

Supplementary Information for

Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics

Yiwei Hou^{1,#}, Shu Gao^{1,#}, Wei Ren¹, Yunzhe Fu¹, Meiqi Li^{2,*}, Peng Xi^{1,*}

¹*Department of Biomedical Engineering, National Biomedical Imaging Center, Peking University, College of Future Technology, Beijing 100871, China,*

²*School of Life Sciences, Peking University, Beijing 100871, China.*

[#]*These authors contributed equally*

^{*}*Correspondence: limeiqi@pku.edu.cn; xipeng@pku.edu.cn*

Supplementary Note 1. A brief review of deep-learning-based fluorescence image restoration literature

Deep learning (DL) has shown marvelous breakthroughs in vision tasks since the introduction of several key technical advancements, such as the convolutional neural network (CNN)¹, nonlinear activation², and skip connection³. In recent years, several DL techniques in natural image restoration have been successfully transferred to the field of fluorescence imaging, accomplishing restoration tasks such as denoising⁴⁻¹⁰, super-resolution^{4, 7, 11-17}, and isotropic resolution correction^{7, 9, 18}. Here, we reviewed some representative works related to our topic.

Weigert et al.⁹ pioneeringly explored a deep learning approach in augmenting fluorescence microscopy. By using a standard 3D-UNet model, they have accomplished denoising and isotropic resolution correction for fluorescence microscopy. Wang et al.¹⁷ explored a generative adversarial network (GAN) in improving the resolution of fluorescence microscopy, namely modality transformation. Ounkomol et al.¹⁹ explored transforming transmitted light microscopy to fluorescence images employing a basic 3D-UNet network. Jin et al.¹² explored utilizing UNet to accomplish reconstruction of structured illumination microscopy, achieving superior results than conventional reconstruction in noisy conditions.

After the initial trials on several basic fluorescence image restoration tasks, researchers began to develop or transfer more advanced models for fluorescence image restoration tasks. Wang et al.¹⁸ developed a global voxel transformer network (GVTNet) that added the attention operation in UNet, outperforming conventional UNet in several image transformation tasks. Bouchard et al.²⁰ proposed to improve the network performance by adding auxiliary tasks such as segmentation and localization. Chen et al.⁴ upgraded RCAN to the three-dimensional version and showed more powerful performance than 3D-UNet in a wide range of volumetric fluorescence image restoration tasks. Qiao et al.²¹ developed a comprehensive single-image super-resolution dataset, termed BioSR. They also devised a deep Fourier channel attention neural network (DFCAN) by adding a Fourier attention mechanism in conventional RCAN, which outperforms previous methods in super-resolving widefield images. They subsequently developed a video super-resolution dataset called BioTISR¹⁶, and a DPA-TISR network that outperforms single-frame super-resolution by leveraging the temporal information. Ma et al.⁷ explored using Swin Transformer as the foundation model for fluorescence restoration tasks, which achieves superior results than previous state-of-the-art CNN architectures in a series of restoration tasks.

The above-mentioned works all focused on supervised restoration tasks. Researchers have also explored self-supervised denoising methods that require no clean images as the ground-truth. Li et al.⁶ developed DeepCAD, which utilizes the adjacent frames in a temporal stack to accomplish noise2noise supervised denoising for calcium imaging. Chen et al.²² developed a neighborhood2neighborhood framework for denoising SIM images. Qu et al.⁸ developed a self-inspired noise2noise framework for general denoising of fluorescence images.

Besides refining the network performance in end-to-end training tasks, some works combine the real physical property of fluorescence microscopy, which also greatly drives progress in this field. For example, Qiao et al.¹⁵ developed a rationalized deep-learning (rDL) framework that preserves the modulation pattern in SIM, which achieves high-fidelity DL-assisted SIM reconstruction. Bilodeau et al.²⁰ developed pySTED, a realistic simulation platform for STED, which assists the deployment of DL methods in real scenarios.

All above-mentioned works shall employ a backbone deep neural network for feature extraction. All employed models could be summarized as variants of UNet, RCAN, and Vision/Swin Transformer. These networks aim at fine-fitting patches in a 128×128 patch region. As from our analysis of the effective receptive field, this is mainly due to that these networks only build an ERF in such a region size. In contrast, our proposed LargePNet aims to gain information effectively for a much larger region (e.g., 512×512) to improve the performance of the above-mentioned fluorescence image restoration tasks. We wish that LargePNet could offer a new insight for developing restoration networks for fluorescence microscopy and inspire similar tasks in other fields.

Supplementary Note 2. Analysis of the effective receptive field

To know how LargePNet differs from conventional restoration deep neural networks, it is essential to conduct some analysis to assist our understanding. Since the idea of LargePNet is primarily aggregating information from a larger area, we employed the effective receptive field²³ (ERF) analysis method proposed by Luo et al. ERF analysis can mathematically estimate how many pixels are engaged in the final decision process of a central pixel in the output image.

Suppose $y_{i,j}$ be the output of the network, with i and j indexing the pixel number in the image. Suppose $x_{i,j}^p$ be the propagated feature map of the p th layer of the network, with $x_{i,j}^0$ being the input image. ERF measures how each $x_{i,j}^0$ determines the decision of $y_{0,0}$. A quantity that can directly reflect this is the partial derivative $\partial y_{0,0} / \partial x_{i,j}^0$. It can be computed via the back-propagation chain rule:

$$\frac{\partial l}{\partial x_{i,j}^0} = \sum_{i,j} \frac{\partial l}{\partial y_{i,j}} \frac{\partial y_{i,j}}{\partial x_{i,j}^0} \quad (2-1)$$

By setting $\partial l / \partial y_{0,0} = 1$, $\partial l / \partial y_{i,j} = 0$ for all $i \neq 0, j \neq 0$, and back-propagating the error gradient, we can obtain the value $\partial y_{0,0} / \partial x_{i,j}^0$ and compose the final ERF map.

Supplementary Note 3. Details of LargePNet

● LargePNet architecture

LargePNet is a pyramid-like architecture that separates the features into different scales. The branch holding the same size as the input image is called the high-frequency feature extractor (HFFE), and the branches processing the down-sampled images are called the low-frequency feature extractor (LFFE). The LFFE is composed of two consecutive U-shaped networks but is modified to fit our goal of training with a larger patch size, which we call InsUNet. Compared with the standard UNet, InsUNet employs instance normalization²⁴ instead of batch normalization to stabilize the training. It also utilizes average pooling instead of maximum pooling, as in the conventional UNet. The reason is that we already handle the processing of high-frequency details to HFFE, without the need to retain such information in LFFE. We typically found that using average pooling in LFFE can help achieve better performance in noisy conditions.

For a batch-channel-height-width feature map, the instance normalization is accomplished as follows:

$$\begin{cases} \mu_{BC} = \frac{1}{HW} \sum_{l=1}^W \sum_{m=1}^H x_{BClm} \\ \sigma_{BC}^2 = \frac{1}{HW} \sum_{l=1}^W \sum_{m=1}^H x_{BClm}^2 \\ y_{BC} = \frac{x_{BC} - \mu_{BC}}{\sqrt{\sigma_{BC}^2 + \varepsilon}} \end{cases} \quad (3-1)$$

where x_{BC} denotes the feature map of the B th in the batch dimension and C th in the channel dimension, y_{BC} denotes the feature map after instance normalization, H and W denotes the height and width of the feature map, μ is the average value of a feature map, σ is the standard deviation of a feature map, ε is a constant and chosen as $1e-5$.

We use the LeakyReLU as the nonlinear activation function in InsUNet:

$$\text{LeakyReLU}(x) = \max(\alpha x, x) \quad (3-2)$$

where α is a parameter controlling the slope in the negative region, and is chosen to be 0.2 in our network.

The HFFE branch is composed of four consecutive reparametrized ultra-large-size convolution layers^{25, 26} (RepLKConv). We also use instance normalization after each convolutional layer. The kernel size of the ultra-large-size convolutional kernel is empirically chosen as 25×25 for all restoration tasks without upscaling, and 15×15 for all restoration tasks with upscaling (i.e., super-resolution) in our experiments. The whole network architecture and the distribution of parameters and computational cost arrangements are shown in Supplementary Fig. 3.

● Efficiency gains of LargePNet over advanced small-patch networks

When checking the computational cost and processing time, we found that LargePNet shows higher efficiency over the state-of-the-art small-patch network. Here we explain the rationale.

In the LargePNet design, the branch that processes the full resolution feature (HF Feature extraction) is the ultra-large-kernel-convolution (LKConv), and we only need a total of 4 LKConv layers to accomplish a good ERF for large-patch (e.g. 512x512) training. This branch takes the major computational cost of the LargeNet. Since it only contains 4 LKConv layers, it is much more efficient than some advanced networks with very deep layers. For example, DFCAN has 292 layers, UniFMIR has 305 layers. In these models, attention operations are repeatedly applied to achieve good performance, further increasing their burden.

LKConv effectively leverages global information and preserves the high-frequency feature in the full-resolution image. While the drawback of LKConv is that it lacks the nonlinear representation capability which is required in image restoration tasks. For example, to remove the random noise in images, the network needs deep layers to gradually transform the images. Therefore, we compensate for this by adding several branches (LF Feature extraction). These branches can provide sufficient nonlinear capability to remove noise/artifacts. Multiple down-sampling operations that average local pixels further aid the removal of noise/artifacts and simultaneously reduce the computational cost.

In summary, LargePNet’s efficiency gains arise from using deep networks to LF branches (where it is most effective and cheapest in computational burden) and using shallow LKConv for high-res HF branches (which builds large ERF and aggregates global information). This scale-adaptive division and rationale choice of network depth achieved these efficiency gains.

Another issue worth discussing is how the efficiency gain of LargePNet varies with different input image sizes compared with the small-patch CNN and Swin Transformer derived model. Here we discuss the example result in Fig. 2f:

- LargePNet versus DFCAN-F (full-size inference mode):

Theoretically, the computational efficiency improvement of LargePNet over DFCAN-Full would be relatively stable with input image size. This is due to that both are CNN architectures, and the convolutional operation complexity would increase proportionally with the square of the image size. In practice, memory footprint also acts as a key factor: LargePNet has a lower memory occupation than DFCAN-Full, thus imposing a smaller memory burden as input size increases, which leads to a slight rise in its fold efficiency improvement. In the provided results in Fig. 2f, LargePNet achieves ~3-fold efficiency gains in 512-size (0.065 sec versus 0.193 sec) inference and ~3.5-fold in 1024-size inference (0.214 sec over 0.736 sec).

- LargePNet versus DFCAN-S (small-size inference and stitching mode)

For DFCAN, when it infers large-size image using the stitching mode, if there is no overlap between the small images of inference, then its computational complexity will also increase proportionally with the square of the image size. However, this can lead to serious edge artifacts. If the inference strategy of overlapping small graphs is adopted, its computational complexity will further increase significantly on this basis. For example, if a 50% overlap strategy is adopted to infer a 512-size image, DFCAN needs to infer 49 128-size images; To infer an image with a size of 1024, DFCAN needs to infer 225 128-size images. This has evidently exceeded the square growth. Moreover, additional cache pressure, parallelism loss, and data transmission will

further reduce its actual speed. As shown in Fig. 2f, LargePNet achieves ~ 12 -fold efficiency gain in 512-size and ~ 30 -fold efficiency gain in 1024-size over DFCAN-stitching.

- LargePNet versus Swin Transformer derived model (UniFMIR and SwinIR)

UniFMIR is derived from Swin Transformer. When inferring a full-size image, its shifted window attention mechanism ensures computational efficiency and allows adjacent 8×8 -size windows to overlap and transmit context, avoiding boundary blocks. As a result, it also follows a square growth with the input image size in the inference stage, just like a CNN adopts stitching mode, but without overlap. Similarly, the memory cost also influences the practical speed. LargePNet has a lower memory occupation than UniFMIR, thus imposing a smaller memory burden as the input size increases, which leads to a slight rise in its fold efficiency improvement. Moreover, the multi-head self-attention operation in UniFMIR/SwinIR causes a very heavy computational burden. LargePNet has a significantly higher efficiency than UniFMIR, whatever the input size. As shown in Fig. 2f, LargePNet achieves ~ 21 -fold efficiency gains in the 512-size input and 26-fold in 1024-size input over UniFMIR.

The fold improvement in efficiency was discussed above. In addition, it is worth noting that as the input size increases, the total time savings brought by LargePNet will also greatly increase. This also constitutes an important consideration for practical deployment.

Supplementary Note 4. Details of LargeP-TISR, 3D-LargePNet, and LargeP-GAN architecture

LargePNet provides a strong backbone for single-image restoration tasks. We further extended it for generative restoration (LargeP-GAN), video super-resolution (LargeP-TISR), and volumetric data restoration (3D-LargePNet) with appropriate modifications of its original architecture.

● LargeP-GAN architecture

Based on the LargePNet architecture, we further devise its generative adversarial form, LargeP-GAN. In the image restoration regime, GAN is composed of a generator G and a discriminator D . The generator G is identical to some non-GAN restoration model that tries to recover the degenerated image. The discriminator D aims to differentiate the GT and the recovered image. The adversary of D and G allows to recover of high-frequency information better than in a non-GAN training scheme.

In conventional GANs, D is generally a classifier. The adversarial loss of G is designed to make D assess the generated image as true. The adversarial loss of D is designed to assess the generated image as false and assess the ground-truth image as true. This scheme was also adopted for some GANs devised for fluorescence image restoration.

However, some later works show that this scheme causes difficulties in training and confines the final outcome. Here, when developing LargeP-GAN, we take the relativistic GAN (RaGAN) training scheme, a more advanced scheme that was adopted in the ESRGAN²⁷ series. Instead of simply assessing true or false, in the RaGAN scheme, the adversarial loss of G is designed to make the recovered image's probability of being true larger than the ground-truth image. The adversarial loss of D is designed to make the ground-truth image's probability of being true larger than the recovered image. The mathematical operation that defines this relativistic probability is as follow:

$$D_{Ra}(x_{gt}, x_{rec}) = \sigma(D(x_{gt}) - E_{x_{rec}}[D(x_{rec})]) \quad (4-1)$$

where D denotes the output from the discriminator, σ denotes the sigmoid operation, x_{gt} denotes the ground-truth image, x_{rec} denotes the image recovered by the generator, and E denotes the calculation of the average value.

The relativistic loss for the generator and discriminator is described as follow:

$$L_D = -E_{x_{gt}}[\log(D_{Ra}(x_{gt}, x_{rec}))] - E_{x_{rec}}[\log(1 - D_{Ra}(x_{rec}, x_{gt}))] \quad (4-2)$$

$$L_G = -E_{x_{gt}}[\log(1 - D_{Ra}(x_{gt}, x_{rec}))] - E_{x_{rec}}[\log(D_{Ra}(x_{rec}, x_{gt}))] \quad (4-3)$$

● LargeP-TISR architecture

LargeP-TISR is devised for video super-resolution tasks, which can leverage the temporal information to improve the performance of single-image super-resolution. To achieve so, we added an optical flow reconstruction network²⁸ (OFRnet) before the super-resolution LargePNet block. We adjusted the propagation

channel number in OFRnet to 64 to save the computational cost for a large patch. The input and output channel number of the consecutive super-resolution LargePNet is adjusted to meet the requirements of video super-resolution. For the 7-frame video super-resolution task in BioTISR, the input channel number and output channel number for the LargePNet in the LargeP-TISR framework should be 25 and 7, respectively. Besides the boundary frames, LargeP-TISR would only preserve the result of the central frame, which leverages adjacent temporal information at best. Then it gradually shifts frame by frame to accomplish the super-resolution of the full video. This processing protocol is the same as the pioneer work. The major innovation in LargeP-TISR is that it adopts the LargePNet for super-resolution with a large-view training paradigm.

● **3D-LargePNet architecture**

To process the volumetric data, we upgrade all the 3×3 convolution operations in LargePNet to the 3D $3\times 3\times 3$ convolution. For the ultra-large-size convolution operation, a direct idea is also to upgrade the 2D 15×15 convolution to 3D $15\times 15\times 15$ operations. However, the connection of biological structures in the axial direction is not as strong as that in the lateral dimensions. Moreover, the sampling rate and resolution of the axial resolution are also much lower than the lateral dimension. Therefore, using a kernel size in the axial dimension similar to the lateral dimensions could cause a massive waste of computational cost. Therefore, we devise a rectangular 3D $15\times 15\times 3$ operation in 3D-LargePNet. This modification turns out to be rather efficient and maintains equivalent performance with the full 3D ultra-large-size convolution.

Supplementary Note 5. Details of loss function, evaluation metric and hyper-parameter selections

(1) Loss function and evaluation metrics

We employed commonly used image metrics as loss functions and quality evaluation metrics. Mathematical descriptions of the employed metrics are as follows:

The mean-square error (MSE) is calculated as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - g_i)^2 \quad (5-1)$$

where N denotes the image pixel volume number, f denotes the network output image, g denotes the ground-truth image.

