Deep adversarial network for super-STED imaging

Mengzhou Li, Hongming Shan, Sergey Pryshchep, Maria M. Lopez and Ge Wang*[a]

Biomedical Imaging Center, Department of Biomedical Engineering, Center for Biotechnology &
Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY USA

ABSTRACT

Stimulated emission depletion (STED), as one of the emerging super-resolution (SR) techniques, defines a state-of-the-art of image resolution. It has been developed into a universal fluorescent imaging tool over the past several years. The currently best available lateral resolution offered by STED is around 20 nanometers but in real live cell imaging applications the regular resolution offered through this mechanism is around 100 nanometers, limited by photo-toxicity. Many critical biological structures are below this resolution level. Hence, it will be invaluable to improve the STED resolution through post-processing techniques. In this project, we propose a Deep Adversarial Network (DAN) for improving the STED resolution significantly, which takes a STED image as an input, relies on physical modeling to obtain training data, train the network with the simulated data, and outputs a self-refined counterpart image at a higher resolution level. In other words, we use the prior knowledge on the STED point spread function, and the structural information about the cells to generate simulated labeled data pairs for network training. Our results suggest that 30-nanometer resolution can be achieved from a 60-nanometer resolution STED image, and in our simulation and experiments, the SSIM values between the label and output result reached around 0.98, much higher than those obtained using the Lucy-Richardson deconvolution (LRD) method.

Keywords: Super-resolution, STED, deep adversarial network, image deconvolution

1. INTRODUCTION

Super-resolution optical microscopy enables quantitative inner vision of subtle structures in cells, which plays an increasingly important role in life science research.[1–3] SR imaging is also referred to as image deblurring or image restoration. With the labeling techniques fluorescent microscopy provides appealing images for studies on targeted biological structures or processes,[4–6] without the interference of irreverent background components which show up in unlabeled imaging results. STED is one of the emerging SR techniques with a state-of-the-art image resolution. It has been developed into a general fluorescent imaging tool over the last several years.[7] The best lateral resolution attained by STED is currently around 20 nanometers.[8–10] However, to avoid photo-toxicity and photo-bleaching in live cell imaging, the depletion and excitation laser beams have limited intensities, which results in a typical image resolution around 100 nanometers. Biologically, numerous critical structures are below this resolution level. Thus, it is of great and immediate significance to improve the STED image resolution for live cell based investigations. Instead of changing the scanning protocols and/or mechanisms for higher resolution,[9, 11, 12] employing deconvolution algorithms is effective and easy-to-implement for SR imaging. This approach has been widely used in many imaging applications. Many classic methods were developed and well known for this purpose, such as Wiener filtration, which is one of the most widely known linear deconvolution methods. It is fast and finds a certain trade-off between sharpness and noise. Superior microscopy image deconvolution methods are usually nonlinear, with the LRD algorithm as an excellent example for Maximum A Priori (MAP) deconvolution. It is an iterative method, which is slow but often generates better SR results than linear deblurring/restoration methods.

Recently, deep-learning based image methods have gained a tremendous attention due to their outstanding performance in solving many challenging imaging problems, such as image-SR, deconvolution, denoising, and compressive imaging. Those end-to-end data-driven methods usually do not require explicit physical models or priors that are used by the traditional model-based methods, but deep learning methods often achieve results surpassing the state-of-the-art results obtained using the traditional methods. Typically, deep learning methods rely on ”big data” which often consist of raw inputs and the corresponding desired outputs known as labels or ground truths. Related to the SR task of interest in our project, several deep-learning-based methods were proposed and demonstrated impressive performance. Given the pairs of low-resolution

---

* Corresponding author: Ge Wang. E-mail: wangg6@rpi.edu.

and high-resolution images, deep neural networks can be trained to successfully transform an image acquired with a low numerical aperture (NA) to one with a high NA,[13] or different modality images can be synergized to improve image resolution,[14] Signal to Noise Ratio (SNR) and throughput in order to push the limit of fluorescent microscopy in terms of SNR, isotropic resolution, and temporal resolution,[15] etc. The key factor to the success of those methods is the availability of big data. However, in biomedical imaging fields, the lack of high-quality datasets is very common because of privacy, cost and other issues. Often times, most imaging data come without appropriate ground-truth images. For example, when you already use the highest resolution imaging system to collect images, we run out of resources to collect higher resolution counterparts for RS imaging. For this reason, some researchers simulated data to train a network, especially in the stochastic SR microscopy filed (i.e. STORM or PALM),[16–18] which is totally a different story from the deterministic SR microscopy field (i.e. STED). The essential difference lies in that the former is a localizing task requiring a time-series of low resolution images to achieve SR, while the latter is a deblurring task processing only a single image. It should be also mentioned that the concepts of SR in the fields of computer vision (CV) and microscopy are also different. First, the input for SR imaging in CV is a down-sampled image from the original high-resolution image. As a result, the goal is to recover the original one. In contrast, the input for microscopic SR imaging is a blurry image due to the diffraction of light. Accordingly, the goal is to restore a high-resolution image of an object being imaged by removing the diffraction effect. In other words, the former is an under-sampling problem, while the latter is a deblurring problem. Second, the nature of SR imaging in CV ensures that an enormous amount of natural images can be used as the training dataset, while the number of good quality data for microscopy SR imaging is limited, as already mentioned.

Here we propose a DAN based SR method which includes a STED image simulation module and uses synthetic data to train the network to overcome the missing data problem. As compared to acquiring real data from microscopes, our method avoids the complex drift correction and image alignment procedures which are extremely critical in high-resolution imaging, especially STED. In addition, a great amount of efforts are saved with respect to collection time, sample preparation and instrumental cost. We rely on a state-of-the-art generative adversarial network (GAN) for SR and noise reduction. Comprehensive experimental results on both simulated and physically acquired STED images demonstrate that the proposed method successfully improves the resolution and effectively suppresses image noise, clearly outperforming the classic LRD method in qualitative and quantitative metrics.

The rest of paper is organized as follows. In the next section, the STED model, sample structure and DAN architecture are introduced. In the third section, The experimental design and results are reported. In the last section, relevant issues are discussed, and then the conclusion is drawn.

2. METHODOLOGY

Our methodology relies on the prior knowledge of the STED PSF (calculated from the parameters specified during the associated imaging experiment) to generate training data. To describe the whole process clearly, let us first review the STED PSF calculation. Then, we cast light on training data generation. Finally, we present our adversarial network architecture. Our goal is to achieve lateral SR for STED images, which means that the training data should be 2D image pairs. But the real STED image formation is a convolution of a 3D sample and a 3D PSF. Thus, in addition to the lateral convolution effects, neighbor slices will contribute to blurring the in-focus image slice. Indeed, it is a complex task for DNNs to learn to separate the lateral blurring from the de-focusing effects from out of focus structure. To avoid complexity, we start from simple cases in which sample structures are constrained to a transverse slice so that the influence of the defocusing effect is avoided. Then the image formation can be taken as the 2D convolution of the sample and the central slices of the 3D PSF.

2.1 STED PSF calculation

The principle of STED is the utilization of a donut-shaped depletion laser spot to suppress fluorescence at the outer part of the excitation focus. For simplicity, STED can be taken as achieved by adding the depletion laser to a confocal microscope. As described by Hell,[8] the effective PSF of STED can be depicted as

$$h_{STED}(\vec{\mathbf{r}}) = \eta(\vec{\mathbf{r}}) h_{confocal}(\vec{\mathbf{r}}), \tag{1}$$

where $\vec{\mathbf{r}}$ stands for a position vector, $h_{confocal}$ represents the PSF of the confocal microscope, and $\eta(\vec{\mathbf{r}})$ is a suppression factor, defined by

$$\eta(\vec{\mathbf{r}}) = \exp[-\ln(2) I_{STED}(\vec{\mathbf{r}})/I_S], \tag{2}$$

where $I_{STED}(\vec{\mathbf{r}})$ is the intensity distribution of the depletion light, while I_S is the effective saturation intensity (at which the probability of fluorescence emission drops to one half of its maximum). For a certain sample with specific fluorophores under illumination with certain polarization, the I_S can be regarded as part of the inherent properties of the sample. We are more interested in the intensity distribution of the depletion laser spot $I_{STED}(\vec{\mathbf{r}})$ (also can be called as the depletion PSF) and the PSF of the confocal microscope $h_{confocal}(\vec{\mathbf{r}})$.

