected in-focus images with the following DV-PSF with 11 defocus levels [9], [18]:

$$h(x, y; z) = \left| \int_0^1 J_0\left(k\frac{NA}{n}\sqrt{x^2+y^2}\rho\right) exp\left(-\frac{1}{2}jk\rho^2 z\frac{NA^2}{n^2}\right)\rho d\rho \right|^2 \tag{2}$$

where J_0 is the Bessel function of the first kind, order zero, $k = 2\pi/\lambda$ is the wavenumber with the wavelength λ of 451 nm, NA is the numerical aperture of 0.5, and n is the refractive index of 1.0 in air. The defocus level has an increment of $2\,\mu$m along the z optical axis, and level 0 corresponds to the in-focus images (i.e., no synthetic degradation). As mentioned above, the synthetic defocused image dataset with 11 defocus levels (from 0 to 10) was obtained by convolving the 381 in-focus images. This resulted in 4191 [$= 381 \times 11$] synthetic images, which were randomly split into training, validation, and test sets with a ratio of 0.75 : 0.15 : 0.1. Each of the 4191 images (size 696×520) was randomly cropped for 20 times to generate image patches of size 84×84. As a result, the total number of images used for training DelpNet was 62865 [$= 4191 \times 20 \times 0.75$]. We assigned some patches the label "bg," whose maximum pixel values are smaller than 230 and the maximum difference among pixel values is smaller than 30 considering 16-bit images, indicating that the patch is almost full of background noise. This special augmentation is helpful because it can partially avoid providing DelpNet with meaningless interference. Hence, after the "bg" augmentation, we had 12 labels of image patches. Linear normalization was applied to all image patches to scale their value ranges to 0–1, which means every pixel value was divided by 65535. The labeled image patches were then used for training DelpNet as well as further analysis.

2.2 DelpNet

As shown in Fig. 2(b), DelpNet includes 7 convolutional layers. The shape of the input layer was set to match the input patch size. Dropout [19] and BatchNormalization [20] layers were added to avoid overfitting. Every BatchNormalization layer was followed by a ReLU activation function. GlobalMaxPooling layer [21] was utilized to transform the convolutional feature maps to a feature vector before the output layer. As mentioned above, we had 12-label outputs consisting of 11 defocus levels together with the extra "bg" label.

There are some special settings in DelpNet. Firstly, we chose a relatively plain CNN architecture instead of residual learning style like ResNet [6] or multi-scale feature fusion design like Inception network [22], as plain architectures surprisingly performed better in our results. Secondly, it is uncommon to put Batch Normalization layers and Dropout layers together densely, but we found this strategy useful for improving the performance and reducing overfitting. In addition, we modified the default momentum value of Batch Normalization [20] layers in Tensroflow from 0.99 to 0.60,

which is also unusual but helpful for convergence for our dataset. All these strategies and results will be described later in Section 3.

As for implementation, DelpNet was built using Keras framework with Tensorflow [23] backend. We set Adam optimizer [24] learning rate of $6e-5$ and decay of $5e-6$. With a batch size of 128, it was then trained and evaluated on a single Nvidia Titan Xp GPU. Categorical Cross-entropy loss was applied to this multiclassification problem.

2.3 Patch-Wise DVD

Assuming Poisson noise for the imaging model, Richardson-Lucy (RL) deconvolution [25], [26] is a widely-used iterative algorithm to restore blurred images with known PSF:

$$O'_{t+1}(x, y; z) = \left[\frac{i(x, y; z)}{h(x, y; z) \otimes o'_t(x, y; z)} \otimes h(-x, -y; -z) \right] o'_t(x, y; z) \tag{3}$$

where t represents the iterative index, $o'_t(x, y; z)$ is the deblurred image in each iteration, and $i(x, y; z)$ is the initial guess of $o'_0(x, y; z)$.

For image patches, the iterative times of 15 were applied. The DV-PSF based on the predicted defocus level by DelpNet was used for deconvolution. No deconvolution was applied to image patches with labels of "bg" and defocus level 0. In our implementation, to avoid edge artifacts due to deconvolution in an image patch, patch size of 84×84 was padded to a larger patch size of 110×110 before deconvolution, and the padded part was removed after deconvolution. The padding values were neighbor pixels in the whole image or just 0 if there are not enough neighbors.