The peak-signal-to-noise ratio (PSNR) is calculated as follows for a normalized image:

$$PSNR(dB) = 10 \lg \frac{1}{MSE} \quad (5-2)$$

The mean absolute error (MAE) is calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |f_i - g_i| \quad (5-3)$$

The normalized root-mean square error (NRMSE) is calculated as follows:

$$NRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - g_i)^2} / (g_{\max} - g_{\min}) \quad (5-4)$$

The structural similarity index (SSIM) is calculated as follows:

$$SSIM = \frac{(2\mu_f\mu_g + c_1)(2\sigma_f\sigma_g + c_2)}{(\mu_f^2 + \mu_g^2 + c_1)(\sigma_f^2 + \sigma_g^2 + c_2)} \quad (5-5)$$

where μ_f and μ_g are their average value, σ_f and σ_g are their standard deviation value, $\sigma_f\sigma_g$ is the covariance of f and g , c_1 and c_2 are two constants to stabilize the result ($c_1=(k_1L)^2$, $c_2=(k_2L)^2$, L is the dynamic range of pixel value, $k_1=0.01$, $k_2=0.03$).

The MS - $SSIM$ value is calculated as the weighted average of SSIM by inspecting the SSIM value at different image resolution scales:

$$MS - SSIM = \left[\frac{2\mu_{f_M}\mu_{g_M} + c_1}{\mu_{f_M}^2 + \mu_{g_M}^2 + c_1} \right]^{\alpha_M} \cdot \prod_{j=1}^M \left[\frac{2\sigma_{f_j}\sigma_{g_j} + c_2}{\sigma_{f_j}^2 + \sigma_{g_j}^2 + c_2} \right]^{\beta_j} \left[\frac{\sigma_{f_j g_j} + c_3}{\sigma_{f_j}\sigma_{g_j} + c_3} \right]^{\gamma_j} \quad (5-6)$$

where f_j represents the image feature level, with f_M being the image with highest resolution level, c_3 is a constant that is chosen as $c_2/2$, α_M , β_j , and γ_j are relative constants used to adjust the weight of different scales. We employed the weight-related coefficients as the default value in the Matlab *multissim* function.

When training the GAN-based model, we added a perceptual loss in the conventional pixel loss term that measures the difference between two images in the feature space:

$$Perceptual\ loss = \frac{1}{N} \sum_{i=1}^N [VGG_{16}^{31}(f)_i - VGG_{16}^{31}(g)_i]^2 \quad (5-7)$$

where $VGG_{16}^{31}()$ denotes the feature map in the 31th layer of the pre-trained VGG_{16} model. The perceptual loss was reported to help the GAN-based model better recover texture information.

We chose the combination of pixel-precision metrics (PSNR, NRMSE) and structural similarity metric (MS-SSIM) based on the dual requirements of fluorescence image restoration for quantitative intensity fidelity and biological structure preservation. Their complementary roles effectively address the multi-dimensional evaluation needs in biological imaging scenarios.

Pixel-precision metric (PSNR and NRMSE) quantifies pixel-level grayscale fidelity that ensures signal discriminability of biological structures. Loss of the completeness of the biological structure or distortion in the recovered gray value compared to the ground-truth image will be directly reflected in PSNR and NRMSE metrics. A higher pixel-precision metric also provides a more reliable grayscale basis for subsequent structural identification, such as segmentation. The pixel-precision metric shall be the premier standard for evaluating the restoration quality in the scientific imaging field, such as fluorescence imaging, that requires high-fidelity.

SSIM focuses on the similarity of two images in brightness (gray value average) and contrast (gray value standard deviation) over a region instead of the pixel-precision difference. MS-SSIM is a multi-scale extension of SSIM that incorporates calculations across multiple scales, addressing the limitations of SSIM's single-scale evaluation. However, since MS-SSIM considers the similarity of different image scales, it is insensitive to the integrity of small, subtle biological structures such as mitochondrial cristae shown in Fig. 3. It mainly serves as an auxiliary evaluation standard for pixel-precision metrics to better reflect the structural distortions.

Besides the three metrics, we added the FID (Fréchet Inception Distance) metric to assist the assessment of the SMLM virtual sampling task. This task differs slightly from other tasks in removing noise or blurring targets, as it involves filling in insufficiently sampled single-molecule emission points. From the gain of the samplings in SMLM, we can more confidently assess the existence of a microtubule filament. For example, in the under-sampled SMLM image, we can hardly assess the existence of a complete microtubule filament, but in a well-sampled SMLM image, the count of microtubules becomes clear. While pixel-precision or local similarity metrics like PSNR, NRMSE, and MS-SSIM can measure the differences, they fail to provide the semantic level comparison, which may also be an important aspect that reflects the restoration quality of the SMLM images. The core idea of FID is to measure the distance between the perception of the distribution of GT images and the restored images on the pre-trained Inception-v3 model. It can evaluate semantic quality information, such as whether the restored microtubules are as complete as the GT SMLM images. Therefore, here we use the FID value as an additional evaluation metric for assessing the quality of the restored SMLM images.

(2) Choice of hyperparameters

The detailed loss function selection and other hyper-parameter choices in various downstream tasks are

as follows:

- **Single image denoising of widefield image (open-source dataset)**

We use $loss\ function = MSE + 0.05 * MAE$ to train the LargePNet and UNet models displayed in Supplementary Fig. 4, since this is a PSNR-oriented task. The initial learning rate of LargePNet and UNet is chosen as 0.001 and 0.0001, respectively. A learning rate of 0.001 for training LargePNet is suitable, but for training UNet would cause a gradient explosion. Both networks were trained with 30 epochs, with the learning rate decaying half in the 10th epoch. Changing the ratio of MSE and MAE generally yields similar results.

- **Single image super-solution (open-source dataset)**

We use $loss\ function = MSE + 0.05 * MAE$ to train the LargePNet for the results displayed in Fig. 2, which was the same as the loss function adopted by the UniFMIR⁷ model. The LargePNet model was trained with 30 epochs with an initial learning rate of 0.001, which was decayed by half after the 10th epoch. The corresponding mini-batch ($1 \times 512 \times 512$) iteration number is $\sim 30,000$ for each structure. Since there are no well-trained DFCAN models for the four structures, we also retrained DFCAN using the parameters provided in their paper. To train a loss function of $loss\ function = MSE + 0.1 * (1 - SSIM)$, and a learning rate of 0.0001 as described in their paper. The DFCAN was trained with 30 epochs, corresponding to a mini-batch iteration ($4 \times 128 \times 128$) number of $\sim 150,000$ for each structure. For UniFMIR, we directly use their fine-tuned models of the four structures provided in the paper⁷.

- **Single image denoising of STED image (self-established dataset)**

We use $loss\ function = MSE + 0.05 * MAE$ to train the LargePNet model for the STED single-image denoising task. The LargePNet model was trained with 30 epochs with an initial learning rate of 0.001, which was decayed by half after the 10th epoch. The corresponding mini-batch ($1 \times 512 \times 512$) iteration number is $\sim 36,000$ for each structure. The competitor UNet-RCAN⁵ model was a proposed network targeting the STED denoising task. Therefore, we trained UNet-RCAN using the loss function and hyperparameters provided in their paper and code. The competitor SwinIR²⁹ model was a strong baseline transformer model for image restoration and was adopted as the major feature extractor in UniFMIR. We trained the SwinIR model with the $loss\ function = MSE + 0.05 * MAE$, an epoch number of 30, and an initial learning rate of 0.0001, which was decayed by half after the 10th epoch. The corresponding mini-batch ($4 \times 128 \times 128$) iteration number is $\sim 150,000$. The parameter setting refers to that adopted by UniFMIR⁷.

- **Single-image deblurring of STED images (self-established dataset)**

We use $loss\ function = MAE + 0.001 * Adversarial\ loss + 0.006 * Perceptual\ loss$ for training the GAN-based STED-deblurring restoration model. The initial learning rate for the LargeP-GAN generator is selected as 0.001, for the RRDB-GAN and UNet-GAN, it is chosen as 0.0001. For the discriminator, we tried a discriminator learning rate of $1e-5$, $5e-6$, $2e-6$, and $2e-7$. We found that using $2e-6$ as the discriminator learning rate can maintain the model training best, generating the best balance in resolution and PSNR in the final image. The training is performed with 50 epochs, with the learning rate decaying at the 10th epoch.

- **Virtual sampling of SMLM (self-established dataset)**

We use $MSE + 0.001 * \text{Adversarial loss} + 0.006 * \text{Perceptual loss}$ to train UNet-GAN, RRDB-GAN, and LargeP-GAN. The generator learning rate for UNet-GAN and RRDB-GAN is chosen as $1e-4$, and larger learning rates were found to cause collapse of the training process. The generator learning rate for LargeP-GAN is chosen as $1e-3$. The learning rate for all the discriminators is chosen as $2e-7$.

- **Self-supervised denoising of widefield image (open-source dataset)**

We use the default SN2N loss function that is devised for self-supervised denoising for a single image to train UNet-SN2N and LargePNet-SN2N. The loss function is composed of a conventional noise2noise loss of a self-constrained loss. The UNet-S2N2 was trained with 100 epochs with an initial learning rate of 0.0002, an epoch number of 100 as suggested in their paper. The corresponding mini-batch ($32 \times 128 \times 128$) iteration number is $\sim 625,000$. The LargePNet-S2N2 was trained with an initial learning rate of 0.001 and an epoch number of 50. The corresponding mini-batch ($1 \times 512 \times 512$) iteration number is $\sim 50,000$.

- **Video super-resolution (open-source dataset)**

We use $loss\ function = MSE + 0.05 * MAE$ to train the LargeP-TISR model for super-resolution of widefield video. The LargeP-TISR model was trained with 30 epochs with an initial learning rate of 0.001, which was decayed by half after the 10th epoch. The corresponding mini-batch ($1 \times 7 \times 512 \times 512$) iteration number is $\sim 30,000$ for each structure. Well-trained DPA-TISR models on the microtubule and mitochondrion were provided in their paper¹⁶, providing objective criteria to assess the performance of LargeP-TISR.

- **Volumetric background removal of widefield image stack (self-established dataset)**

We use $loss\ function = MSE + 0.05 * MAE$ to train the 3D-LargePNet and 3D-RCAN⁴ models for background removal of the volumetric widefield image stack. The initial learning rate of 3D-LargePNet was selected as 0.001. We tried an initial learning rate of 0.0004, 0.0002, and 0.0001 for 3D-RCAN, and found that with 0.0001, they obtain the best performance. A higher learning rate for training the model would cause gradient collapsing. All three models were trained with 200 epochs, with the learning rate decaying half at the 50th epoch. Since 3D volume takes large memory, the batch size for all three models is 1, and the corresponding mini-batch iteration for 3D-LargePNet ($1 \times 16 \times 512 \times 512$) and the two competitor models ($1 \times 16 \times 256 \times 256$) is 3,000. The selection of the patch-size for 3D-RCAN follows the selections in their papers⁴. The validation loss did not vary much after the 100th epoch, as shown in Supplementary Fig. 18.

- **Volumetric confocal to STED transformation (open-source dataset)**

We use $loss\ function = MSE + 0.05 * MAE$ to train the 3D-LargePNet model. The initial learning rate was selected as 0.001, which was decayed half after the 50th epoch. The training was conducted for 300 epochs, corresponding to mini-batch iteration ($1 \times 6 \times 640 \times 640$) of around 25,000. For the 3D-RCAN model, we employed the hyperparameters used in the original paper⁴ tailored for this task, with a $6 \times 320 \times 320$ patch-size.

Supplementary Note 6. Explaining the advantage of LargePNet in different structures

In the fluorescence-image-restoration tasks presented in this paper, LargePNet demonstrated improved performance compared to small-patch training networks. An interesting and noteworthy phenomenon is that LargePNet has greater or smaller advantages in the recovery of certain structures. Here we attempt to analyze the reasons behind this.

The main difference between LargePNet and other networks is whether they aggregate large-view information. Therefore, we examine when LargePNet performs better by analyzing the differences in structural statistical information between large and small view images.

We employ the Gray-Level Co-occurrence Matrix (GLCM) to analyze the structural statistics information in an image. The statistics of GLCM measures the texture information and structural connection in an image. Here we employ the commonly used the Contrast, Energy, and Correlation statistics metrics of GLCM, which can be easily obtained via the MATLAB inner function *graycomatrix* and *graycoprops*. It is worth noting that in image restoration, due to the roughly similar contours between the raw and GT images, the difference in statistics between the large and small images of the structure itself may not be suitable to explain the performance of LargePNet. How the structure is disrupted in the view of large and small images (i.e. the difference between the raw and GT images) is a more reasonable explanatory factor. Based on this rationale, we develop the calculation process for explaining the structure-dependent advantages of LargePNet using GLCM, as shown in the Supplementary Fig. 24.

The first step is to analyze the GLCM statistical metrics of all images belong to a certain structure before and after degradation. This analysis is conducted separately in large-view images and small-view images. The second step is to calculate the changes in GLCM statistical metrics before and after degradation. For GLCM's Contrast and Energy metrics, which are obtained by addition or subtraction, we focus on their fold differences. For Correlation metric, we consider the difference in their absolute values. This analysis was also conducted separately in the large-view images and the small-view images. The third step is to compare the degree of change in GLCM statistics obtained from the large-view and that obtained from the small-view. By comparing the PSNR increments in various experiments, we found that the larger the difference in GLCM statistics changes between the small and large images, the greater the PSNR increment obtained by LargePNet.

Based on this method, we analyze the difference in structural advantage of LargePNet throughout the tasks that appear structural differences in the paper:

- BioSR single-image super-resolution task (Fig. 2).

Analysis results of this dataset are shown in Supplementary Table 11. It can be seen that LargePNet has a relatively lower PSNR gain in the F-actin compared to other structures. The GLCM statistics show that there are more significant differences in statistics between the large and small views of other structures. The differences of the GLCM Correlation metric between the large-small views of CCP, ER, and MT all exceeded

0.03. The difference in Energy statistics of CCP reached 2.8-fold, and MT reached 1.5-fold. The large-small view differences in F-actin on these statistical indicators are not as significant as in other structures. Mapping these statistical indicators onto the image, F-actin is a very complex structure, and its local regions may already contain rich information about the structure and degradation process. Therefore, the structural information lost by small-patch truncation may not be as much as other structural losses. This could be the reason why LargePNet provides a lower PSNR gain in F-actin compared to other structures.

- STED denoising task (Fig. 3).

Analysis results of this dataset are shown in Supplementary Table 12. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito IMM structure. By examining the GLCM statistic, it can be found that Mito IMM has a difference of up to 3.17-fold between the large and small views in the Energy metric; the difference between the large and small views in the Correlation metric is as high as 0.0698. These significant differences in GLCM statistical indicators reflect that it lost more complete structural information about Mito IMM and inter regional connections with small patches. This may explain why LargePNet achieves greater advantages in Mito IMM compared to other structures.

- BioTISR video super-resolution (Fig. 4)

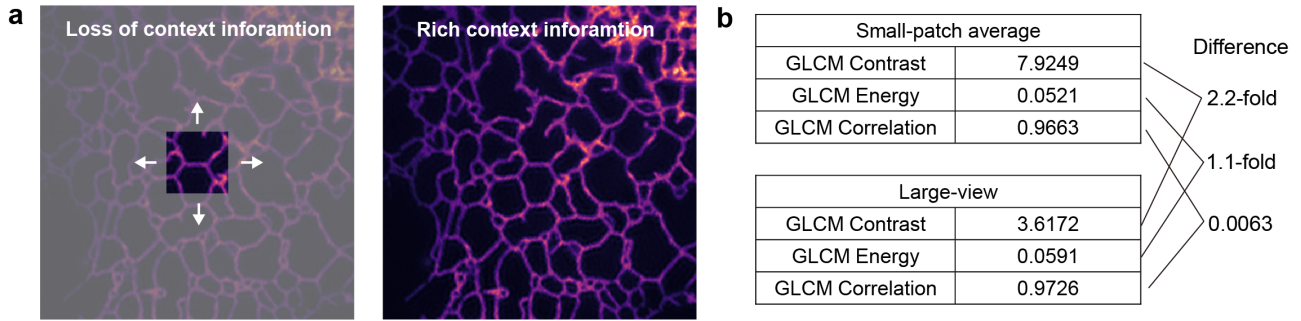
Analysis results of this dataset are shown in Supplementary Table 13. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito OMM structure compared with MT. By examining the GLCM statistic, it can be found that Mito IMM structure has a much more evident large-small view difference compared with the MT structure. Its large-small view difference in the Contrast metric is 6.85-fold, in the Energy metric is 4.16-fold. This means that in small-patch cropping, the Mito OMM structure loses more structural contextual information. It may explain why LargePNet achieves greater advantages in Mito OMM compared to MT.

