

Accelerated Phase Shifting for Structured Illumination Microscopy Based on Deep Learning

Xu Chen^{ID}, Bowen Li, Shaowei Jiang^{ID}, Terrance Zhang^{ID}, Xu Zhang, Peiwu Qin, Xi Yuan, Yongbing Zhang^{ID}, Guoan Zheng, and Xiangyang Ji^{ID}

Abstract—Structured illumination microscopy (SIM) enhances spatial resolution by projecting sinusoidal patterns with various orientations and lateral phase shifts. Here, we report a framework, termed DNN-SIM, powered by a deep neural network that learns the physical relationship between images with different lateral phase shifts. This approach captures one image per sinusoidal pattern orientation and infers the remaining phase-shifted images by the network, reducing the acquisition time to one-third of conventional 2D-SIM. We further extend the DNN-SIM to 3D applications and reduce the acquisition time to one-fifth of the conventional method without sacrificing the resolution. The reported DNN-SIM framework is not sample-specific, and can be used to handle new samples with features that the network has not previously encountered or learned. DNN-SIM is compatible with most existing SIM setups and reconstruction algorithms. It has the potential to address challenges associated with phototoxicity and photobleaching in super-resolution fluorescence microscopy.

Index Terms—Deep learning, generative adversarial network, image reconstruction, structured illumination microscopy, super resolution.

I. INTRODUCTION

STRUCTURED illumination microscopy (SIM) bypasses the diffraction limit of the optical microscope by probing the specimen with spatially modulated illumination [1]. Compared to other super-resolution fluorescence microscopy techniques,

Manuscript received December 30, 2020; revised March 29, 2021 and June 16, 2021; accepted June 22, 2021. Date of publication June 30, 2021; date of current version July 26, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lei Tian. This work was supported in part by the National Natural Science Foundation of China under Grants 61922048, 62031023, 61620106005, and 61827804, in part by the Shenzhen Science and Technology Project (JCYJ20200109142808034 and JS20191129110812708), and in part by Guangdong Special Support under Grant 2019TX05X187. (*Corresponding author: Yongbing Zhang*.)

Xu Chen, Bowen Li, and Xiangyang Ji are with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: chenx18@mails.tsinghua.edu.cn; libwclement@gmail.com; xyji@tsinghua.edu.cn).

Shaowei Jiang, Terrance Zhang, and Guoan Zheng are with the Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: shaowei.jiang@uconn.edu; terrance.zhang@uconn.edu; guoan.zheng@uconn.edu).

Yongbing Zhang is with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: ybzhang08@hit.edu.cn).

Xu Zhang is with the Beijing Institute of Collaborative Innovation, Beijing 100094, China (e-mail: zhangx@bici.org).

Peiwu Qin and Xi Yuan are with the Tsinghua-Berkeley Shenzhen Institute, Tsinghua Shenzhen International Graduate School, iPark Nanshan District, Shenzhen 518055, China (e-mail: pwqin@sz.tsinghua.edu.cn; Xi0218@hotmail.com).

Digital Object Identifier 10.1109/TCI.2021.3093788

SIM employs an efficient widefield detection method that does not require high illumination powers [2]–[5]. This allows the system to be adapted to a variety of dyes and fluorescent proteins. With an increasing interest in SIM, several SIM-related approaches have been reported. Novel approaches may investigate the design of illumination patterns [6]–[12], grazing incidence SIM [13], Hessian SIM [14], 3D-SIM [4], [15], [16], saturated SIM [17], and the incorporation of SIM with light-sheet excitation [18]–[22]. The use of non-uniform illumination patterns has also led to the development of different reconstruction algorithms, including traditional matrix inversion [23]–[27], Bayesian estimation [28], [29], phase retrieval [30], [31], object-pattern joint recovery [11], [12], [32], pixel reassignment [33], [34], and many more.

SIM achieves system bandwidth expansion by disentangling overlapping frequency components and placing them in the correct position in the Fourier space. The inherent operating principle of SIM results in two practical implementation challenges. First, image reconstruction artifacts are introduced by positional errors of the illumination patterns and inaccurate system calibration. Second, multiple images with different sinusoidal illumination patterns are required, leading to low temporal resolution of the acquisition process. 2D SIM implementation requires acquisition of nine images-3 sinusoidal pattern orientations, and 3 lateral phase shifts for each orientation. 3D SIM implementation requires acquisition of 15 images per z plane-3 sinusoidal pattern orientations, and 5 lateral phase shifts for each orientation. Due to the low temporal resolution of the acquisition process, current commercial SIM systems can only monitor slow-paced dynamics of live specimens. The low acquisition speed of conventional SIM also poses the issue of phototoxicity and photobleaching during live-cell observations. In order to accelerate the acquisition process, some solutions have been demonstrated for SIM with a reduced number of raw images, including bayesian estimation [29], frequency domain [35], incoherent Fourier ptychographic [36]. They utilize the redundancy of information and show that, in some conditions, the number of images can be indeed decreased to as low as four.

Deep learning has recently become an emerging image reconstruction tool in microscopy alongside with SIM techniques [37]. As a data-driven methodology, deep learning provides a new perspective for traditional image reconstruction tasks without the need of remodeling image-formation processes or estimating system parameters [38]. Due to its capability of representing hidden mappings between degenerated and high-quality images,

deep learning has been successively implemented in a variety of applications, namely resolution improvement within similar or different imaging modalities [39]–[41].

Augmentation of SIM using a deep-learning approach is a novel task. Here, we introduce a deep-neural-network-based approach, termed DNN-SIM, for super-resolution SIM applications. In the reported approach, we train a deep convolutional neural network (CNN) to construct the underlying manifold mapping between images with different lateral phase shifts under the same sinusoidal pattern orientation. Using this approach, we capture only one image under a single sinusoidal pattern orientation, and use the deep CNN to infer the other two images with two different lateral phase shifts in 2D-SIM implementations. Similarly, we can use the same deep CNN to infer four images with four different lateral phase shifts corresponding to a single sinusoidal pattern orientation in 3D-SIM implementations. Using the statistical properties of the training data, our network can effectively learn the intrinsic connections between different lateral phase shifts. We can then recover a super-resolved image of the sample based on the captured image and the inferred lateral phase-shifted images by utilizing the existing SIM reconstruction algorithms. The reported approach combines the characteristics of deep learning and the reconstruction process of SIM. It is compatible with most existing SIM setups and reconstruction algorithms. It has the potential to address phototoxicity and photobleaching challenges in super-resolution fluorescence microscopy.

This paper is organized as follows: In Section II, we describe the principle of DNN-SIM, including the design of neural network structure and objective function. In Section III, we discuss the DNN-SIM framework for 2D, 3D and live cell super resolution imaging experiments. Conclusions and future research directions are given in Section IV.

II. METHOD

A. Deep Learning for Structured Illumination Microscopy

In the general acquisition procedure of SIM, one needs to shift the illumination pattern with translation spaced by one-third of the sinusoidal period in the 2D-SIM setup (one-fifth of the sinusoidal period for the 3D-SIM setup). This operation is prone to the risk of system calibration error, and it also reduces the temporal resolution of the image acquisition process. We define the α -domain as the set of images captured using the first lateral phase shift. Consequently, the β -domain and γ -domain are two sets of images corresponding to lateral phase shifts of $2\pi/3$ and $4\pi/3$, respectively. There exists a manifold mapping between the different domains (corresponding to the image sets captured under different lateral phase shifts) that can be derived from an appropriate set of training data. We designed the DNN-SIM to provide a new approach to SIM image reconstruction via domain-transformations. DNN-SIM networks can learn the physical space mapping between the α -domain and the β - or γ -domain. Once this neural network is trained to map these spaces, a low-dimensional coupled manifold can be obtained. Thus, we only need to input the α -domain manifold, and the β -domain and γ -domain manifold will be reconstructed via the

fixed latent representation [42], [43]. We implement DNN-SIM with a deep neural network based on U-net [44] and conditional generative adversarial network (cGAN) [45]. U-net is capable of learning the multiscale features and mapping the domain spaces at a pixel-level using hierarchical cross-concatenation. cGAN is a neural network comprised of a generative network (generator) and a discriminative network (discriminator). The generator is used to learn the distribution mapping that produces the desired output distribution from the known input distribution. The discriminator is used to distinguish whether the output of the generator is a true sample or a generated one. The two networks are trained simultaneously to compete against each other. The constant competition between the generator and the discriminator networks allow the generator to produce the desired data distribution while the discriminator becomes more skilled at flagging synthetic images. The generator and the discriminator can then be conditioned on input data if one wishes to add constraints.