Note that DV-PSF can also be spatially variant in a wide-field image. In other words, there exists different DV-PSF within the wide-field image, e.g., due to the surface undulation of a sample. Therefore, it is preferred to apply "patch-wise" LB-DVD to the wide-field image, which better restores the image quality of the wide-field image because local image characteristics are taken into account. Specifically, the two-step LB-DVD was applied to all image patches, as detailed above, and then an overlapping weighting technique was used for image montage to obtain the deconvolved wide-field image. Similar strategies termed as adaptive deconvolution to restore images according to local image characteristics were reported [27]–[29]. The weight function has two types, non-overlapping patch wise and overlapping patch wise, as shown in Fig. 2(d). For the former, the weight function is a rectangular function with the patch stride set the same as the image patch size. This causes severe patch boundary artifacts (demonstrated later). On the other hand, for the latter, the weight function is a triangular function with base width set the same as the image patch size and the patch stride set as half of the image patch size. To be specific, patch-wise LB-DVD was applied to overlapped image patches (as shown in (ii) of Fig. 2(d)), and then, image montage was done by stitching the overlapped deconvolved image patches using the triangular weight function. It can also be regarded as bilinear interpolation of adjacent image patches [30]. As can be seen later in Section 3, patch boundary artifacts can be effectively removed.

3. Results and Analyses

We used the held-out test set to evaluate the defocus level prediction performance and compare strategies in the following Subsection 3.1–3.3. The overall results sequentially combining DelpNet and DV deconvolution were analyzed in Subsection 3.4.

3.1 Results of DelpNet

After 1000 epochs of training, the evaluation results of DelpNet on the test set are shown in Fig. 3, which shows the distribution of prediction results using a confusion matrix.

Generally, our DelpNet method can identify the defocus levels with an accuracy of 98.2%, and the precision, recall as well as f1-score are all above 98%. Besides, from the confusion matrix in Fig. 3, we can easily find that the wrong predictions mostly locate in the neighbors of the ground

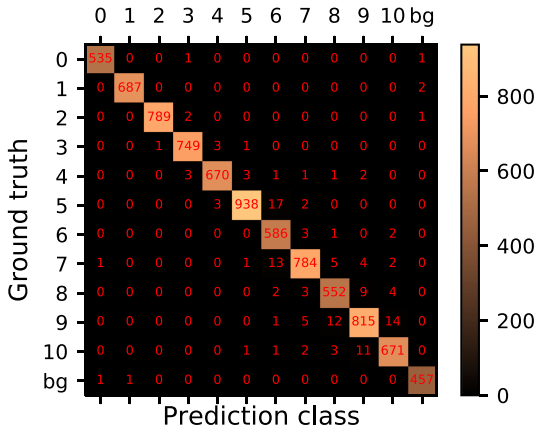


Fig. 3. The evaluation result of DelpNet. The confusion matrix shows the distribution of prediction results of DelpNet on the held-out test set.

TABLE 1
Evaluation Results of the Existence of "bg" Label

Setting	Accuracy	Precision	Recall	F1-score
Without "bg"	96.8%	96.6%	96.9%	96.7%
With "bg"	**98.2%**	**98.2%**	**98.3%**	**98.3%**

truth, which means that most wrong predictions differ from the ground truth slightly, and thus, the subsequent deconvolution with these wrong predictions can still contribute to image restoration to some extent.

3.2 Analysis of Special Settings in DelpNet

As mentioned previously, we set an extra-label "bg" to split image patches full of almost background from normal defocused image patches. To evaluate the value of this strategy, we trained almost the same DelpNet with two settings: with "bg" label and without "bg" label. The only difference is the final dense layer, which has 12 neurons and 11 neurons for cases with "bg" label and without "bg" label, respectively. The evaluation results are listed in Table 1.

As shown in Table 1, the "bg" label strategy worked well with accuracy improvement from 96.8% to 98.2%, which is equivalent to the decrease of top-1 error from 3.2% to 1.8%. In the DelpNet without "bg" label, we found that the image patches full of almost background can be mistakenly predicted to be various defocus levels, which is not desired. We think using "bg" label is very important to prevent CNN from excessively fitting the background. Compared with the DelpNet without "bg" label, that with "bg" label is expected to avoid learning the mapping from image patches full of noise to 11 defocus levels and thus enables better identification of meaningful cell signals, which leads to the enhanced performance.

Different from most works that apply Batch Normalization with the momentum parameter of a relatively large value (e.g., default 0.99 [23]), we assigned this parameter 0.60 because we found large Batch Normalization momentum parameter easily leads to a severe oscillation on the loss curve for validation, as shown in Fig. 4. In contrast, relatively small momentum values (e.g., 0.40 and 0.60) would not ruin the performance and help the steady convergence on the loss curve.