- STED deblurring dataset. (Supplementary Fig. 14)

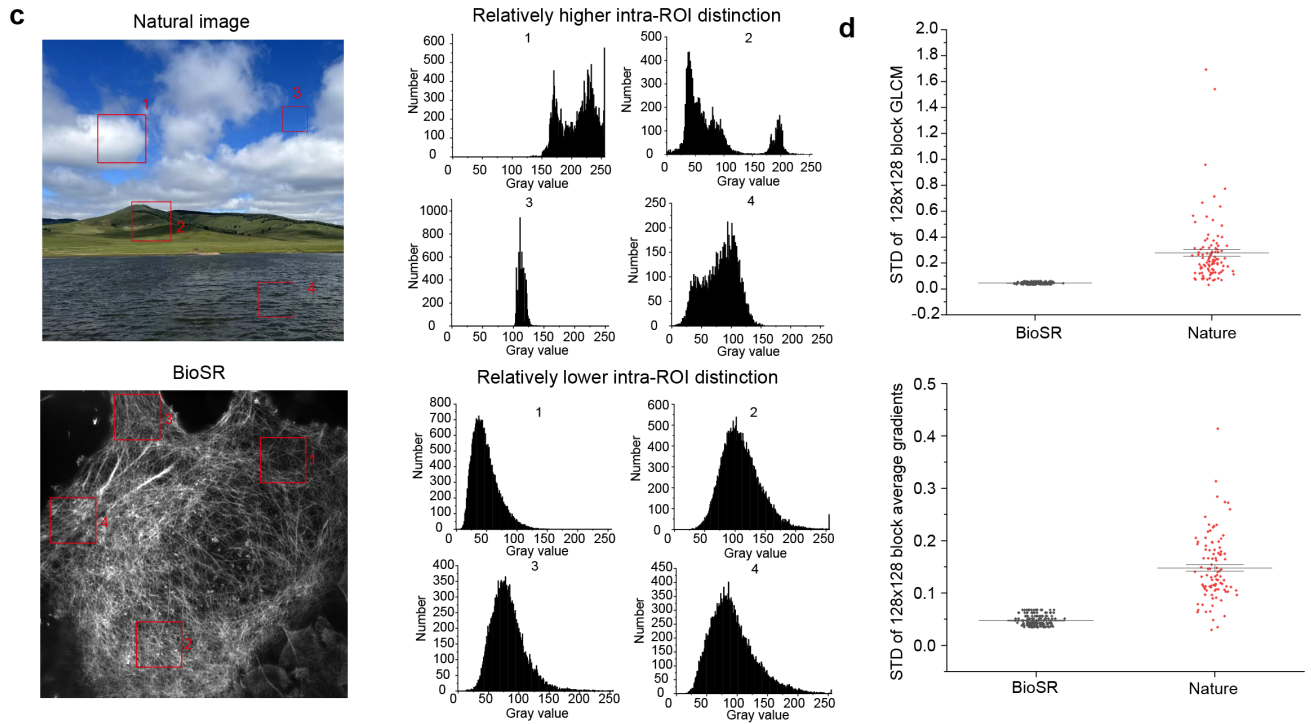
Analysis results of this dataset are shown in Supplementary Table 14. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito IMM structure compared with MT. Inspecting the GLCM statistics, it can be found that the Mito IMM shows an evidently higher large and small views difference (2.02-fold in Energy metric, 0.0652 in Correlation metric) compared with MT (1.47-fold in Energy metric, 0.0184 in Correlation metric). Mapping this indicator to the practical image, this task aims to remove the blurring degeneration in STED images. For MT as a relatively simple structure, small-patch cropping may not disturb the contextual structural information as the complex Mito IMM structure does. While for the Mito IMM structure, which is relatively complex, the advantage of LargeP-GAN's large-view statistics aggregation will exert, and cause a leaping performance difference.

Supplementary Figures

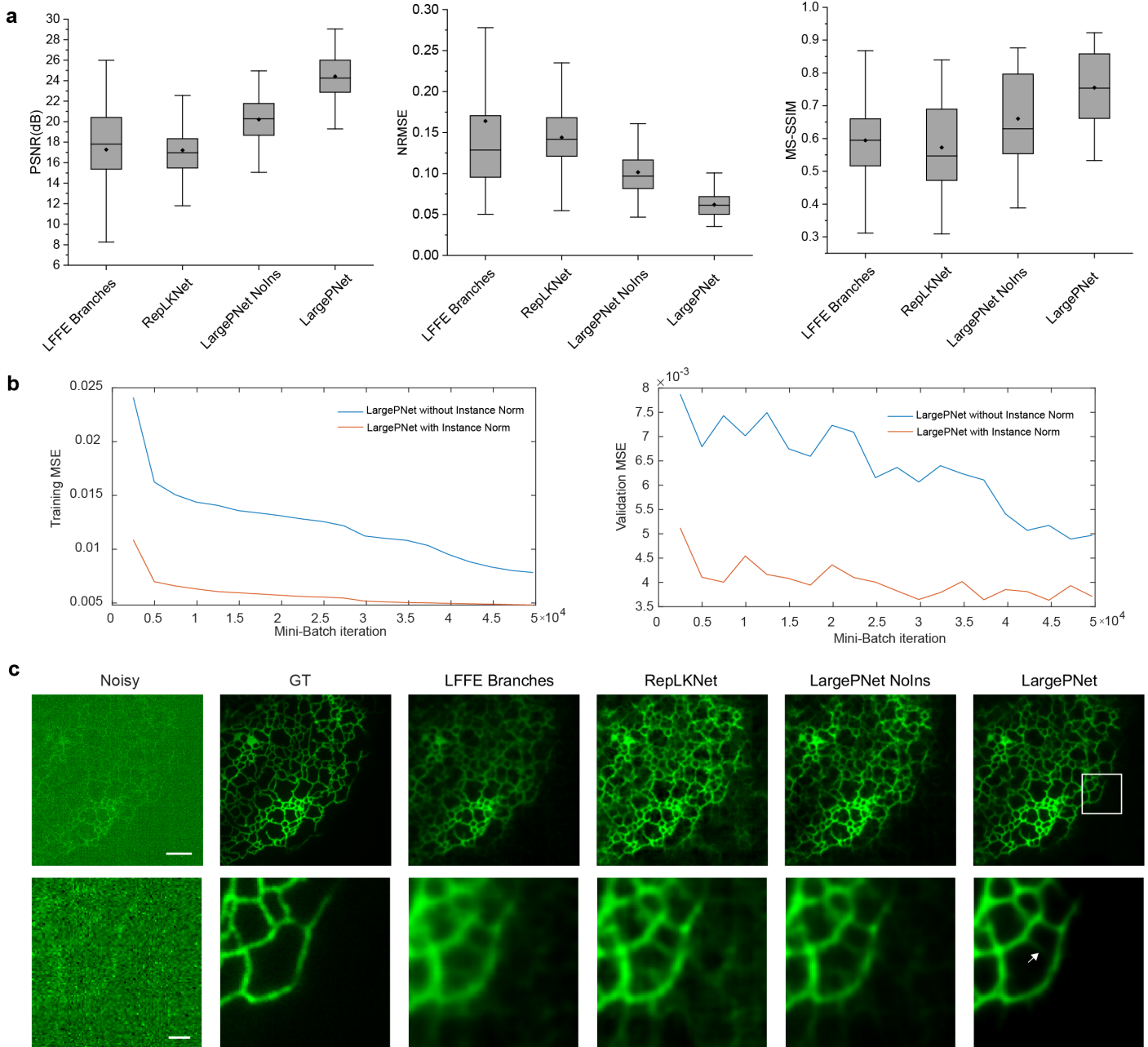
Perils of small-patch training relative to application on large-view images



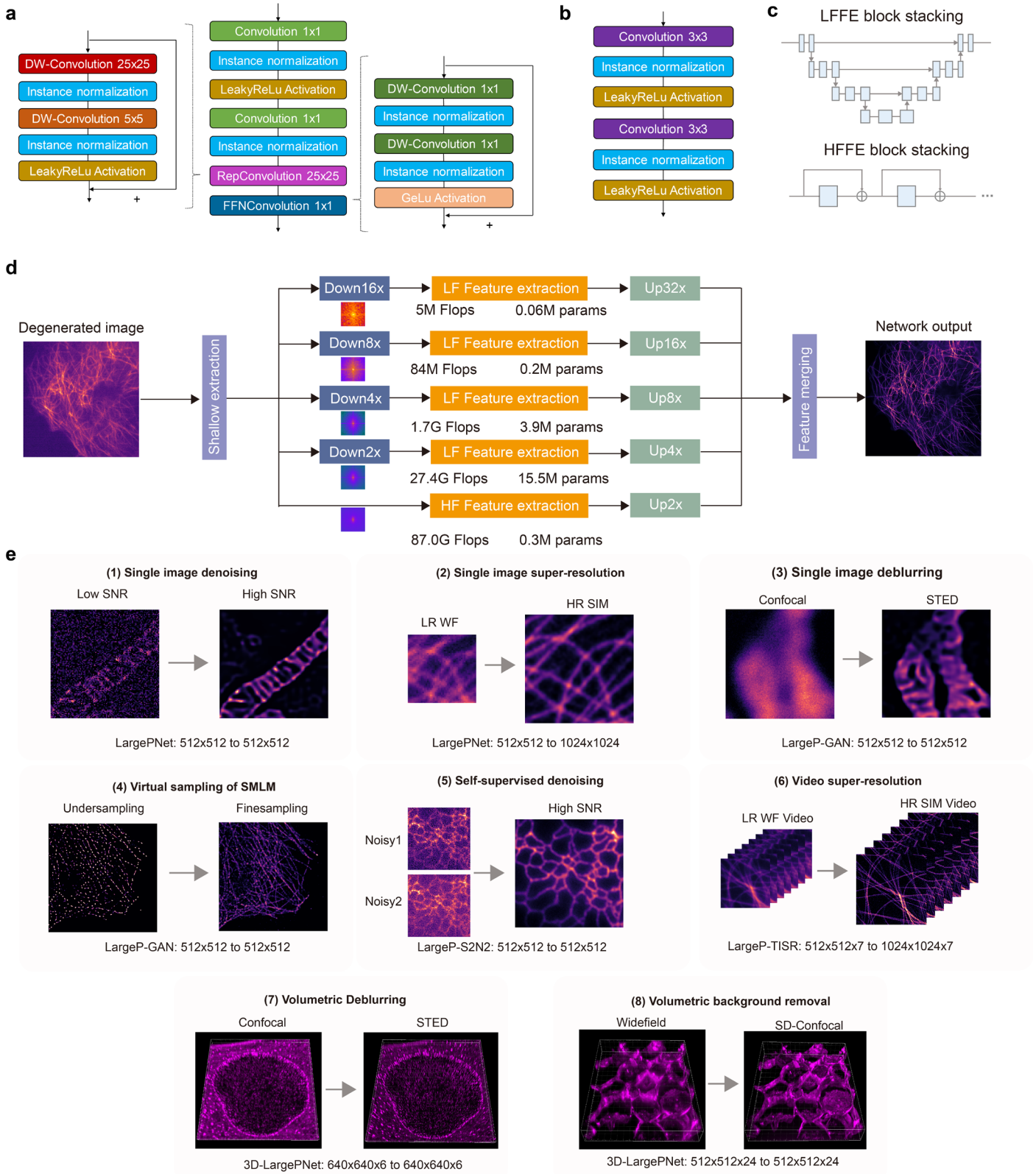
Leveraging large-view statistics is easier in fluorescence images relative to natural images



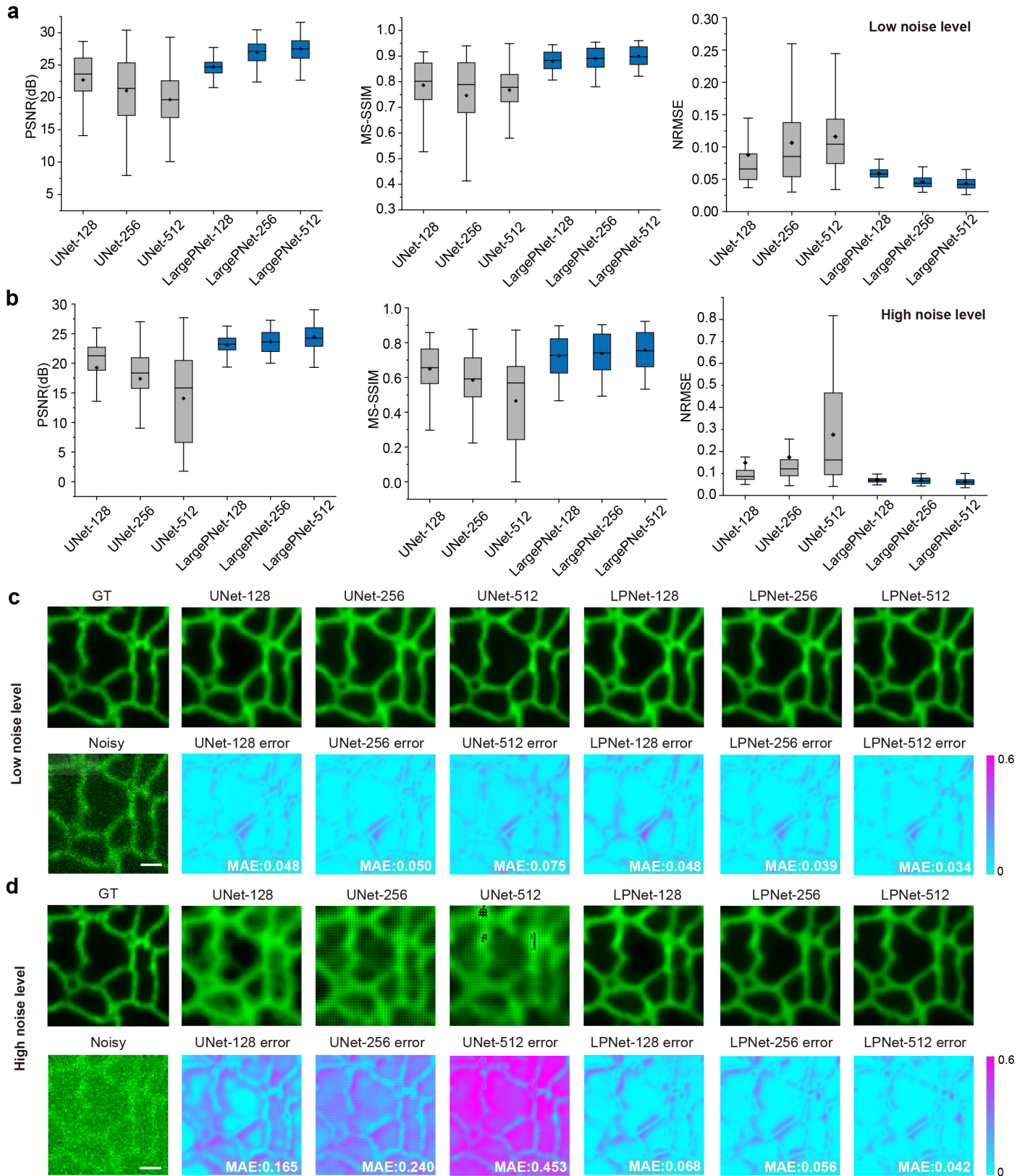
Supplementary Fig. 1. Motivations of developing LargePNet. **a.** Small-patch training forces the network to learn restoration with much truncated contextual structural information, while large-view image provides more contextual clue to benefit restoration. **b.** The inconsistency of feature statistics distribution in training small-patch and final application on large-view image would degrade the model performance. The contrast, energy, and correlation of the gray-level co-occurrence matrix (GLCM) are used to quantify the feature statistics distribution. **c.** A self-captured natural image and a representative image from the BioSR dataset, and the histogram from four selected blocks. **d.** Statistical comparisons of intra-ROI similarity in fluorescence and natural image dataset (DIV2K). The error bar displays the average value $\pm$ SEM. ($n=100$ test images). We divide a full-size image into several 128×128 patches without overlapping. Then we calculated the GLCM energy and average gradient metrics of each 128×128 patch that characterizes the structural features. A spot in the graph represents the standard deviation of the two metrics from a group of 128×128 patches generated from a full-size image. The results in c and d show that fluorescence image has a relatively lower intra-ROI distinction compared with natural images. Thus, leveraging large-view statistics is easier in fluorescence images relative to natural images.