Depletion spot intensity $I_{STED}(\vec{\mathbf{r}})$ equals the modulus squared of the spot's electric field vector $\mathbf{E} = (E_x(\vec{\mathbf{r}}), E_y(\vec{\mathbf{r}}), E_z(\vec{\mathbf{r}}))$ defined in the Cartesian coordinates x, y, z. The donut shape depletion spot is usually generated by focusing a phase modulated circularly polarized laser beam with an objective in common STED configurations. This kind of focusing process can be calculated with Debye-Wolf vector diffraction integral as[19,20]

$$\mathbf{E} = \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \int_{\theta=0}^{\alpha} \int_{\phi=0}^{2\pi} \sqrt{\cos\theta} \begin{bmatrix} \cos\theta + 1 + (\cos\theta - 1)\mathrm{e}^{\mathrm{i}2\phi} \\ \mathrm{i}(\cos\theta + 1) - \mathrm{i}(\cos\theta - 1)\mathrm{e}^{\mathrm{i}2\phi} \\ -2\sin\theta\mathrm{e}^{\mathrm{i}\phi} \end{bmatrix} \cdot T(\theta,\phi) P(\theta,\phi) \exp\left[\mathrm{i}k(\rho\sin\theta\cos(\phi - \xi) + z\cos\theta)\right] \sin\theta \,\mathrm{d}\phi \,\mathrm{d}\theta, \tag{3}$$

where $k = n2\pi/\lambda$, n and λ are the index of fraction and the wavelength respectively. $T(\theta,\phi)$ stands for the beam profile at the pupil of the objective which includes the illumination profile from the laser and the phase modulation for generation of the donut shape spot. $P(\theta,\phi)$ is used to depict the effects of aberrations. The integration is defined on a spherical cap described by the coordinatesθ,ϕ, where θ is the aperture angle while ϕ stands for the azimuthal angle. α is the upper limit of θ, and decided by the numerical aperture of the objective NA as $\alpha = \arcsin(NA/n)$. (ρ, ξ, z) are the cylindrical coordinates to depict the distribution of E_x, E_y and E_z.

To simplify the Eq. (3), if we define $E_l = (E_x - \mathrm{i}E_y)/\sqrt{2}$ and $E_r = (E_x + \mathrm{i}E_y)/\sqrt{2}$, then we obtain

$$\mathbf{E} = \begin{bmatrix} E_l \\ E_r \\ E_z \end{bmatrix} = \int_{\theta=0}^{\alpha} \int_{\phi=0}^{2\pi} \sqrt{\cos\theta} \begin{bmatrix} \sqrt{2}(\cos\theta + 1) \\ \sqrt{2}(\cos\theta - 1)\mathrm{e}^{\mathrm{i}2\phi} \\ -2\sin\theta\mathrm{e}^{\mathrm{i}\phi} \end{bmatrix} \cdot T(\theta,\phi) P(\theta,\phi) \exp\left[\mathrm{i}k(\rho\sin\theta\cos(\phi - \xi) + z\cos\theta)\right] \sin\theta \,\mathrm{d}\phi \,\mathrm{d}\theta, \tag{4}$$

where E_l and E_r are the left and right circularly polarised components.

Assume the system is aberration free and the illumination beam is a plane wave, we have $P(\theta,\phi) = 1$ and $T(\theta,\phi)$ only includes the phase modulation by the STED phase mask. Our initial goal is to achieve lateral SR of STED, thus we will use the 2D STED configurations. For our case, the helicoidal phase pattern of $T(\theta,\phi)$ for 2D STED can be described by

$$T_{2D}(\theta) = \exp(\mathrm{i}\phi). \tag{5}$$

Substituting Eq. (5) into Eq. (4), we can further simplify the equation with the identity $\int_0^{2\pi} \exp[\mathrm{i}w\cos(\beta - \gamma) + \mathrm{i}m\beta]\,\mathrm{d}\beta = \mathrm{i}^m 2\pi \mathrm{e}^{\mathrm{i}m\gamma} J_m(w)$, where J_m is the mth Bessel function of the first kind, and obtain

$$\mathbf{E} = \begin{bmatrix} E_l \\ E_r \\ E_z \end{bmatrix} = 2\pi \int_{\theta=0}^{\alpha} \sqrt{\cos\theta} \begin{bmatrix} \mathrm{i}\sqrt{2}\mathrm{e}^{\mathrm{i}\xi}(\cos\theta + 1) J_1(k\rho\sin\theta) \\ -\mathrm{i}\sqrt{2}\mathrm{e}^{\mathrm{i}3\xi}(\cos\theta - 1) J_3(k\rho\sin\theta) \\ 2\mathrm{e}^{\mathrm{i}2\xi}\sin\theta J_2(k\rho\sin\theta) \end{bmatrix} \cdot \exp(\mathrm{i}kz\cos\theta)\sin\theta\,\mathrm{d}\theta, \tag{6}$$

then the depletion spot intensity $I_{STED}(\vec{\mathbf{r}})$ can be computed as

$$I_{STED}(\rho, \xi, z) = |\mathbf{E}|^2 = |E_x|^2 + |E_y|^2 + |E_z|^2 = |E_l|^2 + |E_r|^2 + |E_z|^2. \tag{7}$$

Based on Eq. (6), Eq. (7), and the conversion formulas between Cartesian coordinates and cylindrical coordinates,

$$\begin{cases} \rho &= \sqrt{x^2 + y^2}, \\ \xi &= \arctan(y/x) \end{cases} \tag{8}$$

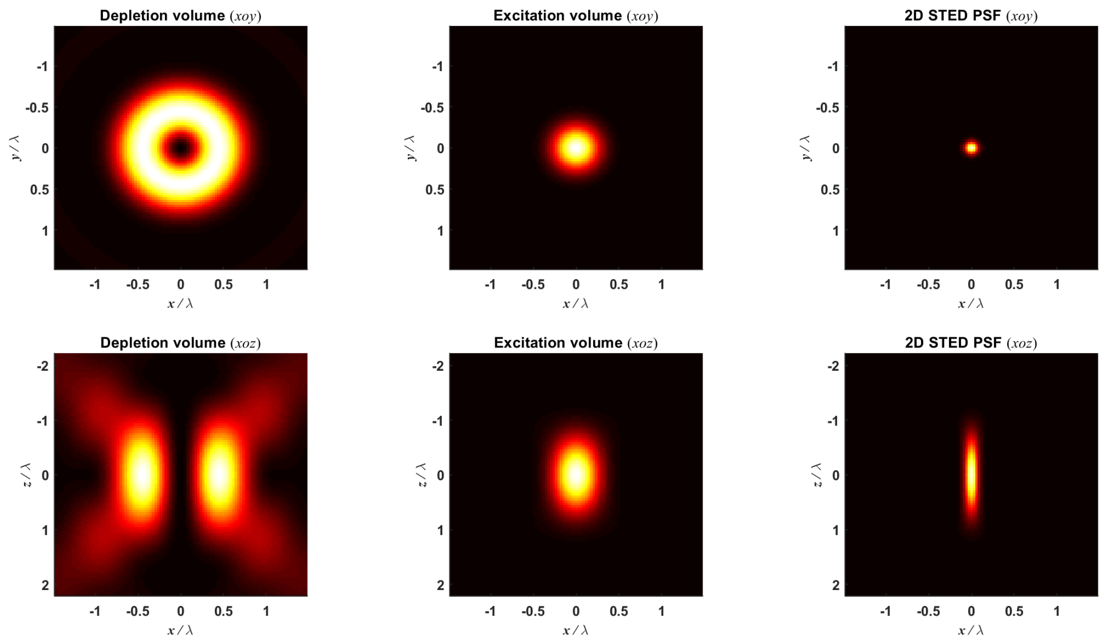


Figure 1: Depletion PSF sections. Colour maps of different images are independent. The parameters for this simulation are $NA = 1.4$, $n = 1.518$, 775 nm depletion laser, 633 nm excitation laser and emission peak around 665 nm. The coordinates are normalized with the wavelengths. The lateral resolution is set around 60 nm (half height full width)

the distribution of $I_{STED}(\vec{\mathbf{r}})$ has been numerically simulated as shown in Fig. 1. The images in upper row and lower row illustrate sections of the depletion PSF, excitation PSF and the effective STED PSF along xy plane and xz plane respectively.