Besides the above two strategies, we also densely used Batch Normalization layers with Dropout layers. For fluorescence microscopy images of cells, there exists much more background noise than meaningful cell signals. As a result, more strategies to mitigate overfitting of the cell images used in this study are needed, compared with processing natural photos. The combination of the strategies

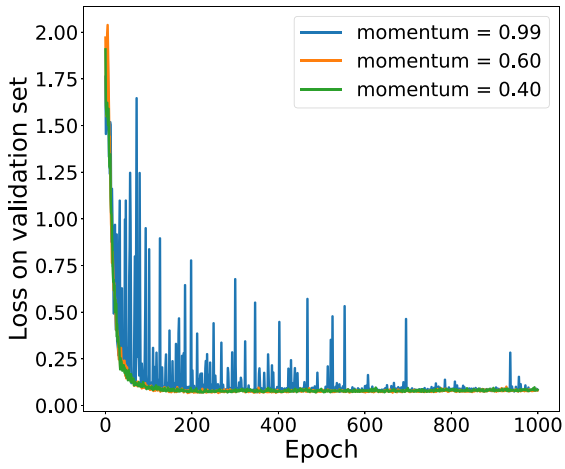


Fig. 4. The loss curves for validation with Batch Normalization momentum parameters of 0.99, 0.60 and 0.40. The blue curve with momentum of 0.99 is oscillatory.

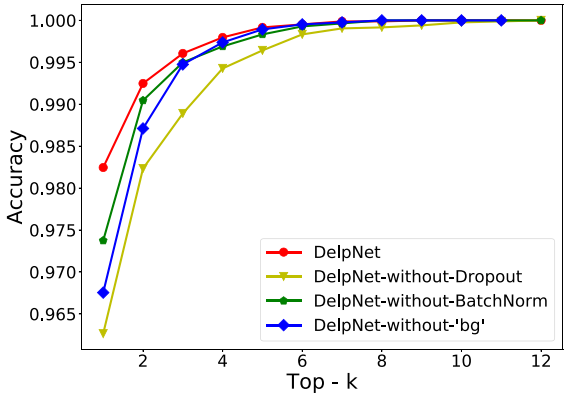


Fig. 5. Cumulative match curves of different settings. Each node represents the corresponding top-k accuracy of a specified training strategy (e.g., the most left blue node indicates the top-1 accuracy on the test set of 96.8% for the DelpNet model trained without "bg" label.). The variable k is in the range of 1–11 for the blue curve and the range of 1–12 for other curves.

to avoid overfitting would be helpful. We thus checked the cumulative match characteristic curves of DelpNet with the combined strategies of using Dropout layers, Batch Normalization layers, and the "bg" label. As a comparison, the curves of DelpNet without each of the above strategies were also studied, as shown in Fig. 5. The results in Fig. 5 indicate that each strategy to avoid overfitting is instrumental in improving performance.

Actually, we compared various image patch sizes as well. We found that larger image patch sizes easily led to higher accuracy. However, too large image patch size is against acquiring local DV-PSF, losing local image characteristics. Thus, we applied a relatively moderate image patch size of 84 $\times$ 84, which was also adopted in another work [9].

3.3 Comparison Among Architectures

To manifest the advantage of the developed DelpNet for fluorescence microscopy images of cells, we also compared the performance of DelpNet with various well-known representative CNN architectures in Fig. 6 and Table 2. Specifically, VGG16 [31] represents the plain CNN architectures without any concatenations or additions between feature maps. Inception_v3 [22] simultaneously applies different convolutional kernels and concatenates the feature maps together

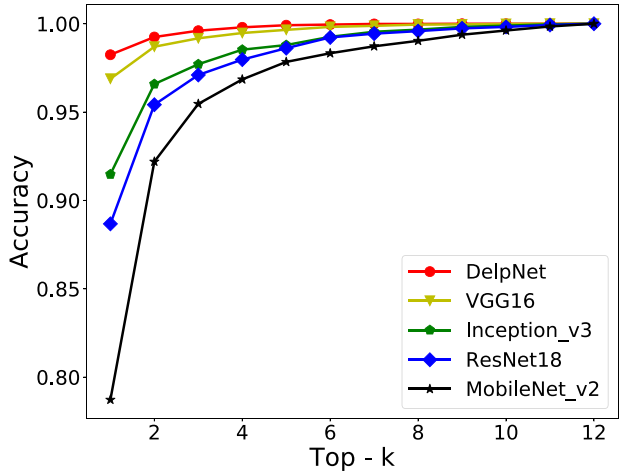


Fig. 6. Cumulative match curves of trained DelpNet and other 4 representative CNNs. Each node represents the corresponding top-k accuracy of a CNN architecture (e.g., the most left blue node indicates the top-1 accuracy on the test set of 88.7% for the trained ResNet18 model.). Only the input and output formats of the 4 evaluated CNNs were adjusted.

TABLE 2
Evaluation Results of Various CNN Architectures

Net	#params	Accuracy	Recall	F1-score
VGG16	39.9M	96.9%	97.0%	96.9%
Inception_v3	21.8M	91.5%	91.7%	91.6%
ResNet18	12.6M	88.7%	89.1%	88.8%
MobileNet_v2	**2.3M**	78.7%	79.6%	79.1%
DelpNet	**2.3M**	**98.2%**	**98.3%**	**98.3%**

for multi-scale fusion. ResNet18 [6] is a typical residual learning network with relatively fair layer numbers compared with its deep versions. Finally, MobelNet_v2 [32] is a light work proposed to implement CNN applications on mobile devices. Compared with their original architectures, we only adjusted the input and output formats to fit our dataset.