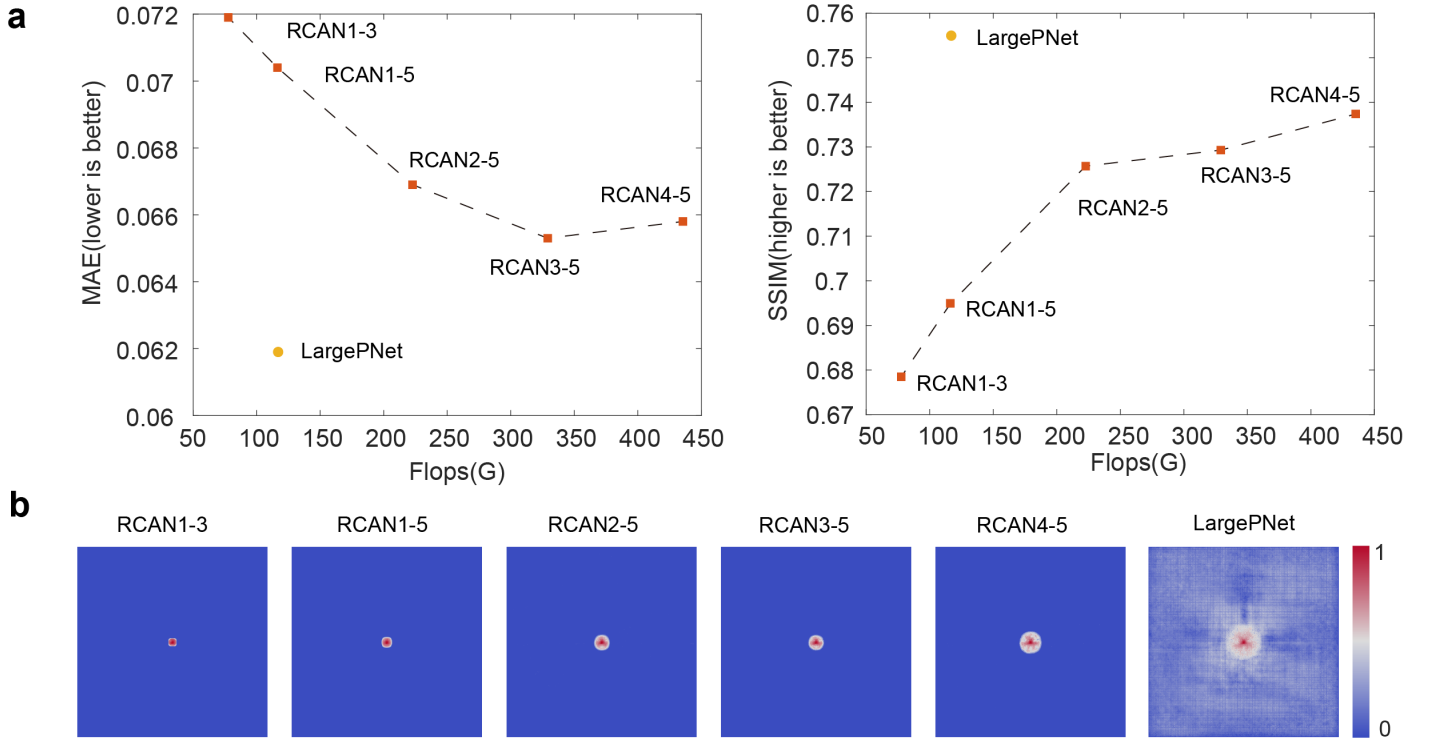


Supplementary Fig. 2. Ablation study on the architecture of LargePNet in denoising the noise-level-1 ER images in the BioSR dataset. **a.** The performance of LFEE branches and the RepLKNet HFE branch when they work alone, and LargePNet without instance normalization and with instance normalization. ($n=126$ test images). **b.** Evolution of the training MSE and validation MSE of LargePNet with and without instance normalization. **c.** Representative recovery results of the noisy ER image. Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: $5 \mu\text{m}$ (c Top), $1 \mu\text{m}$ (c Bottom).

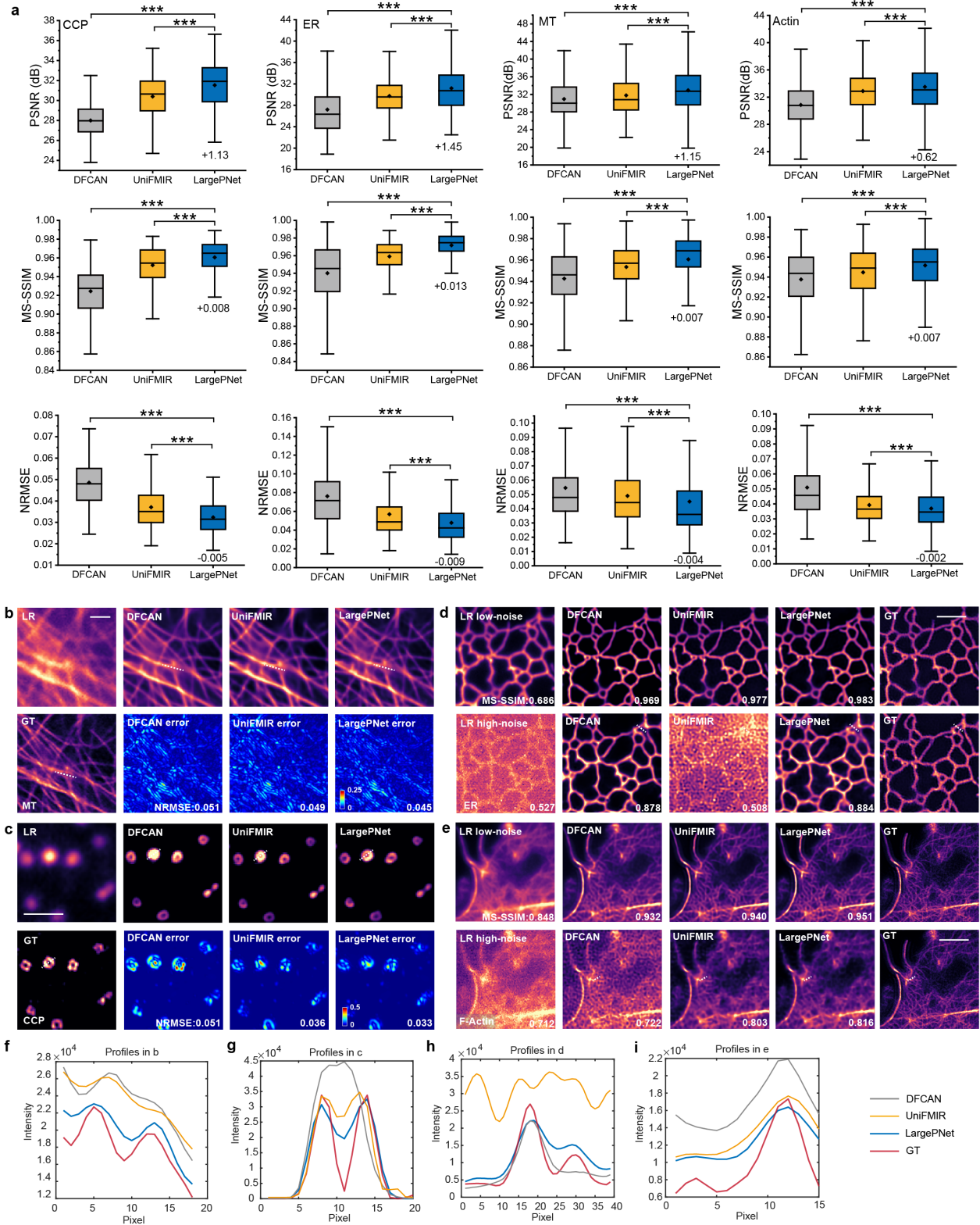


Supplementary Fig. 3. Details of LargePNet architecture and computational costs. **a.** Details of the reparameterized ultra-large-kernel convolutional layer in the high-frequency feature extractor (HFPE) branch. DW: deep-wise, FFN: feed-forward network. **b.** Details of the feature extraction block in U-shaped low-frequency feature extractor (LFPE) branches. **c.** Stacking methods of the basic blocks in LFPE and HFPE. **d.** Distribution of the parameters and computational costs. **e.** The eight common fluorescence restoration tasks explored in this study.





Supplementary Fig. 5. Comparison of the efficacy of LargePNet and RCAN on fitting the 512x512 patches of transforming the noisy ER widefield image to a high-SNR one in the BioSR dataset. a. Comparisons of the image quality metrics of LargePNet and RCAN with different depths. RCANm-n represents an RCAN model with m residual groups and n residual blocks. Deeper RCAN architectures were also tried but were found to collapse in the given data volume. **b.** Analysis of the ERF map of different networks after training.

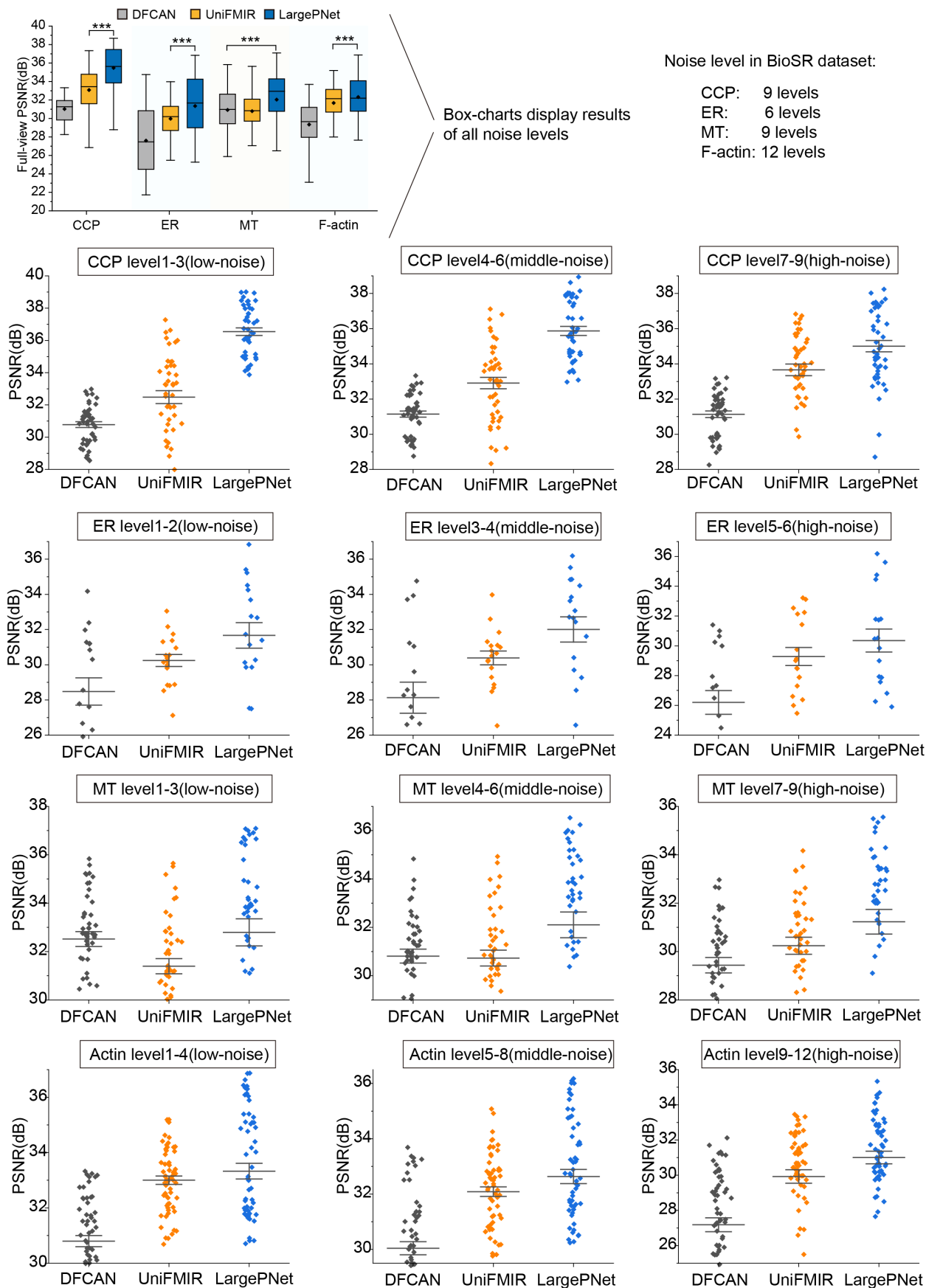


Supplementary Fig. 6. 128×128 Patch-wise metrics on the BioSR dataset. **a.** Statistical comparisons of the PSNR(dB), MS-SSIM, and NRMSE metrics of the images recovered by different networks ($n=400, 703, 1467, 2042$ test images for CCP, ER, MT, and F-actin). The evaluated patches have an average photon count of 25-600. **b, c.** Representative inference regions of microtubule and CCP structures. **d, e.** Representative inference regions of ER and actin structures under different noise conditions. LR, low-resolution. **f-i.** Intensity profiles along the two arrowheads in b-e. Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: $2 \mu\text{m}$ (b, c), $5 \mu\text{m}$ (d, e). (***, $p < 0.001$; Obtained by two-sided t-test; No adjustment for multiple comparisons was applied. The exact p values can be found in Supplementary Table 3).

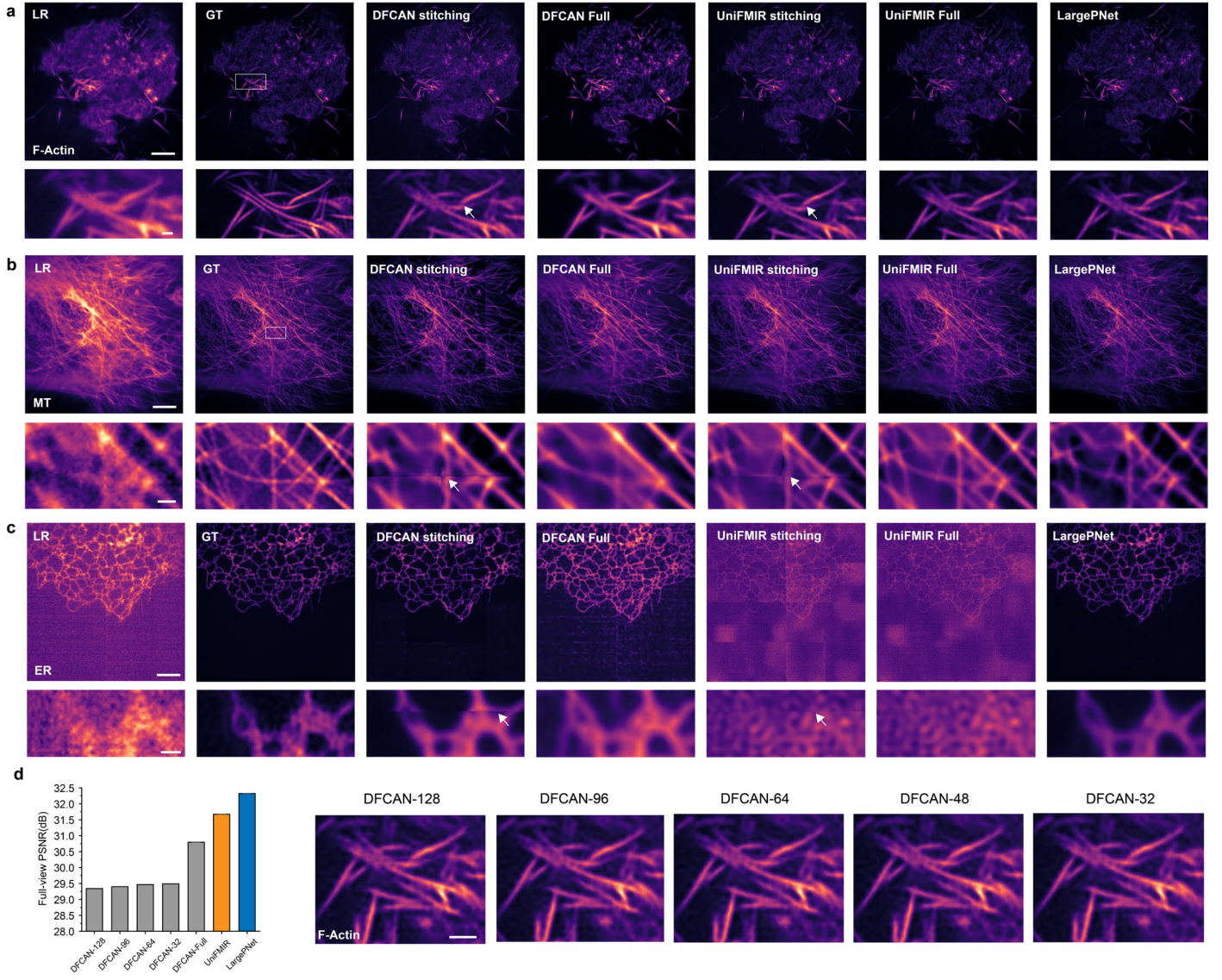
Comparisons of different architectures when inferring full-view image

	Boundary effect	Regional receptive field	Global information utilization	Computational cost
Conventional CNN	Exist	Small	Full-infer: Yes but with mismatch Patch-infer: None	Full-infer: Low Patch-infer(sliding): High
Swin Transformer	None	Large	None	High
LargePNet	None	Large	Yes	Low