The schematic diagram of the proposed DNN-SIM is shown in Fig. 1. The imaging procedure contains a training stage and an inference stage. The training stage is used to derive the mapping between the input image and the two images with different lateral phase shifts. We term the input image under the first lateral phase shift as ‘P1’. The notations ‘P2’ and ‘P3’ represent the two images with $2\pi/3$ and $4\pi/3$ lateral phase shifts sharing the same sinusoidal illumination pattern orientation with ‘P1,’ respectively. Our DNN-SIM network is trained by taking P1 as input and setting the corresponding P2 and P3 as target outputs. During the training process, the output of the generator (called [P2, P3]) is compared with the target images via three loss functions: i) the ‘Spatial-domain’ loss, ii) the ‘Frequency-domain’ loss, and iii) the ‘Adversarial’ loss. Spatial domain loss refers to the difference between the reconstructed images and the target images. We use a combination of distance (L1) and the Structural similarity index (SSIM) as metrics for this loss [46]. Frequency-domain loss measures the difference between the spectrum of the reconstructed images and the spectrum of the target images using L1 distance. The ‘adversarial’ loss attempts to distinguish between the real captured image and the generator’s outputs. We use another CNN as a discriminator in Fig. 1(a). The inputs of this discriminator network are P1 (for condition), and either [P2, P3] or [P2’, P3’]. The output of this discriminator network is a probability value between 0 (synthetic image by the generator) and 1 (real captured image). We use an Adam optimizer [47] to reduce the joint loss function and update the parameters of the generator [47]. In Fig. 1, there are three orientations ($\theta_1, \theta_2, \theta_3$) of the sinusoidal illumination patterns. As a result, we trained three sub-modules for these three orientations. Once trained, the generator can be applied to a newly captured image (P1) and generate two images (P2’ and P3’) with $2\pi/3$ and $4\pi/3$ lateral phase shifts, respectively, as shown in Fig. 1(b). We capture three input images given three respective sinusoidal pattern orientations, and the network generates six synthetic images. All nine images can be fed into a regular SIM reconstruction routine to recover the final super-resolution image (refer to Results and appendix for more details).

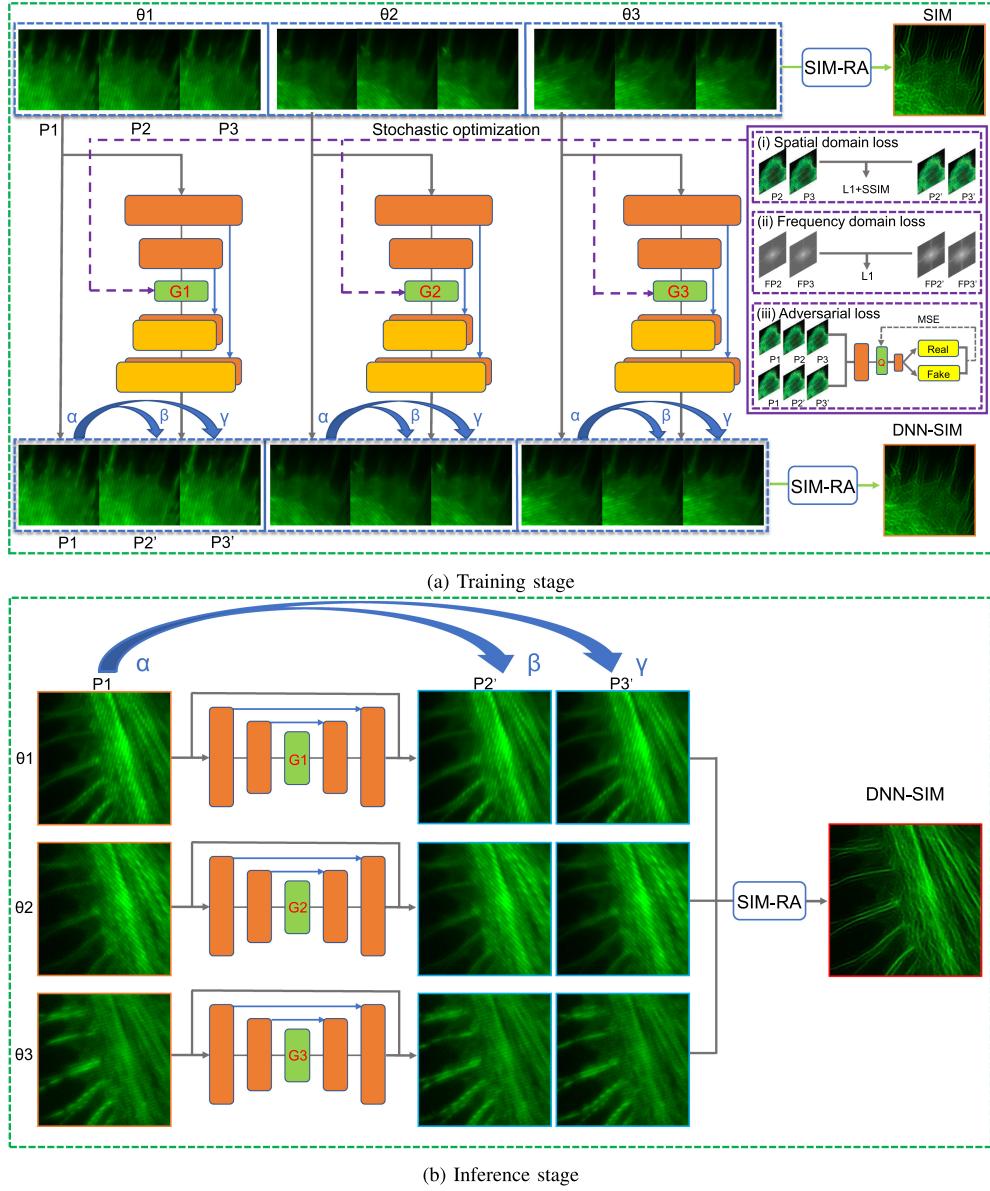


Fig. 1. Schematic of the DNN-SIM for 2D SIM imaging. (a) The training stage for the Deep-SIM. We use three sub-modules (G_1, G_2, G_3) for the three distinct illumination orientations ($\theta_1, \theta_2, \theta_3$). Each orientation has the same setting for loss functions. (i) The ‘spatial domain loss’ measures the distance between the output of the generator [P_2, P_3'] and the ground-truth [P_2, P_3] via L1 norm and SSIM. (ii) ‘The frequency domain loss’ measures the distance between the spectrum of the output of generator [FP_2, FP_3'] and the spectrum of ground-truth [FP_2, FP_3] via L1. (iii) The ‘conditional adversarial loss’ is used to improve the ability of generator to produce the desired output and deceive the discriminator. The parameters of the generator and the discriminator will be updated via stochastic gradient descent. The neural network will learn the latent mapping from the α -domain to the β -domain and the γ -domain by a domain transformation. (b) The inference stage for DNN-SIM. The α -domain data is taken as an input, and manifold mapping will directly generate the β -domain and γ -domain data. In practice, the parameters of the generator will be fixed after the training process. A new image that has not been seen by the generator is fed as the input to the generator. This will then produce reconstructed images with other lateral phase shifts. The final super-resolved image can be obtained via the SIM reconstruction algorithm (SIM-RA).

B. Neural Network Structure

DNN-SIM can learn the latent mapping between images with different lateral phase shifts. In this study, we employed a conditional generative adversarial network (cGAN) that contains a generator G and a discriminator Q . The network structure of the generator G is based on the U-net framework. The input of G is the actually captured image P_1 , which will be used to produce synthetic desired output images P_2 and P_3 . The structure of G is a symmetric encoder-decoder with skip connections, which is illustrated in the Fig. 2(a). The encoder utilizes eight

down-sampling blocks. Each down-sampling block in the encoder consists of three layers: 1) leaky rectified linear unit [48] (‘LRelu’) layer ($f(x) \rightarrow \epsilon x(x < 0) + x(x > 0)$, we set $\epsilon = 0.2$), 2) convolution layer, and 3) batch normalization [49] (‘BN’) layer. The last block has no BN layer. For an input patch sized of $(256n) \times (256n)$, the final size will be $n \times n$ by the encoder operation. The decoder has a mirror-like structure similar to the encoder. It utilizes eight up-sampling blocks. Each up-sampling block consists of three layers: 1) rectified linear unit [50] (‘Relu’) layer ($f(x) \rightarrow 0(x < 0) + x(x > 0)$), 2) deconvolution layer, and 3) BN layer. We have also added a