According to Fig. 6 and Table 2, DelpNet outperforms all the other evaluated CNN architectures. This may be explained as follows. In this study, the features of cells for estimating defocus levels are relatively sparse and homogeneous. The 4 evaluated CNNs employ more complicated networks and may not be suitable for cell images. For example, functions such as feature fusion and residual learning easily bring more but noisy features. The results also justify the necessity to design and develop a suitable CNN for cell images, rather than directly adopting an existing well-known CNN architecture. Besides, according to Table 2, there are only 2.3 M parameters in DelpNet, which is helpful for integrating DelpNet into embedded systems like an intelligent microscope.

In addition, we tried transfer learning by using the CNN architectures (VGG16, ResNet, InceptionNet, and MobileNet) pretrained on the ImageNet dataset [33]. The results show that transfer learning does not perform better than training from scratch.

3.4 Results and Comparisons of Patch-Wise LB-DVD

The trained DelpNet was then used for defocus level prediction and the subsequent deconvolution. The results are shown in Fig. 7. To show the applicability of the DV deconvolution based on the trained DelpNet (using the dataset mentioned in Subsection 2.1) to more shapes, created phantoms including beads (denoted as O1) and a ring (denoted as O2) were tested first. Then, a realistic image patch with 3 adjacent U2OS cells (denoted as O3) was used. Each cell was about