Similarity, we can compute the distribution of $h_{confocal}(\vec{\mathbf{r}})$. If we remove the phase modulation term $T(\theta, \phi)$ from Eq. (3) or Eq. (4), we can obtain the focusing spot $\mathbf{E}'(\vec{\mathbf{r}})$ of a circularly polarised plane wave beam. Then the PSF of the objective $h_{lens}(\vec{\mathbf{r}})$ will be the modulus squared of the $\mathbf{E}'(\vec{\mathbf{r}})$,

$$\mathbf{E}' = \begin{bmatrix} E'_l \\ E'_r \\ E'_z \end{bmatrix} = 2\pi \int_{\theta=0}^{\alpha} \sqrt{\cos\theta} \begin{bmatrix} \sqrt{2}(\cos\theta + 1)J_0(k\rho\sin\theta) \\ -\sqrt{2}\mathrm{e}^{\mathrm{i}2\xi}(\cos\theta - 1)J_2(k\rho\sin\theta) \\ -2\mathrm{i}\mathrm{e}^{\mathrm{i}1\xi}\sin\theta J_1(k\rho\sin\theta) \end{bmatrix} \cdot \exp(\mathrm{i}kz\cos\theta)\sin\theta\,\mathrm{d}\theta, \tag{9}$$

$$h_{lens}(\rho, \xi, z) = |\mathbf{E}'|^2 = |E'_l|^2 + |E'_r|^2 + |E'_z|^2. \tag{10}$$

For a conventional confocal microscope, its PSF will be the product of the PSF of focusing lens and the PSF of the collecting lens. In our case, we use the same objective to focuse the excitation beam and to collect the fluorescence signals, thus the PSF of our confocal microscope $h_{confocal}(\vec{\mathbf{r}})$ can be depicted as

$$h_{confocal}(\vec{\mathbf{r}}) = h_{lens}(\vec{\mathbf{r}}) \cdot h_{lens}(\vec{\mathbf{r}}). \tag{11}$$

Combining Eqs. (1) to (11), we will obtain the 3D PSF of STED. As mentioned above, to simplify the image formation model, we use the average of several central slices of 3D STED PSF as a simplified 2D PSF that will be used to generate training data.

2.2 Training data generation

To generate image pairs for training a neural network, common structures of cell specimens should be modeled, which can be classified into three categories, dots, lines and planes. Small particle-like structures like nuclear pores, small vesicle and single fluorophores can be modeled as dots, while microtubules and cytoskeletal filaments are line-type structures which can be modeled by lines, and plane type structures like membrane and golgi apparatus will be modeled by ellipsoid shells or curved planes.

Based on literature, the thickness of the cell membrane is approximately 7 to 10 nm,[21] the diameters of cytoskeletal filaments (or microfilaments) range from 8 nm (actin filaments)[22] to 24 nm.[23] The nuclear pores are usually dense in nuclear envelope and the channel diameter ranges from 5.2 nm (human)[24] to 10.7 nm. In addition, the fluorescent protein size taking the GFP as an example will be around 4.2 nm (4.2 nm long and 2.4 nm in diameter).[25] The structural size of those nuclear pores and fluorophores are too small, so we will model them as dots. Most of structures are imaged with the aid of fluorophores under microscopes since most structures do not generate fluorescence themselves. In other words, what we see in fluorescent microscopes is the image of the fluorophores which form the shape of target structures when they are bound to the structures. Thus, we need to add the size of fluorophores to the size of structures to correctly model the real size of the structures we will see under the microscope with ultra-high resolution like STED.

2.2.1 Sample structure simulation

Since we limit our scope to the 2D case, we will not deal with the membrane structures which usually couple the defocused blurring and lateral blurring together and represent 3D cases. Thus, we just use lines and dots to mimic the structures in cells, especially ellipses and dots because the curved structures and dot structures are very common in cells. In addition, we add a random value (0, 1) for every pixel instead of all 1 to mimic the discrete property and intensity difference of fluorophores. As for the details, we set the pixel size of the 2D structure image slice as 10 nm, and the image size as 512 by 512. Each slice contains randomly generated structures including:

a) 3 actin filaments (straight line segments with lengths in [0, 10μm], random positions and directions in plane, line width 0.02μm);

b) 3 microtubules (straight line segments with lengths in [0, 10μm], random positions and directions in plane, line width 0.035μm);

c) 8 curved structures (ellipse with random orientation and random position in plane, radii in [0.4μm, 3μm], line width around 0.035μm);

d) 200 nuclear pores (dots with random positions in plane, the radii 0.01μm).

Finally, the average mean of the slice (all effective pixels are set 1 in value) is around 0.12. Those structure slices will serve as the ground truth images. Low resolution STED images can be generated with the structures and the simplified 2D PSF of STED, which will be introduce in details in 2.2.2.

2.2.2 STED imaging simulation

As mentioned before, we only use the average of several central slices of the 3D PSF to simply the model to 2D based on the assumption that samples are thin enough. This is a reasonable assumption when we are looking at the tiny structures on the outer layer of the sample, like nano-beads sample on a slide or nuclear pores on the nuclear membrane.

Without loss of generality, we choose the NA of objective as 1.4, the refractive index as 1.518, a depletion laser of 775 nm, an excitation laser of 633 nm, and the emission peak around 665 nm, which are the common settings used in STED. The lateral resolution is set around 60 nm (half height full width) with a certain saturation factor, and the axial resolution remains the same with that of the confocal microscopy around 1 micron. The pixel size of the PSF is set as 10 nm, resulting a 3D volume with the size 15*15*101. We select the central five slices and generate the simplified 2D PSF after averaging operation along slice dimension.

The ideal STED images are then generated through convolution between ground truth structures and the 2D PSF. Then a low pass filter is applied on the ideal STED images to generate the corresponding backgrounds which mimic the auto-fluorescence. Finally, mixed Poisson-Gaussian noise is added on the sum of background images and the ideal STED images to generate the simulated noisy STED images, which follows the following model

$$I_{noisy} = \alpha P\left(\frac{I_{ideal}+I_{bg}}{\alpha}\right) + N\left(0,\sigma^2\right) \tag{12}$$

where I_{noisy}, I_{ideal} and I_{bg} are the simulated noisy STED images, the ideal STED images and the background images; $P(\cdot)$ and $N(\cdot)$ stand for Poisson and Normal distribution respectively; α and σ control the Poisson noise level and the Gaussian noise level respectively.

The noisy STED images and their corresponding structure images are paired as low resolution input and high resolution labels for network training.

2.2.3 Deep adversarial network

The training network is adapted from the Conveying Path based Convolutional Encoder-decoder (CPCE) network,[26,27] which is under the Wasserstein Generative Adversarial Network (WGAN)[28] framework as shown in Fig. 2. The GAN[29] generally consists of two networks, a generator G and a discriminator D. The generator takes the input and tries to generate samples that meet our expectation, while the discriminator takes the role as a critic which receives both the generated samples from the generator and the real good samples which stand for the golden standards, and tries to distinguish between the two sources. It is a game relationship between the two networks. During the training, the generator improves its generation ability while the discriminator improves its discrimination ability. The final goal is that the generator generates so realistic samples that makes the discriminator fail to distinguish them from the real data.