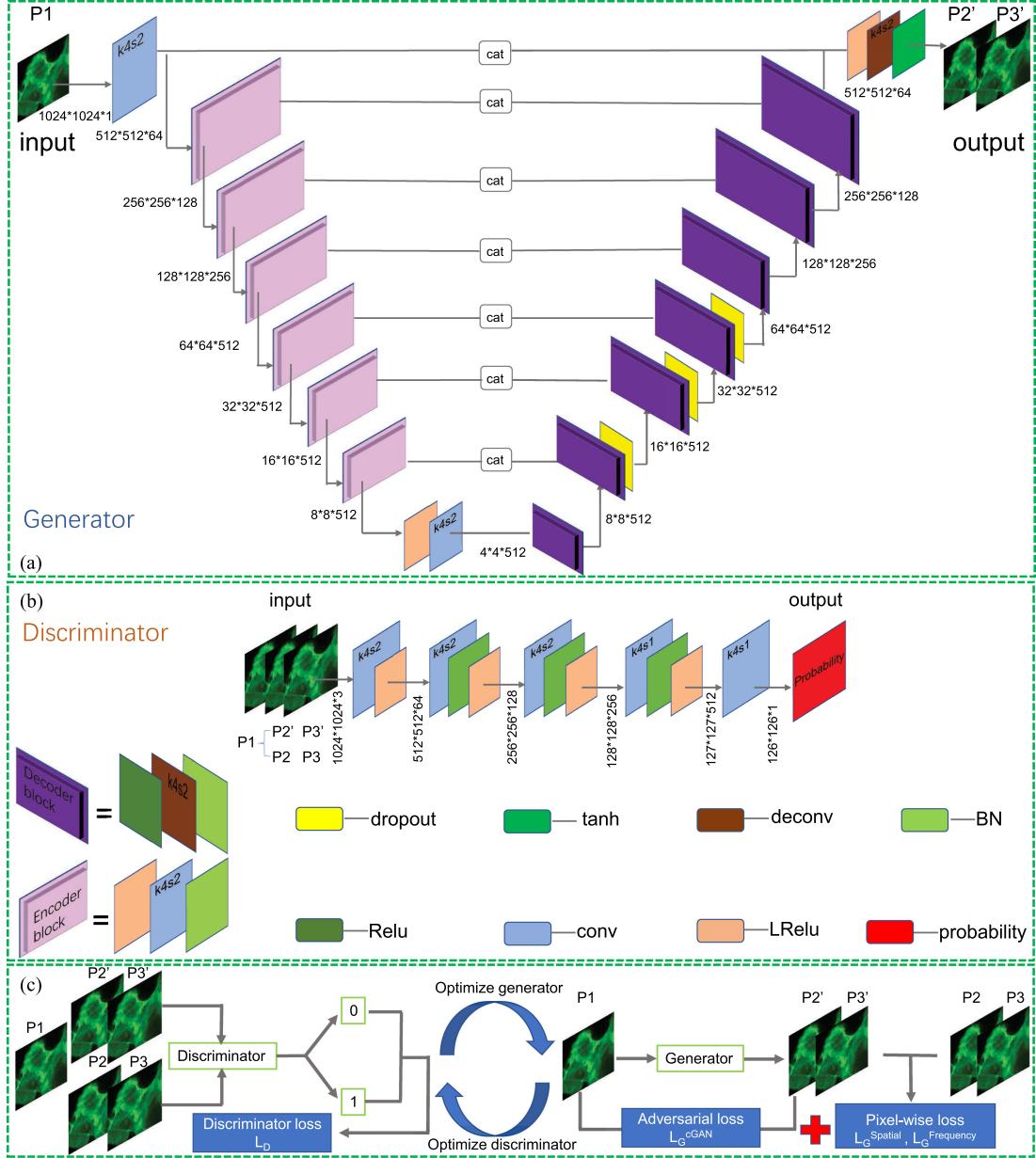


Fig. 2. The architecture of the conditional generative adversarial network (cGAN) and the alternate training process. (a) The architecture of the generator. (b) The architecture of the discriminator. Note that the k_{4s2} identification in the convolution layer concludes that the size of the convolution kernel is 4 and the size of the stride is 2. (c) The training process of our network. We first fix the generator and perform iterative optimization of the discriminator (as shown in the left half of c). We then fix the discriminator and perform iterative optimization of the generator (as shown in the right half of c). The parameters of the generator and discriminator are alternatively updated.

dropout layer [51] with a dropout probability of 0.5 to prevent overfitting in the second, third, and the fourth blocks. In the last block, we use the hyperbolic tangent activation function ('tanh') to replace the BN layer. For the input patch sized of $n \times n$, the output size becomes $(256n) \times (256n)$ by the decoder operation. Therefore, feature maps at the same depth are located symmetrically and will have the same size in both the encoder and the decoder. We concatenate the two feature maps as the input of the next decoder block. The skip connection is used to ensure that the low-level information can be shared between the input and output. There are about 54.4 million parameters in network G.

The discriminator Q competes with the generator G to distinguish whether the output of G is real or fake. Fig. 2 b shows the network structure of the discriminator Q with five blocks. The middle three blocks contain a convolution layer, a BN layer, and an LReLU layer. The first block removes the BN layer and the last block only contains a convolution layer. For an input image sized $(8n+16) \times (8n+16)$ pixels, the final output size becomes $n \times n$ via the discriminator. We note that the input of the discriminator Q also contains the input of the generator G (i.e., ' P_1 '). If the network G is fixed and only the network Q is trained, we can divide the input of discriminator Q into two categories: 1) The input contains the ground-truth image [P_2 ,

P3], the output is compared to 1 (real), 2) The input contains the output of generator [P2', P3'], the output is compared to 0 (fake). The total loss is the sum of the above two categories via the least-square loss [52]. When network Q is fixed and only network G is trained, network Q takes three images [P1,P2',P3'] as input. The output of Q is a small feature map whose value represents the probability. The larger the probability value, the stronger the capacity of G. This means that network Q believes the image synthesized by the network G is close to the ground-truth. There are about 2.8 million parameters in the network Q.

C. Training Objectives and Schedules

As shown in Fig. 1, we use three loss functions for the network training process. The first part of the loss function is spatial domain loss $L_{Spatial}(G)$ with a weighted combination of L_1 norm and structural similarity index (SSIM). This term penalizes the differences between the synthesized images and the target ones.

$$\begin{aligned} L_{Spatial}(G) &= E_{(P1, P2, P3) \sim P_{data}(P1, P2, P3)}(L_1(P, P') \\ &\quad + \lambda * SSIM(P, P')) \\ &= E_{(P1, P2, P3) \sim P_{data}(P1, P2, P3)}(L_1([P2, P3], \\ &\quad G(P1)) + \lambda * SSIM([P2, P3], G(P1))), \end{aligned} \quad (1)$$

where E denotes expectation, $P_{data}(P1, P2, P3)$ is the joint probability density of three images with different lateral phase shifts, $P = [P2, P3]$ denotes the ground-truth images, $P' = [P2', P3'] = G(P1)$ is the synthesized image(s) by the generator G , constant λ is used to weight the L_1 norm and SSIM. We use L_1 distance instead of the mean square error (also called L_2 loss) which will bring a blur phenomenon.

$$L_1(P, P') = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (|P(i, j) - P'(i, j)|), \quad (2)$$

where i and j are the indexes of an individual pixel, M and N is the size of the image. The structural similarity index is defined as:

$$SSIM(P, P') = \frac{(2\mu_p\mu_{p'} + c_1)(2\sigma_{pp'} + c_2)}{(\mu_p^2 + \mu_{p'}^2 + c_1)(\sigma_p^2 + \sigma_{p'}^2 + c_2)}, \quad (3)$$

where $\mu_p, \mu_{p'}$ are the averages of P and P' , $\sigma_p, \sigma_{p'}$ are the variances of P and P' , $\sigma_{pp'}$ is the covariance of P and P' , and c_1, c_2 are constants to avoid the instability caused by the denominator in the above equation. The value of $SSIM(P, P')$ should be between $0 \sim 1$. The larger the value, the closer the synthesized image P' is to the reference image P .

The second part of the loss function is the Fourier-domain loss $L_{Frequency}(G)$, which measures the consistency between the Fourier spectrum of the generated images (denoted as $[FP2', FP3']$) and that of the target image (denoted as

$[FP2, FP3]$).

$$\begin{aligned} L_{Frequency}(G) &= E_{(P1, P2, P3) \sim P_{data}(P1, P2, P3)}(L_1([FP2, FP3], [FP2', FP3'])) \\ &= E_{(P1, P2, P3) \sim P_{data}(P1, P2, P3)}(|fft([P2, P3]) \\ &\quad - fft([P2', P3'])|). \end{aligned} \quad (4)$$

This term is used to suppress the periodic highlights that appear in the low-frequency region and to avoid the large deviation of frequency components.

The third part of the loss function is the adversarial loss, which is derived from cGAN. In a cGAN, the generator G and discriminator D have an extra conditional variable, referred as $P1$ in our case. It provides a constraint to ensure that the output of the network is controllable. The generator G learns to map from the image sets of one lateral phase shift to the image sets of other lateral phase shifts. Discriminator D is trained to simultaneously discriminate the true sample from the generated image by G . Thus, it is called a two-player minmax game with conditional GAN loss $L_{cGAN}(G; D)$.

$$\begin{aligned} \min_G \max_D L_{cGAN}(G; D) &= E_{x \sim P_{data}(x)}[\log(D(x|y))] \\ &\quad + E_{z \sim P_z(z)}[\log(1 - D(G(z|y)))]. \end{aligned} \quad (5)$$

In the alternate training process, the optimization objective function of G and D can also be written as:

$$\begin{aligned} D^* &= \operatorname{argmax}_D(E_{x, y \sim P_{data}(x, y)}[\log(D(x, y))] \\ &\quad + E_{z \sim P_z(z)}[\log(1 - D(G(y, z)))]), \\ G^* &= \operatorname{argmin}_G(E_{z \sim P_z(z)}[\log(1 - D(y, G(y, z)))]), \end{aligned} \quad (6)$$

where $P_{data}(x)$ indicates the distribution of the true sample, $P_z(z)$ is the probability density of input data sample, and y is the conditional variable. It should be noted that sigmoid cross entropy loss function can easily reach the saturation state and will cause the problem of vanishing gradients in the training process. Though the fake samples are on the correct side of the decision boundary, but are still far from the real data distribution. In our case, in order to stabilize the training process, we replaced the logarithmic loss by the least-square loss [52], owing to the least square loss only reaches saturation at one point and relieves the problem of vanishing gradients. This allows Least-Square GAN (LSGAN) to perform more stable during the learning process. We rewrote the formula as follows:

$$\begin{aligned} EL_D(D; G) &= E_{(P1, P2, P3) \sim P_{data}(P1, P2, P3)}(D(P1, P2, P3) - 1)^2 \\ &\quad + D(P1, G(P1))^2, \\ L_{cGAN}(G; D) &= E_{P1, P2, P3 \sim P_{data}(P1, P2, P3)}(D(P1, G(P1)) - 1)^2. \end{aligned} \quad (7)$$

By combining the three items above, we get the loss function shown below:

$$D^* = \text{argmax}_D(L_D(D; G)),$$

$$G^* = \text{argmax}_G[\gamma * L_{\text{Spatial}}(G) + \eta * L_{\text{Frequency}}(G) + L_{cGAN}(G; D)]. \quad (8)$$

The weights γ , η , λ are hyperparameters, and we have $\gamma=50$, $\eta=1$, $\lambda=0.1$ in our experiment.