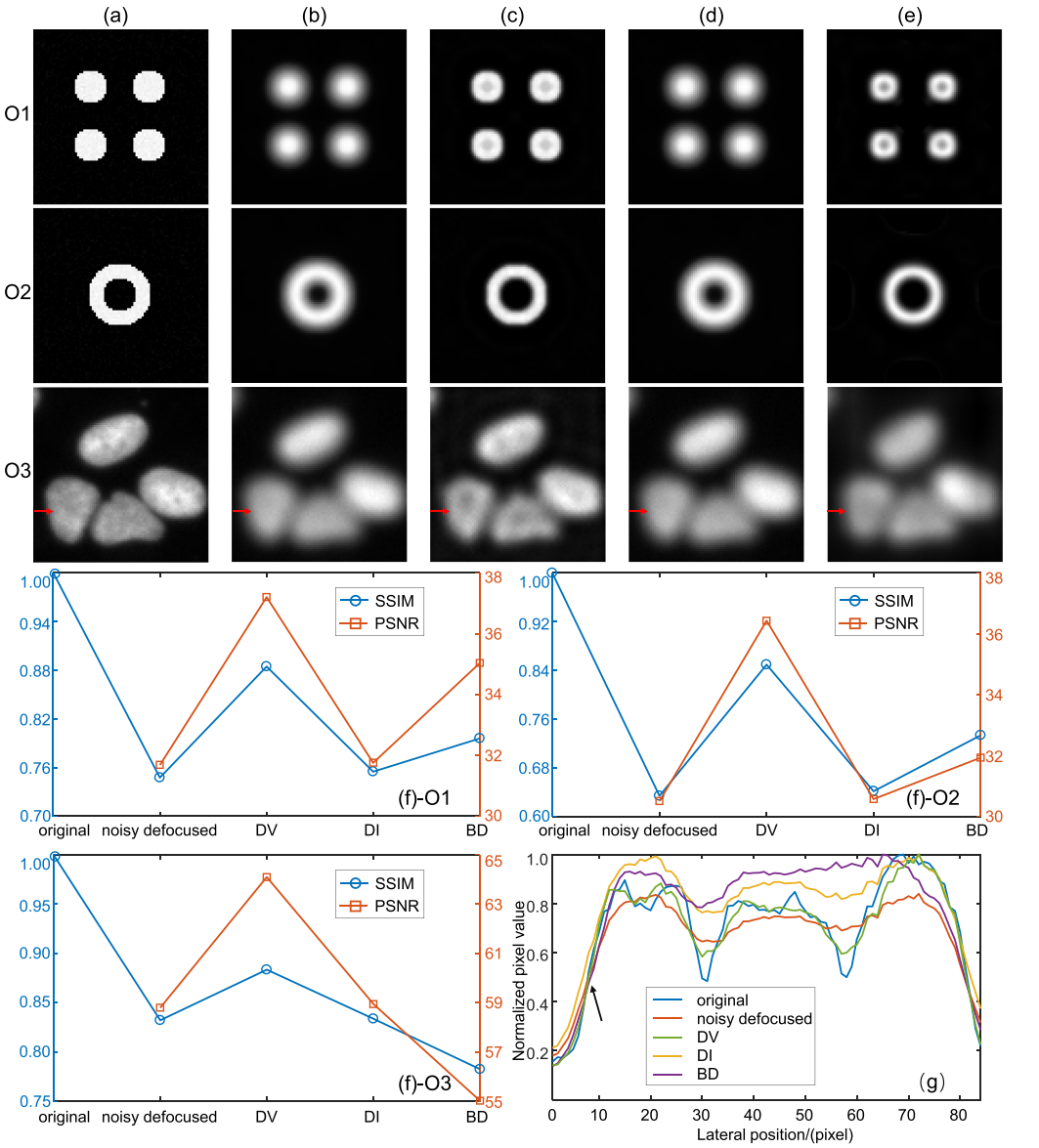


Fig. 7. Image restoration of created phantoms (O1 for beads and O2 for a ring) and realistic image (O3). (a) original in-focus images, (b) noisy defocused images, (c) DV deconvolved images, (d) DI deconvolved images, and (e) BD deconvolved images. (f) The SSIM and PSNR values of the above images. The PSNR values of (a) are infinite. (g) The 1D profiles of the (a3)–(e3) images along the red arrow direction.

10 μm in diameter. The scale bar was not shown as it is not provided in the public dataset. The original in-focus images are shown in Fig. 7(a). The DV-PSF with a random defocus level blurred the in-focus images, resulting in defocused images. Note that for a fair comparison, the same defocus level was applied to O1–O3. Poisson noise was then added to the defocused images, named as noisy defocused images, as shown in Fig. 7(b). Then, the noisy defocused images were DV deconvolved using the DV-PSF of the predicted defocus level (e.g., defocus level 4 for O3) and DI deconvolved using the DI PSF of defocus level 0 (i.e., in-focus PSF), as shown in Figs. 7(c) and 7(d), respectively. Besides, the noisy defocused images were deconvolved using the Matlab built-in function “deconvblind” (denoted as BD deconvolution hereinafter) [34] for further comparison, as

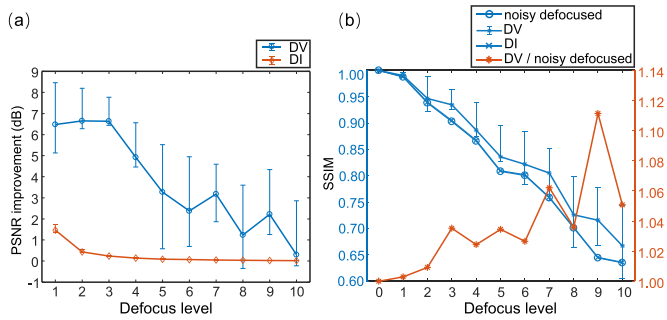


Fig. 8. (a) PSNR improvement for DV and DI deconvolved images. In the curves, the circles or diamonds represent mean values, and the vertical bars indicate the range between the 25th and 75th percentiles. (b) SSIM of the noisy defocused, DV deconvolved, and DI deconvolved images; the SSIM ratio (the orange curve and text). In the SSIM curve of the DV deconvolution, the stars represent mean values, and the vertical bars indicate the range between the 25th and 75th percentiles.

shown in Fig. 7(e). For quantitative comparison, the PSNR and SSIM [14] were plotted in Fig. 7(f) for O1–O3. In Fig. 7(g), we compared the one-dimensional (1D) profiles along with the red arrows (the whole 84 pixels) in Figs. 7(a)–7(e) for O3. Overall, the proposed LB-DVD performs better than the DI and BD deconvolution methods in terms of both visual effects (Figs. 7(c)–7(e) and 7(g)) and quantitative metrics (Fig. 7(f)) for O1–O3. In Fig. 7(f) for O3, the PSNR (SSIM) in the DV deconvolved image is improved to 63.8 dB (0.86) from about 59.0 dB (0.83) in the noisy defocused and DI deconvolved images. Also, in Fig. 7(g), the 1D profile by DV deconvolution can restore the original fine structure much better than those by DI and BD deconvolution. Besides, by using DV deconvolution, the contrast is improved compared with the noisy defocused 1D profile, and the edge is sharpened, as indicated by the arrow in Fig. 7(g). Note that in Fig. 7(d) and 7(e), BD deconvolution performs better than DI deconvolution for O1 and O2, while it is not the case for O3. This can also be observed in Fig. 7(f) for O1–O3, where BD deconvolution presents higher PSNRs and SSIMs than DI deconvolution for O1 and O2, but not for O3. This might be because the performance of the BD deconvolution method highly depends on the choice of the initial PSF used in the iteration algorithm.