The adopted network mainly contains three parts, a generator which uses a CPCE structure, a perceptual loss calculator utilized with a pre-trained VGG-19 network,[30] and a discriminator consisting of 6 convolutional layers. The network is trained with image patches and applied on entire images. The simulated noisy STED images are used as inputs while corresponding structure images are used as the high resolution label images during the training. During the testing procedure, the noisy low resolution images are sent to the generator and high resolution output will be generated.

3. EXPERIMENTS

3.1 Experimental setup

Dataset: A total of 4000 pairs of 512 by 512 simulated pictures are used to train the network, 100 pairs serve for validation, and another 100 pairs for testing. The images pairs for training are sliced into patches, and the patches are then randomly ordered for training. The patch size is 64 by 64. In addition, two real STED images of nano-beads and nuclear pores are acquired to evaluate the proposed method.

Protocol: The Adam optimization method[31] was used to train the network with a learning rate α 1.0×10^{-4}, two exponential decay rates β_1 and β_2 were set to 0.9 and 0.999 respectively. The trade-off parameter λ set to 10, as suggested in,[32] and λ_1 was empirically set to 0.1. The mini-batch size was 128. Figure 3 demonstrates the convergence of the model after 24 epochs.

3.2 Results

The performance of the method is first validated on the simulated testing data and benchmarked against the classic LRD method. The 2D PSF used during image generation was provided to 'deconvlucy' function, which is a build-in function in MATLAB, to implement the deconvolution, and the iteration number was set as the default value 10. The restoration results of different methods on a randomly selected testing image are shown in Fig. 4. From the magnified views of the red region, it clearly demonstrates the effectiveness of the proposed method. The result is almost the same with the ground truth, and high frequency structures are all recovered from the blurry cloud in the input despite slight difference in the intensity distribution which might be induced by the presence of noise in the input image. The LRD method also restores part of the structures and it is obvious it improves the result. But it still misses the two holes in the center and the result is not sharp and also seems noisy. For the orange region, two adjacent lines evidenced by the ground truth merge to one thick line in the noisy input, and input also contains lot of noises. The proposed method successfully distinguished the two lines but the right line wrongly appears brighter than the left one. In addition, most of noises are removed but fine structures are kept which can be evidenced by the recovered small dot near the lines which is extracted from the noisy background in the input. The LRD method failed to distinguish the two lines or keep the dot, and the result is not clean but line width is reduced. We also conducted the width projection along the line in the yellow region as shown in Fig. 4m. The result obtained from the proposed method almost overlaps with the ground truth while the LRD result is slightly wider than the ground truth but still demonstrates the SR effect.

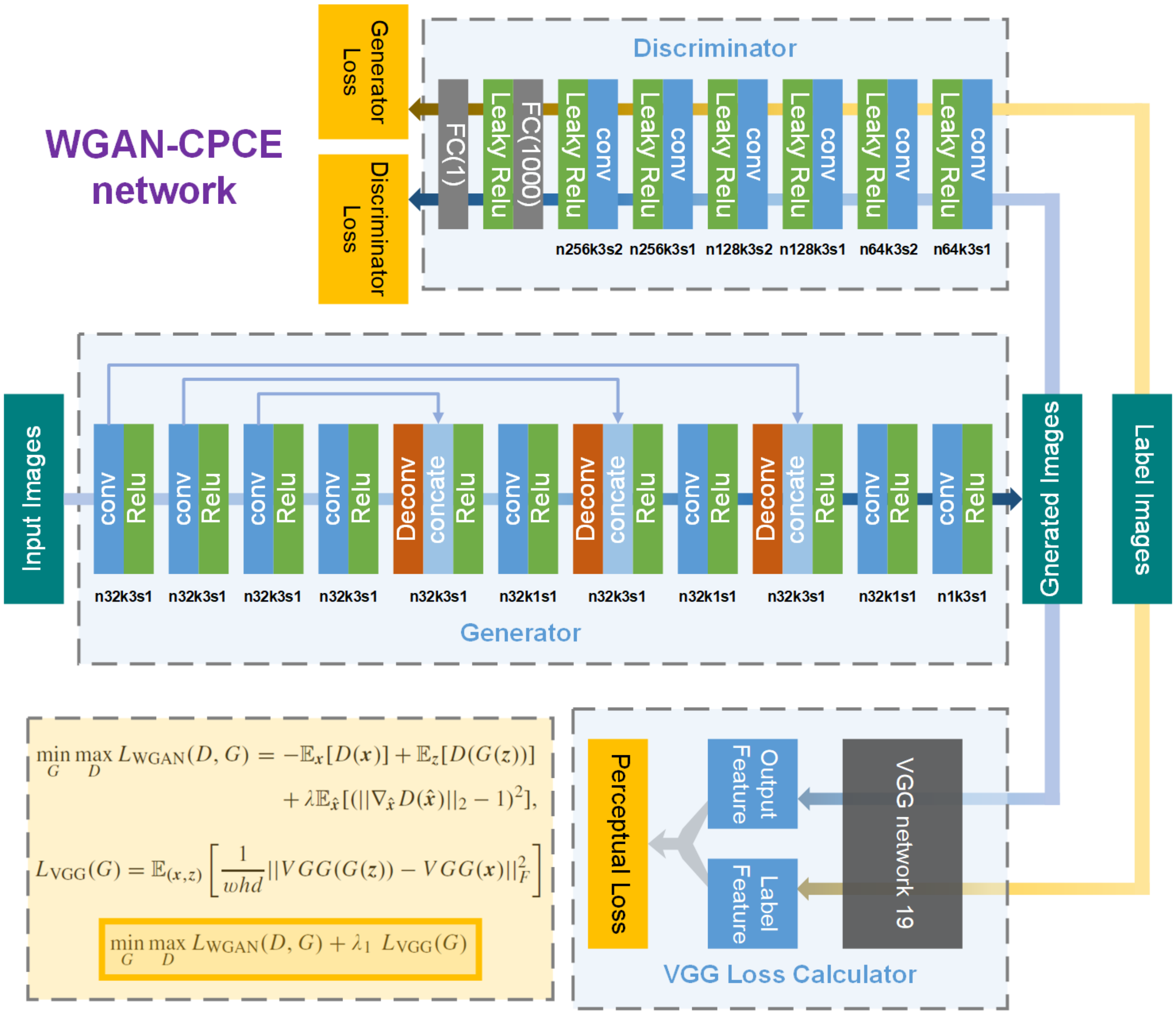


Figure 2: The adopted deep neural network structure.n, k and s stand for the convolutional kernel number, kernel size and convolutional stride respectively. z denotes a noisy STED image and x denotes the corresponding high resolution structure image. λ and λ_1 are constant weighting parameters. $\|\cdot\|_F$ denotes the Frobenius norm.