During the training process, we directly use the whole image as the input of network G. The image size is 1024×1024 , and the output size of network D is 126×126 , as shown in Fig. 2 b. Both the weights of generator G and discriminator D are initialized randomly from Gaussian distribution with a fixed mean (0) and a fixed standard deviation (0.02). In each iteration, we first fix the generator G and update parameters of the discriminator D. We then fix D and update the parameters of G. We used adaptive moment estimation (Adam [45]) with the hyperparameters $\beta_1=0.5$ and $\beta_2=0.999$ in the optimization process. For the first 100 epochs, we use unchanged learning rate (which is set to 0.002). In epochs 101-200, we reduced learning rate linearly to 0. We used a Geforce GTX 1080 graphical processing unit for the training process. The batch size was set to 1, and the final model was trained for ~ 40000 iterations. The training time was ~ 5 hours in our experiments. Since the stripes on each image had exact orientations (directions), we didn't perform data augmentation. The domain transformation was learned by employing only hundreds of training image pairs.

D. Fourier-Domain Loss Enables Artifact Suppression

SIM improves the spatial resolution by expanding the Fourier spectrum. Misalignment of the spectral information in the Fourier space leads to obvious reconstruction artifacts. To address this problem, we introduced Frequency domain loss to suppress artifacts in the spatial domain. In order to explain that frequency loss is critical to the training of the reported generative network, we compare the performance with and without adding the frequency loss in Fig. 3. Figures 3 d, j, m show the results with the spatial domain loss only. We can see that there are periodic highlights in spectrum and stripes in reconstructed images. Figures 3 e, k, n show the results with both the spatial domain and frequency domain loss. In these cases, no artifacts appear in the reconstructed images. This experiment validates the importance of adding frequency domain loss in the reported network.

III. RESULTS

A. DNN-SIM for 2D Super-Resolution Imaging

In the 2D mode, a typical SIM implementation captures nine images with three sinusoidal pattern orientations, and three lateral phase shifts for each orientation. We used 100 nm fluorescent beads as the test sample to demonstrate the resolution enhancement of the DNN-SIM. We used an oil-immersion objective (100×1.49 NA) in this experiment. The excitation light wavelength was 480 nm and the corresponding emission light wavelength was 522.5 nm. Fig. 4 shows images of the beads. The widefield image in Fig. 4 was obtained by summing all nine of the captured images. Fig. 4 b shows the deconvolved image using the

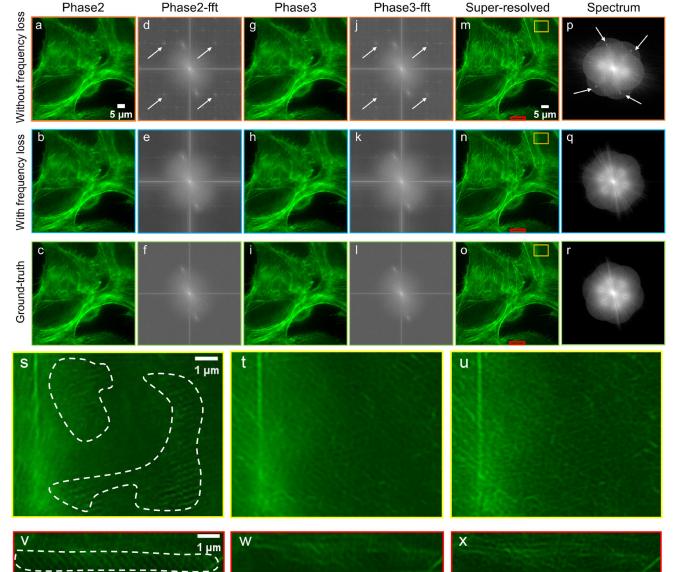


Fig. 3. Frequency loss is critical to the training of the generative network. The first row shows the results of the images inferred by generative model trained without adding frequency loss in the training process. The second row contains the results of images inferred by a generative model trained with the frequency loss. The third row shows the ground-truth images. (a-c) represent the images with $2\pi/3$ lateral phase shift. (d-f) represent the spectrum of the corresponding images a-c. (g-i) represent the images with $4\pi/3$ lateral phase shift. (j-l) represent the spectrum of corresponding images g-i. Note that the images above correspond to the third sinusoidal pattern orientation ($\theta/3$). (m-o) are the super-resolved images. (p-r) are the spectrum of super-resolved images. The white arrowheads in (d, j, p) indicate the periodic highlights in the spectrum, which carry stripe artifacts in super-resolved images. (s-u) are magnified views of the yellow boxes in (m-o). (v-x) are magnified views of red boxes in (m-o). We can see that there are periodic stripes in (s) and (v), indicated by white dotted curves. These periodic stripes do not appear in (t) and (w). Adding frequency loss in the training process is vital to suppress frequency-domain artifacts for SIM.

Lucy-Richardson algorithm. Fig. 4 c shows the super-resolution recovery using the reported DNN-SIM. The ground-truth image using 9-frame SIM is shown in Fig. 4 d. There is no visual difference between DNN-SIM and ground-truth SIM. It is difficult to resolve the closely spaced beads in both the widefield and deconvolved images. However, they are readily discernible in the DNN-SIM and the ground-truth SIM. Figures 4k-m show two beads at three distinct regions surrounded by the white dashed box. The line traces of the adjacent beads illustrate the lateral resolution enhancement of DNN-SIM method.

We also imaged the F-actin structure of MCF-7 cells in Fig. 5. Figure 5 a shows the widefield image and Fig. 5 b shows the deconvolved image. Fig. 5 c shows the image recovered by DNN-SIM, and Fig. 5 d shows the reconstruction of ground-truth SIM. We can see that the DNN-SIM method reveals details similar to that of the ground-truth SIM. Figures 5 e-h show the zoomed-in Region of Interest (ROI) of Fig. 5 a-d. Fig. 5 i, j show the overlay of the DNN-SIM reconstruction (red) and the ground-truth SIM reconstruction (green). We can see that they are in good agreement with each other. The lines traced across the filaments in Fig. 5 e-h are plotted in Fig. 5 k and Fig. 5 l. For a single filament shown in Fig. 5 k, the line corresponding to full width half maximum (dashed horizontal line) reveals that the

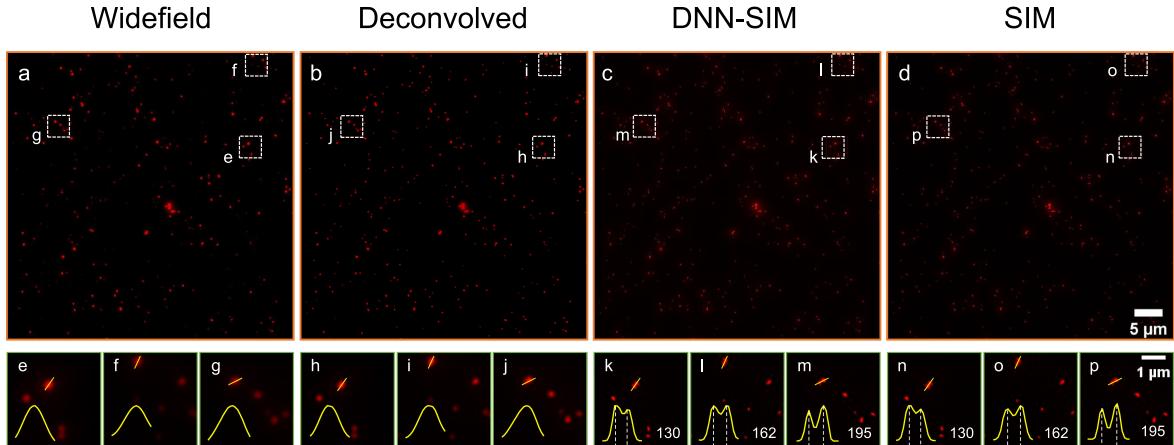


Fig. 4. DNN-SIM imaging of 100 nm fluorescent beads. (a) Widefield image. (b) Widefield image after deconvolution. (c) Reconstructed result by our DNN-SIM. (d) Reconstructed result by ground-truth SIM. The images in the second row are magnified views of white box regions in (a-d). We can easily see that some closely spaced beads cannot be resolved in widefield mode or deconvolved version. DNN-SIM is capable of resolving the individual beads (k-m), achieving a similar resolution when compared to the ground-truth (n-p). The cross-sectional profiles reported in (e-p) are shown in the lower-left corner. Peak-to-peak distance (nm) is reported in (k-p).