To validate the efficiency of DV deconvolution on image patches, the PSNR and SSIM improvements were calculated for each of the 11 defocus levels using about 45 image patches for each defocus level for statistical purposes. The results are shown in Fig. 8(a) and 8(b) for the PSNR and SSIM, respectively. The PSNR improvement was compared with the noisy defocused images. The DV deconvolution and DI deconvolution attained PSNR improvement of 0.3–6.6 dB and 0–1.5 dB, respectively. That is, the DV deconvolution performs much better in PSNR improvement. As the defocus level grows, the PSNR improvements show a decreasing trend for both DV and DI deconvolution, which can be explained as follows. The larger the defocus level, the lower the SNR, leading to more difficulty for image recovery. The SSIM was calculated by choosing the original in-focus image patches as reference. The SSIM of the DV deconvolution ranged from 0.66 to 1.0, which is higher than that of the noisy defocused and DI deconvolution, especially for larger defocus level. Interestingly, the SSIM curves of the noisy defocused and the DI deconvolution are almost the same (Fig. 8(b)), which means DI deconvolution contributes little to restoring out-of-focus images. The SSIM of all three cases decreases as the defocus level increases, which is due to reduced SNR and thus leads to more difficulty for image restoration, as also explained above in the case of PSNR improvement. Furthermore, SSIM improvement was evaluated by checking the SSIM ratio, which is defined as the SSIM of the DV deconvolved image over that of the noisy defocused one. Overall, the SSIM ratio increases with the defocus level, achieving an improvement of 0–11%.

Fig. 9 shows two realistic wide-field images from the BBBC006 dataset, which were processed by patch-wise LB-DVD with two different weight functions (i) and (ii), as illustrated in Fig. 2(d). The three rows in Fig. 9 show the realistic image, the image by LB-DVD with weight function (i),

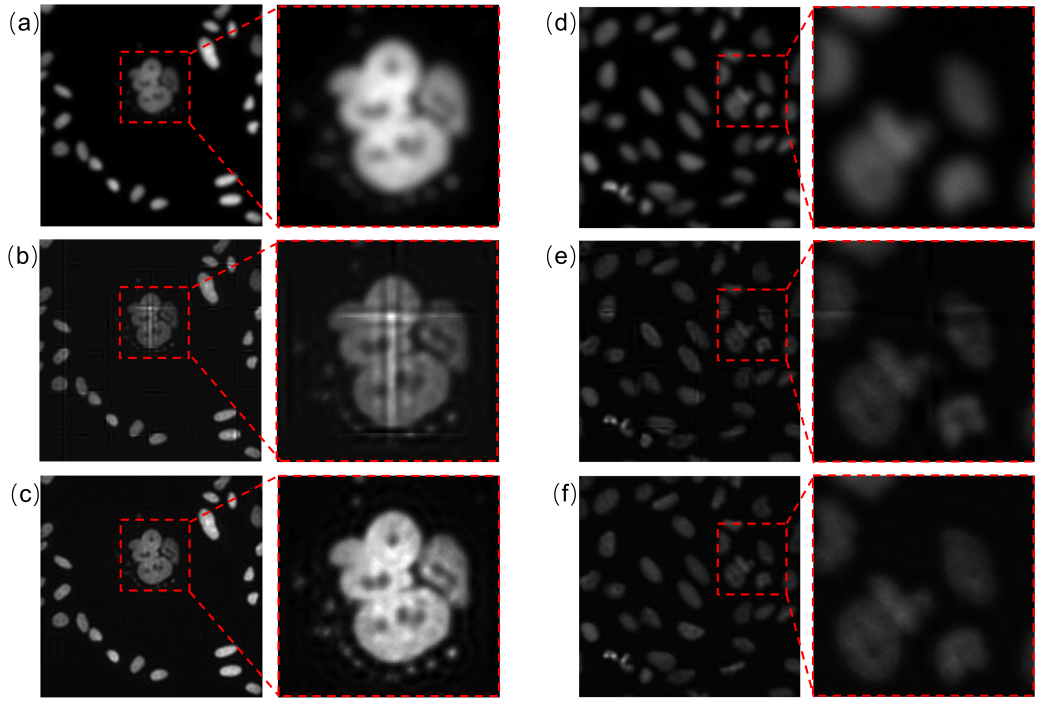


Fig. 9. Wide-field image results for (a, d) realistic image, (b, e) LB-DVD with weight function (i), and (c, f) LB-DVD with weight function (ii). The zoom-in images corresponding to the dashed boxes are also displayed.

and that by LB-DVD with weight function (ii), respectively. As can be seen in Fig. 9(b) and 9(e), non-overlapping patch-wise LB-DVD suffers from severe patch boundary artifacts. In contrast, in Fig. 9(c) and 9(f), patch boundary artifacts are almost disappeared. The zoom-in images indicated by the dashed boxes were also plotted for close comparison. Note that an alternative method using PSF interpolation would also be useful to remove patch boundary artifacts [35].