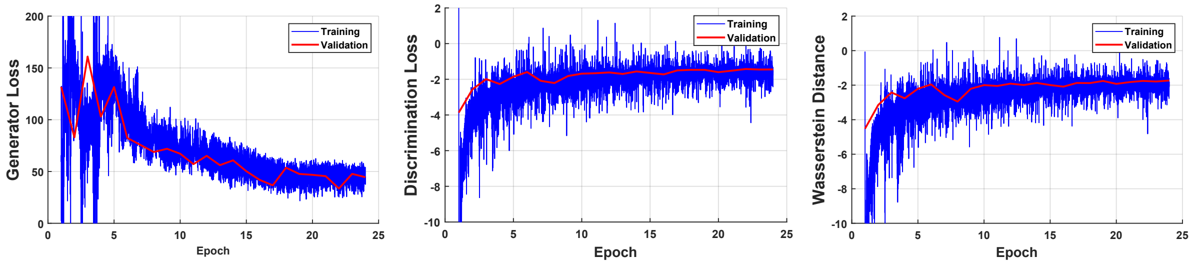
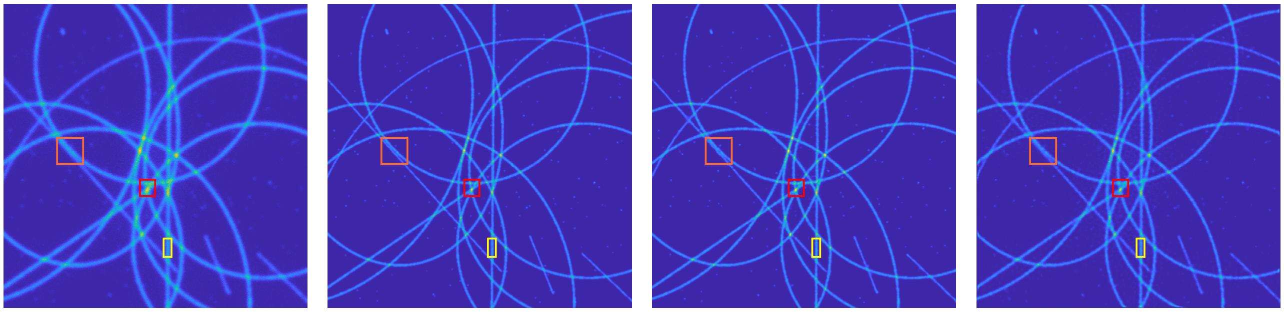
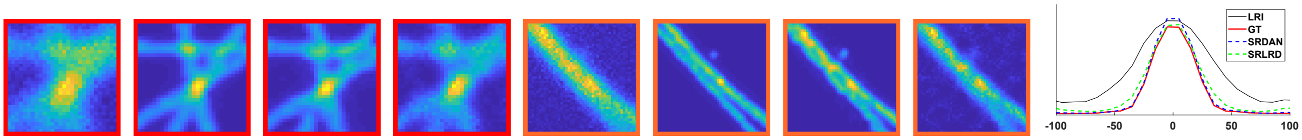


Figure 3: Plots of generator loss, discriminator loss, and Wasserstein distance versus the epoch number during training.



(a) LRI (b) GT (c) SRDAN (d) SRLRD



(e) LRI (f) GT (g) SRDAN (h) SRLRD (i) LRI (j) GT (k) SRDAN (l) SRLRD (m) Width Projection

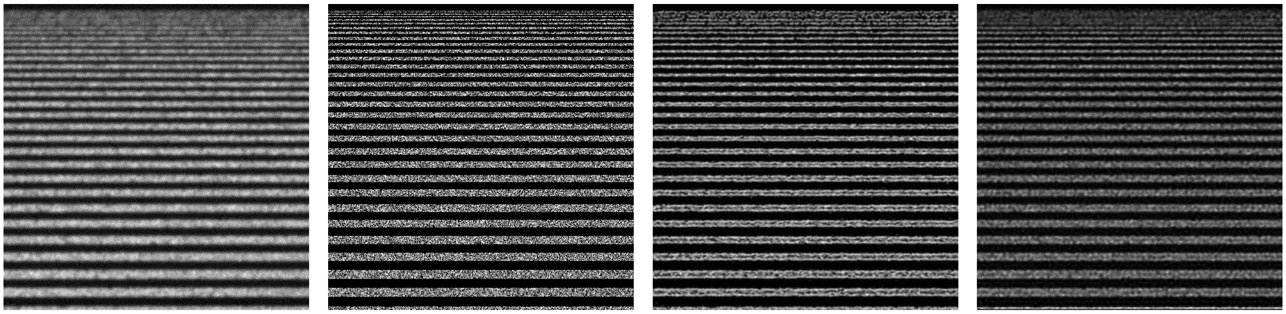
Figure 4: Testing on a simulated image. (a) Input low-resolution noisy STED image (LRI). (b)Ground truth (GT). (c) Super-resolution result with our deep adversarial network (SRDAN). (d) Super-resolution result with LRD method (SRLRD). (e) to (h) Magnified views of red region. (i) to (l) Magnified views of orange region. (m) Width projections of the yellow region.

In order to estimate the resolution improvement through the network, a set of horizontal stripes are synthesized with decreasing separations (resolution phantom), as shown in Fig. 5b. Then following the procedure to generate the noisy STED images, it was transformed into a low resolution noisy image as shown in Fig. 5a, and the noisy result was used to test the proposed method as shown in Fig. 5c. The results are summed along horizontal direction with range normalized to $[0, 1]$ to assess the resolution improvement, and lines corresponding to the top part is shown in Fig. 5e. From those restoration results in Fig. 5, we can see that the result of the proposed method is much cleaner than the LRD result and the input, but there are also tiny artifacts around the edges, and textures seem not similar to those of the ground truth. This may be attributed to the fact that the phantom structures are too sharp and none of similar images appears in the training dataset. The texture of the phantom are generated by random noise which has a resolution up to one pixel which is out of the capability of the proposed method. But after the sum operation along horizontal direction, this randomness can be kind of removed and the resulting line will reflect the SR ability of different methods.[16] In Fig. 5e, based on the contrast calculated in Tab. 1, the line of the input shows the resolution of the input image is around around 60 nm (using 50% modulation criterion), while that of the network output is around the third gap which corresponds to 30 nm. This demonstrates that the proposed method can improve the STED resolution from 60 nm to 30 nm.

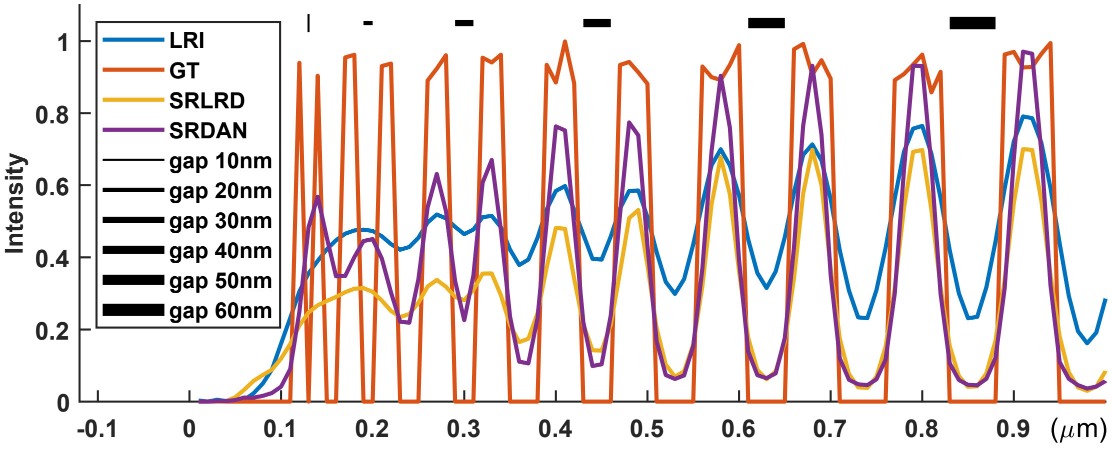
Table 1: Contrast Measurements on the Resolution Phantom

Gap width	20 nm	30 nm	40 nm	50 nm	60 nm
LRI	0.0019	0.0537	0.2006	0.3832	**0.5418**
SRDAN	0.0030	**0.4846**	0.7725	0.8680	0.9095
SRLRD	0.0083	0.1037	**0.5630**	0.8348	0.8894