DNN-SIM can achieve excellent resolution enhancement compared to widefield and deconvolved images. The resolution of the DNN-SIM reconstruction is similar to that of the ground-truth SIM. Fig. 5*l* shows that two adjacent filaments (corresponding to the red lines in Fig. 5 *e-h*) can readily be resolved by DNN-SIM. In contrast, widefield and deconvolved methods can each only resolve a single peak. These results validate the performance and efficacy of the reported DNN-SIM.

We have also tested whether the reported DNN-SIM has good generalization performance when imaging objects with different image features. In this experiment, we used F-actin images of MCF-7 cells in the training stage.

We then applied the trained network to F-actin, mitochondria, and nuclei of MCF-10 A cells for the testing stage. Fig. 6 *a* shows deconvolved images of the three cell structures. The corresponding DNN-SIM and ground-truth SIM results are shown in Fig. 6 *b* and Fig. 6 *c*. From the magnified views and the line traces in Fig. 6 *d-i*, we can see that the DNN-SIM approach achieves resolution improvement similar to that of ground-truth SIM. It can resolve finer details when compared to that of the deconvolved image. Fig. 6*n-q* show the DNN-SIM recovered images of three separate channels that can be used to visualize the structures of the actincytoskeleton, mitochondria, and the nucleus. We then further overlaid the DNN-SIM reconstruction (red) and the ground-truth SIM reconstruction (green) in Fig. 8*i-l* (shown in appendix). The corresponding difference is shown in Fig. 8*m-p* (shown in appendix). Our results show that the Deep-SIM network has good generalization performance when imaging different cell features.

B. DNN-SIM for Live Cell Imaging

For live-cell imaging, prolonged exposure to light can damage cells due to phototoxicity and photobleaching. It is necessary to test the capability of DNN-SIM framework for live-cell imaging. Therefore, we imaged the mitochondria structure in Hela cell with Mito-Tracker and tested the DNN-SIM in the 2D-SIM mode. To obtain sufficient training data, we implemented two

acquisition modes. In the first acquisition mode, we captured a total of 25 images spaced 10 seconds for each sample. In the second acquisition mode, we captured a total of 41 images spaced 3 seconds for each sample. The training data consists of 7 samples (a total of 175 images) in the first acquisition mode and 33 samples (a total of 1353 images) in the second acquisition mode. The testing is conducted over two samples, which are captured in the first acquisition mode. The parameter setting is the same as the previous corresponding experiments. From Visualization 1-2, we can see that the DNN-SIM framework is also effective for live-cell imaging. The resolution has been improved compared with the deconvolved widefield images. The results of DNN-SIM are similar to those of the ground-truth SIM images, with a 3-times shorter acquisition time.

C. DNN-SIM for 3D Super-Resolution Imaging

For the 3D super-resolution, we applied 3 order SIM, where 15 images were captured with 3 sinusoidal pattern orientations and 5 lateral phase shifts per z plane for each orientation. A new training was performed for the 3D-SIM applying the schematic diagram similar to Fig. 1. The input image under the first lateral phase shift was noted as ‘P1’, and the other four images with $2\pi/5$, $4\pi/5$, $6\pi/5$ and $8\pi/5$ were noted as ‘P2’, ‘P3’, ‘P4’, and ‘P5’, respectively. Our DNN-3D-SIM network is trained by taking P1 as input and setting the corresponding P2, P3, P4 and P5 as target outputs. During the training process, the output of the generator (called [P2, P3, P4, P5]) is compared with the target images.

To test 3D-SIM imaging, we imaged F-actin structures in MCF-10 A cell stained with Alexa Fluor 488. We scanned the sample over 12 axial positions with a step size of 100 nm. For each axial plane, we only captured 3 images corresponding to the 3 respective sinusoidal pattern orientations. We then generated the remaining 12 images via the trained neural network. The whole 3D volume was $66.6 \times 66.6 \times 1.2 \mu\text{m}^3$. Fig. 7 *a* and Fig. 7 *d* show the deconvolved images at the axial positions of $z = 0.1 \mu\text{m}$ and $z = 0.9 \mu\text{m}$. The second and third columns

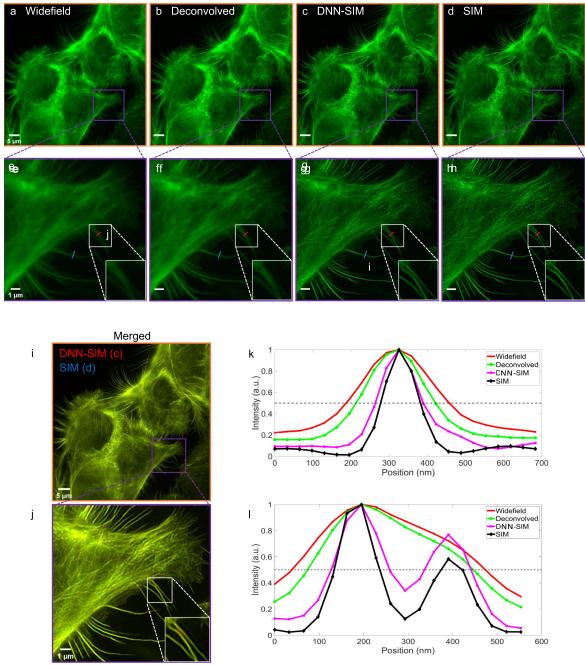


Fig. 5. DNN-SIM imaging of MCF-7 cell stained with Alexa Fluor 488-phalloidin for F-actin. In contrast to the widefield image (a) and the deconvolution result (b), the 2D reconstruction by DNN-SIM (c) resolves great details similar to the ground-truth image (d). The cross-sectional profiles reported in e-h are magnified views of the purple boxed regions in a-d. The insets in e-h show further zoomed-in views of the white boxed region with greater details. (i) Overlay image of DNN-SIM reconstruction image (c) in red, and ground-truth SIM image (d) in green. It can be seen that there are not too many red or green color blocks, which shows that there are no obvious differences between the DNN-SIM reconstruction image and the ground-truth image. (k) The line traced across the single filament labeled by blue segment in (e-h). (l) The line traced across two adjacent filaments labeled by red segment in (e-h) (indistinguishable in widefield and deconvolved image). It can be observed that the discernibility of DNN-SIM is very close to the ground-truth SIM.

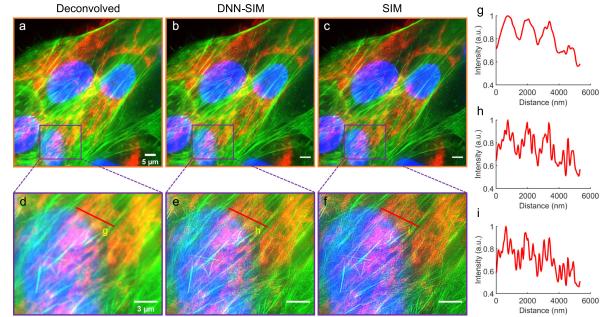


Fig. 6. DNN-SIM imaging of MCF-10 A cell stained with Alexa Fluor 488-phalloidin for F-actin (green), TOM20 conjugated with Alexa Fluor 555 for mitochondria (red), and DAPI for the nucleus (blue). Panels (a-c) show the multi-color imaging results of the deconvolved widefield, DNN-SIM, and ground-truth SIM (from left to right). Panels (d-f) are magnified views of panels (a-c). (g-i) are intensity profiles of red segments of panels (d-f). Additionally, we present the results of F-actin (panels j-l), mitochondria (panels m-o) and nuclei (panels p-r). The first, second, and third columns exhibit the results of deconvolved widefield, DNN-SIM, and ground-truth SIM imaging, respectively. Panels j-r are presented with insets that zoom in on a section of F-actin, mitochondria, and nuclei. Intensity profiles are provided in panels j-o under the magnified insets of the F-actin and mitochondria experiments.

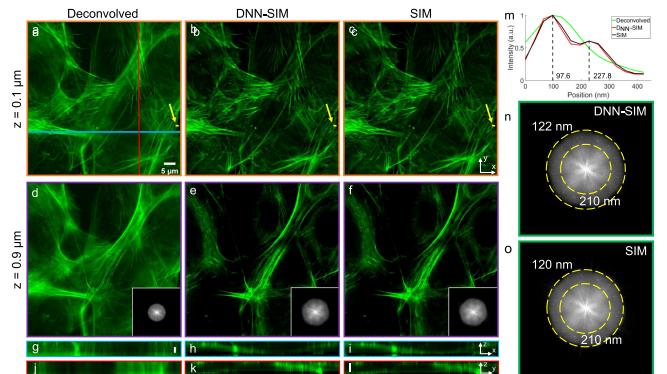


Fig. 7. 3D-SIM imaging of MCF-10 A cell stained with Alexa Fluor 488-phalloidin expressing F-actin. (a, d) Deconvolution of widefield (the left column). (b, e) Super-resolved images by Deep-SIM (the middle column). (c, f) Super-resolved images of the ground-truth 3D-SIM (the right column). The images in the first and second rows correspond to the axial positions $z=0.1 \mu\text{m}$ and $z=0.9 \mu\text{m}$, respectively. The boxed lower right images of (d, e, f) were obtained by Fourier transform. (g-i, j-l) A single x-z and y-z cross-section cut through the solid horizontal and vertical lines shown in (a). The data has been interpolated along the axial direction to match the 3D volume. (m) Line profiles across the yellow line (labeled by yellow arrow) in the right part of the first row of images (a, b, c) are plotted. Panel (n) and (o) are magnified views of the power spectrum of (e) and (f). The yellow dotted circular lines show the estimated resolution, where the inner circle is the maximum resolution that can be achieved by deconvolution. Scale bars: g-l $0.4 \mu\text{m}$.

scan of all the layers. These results validate the effectiveness of the DNN-SIM framework for 3D super-resolution imaging.