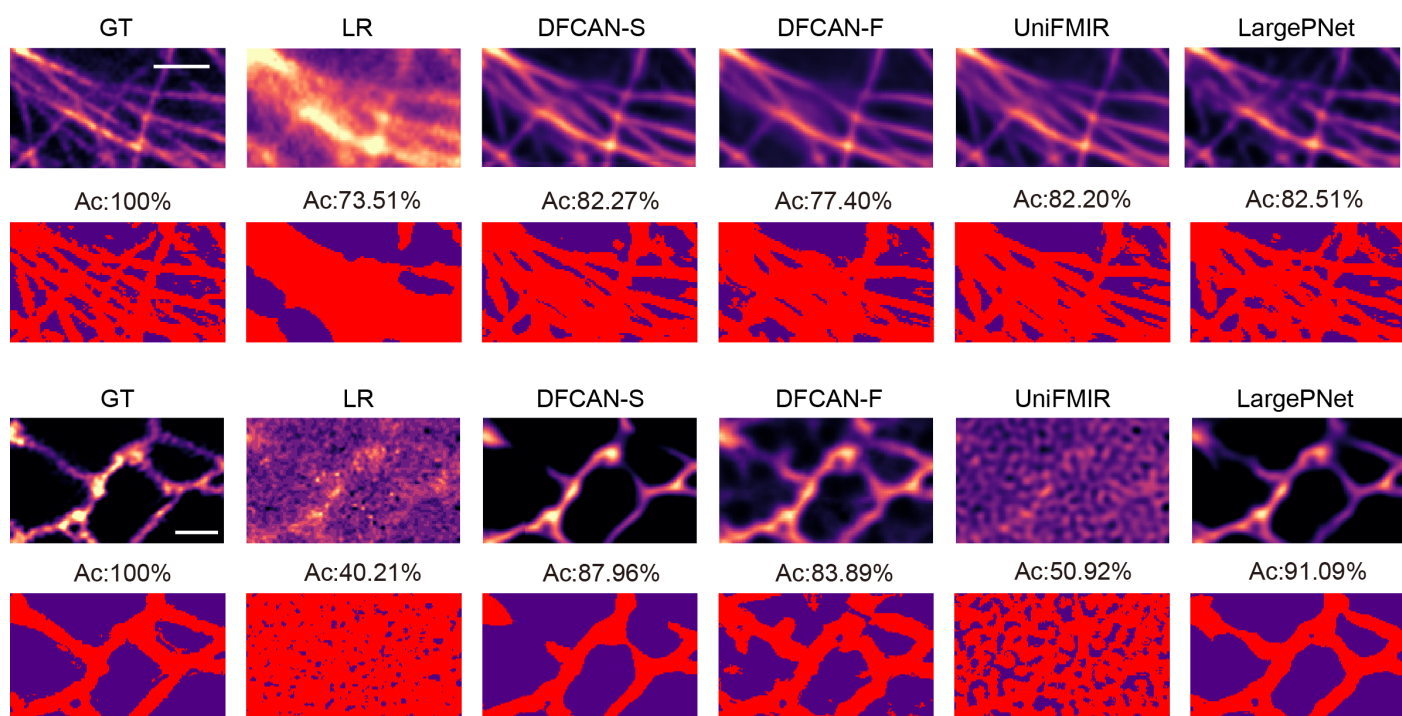
Supplementary Fig. 7. Comparisons of LargeNet with mainstream CNN and Swin Transformer based fluorescence image restoration architectures when inferring full-view images.



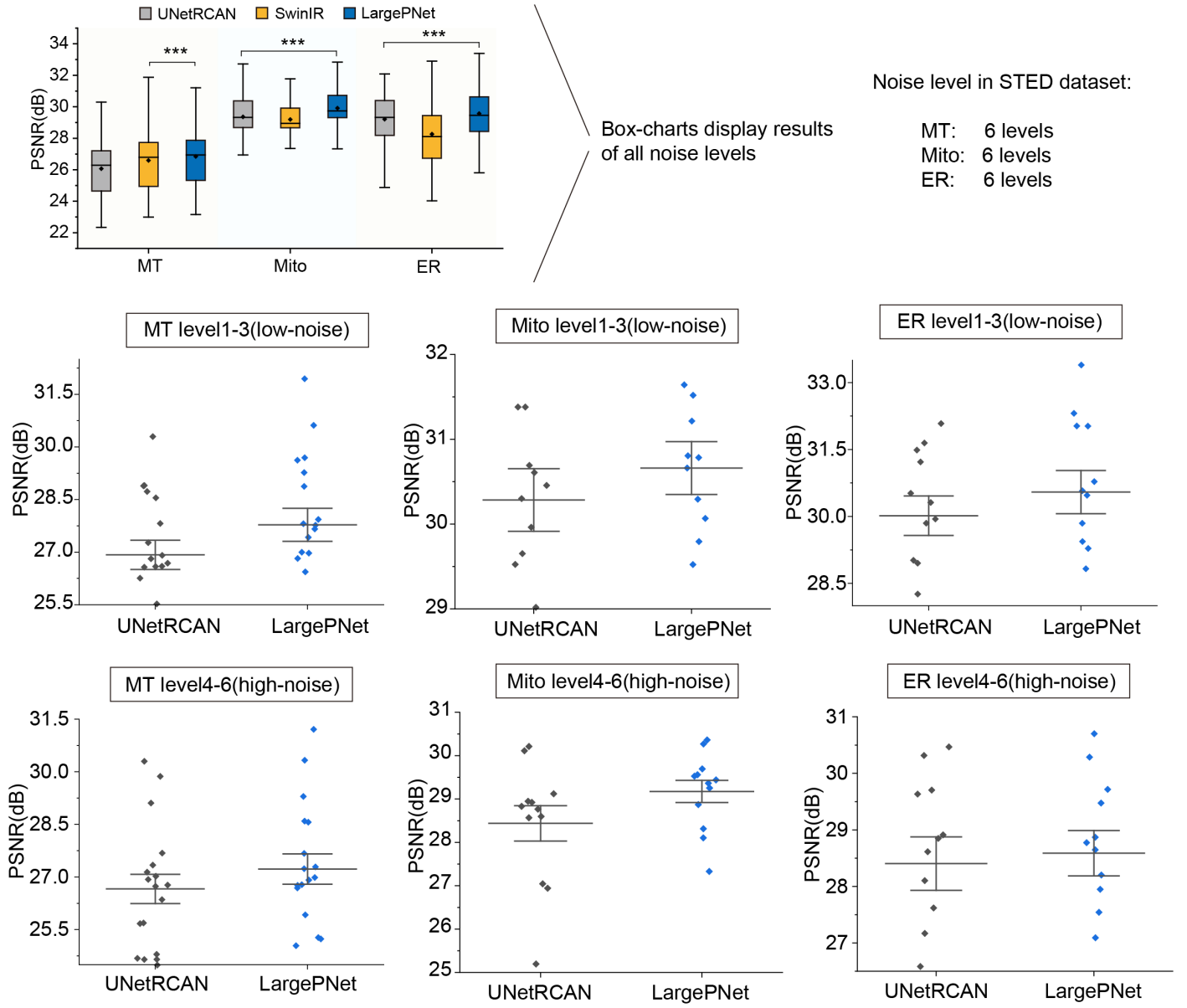
Supplementary Fig. 8. Statistical comparisons in the BioSR super-resolution dataset with separated noise levels. Each point displays the metric of an image. The error bar displays the average value $\pm$ SEM. ($n = 45, 38, 45, 60$ test images in the dataset for CCP, ER, MT, and F-actin in each separated noise level).



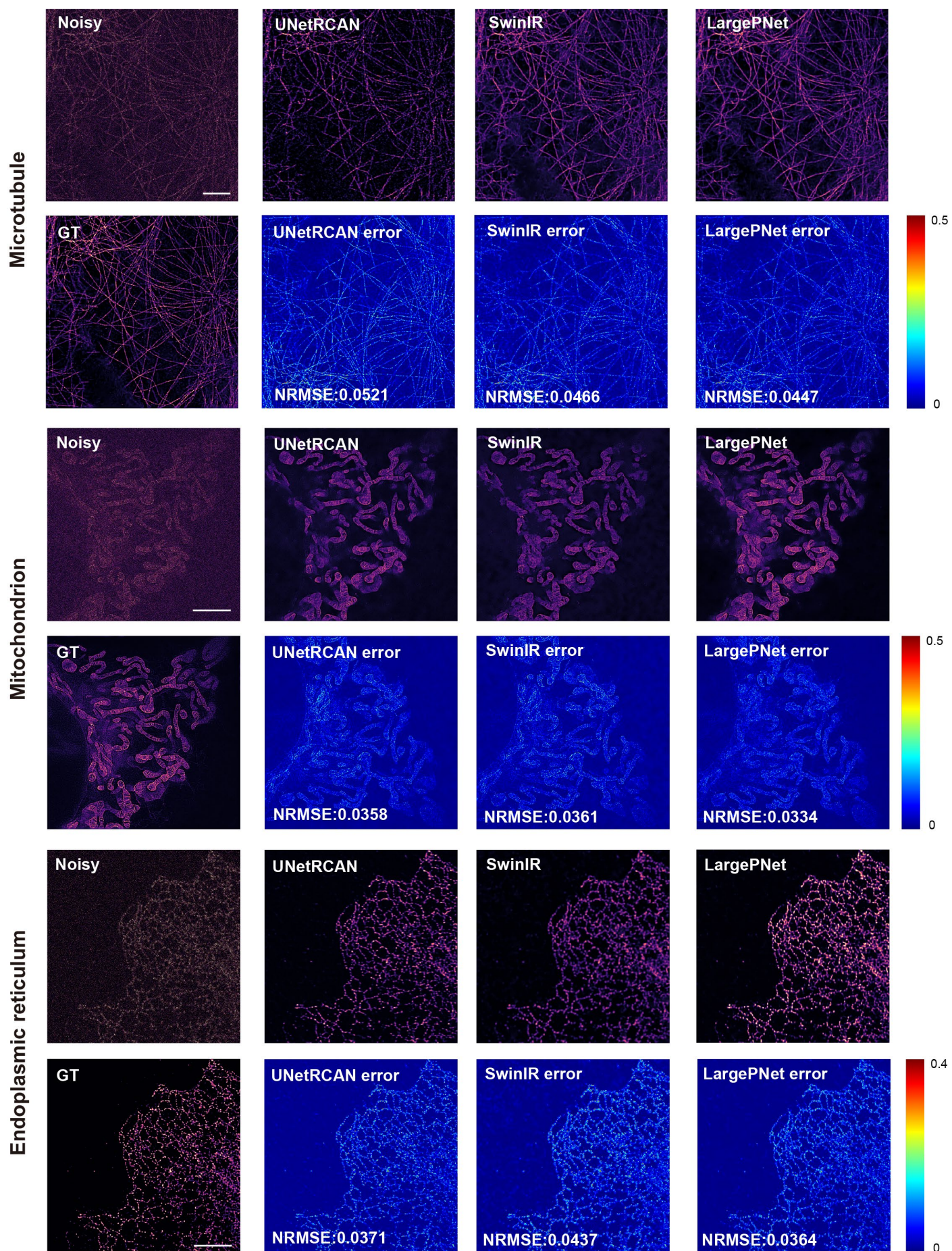
Supplementary Fig. 9. Additional examples of the BioSR dataset results. **a, b, c.** Additional inference results of DFCAN and UniFMIR with the stitching and full-inference schemes. **d.** The PSNR value of DFCAN adopting different overlapping methods using the stitching scheme to merge a full-size image. Scale bars: 5 μm (a, b, c Top), 1 μm (a, b, c Bottom), 2 μm (d).



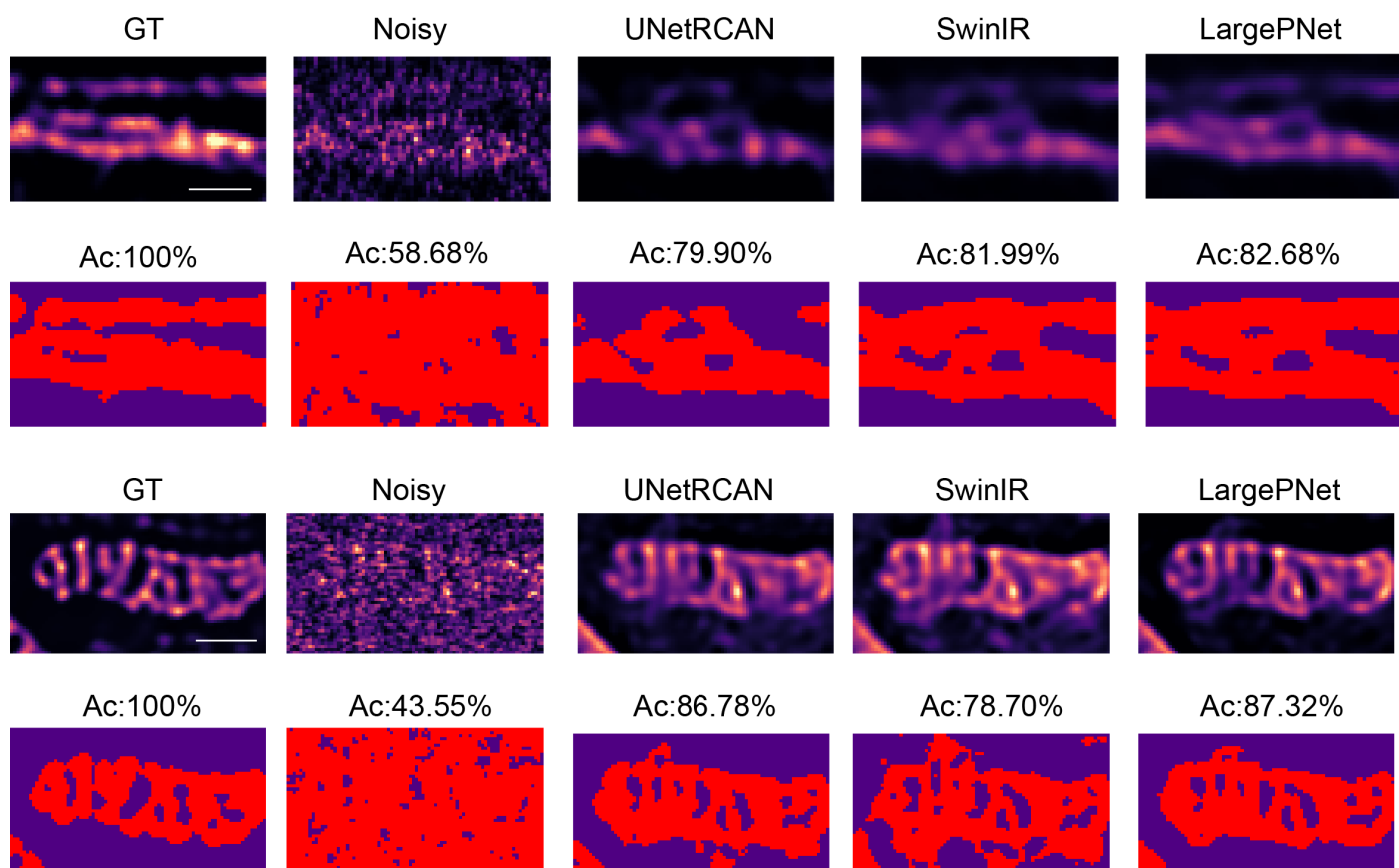
Supplementary Fig. 10. Segmentation accuracy of the single-image super-resolution results. Scale bars: 1 μm .