Structural similarity index (SSIM) and peak signal-to-noise ratio metric have also been evaluated on randomly selected 8 images in the test dataset between the ground truth and the low-resolution input, the proposed method SR result, and the LRD method SR result, as shown in Tab. 2. The selected images are attached in the appendix. From Tab. 2 we can see that both the proposed method and the deconvolution method improve the original input significantly in perspective of SSIM and PSNR, and the proposed method performs more stable and better than the other one in evidence of bad performance of deconvolution method on sample 6 and the leading scores of the proposed method on all images. This suggests the



(a) LRI (b) GT (c) SRDAN (d) SRLRD



(e) resolution test on simulated phantom

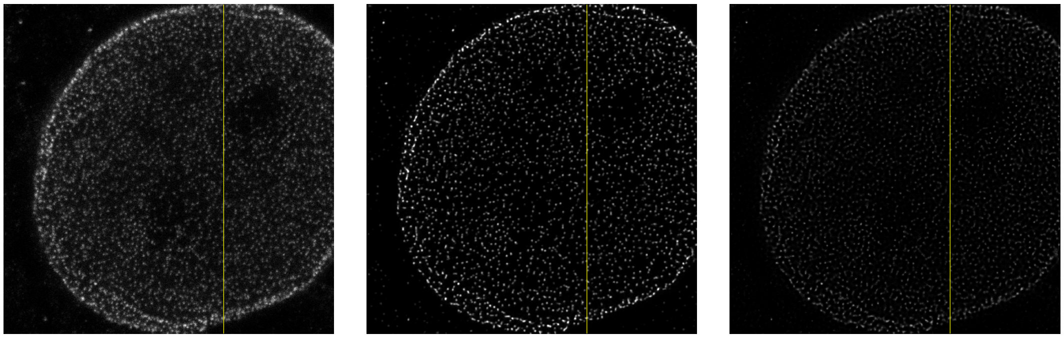
Figure 5: Testing on a simulated resolution phantom. (a) Low resolution noisy STED image of the phantom. (b)Ground truth. (c) Super-resolution result with our deep adversarial network . (d) Super-resolution result with the LRD method. (e) Sum along the horizontal direction of phantom images mentioned above.

feasibility of the idea behind the proposed method that we can meet the demand of big data with simulation supported by our professional knowledge.

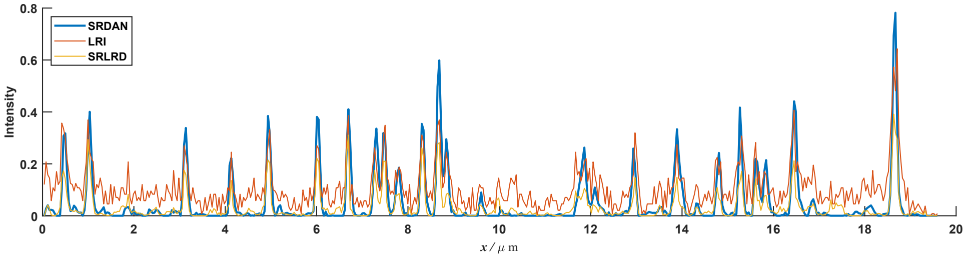
Table 2: Quantitative metrics on Testing Images

GT as ref.	sample 1	sample 2	sample 3	sample 4	sample 5	sample 6	sample 7	sample 8
LRI (SSIM)	0.77402	0.77179	0.73163	0.75592	0.70309	0.68028	0.72022	0.67642
SRDAN (SSIM)	**0.98534**	**0.97354**	**0.98380**	**0.98226**	**0.98587**	**0.97556**	**0.98637**	**0.98100**
SRLRD (SSIM)	0.94555	0.94686	0.93059	0.95499	0.93754	0.87277	0.94117	0.94424
LRI (PSNR)	24.5195	26.1955	25.3600	24.1606	23.5735	25.9906	23.9409	24.3680
SRDAN (PSNR)	**35.1704**	**35.4851**	**36.5990**	**35.1378**	**36.1270**	**36.6259**	**36.1870**	**36.4868**
SRLRD (PSNR)	30.9748	32.5516	32.0144	31.4760	31.3604	30.9036	31.3858	33.1481

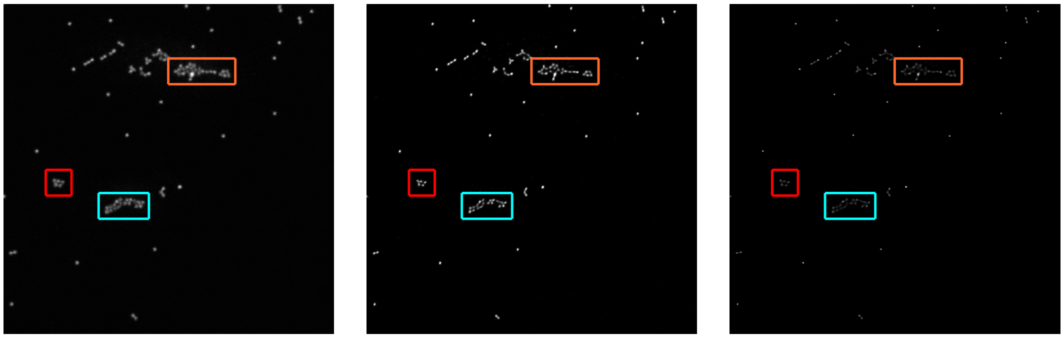
We also tested the trained network on experimental STED images of nuclear pores and beads. The parameters for STED images acquisition are similar to those used during simulation, namely, objective NA, wavelengths and resolution. The results are shown in Fig. 6. From Figs. 6a to 6c, we can see that both methods can remove most of the noises and make the images sharper. Visually, the network output is slightly brighter than the LRD result, and at the similar level to the input. Moreover, in the network output, the dots seem more regular and symmetric than those in the LRD result, and the dots in LRD result are in bar-shapes which are less appealing. Figure 6d shows that both methods remove the background noises but still keep the tiny structures. As for the nano-beads image, it can be clearly observed that the resolution of the



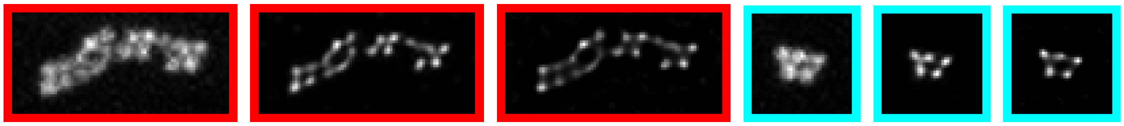
(a) Nuclear Pores (b) SRDAN (c) SRLRD



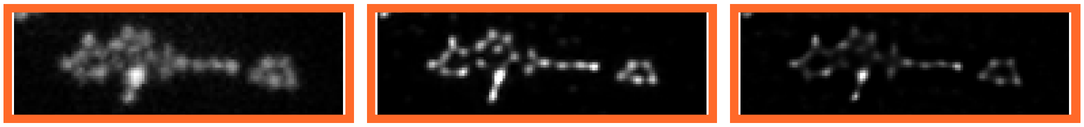
(d) Intensity Profile



(e) Beads (f) SRDAN (g) SRLRD



(h) LRI (i) SRDAN (j) SRLRD (k) LRI (l) SRDAN (m) SRLRD



(n) LRI (o) SRDAN (p) SRLRD

Figure 6: Restoration results on experimentally measured nuclear pores and nano-beads images. (a) Low resolution noisy STED nuclear pores image. (b)SRDAN of the nuclear pores image. (c) SRLRD of the nuclear pores image. (d) Line intensity profile across the images. (e) Noisy STED nano-beads image. (f) SRDAN of the beads image. (g) SRLRD of the beads image. (h) to (j) Magnified views of the red region. (k) to (m) Magnified views of the blue region. (n) to (p) Magnified views of the orange region.

original images are significantly enhanced through both methods, but the results of the proposed method are more likely to be the truth than the LRD method for the following reason. We know the beads generally should be in round shape and they should be very close to each other if they gather to a block as shown in Figs. 6h and 6n. Comparing Figs. 6i and 6o with Figs. 6j and 6p, we found that the beads in LRD results seem to be escaping from each other and have obvious gaps in between while the beads' centers in the network outputs are closer and the gaps are very small. Moreover, beads in Fig. 6l seem closer and more symmetric than in Fig. 6m.