D. Structured Illumination Microscopy Reconstruction Algorithm

We used the matrix inversion algorithm [26] for 2D-SIM experiments. The parameters are consistent with our microscope system, with a Numerical Aperture (NA) of 1.49 and a pixel size of 65 nm. When imaging the nucleus, F-actin, and mitochondria channels, the fluorescence emission wavelengths were set to be 460 nm, 522.5 nm, and 605 nm, respectively. The deconvolution was performed by the Richardson-Lucy algorithm. For 3D-SIM implementation, we used the fairSIM [24], which is an open-source Fiji plugin, to reconstruct the images. All the system parameters are the same as those in 2D-SIM.

E. Sample Preparation

MCF-10 A cells were obtained from ATCC (CRL-10317) and maintained in a complete BMEM medium (Lonza, CC-3150). MCF-7 cells were obtained from ATCC (HTB-22) and maintained in DMEM medium (GIBCO) supplemented with 10% FBS (GIBCO) and 1% Pen/Strep antibiotics (GIBCO).

For imaging samples, 2×10^4 cells were seeded on $\Phi 12$ mm circular coverslips (Marienfeld 0117520) pre-treated for 1 h with 0.1% PDL (Sigma P6407) at RT, and returned to a CO₂ cell incubator overnight. Cells were then washed with PBS (GIBCO), fixed in 4% PFA (Solarbio) for 15 minutes, and permeated and blocked with 0.2% TritonX-100 (Sigma T8787) in 3% BSA (Amresco 0332) PBS buffer at RT for 30 minutes. After that, cells were incubated with 1:1000 diluted TOM20 antibody (Santa Cruz sc-11451) at 4 °C overnight. The next day, cells were washed with 0.05% PBST (0.05% Tween-20 (Sigma P7949) in PBS) and stained for 2 hours with a 1:200 dilution of Alexafluo555-labeled secondary antibody (CST 4413S) and a 1:200 dilution of Alexafluo488-labeled phalloidin (CST 8878S) at RT. After being thoroughly washed with PBS, the cells were mounted using a Mowiol mounting medium containing 2 $\mu\text{g}/\text{mL}$ DAPI and stored in a dark environment at 4 °C to avoid light exposure.

Tetraspeck fluorescent beads (ThermoFisher, T7279) were dissolved in a PBS solution at a ratio of 1:50 (v / v). We then pipetted 10 μl of this solution onto a coverglass and covered the setup with foil to prevent light contamination. The beads in the solution were incubated for 30 minutes at room temperature to allow them to adhere to the coverglass. The aqueous solution was carefully removed, and the coverglass was gently washed with ddH₂O three times. The coverglass was mounted to a slide using PBS as the mounting medium, and the coverglass was sealed using nail polish.

F. Image Acquisition

We performed all the experiments on a commercial N-SIM S super-resolution microscope, a motorized inverted microscope (Eclipse TI2-E) equipped with a 100 \times , 1.49 NA oil immersion objective (CFI SR HP Apochromat TIRF 100 \times). We used

TABLE I
THE SIZE OF THE DATASETS FOR DIFFERENT IMAGING MODE

mode	Training	Validation	Test	Total
2D-SIM	756	231	138	1285
3D-SIM	325	116	75	516
Live cell-SIM	945	312	271	1528

an ORCA-Flash 4.0 sCMOS camera (Hamamatsu Photonics K.K) to acquire images having 1024 \times 1024 pixels. The effective pixel size was 65 nm, resulting a field of view of 66.6 $\mu\text{m} \times$ 66.6 μm . In the F-actin illumination experiment, the excitation light wavelength and the emission light wavelength was 480 nm and 522.5 nm. While imaging the nucleus and mitochondria, the emission light wavelengths were 460 nm and 605 nm, respectively. For the live cell imaging experiment, the training data was captured in two shooting modes: 1. A total of 25 images spaced 10 seconds during 4 minutes per sample; 2. A total of 41 images spaced 3 seconds during 2 minutes per sample. The exposure time for image acquisition in all the experiments was 90 milliseconds. All images were saved as 16-bit grayscale images. we also count the size of the datasets for different imaging mode (shown in table I).

IV. CONCLUSION

We report a unique framework, termed DNN-SIM, powered by a deep neural network that enables rapid SIM without the need to capture all of the images corresponding to different lateral phase shifts. DNN-SIM is capable of learning the latent mapping between different domains related to images with different lateral phase shift. The reported DNN-SIM approach brings two benefits to the existing SIM system: (i) When compared to the conventional SIM methods, DNN-SIM enables significant improvements in imaging speeds without sacrificing spatial resolution of the final reconstructed image. The acquisition time can therefore be reduced to one-third or one-fifth of regular 2D SIM and 3D SIM implementations, respectively. (ii) Since only one image is captured per sinusoidal pattern orientation, the use of hardware to operate complicated, precise, and accurate phase shifting can be omitted. This significantly reduces the system complexity and the overall cost of SIM. Our results show that DNN-SIM can achieve the same resolution as that of traditional SIM imaging methods without the tedious operations that involve phase shifting. The DNN-SIM is also compatible with most existing SIM setups and reconstruction algorithms.

One significant feature of this DNN-SIM framework is that it introduces Fourier-domain loss to enable artifact suppression. For the objective function during the training stage, we defined a term to measure the consistency between the spectrums of the synthesized image and the target image. The Fourier-domain loss removes a significant number of artifacts in the reconstructed images when compared to an application without frequency loss, since Fourier-domain loss enables accurate alignment of spectral information in the Fourier space. Fig. 3 provides the comparison results with and without the Fourier-domain loss.

We demonstrated good generalization performance by utilizing input testing images that displayed features different from

input images used during the training stage. We trained the model using F-actin structures of MCF-7 cells and tested the images of F-actin, mitochondria, and nuclei of MCF-10 A cells. Experiments revealed that resolution improvements were similar to those seen in ground-truth SIM (Fig. 6). This is mainly due to the fact that DNN-SIM is capable of learning the physical relationship between images with different lateral phase shifts. This desirable property guarantees superior performance when testing multiple images with several different features.

We also carried out an experiment for live cell imaging in 2D-SIM. We tested the performance over two samples with the mitochondria structure in Hell Cells with Mito-Tracker. We can observe much finer and fast movement of cells without the occurrence of motion artefact (Visualization 1-2). This reveals the great potential application of live cell imaging by DNN-SIM framework.

In addition, our DNN-SIM framework also exhibits excellent performance when implemented into 3D imaging. For each axial plane, 3 true captured images and 12 images generated by the trained model were fed into the 3D-SIM reconstruction software. As illustrated in Fig. 7, the reported DNN-SIM was able to achieve similar lateral and axial resolution improvements when compared to the ground-truth. We also provided the volumetric views of the 3D structures in Visualization 3-4 and axial scans of all the layers in Visualization 5-6. The excellent performance reveals that DNN-SIM has the capacity to learn the latent physical mappings for 3D-SIM applications despite having significantly complicated and numerous phase shift mappings ($0 \rightarrow 2\pi/5$, $4\pi/5$, $6\pi/5$, $8\pi/5$) when compared to 2D-SIM ($0 \rightarrow 2\pi/3$, $4\pi/3$) per each sinusoidal pattern orientation. Thus, we can accurately infer the necessary phase-related images based solely on the learned network and successfully remove the out-of-focus information by coupling a deep CNN with existing 3D-SIM reconstruction methods. Recently, novel studies [53], [54] have been shown that deep neural network could fit the reconstruction algorithm. And it could directly restore super resolved sample by captured low resolution frames as network input. However, these methods cannot well generalize to samples with features different from training datasets or under additional assumption like ergodicity. Therefore, though our study utilizes two-stage pipeline that demand more workload comparing to end-to-end networks, it generalizes better and shows more robustness.

It is important to mention that while our DNN-SIM method was implemented in linear SIM systems, it can also be applied to nonlinear SIM systems [17] to reduce the number of required input images. Nonlinear SIM requires the utilization of fluorescent dyes with longer fluorescence lifetime and illumination lights with greater energy, thereby limiting its applications in the life sciences. Thus, the DNN-SIM framework should be applied in nonlinear SIM in the future to reduce phototoxicity and photobleaching associated with these fluorescent dyes and illumination lights.

This study was limited to imaging biological samples. Future applications of DNN-SIM should implement other samples to further evaluate its efficacy. Our current work is limited by the system, as all networks must be trained over the dataset

captured from a specific imaging platform. Therefore, future studies involving DNN-SIM should explore potential issues regarding the system's generalization performance under variable external physical environments and factors, such as switching the objective lens or changing the microscope imaging platform. This problem can be addressed through fine-tuning of the neural network or transfer learning to achieve optimal performance levels when collecting new data while changing the physical imaging devices. Furthermore, since the proposed method is highly ill posed, there is possible ambiguities and an artefact could be misinterpreted as a real signal. A detailed study might be needed in future to demonstrate the generalization of the network for new samples with features that the network has not encountered before. Meantime, in order to further speed up the data acquisition, we may speed up the two-stage pipeline by using lightweight network like the Mobile-Net [55] or ShuffleNet [56] and so on.