4. Discussion

Compared with [9], although part of our work (specifically, DelpNet) and [9] are similar, both aiming at predicting defocus level, our work has two further advances. Firstly, as for DelpNet, we newly adopted the special strategies (e.g., splitting "bg" label; densely using Batch Normalization layers and Dropout layers) to improve the performance of predicting defocus level. The strategies may be useful for dealing with similar sparse images in the future. Secondly, we took a further step rather than just investigating DelpNet for estimating DV-PSF. In our work, the DV deconvolution using the estimated DV-PSF was conducted to restore out-of-focus microscopy images.

As mentioned previously, a learning-based SV deconvolution was demonstrated to remove SV image blurs within a wide-field image [12]. On the other hand, the goal of our LB-DVD is to remove DV image blurs, which were not in the scope of the work in [12]. That is, images blurred by DV-PSF cannot be restored by the method in [12]. Moreover, for image montage, our study utilizes the overlapping weighting technique, as detailed previously, which performs better in eliminating patch boundary artifacts than the simple median filtering used in [12]. Even by setting the patch stride as smaller values, e.g., 1, our weighting technique has the potential to realize pixel-wise LB-DVD to extract pixel-wise DV-PSF, like some biomedical segmentation methods [36]–[38]. This is another interesting future work.

Compared with model-based methods for DV deconvolution [15], [16], one advantage of the proposed LB-DVD may lie in simplicity. The former requires complex mathematical modeling, while

the latter adopts common deconvolution once the DV-PSF can be accurately predicted by the DelpNet, which relies more on the collection of a sufficient dataset. Although it requires high computational power for training the DelpNet, once the training is done, the prediction process can be speedy. In our case, the defocus level prediction of 500 image patches (size 84×84) is within a second using a cheap Nvidia 1060 GPU.

In this work, for simplicity, the DV-PSF model in Eq. (2) was used to obtain the synthetic defocus image dataset. The model was also used in another work [9]. An improved model of the DV-PSF for producing synthetic datasets may be tested in the future [39]. Besides, in this work, without loss of generality, the DV-PSF with fixed values of parameters such as the wavelength λ, the numerical aperture *NA*, and the refractive index n were used. For practical applications, one needs to use proper values of parameters depending on the specific microscope (e.g., $n = 1.33$ for a water immersion objective).

There are some limitations of the used dataset [17]. In this study, we took the image with the maximum standard deviation among a realistic image stack as the in-focus image, which was then used to generate synthetic defocused images. However, there are still some out-of-focus cells in the reference in-focus image in part because of the wide field. As a result, the performance of DelpNet for estimating DV-PSF would be impaired. If more accurate reference in-focus image (i.e., the patterns in focus over the whole image) can be obtained (e.g., more accurate dataset available), the performance could be further enhanced.

As can be seen in Fig. 8, images with relatively large defocus levels can hardly be recovered to high-quality ones (i.e., low PSNR improvement and low SSIM), which shows the limit of the proposed DV deconvolution (based on RL deconvolution). Therefore, it seems not to make much sense to account for the samples that are beyond the largest defocus level (i.e., defocus level 10 in this study). On the other hand, to account for the samples beyond the largest defocus level for selected applications, one may consider the implementation of the DelpNet using more defocus levels or a larger increment for defocus levels. Alternatively, the samples beyond the largest defocus level can be classified as the largest defocus level, which can also contribute to image quality improvement to some extent.

5. Conclusion

In this paper, we proposed a semi-blind LB-DVD to handle out-of-focus fluorescence microscopy images. To our knowledge, this is the first attempt to deal with DV image blurs using a learning-based approach. DelpNet was developed to estimate DV-PSF, whose performance was optimized by exploring several special strategies such as splitting "bg" label, etc. The strategies used in this work to avoid overfitting may also be useful for processing the images with similar sparse and homogeneous characteristics. Further, our study suggested that plain CNN is more suitable for a specific dataset. The DelpNet achieved high accuracy of 98.2%. Then, DV deconvolution based on the estimated DV-PSF achieved maximum PSNR improvement of 6.6 dB and SSIM improvement of 11%, which performed better than DI deconvolution overall. Overlapping patch-wise LB-DVD applied to wide-field images with good image quality was also demonstrated. The results show that our LB-DVD is promising in restoring out-of-focus fluorescence microscopy images.