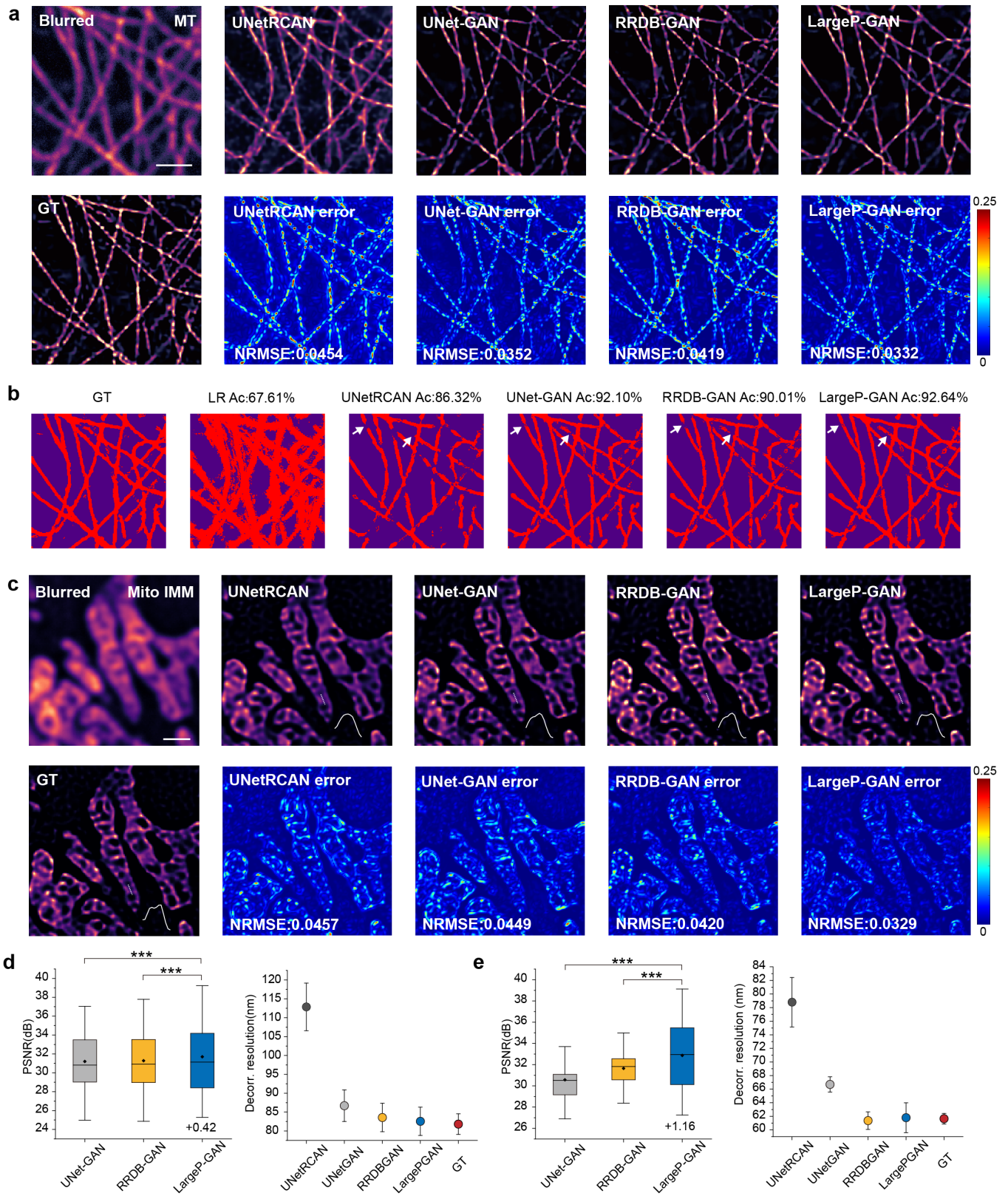
Supplementary Fig. 11. Statistical comparisons in the STED denoising dataset with separated noise levels. Each point displays the metric of an image. The error bar displays the average value $\pm$ SEM. ($n = 36, 12, 12$ test images in the dataset for MT, Mito, and ER in each separated noise level).



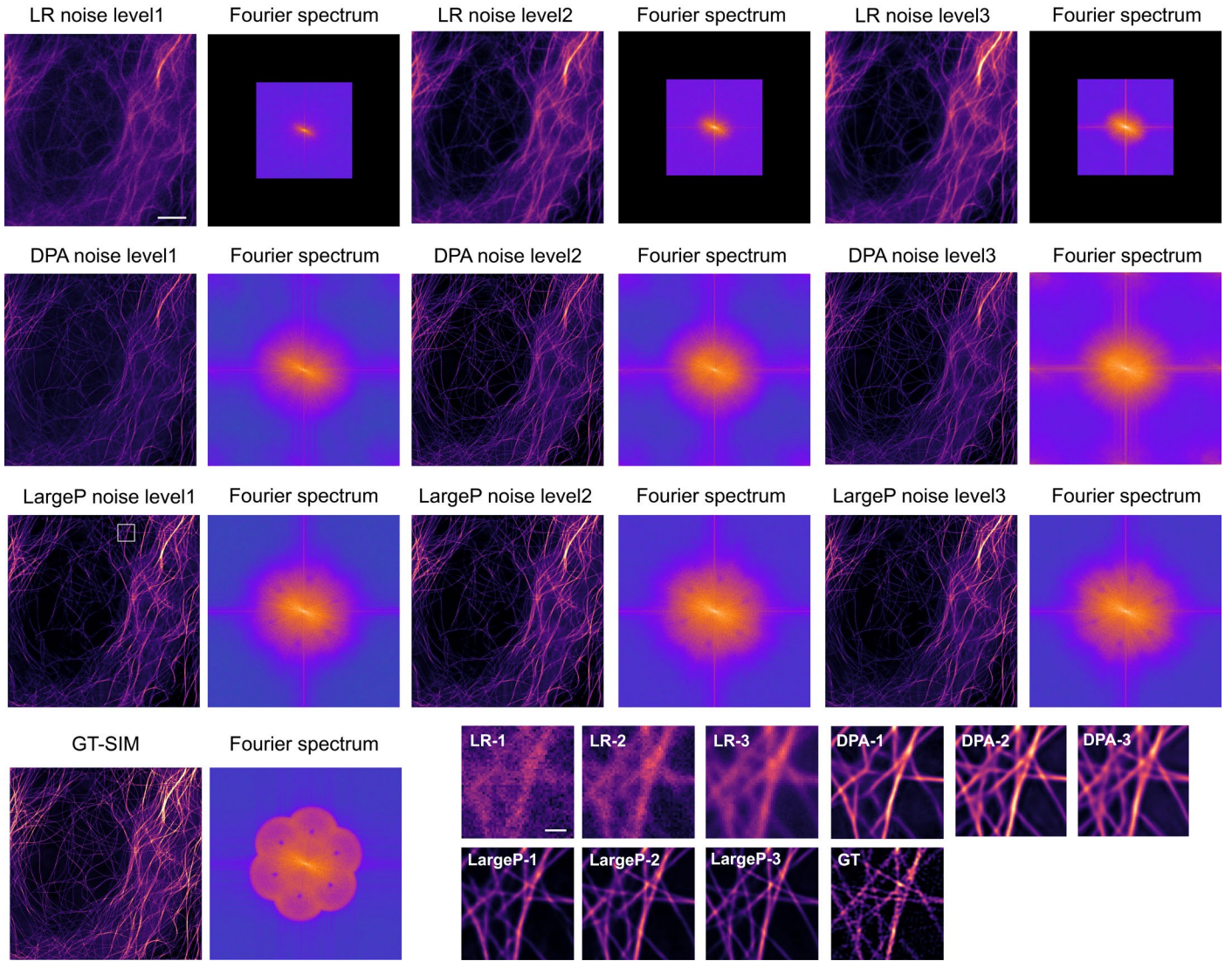
Supplementary Fig. 12. Additional examples of the self-established STED denoising dataset. Scale bars: $5\ \mu\text{m}$.



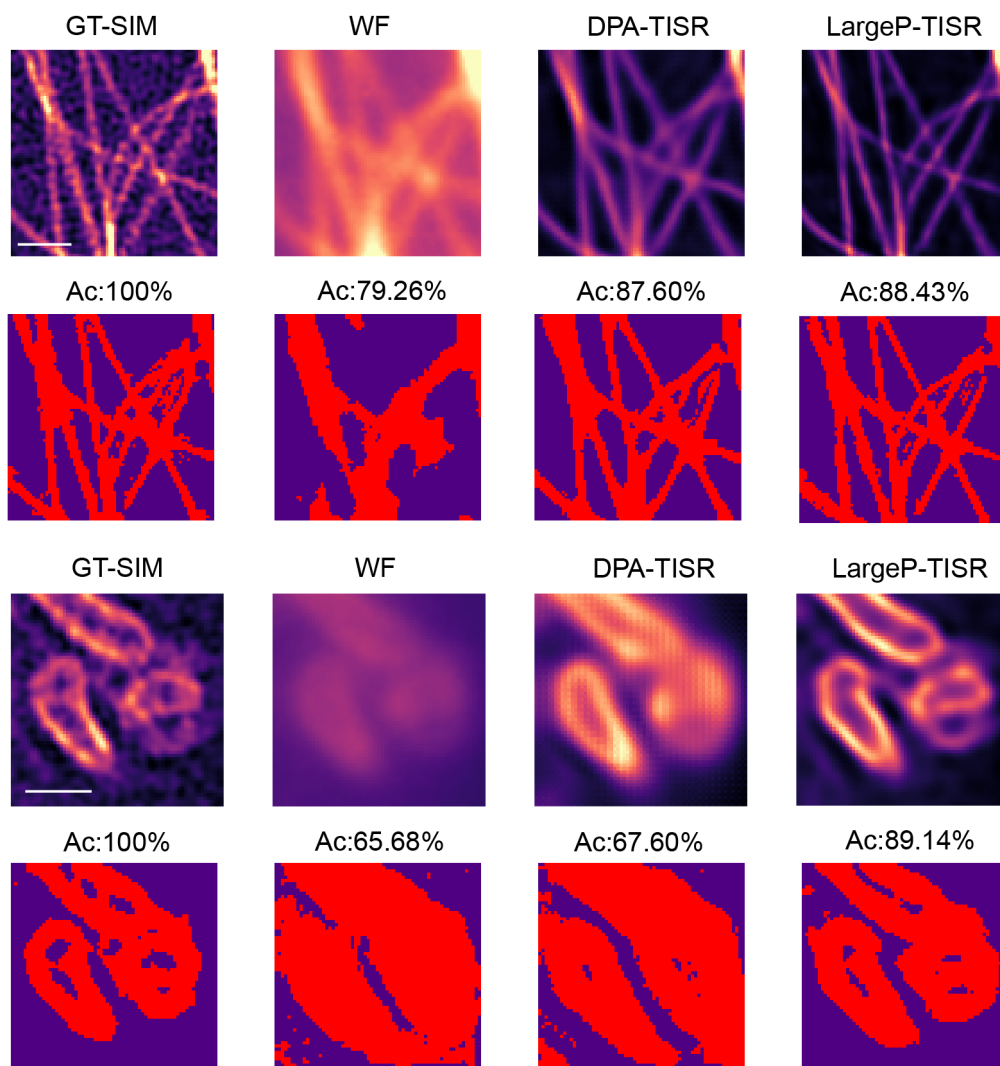
Supplementary Fig. 13. Segmentation accuracy of the STED denoising results. Scale bars: 0.5 μm .



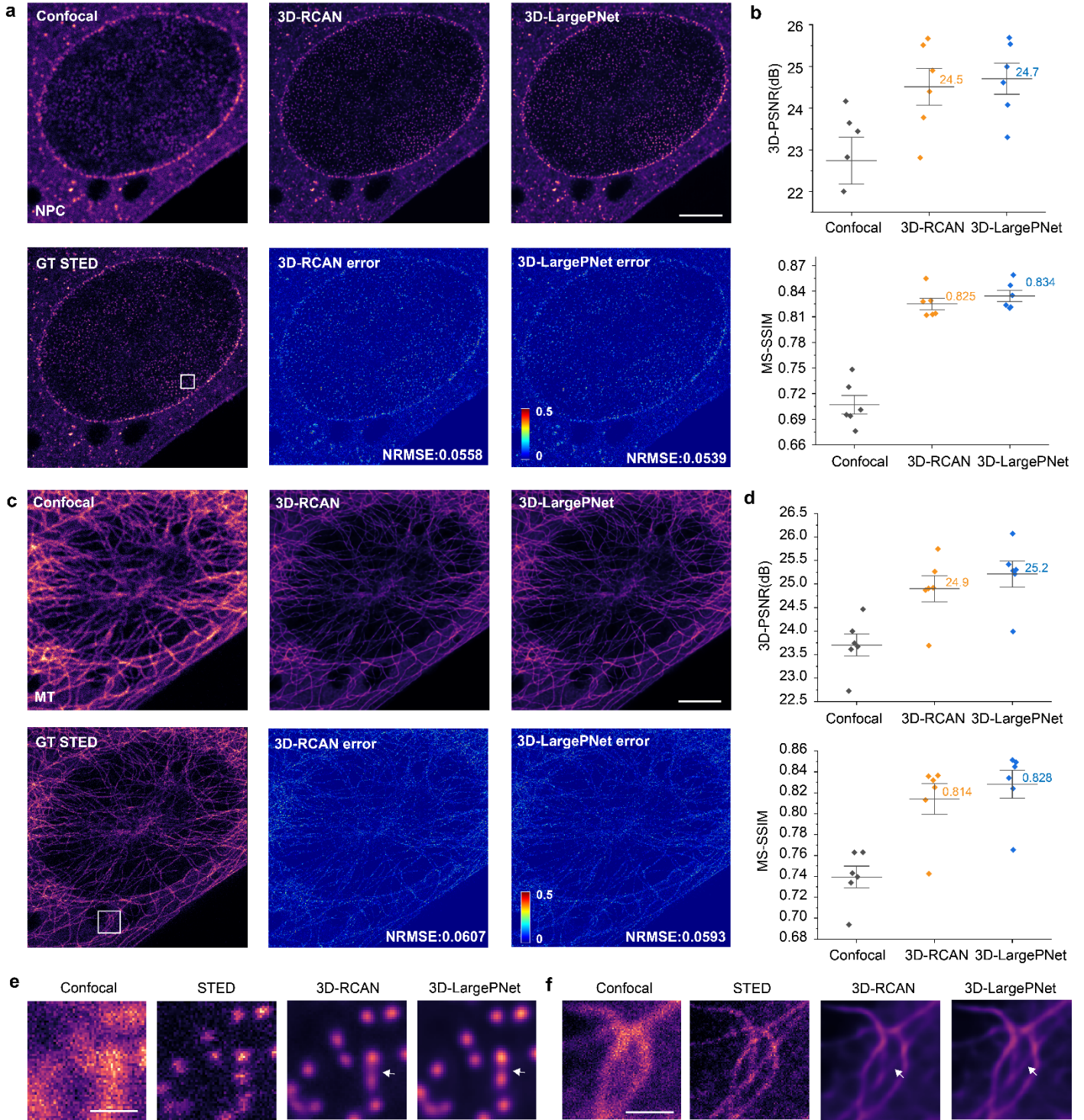
Supplementary Fig. 14. Application of LargeP-GAN for transforming a single blurred STED image to a super-resolution STED image. **a.** A representative inference result of microtubule structure. **b.** Segmentation results of a using the TWS toolbox. **c.** A representative inference result of the mitochondria inner membrane structure (IMM). **d, e.** The PSNR and decorrelation resolution of the two structures ($n = 144$ for MT, $n = 48$ for Mito IMM). Full statistics, together with the ER structure results, are shown in Supplementary Table 5. Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: $1 \mu\text{m}$ (a, c). (***, $p < 0.001$; Obtained by two-sided t-test; No adjustment for multiple comparisons was applied. The exact p values can be found in Supplementary Table 5).



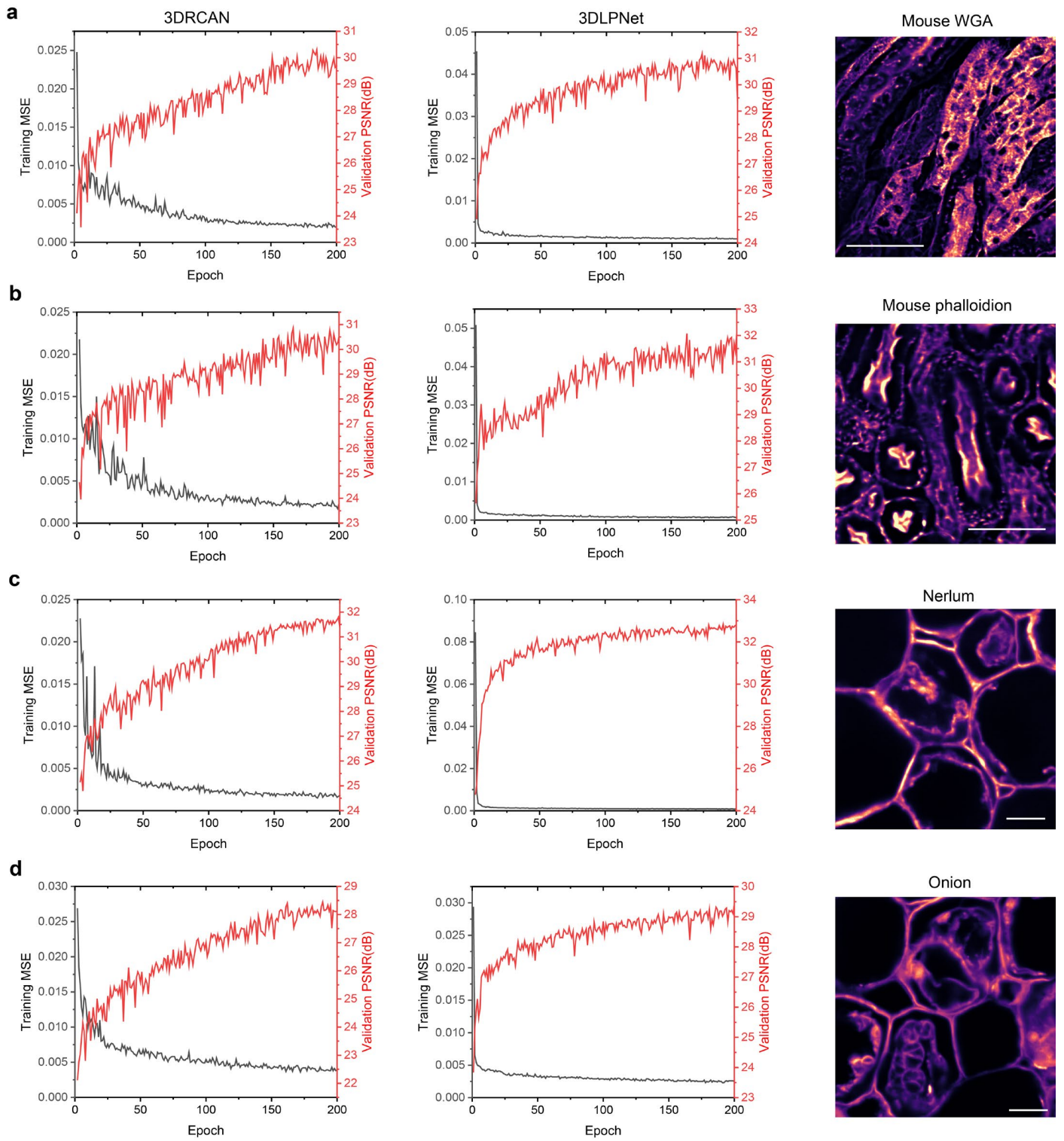
Supplementary Fig. 15. Inference results of DPA-TISR and LargeP-TISR under different noise levels of the Bio-TISR dataset. The results show that LargeP-TISR shows higher robustness under different noise levels. LR, low-resolution. Scale bars: 5 μm (full-view graph), 1 μm (sub-graph).