DNN-SIM is a computational framework that can be implemented in most existing SIM systems with little to no need of system adaptation. Our work provides great contribution to the development of SIM, as applications utilizing the DNN-SIM framework have delivered promising results that highlight its potential in high-speed dynamic imaging and high-throughput super-resolution imaging.

APPENDIX I THE DIFFERENCES BETWEEN THE DNN-SIM IMAGE AND THE CORRESPONDING GROUND-TRUTH SIM IMAGE

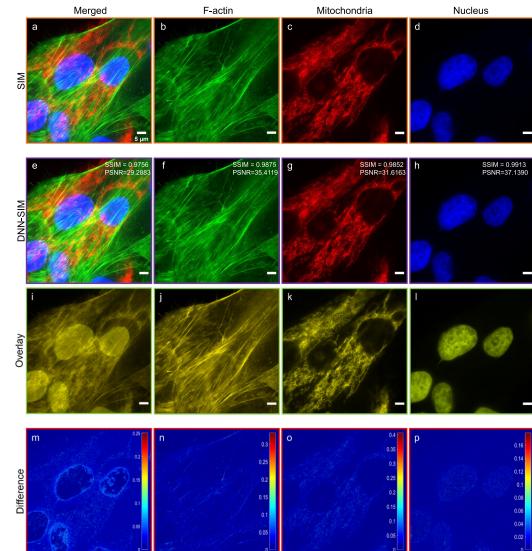


Fig. 8. The images are adapted from the manuscript (see Fig. 6 of the main text). The four rows of images correspond to the results of ground-truth SIM, DNN-SIM, overlay of ground-truth SIM and DNN-SIM images, and the error map. The first column corresponds to the merged images - images of F-actin, images of mitochondria, and images of the nucleus. The SSIM and PSNR indexes are shown in top right corner of (e-h). The results explain the similarity between the DNN-SIM and ground-truth SIM. The more results of PSNR and SSIM are shown in table II. (i-l) The merged DNN-SIM image is in red, and the ground-truth SIM image is in green. The results show a good agreement between the two methods. (m-p) Note that the difference between the two has been normalized. This also illustrates the high similarity between the DNN-SIM reconstruction and ground-truth SIM result.

APPENDIX II COLOR-CODED DEPTH OF THE 3D-SIM IMAGING OF MCF-7 CELL STAINED WITH ALEXA FLUOR 488-PHALLOIDIN EXPRESSING F-ACTIN.

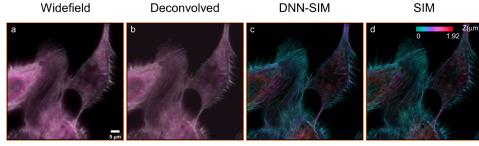


Fig. 9. (a) Widefield image. (b) Deconvolved image. (c) Super-resolved image by DNN-SIM. (d) Super-resolved image by ground-truth SIM. We can readily observe the enhancement of resolution and optical section ability of DNN-SIM when compared to Widefield and Deconvolved mode. DNN-SIM produces images that are nearly identical to ground-truth SIM images. .

APPENDIX III THE DIFFERENCE OF CAPTURED IMAGES WITH A DIFFERENT LATERAL PHASE SHIFT.

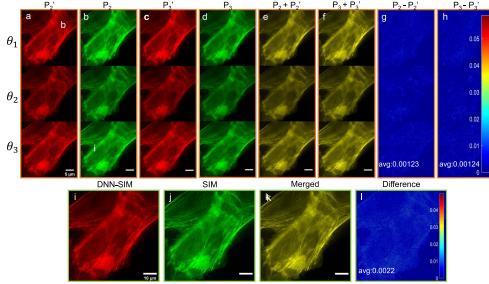


Fig. 10. The first three rows correspond to three different illumination angles θ_1 , θ_2 , θ_3 . (a, c) The output images generated by the neural network that correspond to the second and the third lateral phase shifts, respectively. (b, d) The ground-truth images corresponding to the second and the third lateral phase shifts, respectively. (e-f) Overlay of ground-truth image and inference image. (g-h) The error map of ground-truth image and inference image, which is derived by normalizing the difference image. The average error values are 0.00123 and 0.00124, respectively. (i) Super-resolved image by neural network. (j) Super-resolved ground-truth image. (k) The merged version of (i) and (j). (l) The error map between (i) and (j). It can be seen that there is high consistency between the neural-network-predicted and the ground-truth images for raw images and super-resolved images.

APPENDIX IV MAX INTENSITY PROJECTION RESULT OF 3D-SIM IMAGING OF MCF-10 A CELL STAINED WITH ALEXA FLUOR 488-PHALLOIDIN EXPRESSING F-ACTIN.

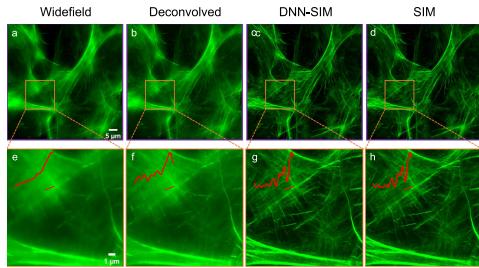


Fig. 11. (a) Widefield image. (b) Deconvolved image. (c) Super-resolved image by DNN-SIM. (d) Super-resolved image by ground-truth SIM. Panel e-h is the magnified view of the orange box region in a-d. We also plotted the line of the shown cross-section. These results show that our method can achieve superior performance when compared to conventional SIM methods.

APPENDIX V TEST DIFFERENT STRUCTURES USING MODEL PRE-TRAINED WITH ONE STRUCTURE.

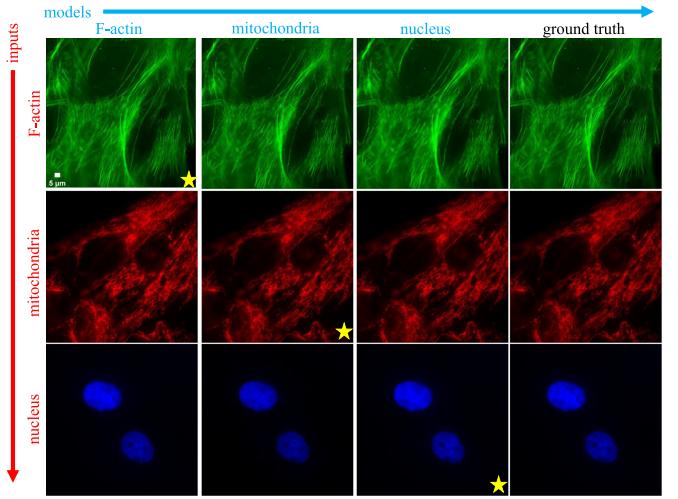


Fig. 12. Each column shows the model trained with the indicated structures. Each row shows the test datasets of different structures. And the last column shows the SIM reconstruction of the corresponding dataset. The panels marked with a yellow asterisk were tested using the model pretrained with same structure. Scale bar: 5 μ m.

APPENDIX VI ARTIFACT ANALYSIS USING THE NANOJ-SQUIRREL PLUGIN.

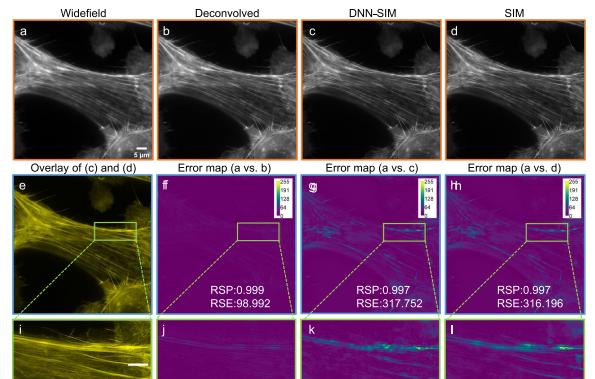


Fig. 13. (a) Widefield image. (b) Deconvolved image. (c) The result of DNN-SIM imaging. (d) The result of ground-truth SIM. (e) Overlay images of (c) in red and (d) in green. There are no obvious red or green blocks, which represents a good feature match between the two. (f-h) Error maps between the widefield and the deconvolved image (f), the widefield and the DNN-SIM (g), the widefield and the ground-truth SIM (h), calculated by NanoJ-Squirrel [57]. The reference image was the widefield result (a), and the compared image was set to the deconvolved, the DNN-SIM, and the ground-truth SIM, respectively. The value is indicated in (f-h). The results show that DNN-SIM doesn't produce noticeable super-resolution artifacts and has nearly the same error score as that of ground-truth (SIM) image with respect to the widefield image (Resolution Scaled Error $\sim=0.997$ and Resolution Scaled Pearson coefficient $\sim=317$). (i-l) Magnified view of (e-h) to further see more details. Scale bar: 3 μ m.

APPENDIX VII
EVALUATION OF THE PERFORMANCE OF DNN-SIM FOR DIFFERENT IMAGING MODE.