References

[1] Z. Liu, L. D. Lavis, and E. Betzig, "Imaging live-cell dynamics and structure at the single-molecule level," *Mol. Cell*, vol. 58, no. 4, pp. 644–659, 2015.
[2] J. M. Murray, "Methods for imaging thick specimens: Confocal microscopy, deconvolution, and structured illumination," *Cold Spring Harbor Protocols*, vol. 2011, no. 12, pp. 1399–1437, 2011.
[3] J. W. Shaevitz and D. A. Fletcher, "Enhanced three-dimensional deconvolution microscopy using a measured depth-varying point-spread function," *J. Opt. Soc. Amer. A*, vol. 24, no. 9, pp. 2622–2627, 2007.
[4] J. Arines, "Partially compensated deconvolution from wavefront sensing images of the eye fundus," *Opt. Commun.*, vol. 284, no. 6, pp. 1548–1552, 2011.

[5] B. Kim and T. Naemura, "Blind depth-variant deconvolution of 3d data in wide-field fluorescence microscopy," *Scientific Rep.*, vol. 5, 2015, Art. no. 9894.
[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.
[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 779–788.
[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 2961–2969.
[9] S. J. Yang *et al.*, "Assessing microscope image focus quality with deep learning," *BMC Bioinf.*, vol. 19, no. 1, 2018, Art. no. 77.
[10] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1790–1798.
[11] X. Fei, J. Zhao, H. Zhao, D. Yun, and Y. Zhang, "Deblurring adaptive optics retinal images using deep convolutional neural networks," *Biomed. Opt. Express*, vol. 8, no. 12, pp. 5675–5687, 2017.
[12] A. Shajkofci and M. Liebling, "Semi-blind spatially-variant deconvolution in optical microscopy with local point spread function estimation by use of convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 3818–3822.
[13] R. Morin, S. Bidon, A. Basarab, and D. Kouamé, "Semi-blind deconvolution for resolution enhancement in ultrasound imaging," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 1413–1417.
[14] 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, Apr. 2004.
[15] C. Preza and J.-A. Conchello, "Depth-variant maximum-likelihood restoration for three-dimensional fluorescence microscopy," *J. Opt. Soc. Amer. A*, vol. 21, no. 9, pp. 1593–1601, 2004.
[16] S. Ghosh and C. Preza, "Three-dimensional block-based restoration integrated with wide-field fluorescence microscopy for the investigation of thick specimens with spatially variant refractive index," *J. Biomed. Opt.*, vol. 21, no. 4, pp. 1–14, 2016.
[17] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation." *Nature Methods*, vol. 9, no. 7, pp. 637–637, 2012.
[18] M. Born and E. Wolf, "Elements of the theory of diffraction," in *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge, U.K.: Cambridge Univ. Press, 1999, pp. 412–512.
[19] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.
[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
[21] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 2818–2826.
[23] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Des. Implementation*, 2016, pp. 265–283.
[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[25] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, 1974, Art. no. 745.
[26] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, 1972.
[27] L. Yan, H. Fang, and S. Zhong, "Blind image deconvolution with spatially adaptive total variation regularization," *Opt. Lett.*, vol. 37, no. 14, pp. 2778–2780, 2012.
[28] L. Yan, H. Liu, S. Zhong, and H. Fang, "Semi-blind spectral deconvolution with adaptive tikhonov regularization," *Appl. Spectroscopy*, vol. 66, no. 11, pp. 1334–1346, 2012.
[29] H.-C. Tsai and J.-L. Wu, "An improved adaptive deconvolution algorithm for single image deblurring," *Math. Problems Eng.*, vol. 2014, pp. 1–11, 2014.
[30] É. Thiébaut, L. Denis, F. Soulez, and R. Mourya, "Spatially variant PSF modeling and image deblurring," in *Proc. Adaptive Opt. Syst. V*, vol. 9909, 2016, Art. no. 99097N.
[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 248–255.
[34] D. S. C. Biggs and M. Andrews, "Acceleration of iterative image restoration algorithms," *Appl. Opt.*, vol. 36, no. 8, pp. 1766–1775, 1997.
[35] E. Maalouf, B. Colicchio, and A. Dieterlen, "Fluorescence microscopy three-dimensional depth variant point spread function interpolation using zernike moments," *J. Opt. Soc. Amer. A*, vol. 28, no. 9, pp. 1864–1870, Sep. 2011.
[36] A. Camino *et al.*, "Deep learning for the segmentation of preserved photoreceptors on en face optical coherence tomography in two inherited retinal diseases," *Biomed. Opt. Express*, vol. 9, no. 7, pp. 3092–3105, 2018.
[37] S. K. Devalla *et al.*, "A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head," *Invest. Ophthalmol. Vis. Sci.*, vol. 59, no. 1, pp. 63–74, 2018.
[38] X. Liu *et al.*, "Automated layer segmentation of retinal optical coherence tomography images using a deep feature enhanced structured random forests classifier," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1404–1416, 2019.
[39] S. F. Gibson and F. Lanni, "Diffraction by a circular aperture as a model for three-dimensional optical microscopy," *J. Opt. Soc. Amer. A*, vol. 6, no. 9, pp. 1357–1367, 1989.