4. DISCUSSIONS AND CONCLUSION

As mentioned earlier, this pilot study is by design a 2D SR imaging task, and has been intended to deal with thin samples only. In the next step, we would like to extend our approach to a 3d setting. Our hypothesis is that both 3D cellular structures and the system PSF are spatially correlated, and the defocusing effect can be recognized by a deep neural network. Actually, experts' eyes can perceive the defocusing effect and infer underlying structures, and hopefully a data-driven deep learning process could, to a good degree, simulate the human perception. If this is proven successful, our SR work will be much more significant than 2D SR imaging. Also, our results suggest the feasibility of using simulation data to train a deep network and achieve resolution enhancement in other fields.

In conclusion, we have presented a DAN-based method for improving the STED resolution to the next level. The proposed method overcomes the unavailability of missing paired training data with physically realistic modeling. The synthetic data generated according to the analytic PSF model has been cost-effective in training our deep neural network and delivered promising SR results. In comparison, the testing results on both simulated images and experimental images suggest that the proposed method performs much better than the classic LRD method in resolution enhancement and noise reduction, and quantitatively achieves 30 nm resolution from 60 nm-resolution STED images.

5. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (Award MRI-1725984).

APPENDIX A. MORE TESTING RESULTS ON SIMULATED IMAGES

Testing results on 8 randomly selected samples from the testing dataset are shown in Figs. A.1 and A.2, on which SSIM and PSNR are also evaluated to obtain Tab. 2.

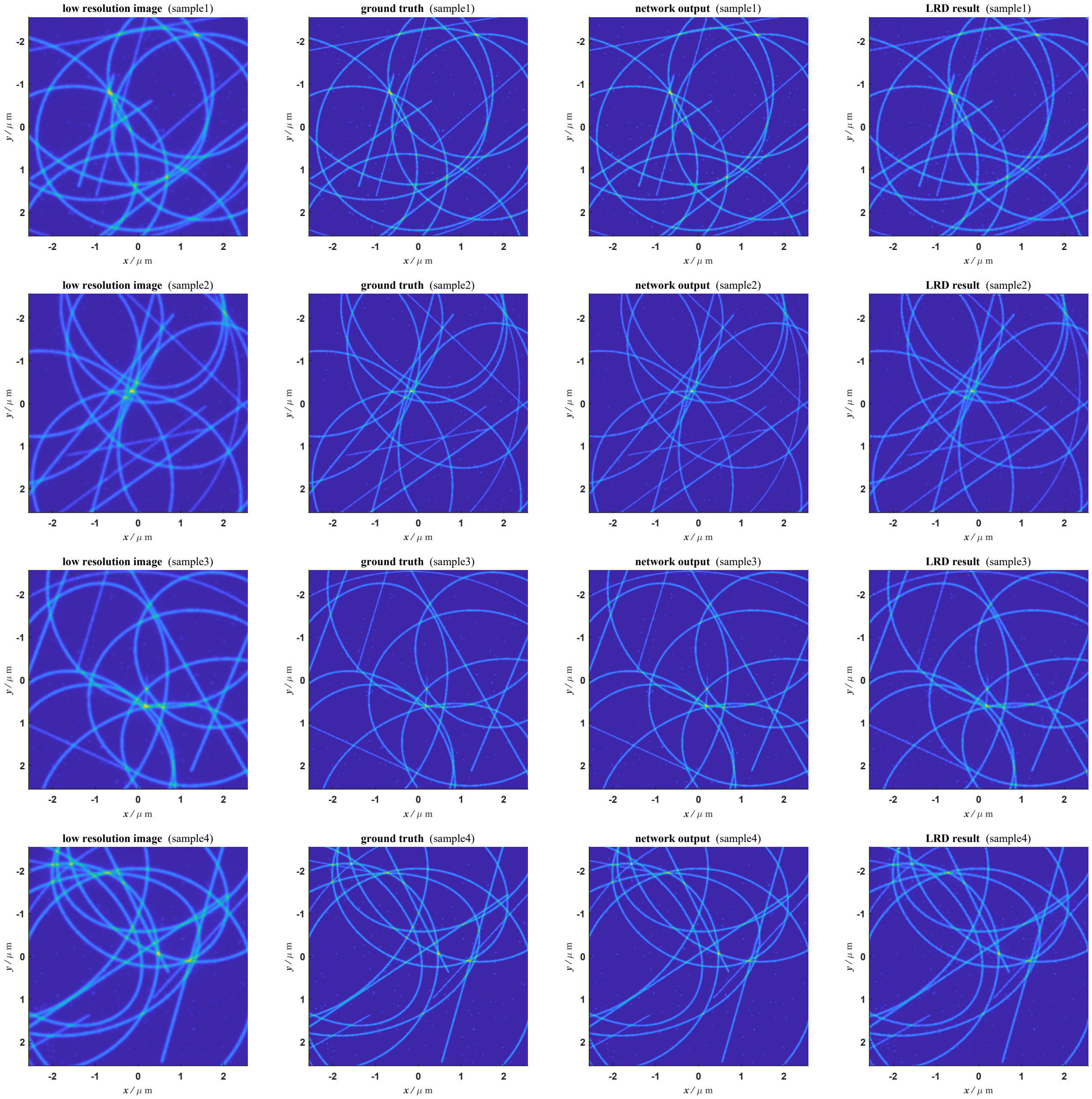


Figure A.1: Testing results on simulated sample images 1 to 4.