TABLE II
 THE RESULT OF PSNR AND SSIM FOR DIFFERENT IMAGING MODE

Mode	Index	Modality		
		Widefield	Widefield+Deconvolved	DNN-SIM
2D-SIM	PSNR	27.66±4.85	27.71±4.87	41.81±4.47
	SSIM	0.89±0.03	0.91±0.036	0.98±0.02
3D-SIM	PSNR	14.88±2.55	14.90±2.54	30.99±3.44
	SSIM	0.36±0.16	0.54±0.17	0.89±0.06
Live cell-SIM	PSNR	21.83±1.37	22.30±1.40	32.83±1.08
	SSIM	0.87±0.02	0.88±0.02	0.91±0.02

REFERENCES

- [1] M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *J. Microsc.*, vol. 198, no. Pt 2, pp. 82–87, 2010.
- [2] P. Kner, B. B. Chhun, E. R. Griffis, L. Winoto, and M. G. L. Gustafsson, "Super-resolution video microscopy of live cells by structured illumination," *Nature Methods*, vol. 6, no. 5, p. 339, 2009.
- [3] L. M. Hirvonen, K. Wicker, O. Mandula, and R. Heintzmann, "Structured illumination microscopy of a living cell," *Eur. Biophys. J.*, vol. 38, no. 6, pp. 807–812, 2009.
- [4] L. Shao, P. Kner, E. H. Rego, and M. G. L. Gustafsson, "Super-resolution 3D microscopy of live whole cells using structured illumination," *Nature Methods*, vol. 8, no. 12, pp. 1044–1046, 2011.
- [5] W. Liu, Q. Liu, Z. Zhang, Y. Han, and X. Liu, "Three-dimensional super-resolution imaging of live whole cells using galvanometer-based structured illumination microscopy," *Opt. Exp.*, vol. 27, no. 5, p. 7237, 2019.
- [6] A. G. York *et al.*, "Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy," *Nature Methods*, vol. 9 no. 7, pp. 749–754, 2012.
- [7] M. Schropp, C. Seebacher, and R. Uhl, "XI-sim: Extending superresolution into deeper layers," *Photonics*, vol. 4, no. 2, p. 33, 2017.
- [8] M. Ingaramo, A. G. York, P. Wawrzusin, O. Milberg, and G. H. Patterson, "Two-photon excitation improves multifocal structured illumination microscopy in thick scattering tissue," *Proc. Nat. Acad. Sci. USA.*, vol. 111, no. 14, pp. 5254–5259, 2014.
- [9] A. G. York *et al.*, "Instant super-resolution imaging in live cells and embryos via analog image processing," *Nature Methods*, vol. 10, no. 11, pp. 1122–1126, 2013.
- [10] M. Schropp and R. Uhl, "Two-dimensional structured illumination microscopy," *J. Microsc.*, vol. 256, no. 1, pp. 23–36, 2015.
- [11] Siyuan *et al.*, "High-resolution fluorescence imaging via pattern-illuminated fourier ptychography," *Opt. Exp.*, vol. 22, no. 17, pp. 20856–20870, 2014.
- [12] K. Guo, Z. Zhang, S. Jiang, J. Liao, and G. Zheng, "13-fold resolution gain through turbid layer via translated unknown speckle illumination," *Biomed. Opt. Exp.*, vo. 9, no. 1, p. 260, 2017.
- [13] Y. Guo *et al.*, "Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales," *Cell*, vol. 175, no. 5, pp. 1430–1442, 2018.
- [14] L. Chen, X. Huang, J. Fan, L. Li, and S. Tan, "Fast, long-term super-resolution imaging with hessian structured illumination microscopy," *Nature Biotechnol.*, vol. 36, no. 5, pp. 451–459, 2018.
- [15] M. G. L. Gustafsson *et al.*, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," *Biophys. J.*, vol. 94, no. 12, pp. 4957–4970, 2008.
- [16] L. Schermelleh *et al.*, "Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy," *Science*, vol. 320, no. 5881, pp. 1332–1336, 2008.
- [17] M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution," *Proc. Nat. Acad. Sci.*, vol. 102, no. 37, pp. 13081–13086, 2005.
- [18] T. A. Planchon *et al.*, "Rapid three-dimensional isotropic imaging of living cells using bessel beam plane illumination," *Nature Methods*, vol. 8, no. 5, pp. 417–423, 2011.
- [19] P. J. Keller *et al.*, "Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy," *Nature Methods*, vol. 7, no. 8, pp. 637–642, 2010.
- [20] L. Gao *et al.*, "Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens," *Cell*, vol. 151, no. 6, pp. 1370–1385, 2012.
- [21] B. C. Chen *et al.*, "Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution," *Science*, vol. 346, no. 6208, 2014, Art. no. 1257998.
- [22] D. Li *et al.*, "Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics," *Science*, vol. 349, no. 6251, 2015.
- [23] A. Lal, C. Shan, and P. Xi, "Structured illumination microscopy image reconstruction algorithm," *IEEE J. Sel. Topics Quantum Electron.*, vol. 22, no. 4, pp. 50–63, Jul./Aug. 2016.
- [24] M. Müller, V. Mönkemöller, S. Hennig, W. Hübner, and T. Huser, "Open-source image reconstruction of super-resolution structured illumination microscopy data in imagej," *Nature Commun.*, vol. 7, no. 1, pp. 1–6, 2016.
- [25] P. Krížek, T. Lukeš, M. Ovesný, K. Fliegel, and G. M. Hagen, "Simtoolbox: A MATLAB toolbox for structured illumination fluorescence microscopy," *Bioinformatics*, vol. 32, no. 2, pp. 318–320, 2016.
- [26] R. Cao *et al.*, "Inverse matrix based phase estimation algorithm for structured illumination microscopy," *Biomed. Opt. Exp.*, vol. 9, no. 10, pp. 5037–5051, 2018.
- [27] J. Pospíšil, T. Lukeš, J. Bendesky, K. Fliegel, K. Spendier, and G. M. Hagen, "Imaging tissues and cells beyond the diffraction limit with structured illumination microscopy and Bayesian image reconstruction," *GigaScience*, vol. 8, no. 1, 2019, Art. no. giy 126.
- [28] J. Min *et al.*, "Fluorescent microscopy beyond diffraction limits using speckle illumination and joint support recovery," *Sci. Rep.*, vol. 3, no. 1, pp. 1–6, 2013.
- [29] F. Orioux, E. Sepulveda, V. Loriette, B. Dubertret, and J. C. Olivo-Marin, "Bayesian estimation for optimized structured illumination microscopy," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 601–614, Feb. 2012.
- [30] H. Yilmaz, E. G. van Putten, J. Bertolotti, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Speckle correlation resolution enhancement of wide-field fluorescence imaging," *Optica*, vol. 2, no. 5, pp. 424–429, 2015.
- [31] O. Wagner, A. Schwarz, A. Shemer, C. Ferreira, J. García, and Z. Zalevsky, "Superresolved imaging based on wavelength multiplexing of projected unknown speckle patterns," *Appl. Opt.*, vol. 54, no. 13, pp. D 51-D60, 2015.
- [32] E. Mudry *et al.*, "Structured illumination microscopy using unknown speckle patterns," *Nature Photon.*, vol. 6, no. 5, pp. 312–315, 2012.
- [33] G. P. J. Laporte, N. Stasio, C. J. R. Sheppard, and D. Psaltis, "Resolution enhancement in nonlinear scanning microscopy through post-detection digital computation," *Optica*, vol. 1, no. 6, pp. 455–460, 2014.
- [34] I. J. Cox *et al.*, "Super-resolution by confocal fluorescent microscopy," *SPIE Milestone Series MS*, vol. 131, pp. 178–181, 1996.
- [35] L. Amit *et al.*, "A frequency domain sim reconstruction algorithm using reduced number of images," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4555–4570, Sep. 2018.
- [36] Siyuan *et al.*, "Resolution doubling with a reduced number of image acquisitions," *Biomed. Opt. Exp.*, vol. 6, no. 8, pp. 2946–2952, 2015.
- [37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [38] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, vol. 4, no. 11, pp. 1437–1443, 2017.
- [39] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnol.*, vol. 36, no. 5, pp. 460–468, 2018.
- [40] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-storm: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018.
- [41] H. Wang *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, 2019.
- [42] N. Lei *et al.*, "A geometric understanding of deep learning," *Engineering*, 2020, pp. 361–374.
- [43] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.
- [44] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
- [45] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [46] 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.

- [47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [48] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, 2013, vol. 30, p. 3.
- [49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
- [50] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
- [51] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [52] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S.P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2794–2802.
- [53] C. Ling, C. Zhang, M. Wang, F. Meng, L. Du, and X. Yuan, "Fast structured illumination microscopy via deep learning," *Photon. Res.*, vol. 8, no. 8, pp. 1350–1359, 2020.
- [54] L. Jin *et al.*, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," *Nature Commun.*, vol. 11, no. 1, pp. 1–7, 2020.
- [55] A. G. Howard *et al.*, "Mobilennets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [56] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vision Patternrecognition*, 2018, pp. 6848–6856.
- [57] S. Culley *et al.*, "Quantitative mapping and minimization of super-resolution optical imaging artifacts," *Nature Methods*, vol. 15, no. 4, p. 263, 2018.