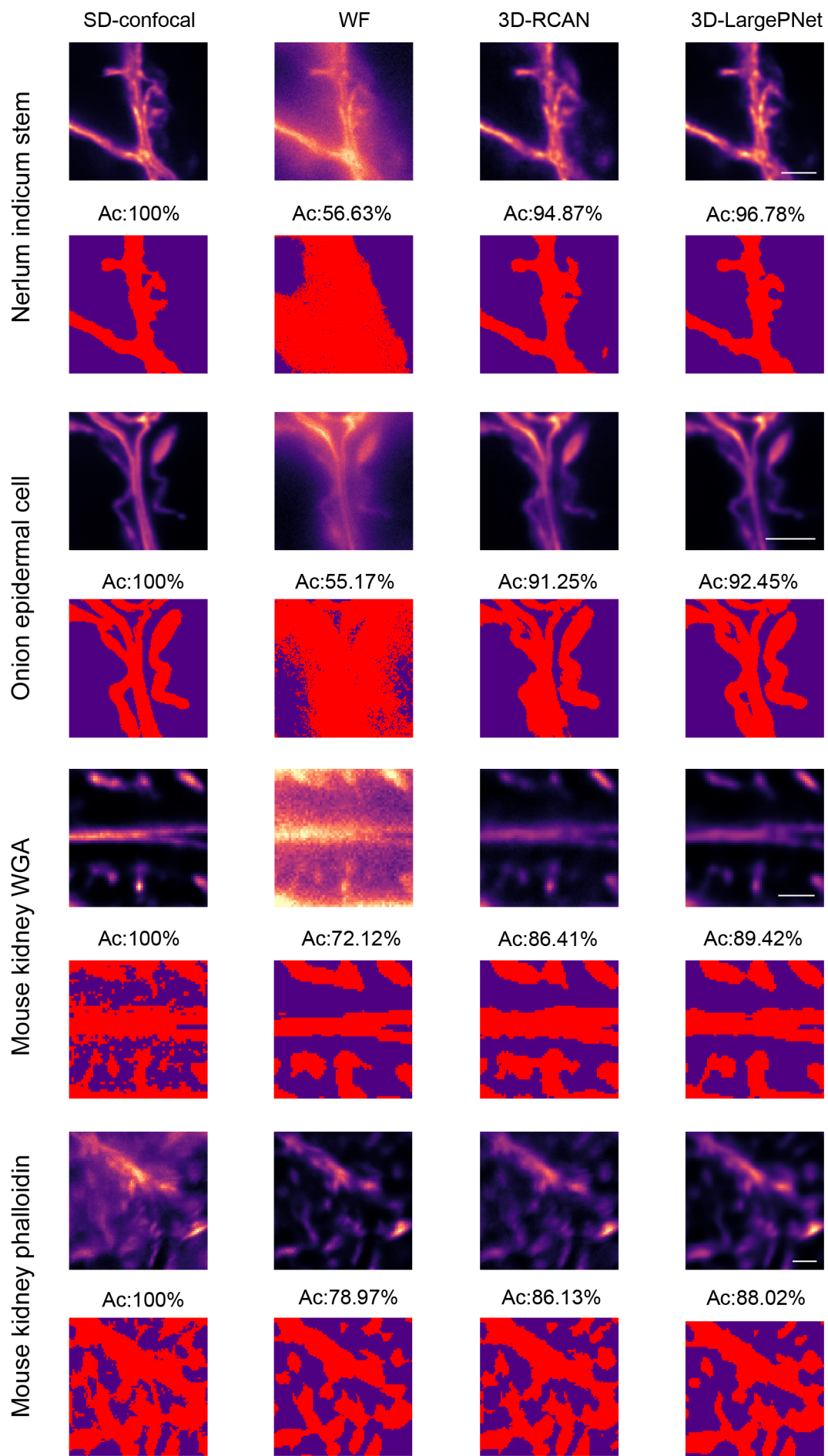
Supplementary Fig. 16. Segmentation accuracy of the video super-resolution results. Scale bars: 1 μm .



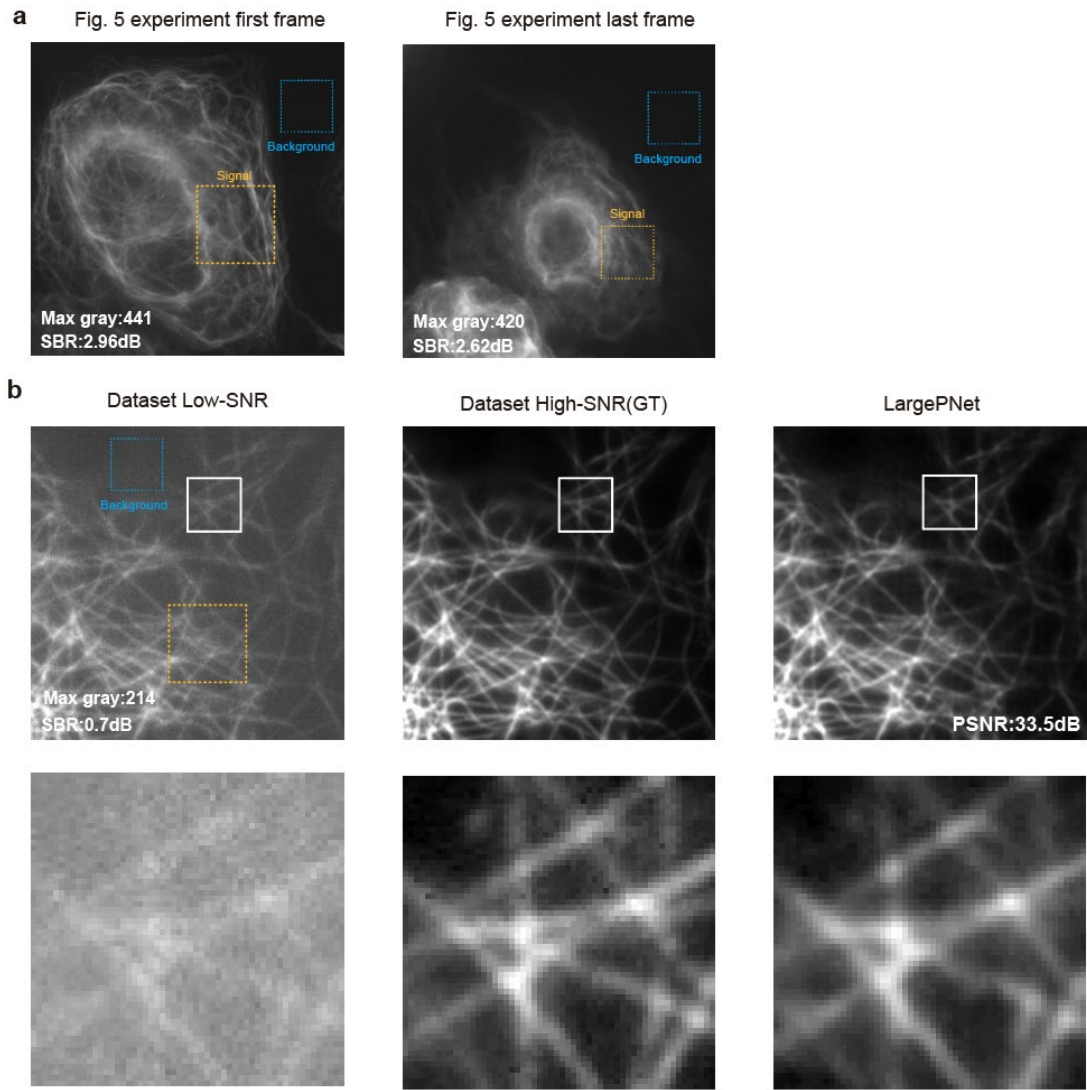
Supplementary Fig. 17. Comparison of 3D-LargePNet and 3D-RCAN on the volumetric transformation of confocal stack to STED stack. **a. c.** Representative deburring results of the confocal images of the nuclear pore complex (NPC) and microtubule structure. **b. d.** The quantitative metrics of the NPC and microtubule structure. The error bar displays the average value $\pm$ SEM. ($n=6$ test images). **e. f.** Magnified white boxed region in **a** and **c**, respectively. Scale bars: $5\ \mu\text{m}$ (**a, c**), $0.5\ \mu\text{m}$ (**e**), $1\ \mu\text{m}$ (**f**).



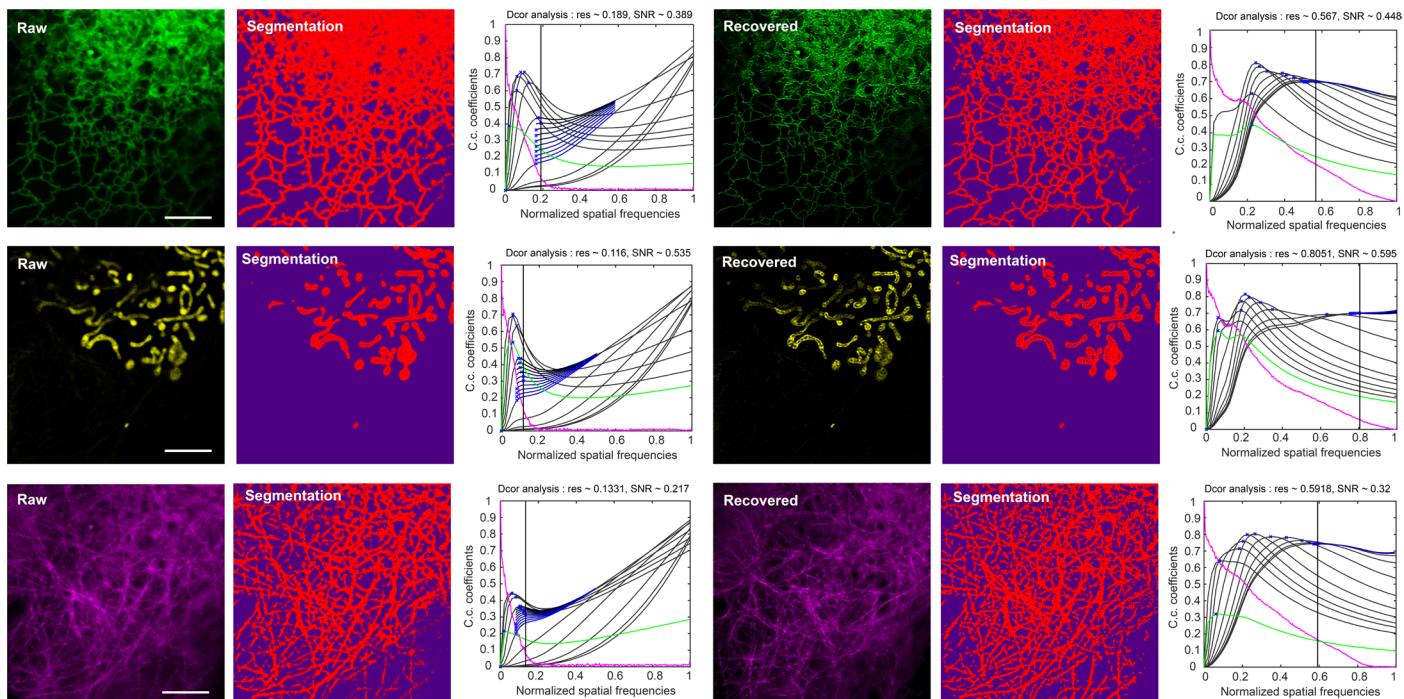
Supplementary Fig. 18. Evolution of the training MSE and validation PSNR of 3D-RCAN and 3D-LargePNet in the widefield to confocal transforming tasks of a. Mouse kidney WGA, b. Mouse kidney phalloidin, c. Nerlun indicum stem and d. Onion epidermal cell. Scale bars: 20 μm .



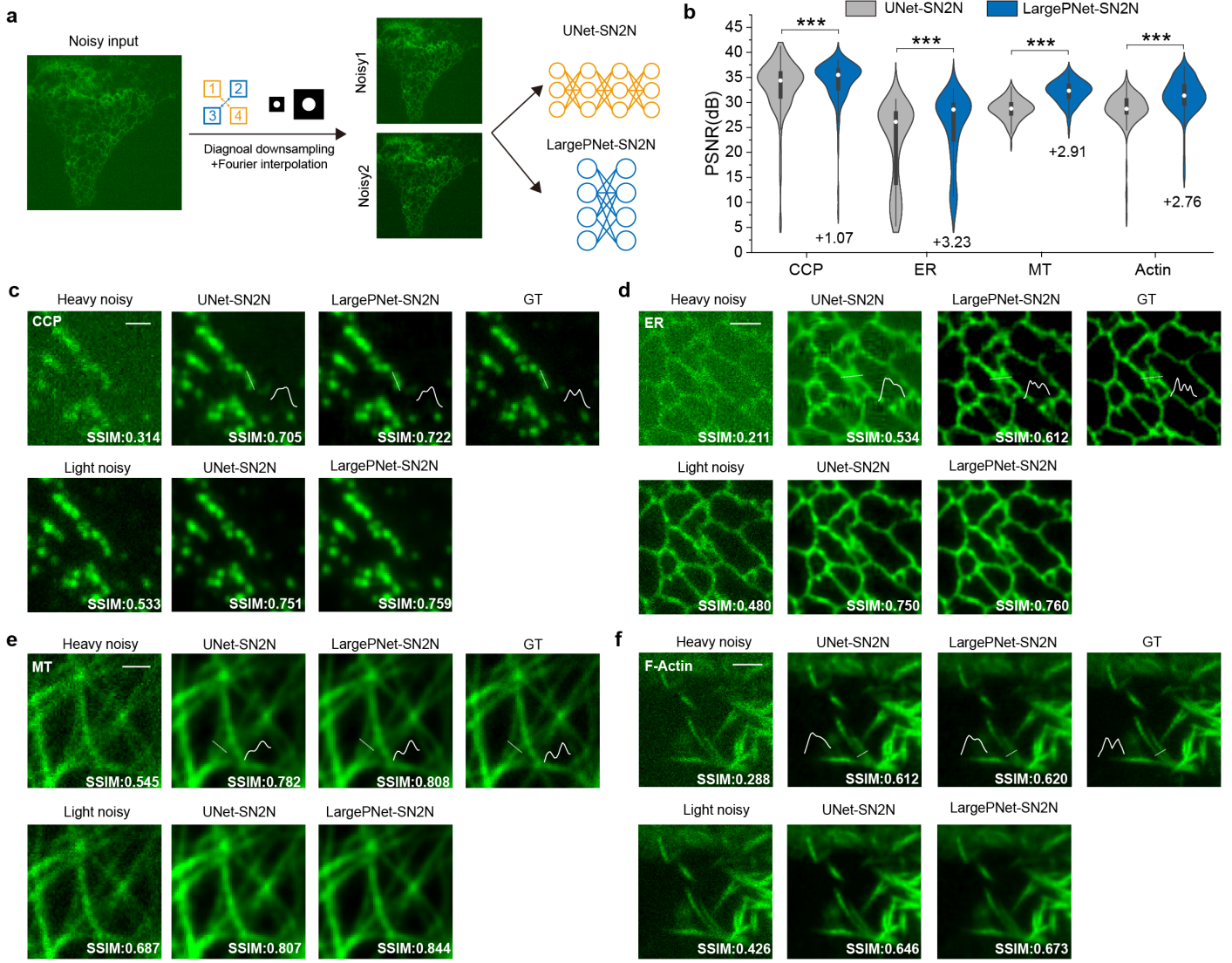
Supplementary Fig. 19. Segmentation accuracy of the background removal results. Scale bars: 5 μm .