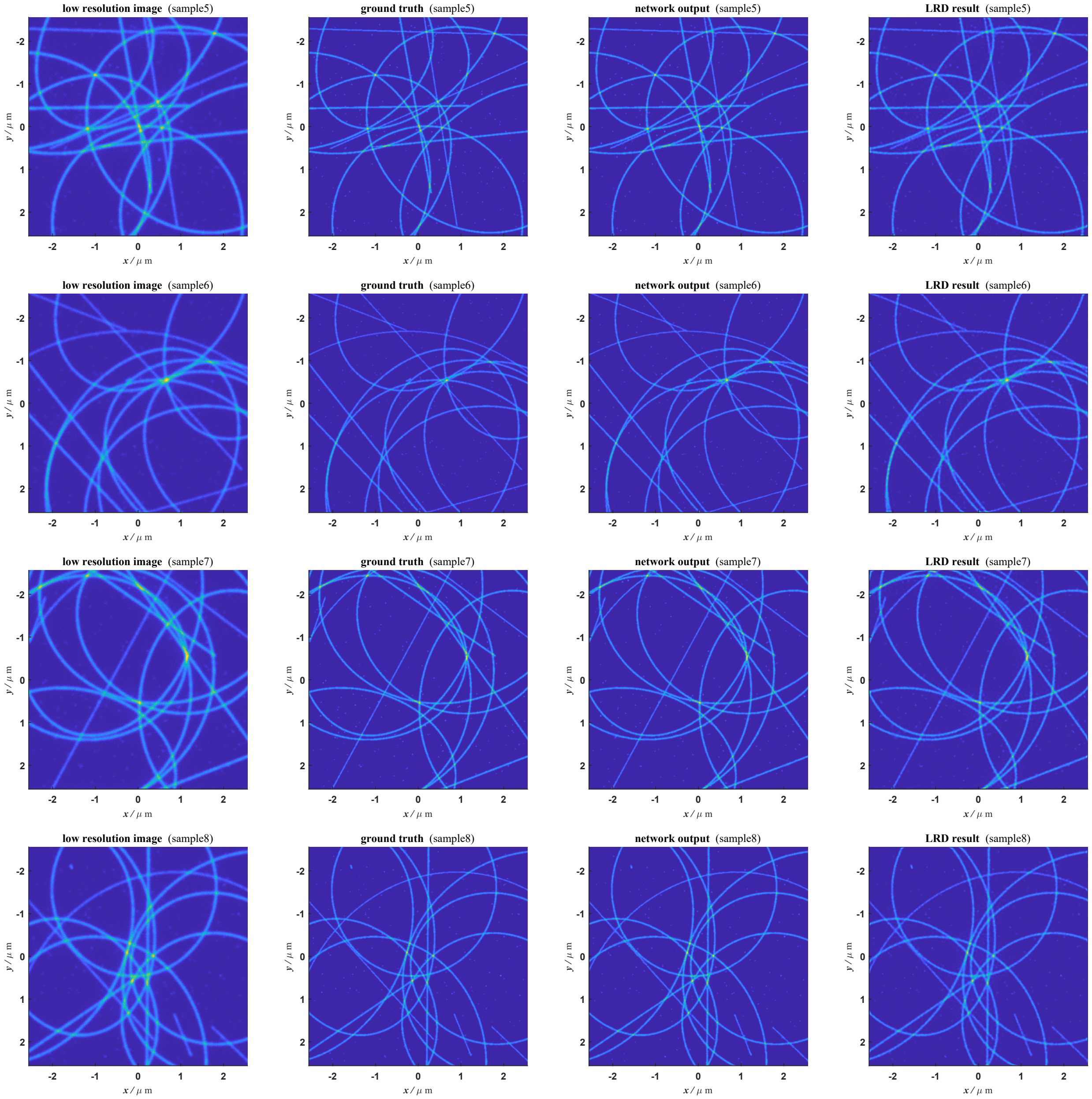


Figure A.2: Testing results on simulated sample images 5 to 8.

REFERENCES

[1] Sahl, S. J., Hell, S. W., and Jakobs, S., "Fluorescence nanoscopy in cell biology," *Nature reviews Molecular cell biology* **18**(11), 685 (2017).

[2] Phan, N. T., Li, X., and Ewing, A. G., "Measuring synaptic vesicles using cellular electrochemistry and nanoscale molecular imaging," *Nature Reviews Chemistry* **1**(6), 0048 (2017).

[3] Leterrier, C., Dubey, P., and Roy, S., "The nano-architecture of the axonal cytoskeleton," *Nature Reviews Neuroscience* **18**(12), 713 (2017).

[4] Sahoo, H., "Fluorescent labeling techniques in biomolecules: a flashback," *Rsc Advances* **2**(18), 7017–7029 (2012).

[5] Szent-Gyorgyi, C., Schmidt, B. F., Creeger, Y., Fisher, G. W., Zakel, K. L., Adler, S., Fitzpatrick, J. A., Woolford, C. A., Yan, Q., Vasilev, K. V., et al., "Fluorogen-activating single-chain antibodies for imaging cell surface proteins," *Nature biotechnology* **26**(2), 235 (2008).

[6] Weil, T. T., Parton, R. M., and Davis, I., "Making the message clear: visualizing mrna localization," *Trends in cell biology* **20**(7), 380–390 (2010).

[7] Blom, H. and Widengren, J., "Sted microscopytowards broadened use and scope of applications," *Current opinion in chemical biology* **20**, 127–133 (2014).

[8] Harke, B., Keller, J., Ullal, C. K., Westphal, V., Schönle, A., and Hell, S. W., "Resolution scaling in sted microscopy," *Optics express* **16**(6), 4154–4162 (2008).

[9] Vicidomini, G., Bianchini, P., and Diaspro, A., "Sted super-resolved microscopy," *Nature methods* **15**(3), 173 (2018).

[10] Göttfert, F., Wurm, C. A., Mueller, V., Berning, S., Cordes, V. C., Honigmann, A., and Hell, S. W., "Coaligned dual-channel sted nanoscopy and molecular diffusion analysis at 20 nm resolution," *Biophysical journal* **105**(1), L01–L03 (2013).

[11] Göttfert, F., Pleiner, T., Heine, J., Westphal, V., Görlich, D., Sahl, S. J., and Hell, S. W., "Strong signal increase in sted fluorescence microscopy by imaging regions of subdiffraction extent," *Proceedings of the National Academy of Sciences* **114**(9), 2125–2130 (2017).

[12] Oracz, J., Westphal, V., Radzewicz, C., Sahl, S. J., and Hell, S. W., "Photobleaching in sted nanoscopy and its dependence on the photon flux applied for reversible silencing of the fluorophore," *Scientific reports* **7**(1), 11354 (2017).

[13] Rivenson, Y., Göröcs, Z., Günaydin, H., Zhang, Y., Wang, H., and Ozcan, A., "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).

[14] Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., Bentolila, L. A., Kural, C., and Ozcan, A., "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods* **16**, 103–110 (2019).

[15] Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., et al., "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nature methods* **15**(12), 1090 (2018).

[16] Nehme, E., Weiss, L. E., Michaeli, T., and Shechtman, Y., "Deep-storm: super-resolution single-molecule microscopy by deep learning," *Optica* **5**(4), 458–464 (2018).

[17] Ouyang, W., Aristov, A., Lelek, M., Hao, X., and Zimmer, C., "Deep learning massively accelerates super-resolution localization microscopy," *Nature biotechnology* **36**(5), 460 (2018).

[18] Li, Y., Xu, F., Zhang, F., Xu, P., Zhang, M., Fan, M., Li, L., Gao, X., and Han, R., "Dlbi: deep learning guided bayesian inference for structure reconstruction of super-resolution fluorescence microscopy," *Bioinformatics* **34**(13), i284–i294 (2018).

[19] Richards, B. and Wolf, E., "Electromagnetic diffraction in optical systems, ii. structure of the image field in an aplanatic system," *Proc. R. Soc. Lond. A* **253**(1274), 358–379 (1959).

[20] Antonello, J., Burke, D., and Booth, M. J., "Aberrations in stimulated emission depletion (sted) microscopy," *Optics communications* **404**, 203–209 (2017).

[21] Hine, R. et al., [*The facts on file dictionary of biology*], Infobase Publishing (2009).

[22] Egelman, E. H., "The structure of f-actin," *Journal of Muscle Research & Cell Motility* **6**(2), 129–151 (1985).

[23] Pilhofer, M., Ladinsky, M. S., McDowall, A. W., Petroni, G., and Jensen, G. J., "Microtubules in bacteria: ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton," *PLoS biology* **9**(12), e1001213 (2011).

[24] Mohr, D., Frey, S., Fischer, T., Güttler, T., and Görlich, D., "Characterisation of the passive permeability barrier of nuclear pore complexes," *The EMBO journal* **28**(17), 2541–2553 (2009).

[25] Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M., and Remington, S. J., "Structural basis for dual excitation and photoisomerization of the aequorea victoria green fluorescent protein," *Proceedings of the National Academy of Sciences* **94**(6), 2306–2311 (1997).

[26] Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M., Sun, L., Cong, W., and Wang, G., "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans Med Imaging* **37**(6), 1522–1534 (2018).

[27] Shan, H., Padole, A., Homayounieh, F., Kruger, U., Khera, R. D., Nitiwarangkul, C., Kalra, M. K., and Wang, G., "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction," *Nature Machine Intelligence* **1**(6), 269–276 (2019).

[28] Arjovsky, M., Chintala, S., and Bottou, L., "Wasserstein generative adversarial networks," in [*International Conference on Machine Learning*], 214–223 (2017).

[29] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*Advances in Neural Information Processing Systems*], 2672–2680 (2014).

[30] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," in [*Int. Conf. Learn. Representations*], (2015).

[31] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," in [*International Conference on Learning Representations*], (2015).

[32] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C., "Improved training of Wasserstein GANs," in [*Advances in Neural Information Processing Systems*], 5769–5779 (2017).