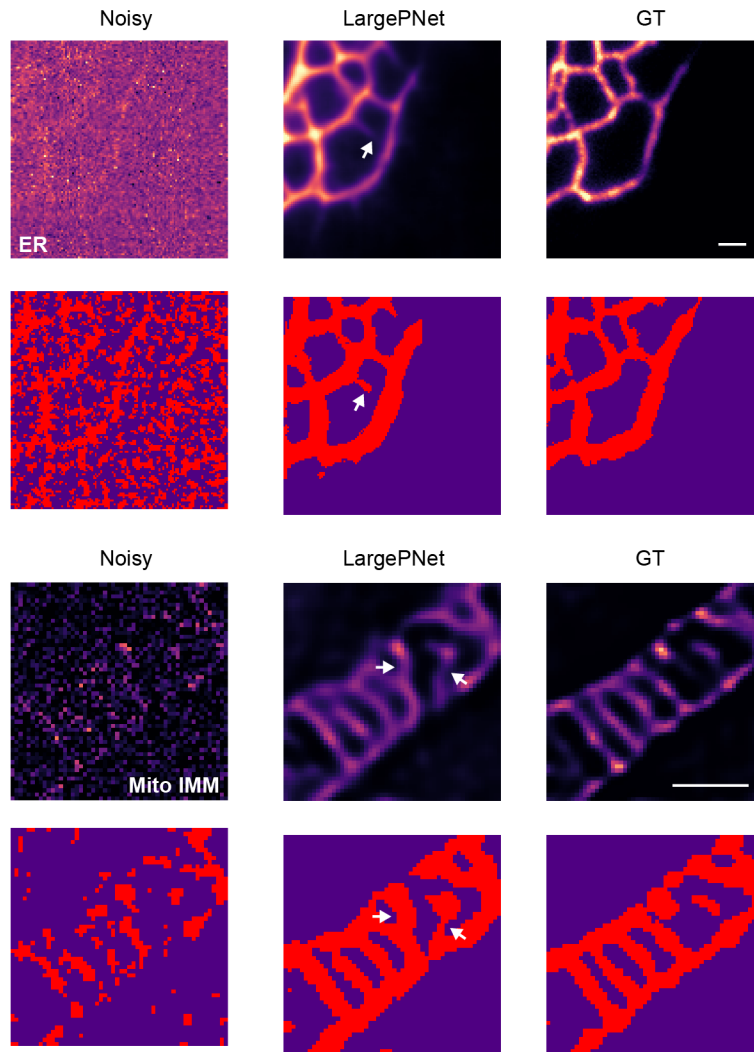
Supplementary Fig. 20. Validation of the model recovery capability in Fig. 5 by inspecting the model performance on the test dataset with a lower signal intensity. a. The image in the experiment in Fig. 5. The orange box is used as signal sampling region, the blue box is used as the background sampling region. **b.** A test image in the BioSR dataset. The magnified boxed regions are shown below.



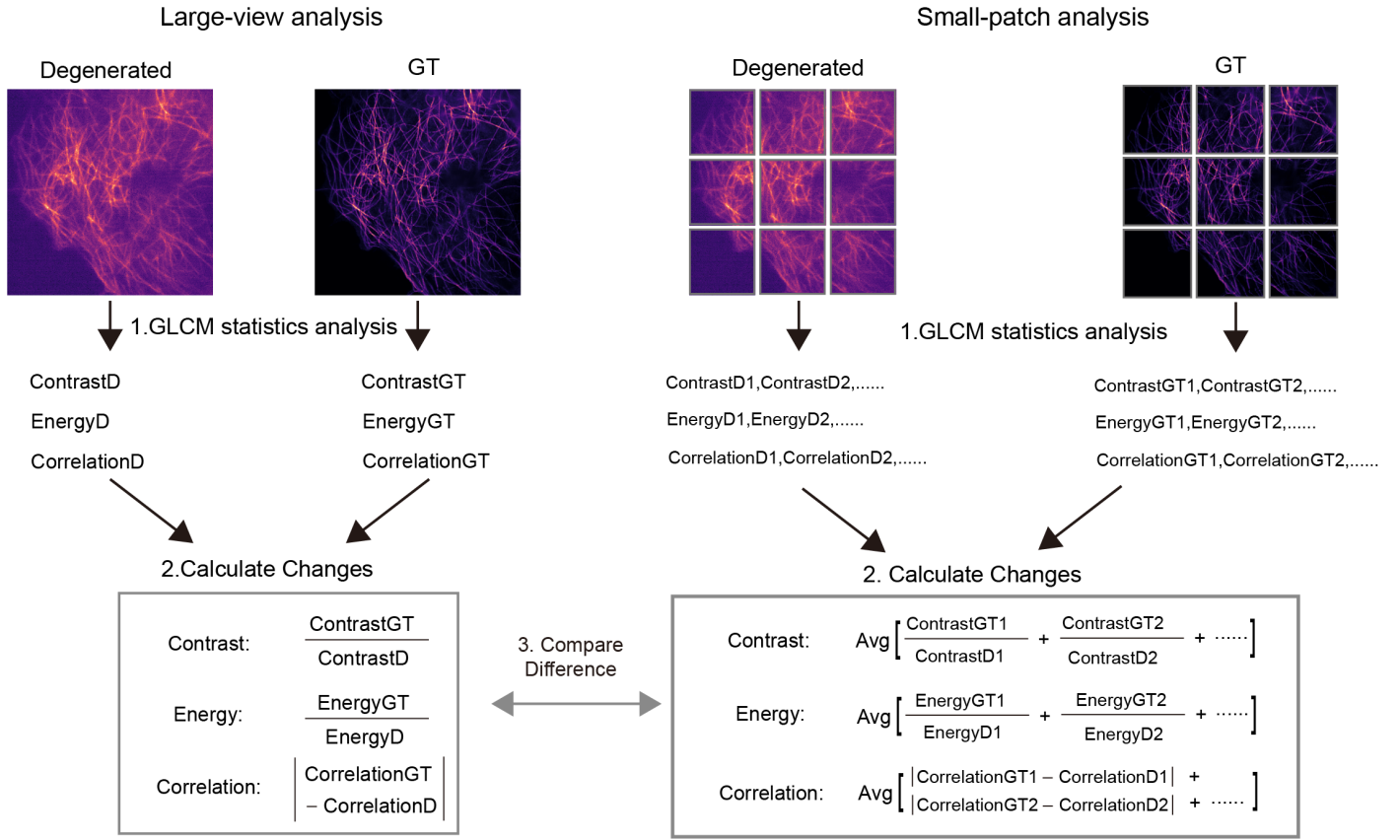
Supplementary Fig. 21. Segmentation and decorrelation analysis results of the LargeP-GAN-recovered STED ER, mitochondria, and microtubule images. Scale bars: 5 μm .



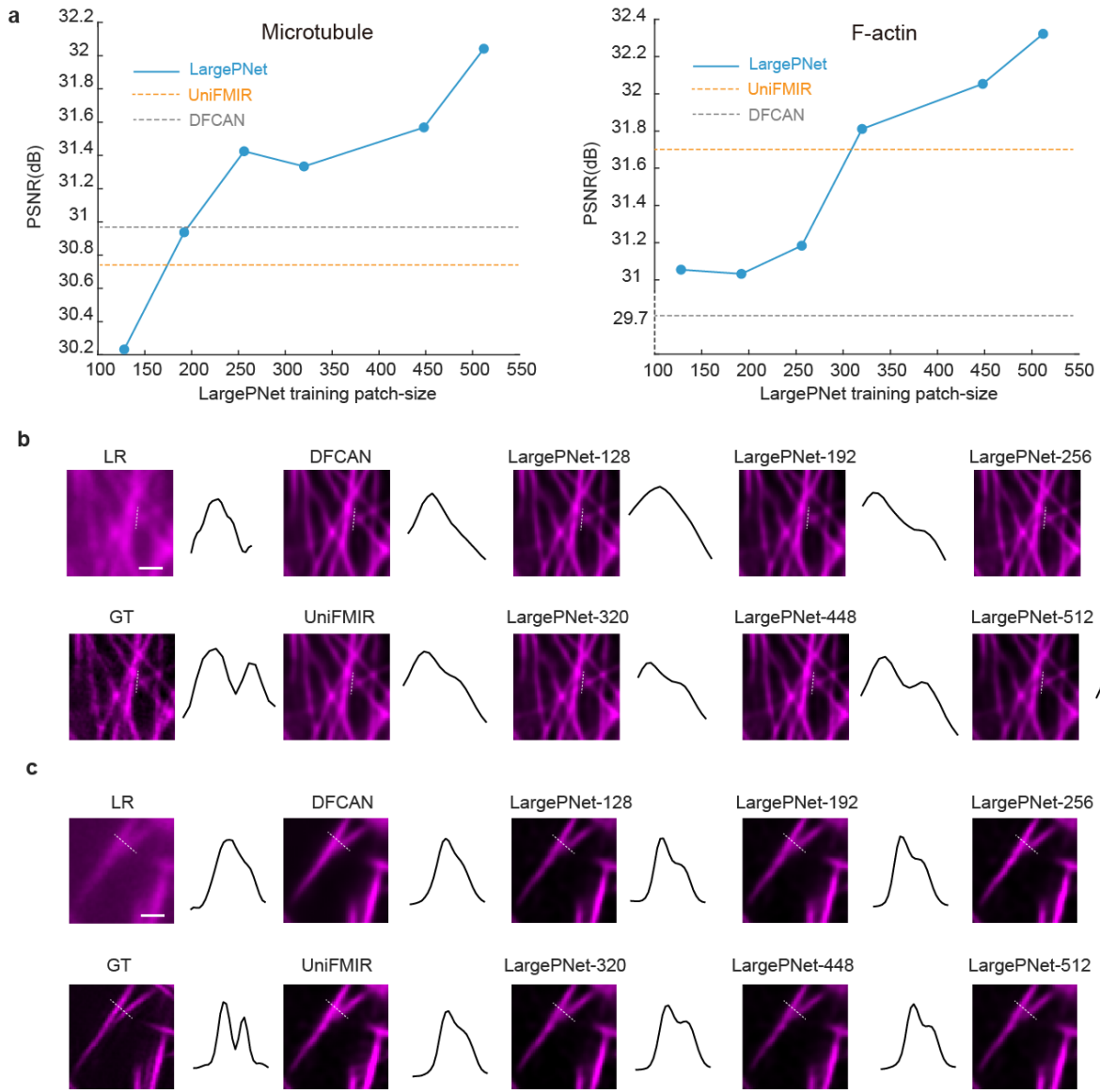
Supplementary Fig. 22. LargePNet improves the performance of the self-supervised image denoising framework. **a.** Overview of the principle for self-inspired noise2noise denoising framework. **b.** Performance of UNet-SN2N and LargePNet-SN2N on denoising the widefield images of four typical structures in the BioSR dataset ($n = 135, 114, 135, 180$ test images for CCP, ER, MT, and F-actin). **c-f.** Example instances of the denoising performance on CCP, ER, MT, and F-actin at two SNR levels. Violin plot: shape, data density distribution; centerline, median; point, mean; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: $1 \mu\text{m}$ (c, e), $2 \mu\text{m}$ (d, f). (***, $p < 0.001$; Obtained by two-sided t-test; No adjustment for multiple comparisons was applied. The exact p values can be found in Supplementary Table 9).



Supplementary Fig. 23. When the noise is too strong, there may be hallucinations in LargePNet restoration. The white arrows denote the example hallucination region. Scale bars: 0.5 μm .



Supplementary Fig. 24. Calculation process for explaining the structure-dependent advantages of LargePNet. GLCM: Gray Level Co-occurrence Matrix.



Supplementary Fig. 25. Investigation of the LargePNet performance when only small ROIs are available for training. a. The PSNR metrics of microtubule and F-actin structures in BioSR datasets. **b.** Instance inference results of the microtubule structures. **c.** Instance inference results of the F-actin structures. Scale bars: $0.5\ \mu\text{m}$.

Supplementary Tables

Supplementary Table 1. Comparisons of the training patch-size of LargePNet and other competitor models in different restoration tasks.

Task	Our model	Patch-size	Competitor models	Patch-size
WF image denoising	LargePNet	512×512	UNet	128×128
			RCAN	128×128
WF image SR 2×	LargePNet	512×512	DFCAN	128×128
			UniFMIR	128×128
STED image denoising	LargePNet	512×512	UNetRCAN	256×256
			SwinIR	128×128
SN2N denoising	SN2N-LargePNet	512×512	SN2N-UNet	128×128
WF Video SR 2×	LargeP-TISR	7×512×512	DPA-TISR	7×128×128
3D CF to STED	3D-LargePNet	6×640×640	3D-RCAN	6×320×320
3D WF to CF	3D-LargePNet	16×512×512	3D-RCAN	16×256×256
GAN-based restoration	LargeP-GAN	512×512	UNetGAN	128×128
			RRDBGAN	128×128

Supplementary Table 2. Metrics of UNet and LargePNet trained with different patch-sizes when denoising the ER images in the BioSR dataset

Model-Patch size	Metric	Low noise level	Heavy noise level
UNet-128	PSNR (dB)	22.7026±5.8357	19.2566±8.8340
	MS-SSIM	0.7862±0.1331	0.6491±0.1787
UNet-256	PSNR (dB)	21.0630±6.8150	17.3831±9.5159
	MS-SSIM	0.7464±0.2598	0.5848±0.2022
UNet-512	PSNR (dB)	19.6604±5.2276	14.0635±10.4483
	MS-SSIM	0.7674±0.1398	0.4653±0.3504
LargePNet-128	PSNR (dB)	24.7223±1.7811	23.0972±2.2451
	MS-SSIM	0.8807±0.0588	0.7238±0.1439
LargePNet-256	PSNR (dB)	26.9687±2.6767	23.6079±2.4132
	MS-SSIM	0.8908±0.0575	0.7382±0.1436
LargePNet-512	PSNR (dB)	27.5225±3.0077	24.4194±2.8614
	MS-SSIM	0.8991±0.0568	0.7549±0.1339

Supplementary Table 3. Metrics of DFCAN, UniFMIR, and LargePNet on BioSR single image 2x super-resolution task

128×128 Patch-wise metrics					
Model	Metric	CCPs	ER	Microtubules	F-Actin
DFCAN	PSNR (dB)	27.9994±2.0198	27.2007±5.9145	30.9399±6.1196	30.8532±4.2354
	MS-SSIM	0.9244±0.0322	0.9402±0.0481	0.9427±0.0383	0.9376±0.0388
	NRMSE	0.0485±0.0140	0.0761±0.0394	0.0546±0.0328	0.0509±0.0269
UniFMIR	PSNR (dB)	30.4001±2.9447	29.7500±3.8662	31.7795±6.2126	32.8834±3.6482
	MS-SSIM	0.9522±0.0264	0.9591±0.0235	0.9535±0.0298	0.9447±0.0350
	NRMSE	0.0371±0.0136	0.0569±0.0311	0.0490±0.0250	0.0392±0.0168
LargePNet	PSNR (dB)	31.5318±3.4794	31.1994±5.7858	32.9335±6.8884	33.5019±4.1718
	MS-SSIM	0.9606±0.0255	0.9717±0.0176	0.9606±0.0405	0.9516±0.0313
	NRMSE	0.0323±0.0105	0.0478±0.0255	0.0450±0.0347	0.0369±0.0171
Significance of LargePNet improvement through t-test					
Model	Metric	CCPs	ER	Microtubules	F-Actin
Versus DFCAN	PSNR (dB)	$p=1.49\text{e-}140(***)$	$p=3.69\text{e-}167(***)$	$p=1.17\text{e-}115(***)$	$p=4.57\text{e-}235(***)$
	MS-SSIM	$p=2.33\text{e-}91(***)$	$p=8.67\text{e-}129(***)$	$p=8.39\text{e-}215(***)$	$p<2.2\text{e-}308(***)$
	NRMSE	$p=4.05\text{e-}125(***)$	$p=3.05\text{e-}150(***)$	$p=7.46\text{e-}57(***)$	$p=1.22\text{e-}189(***)$
Versus UniFMIR	PSNR (dB)	$p=3.26\text{e-}24(***)$	$p=2.22\text{e-}33(***)$	$p=1.13\text{e-}37(***)$	$p=1.37\text{e-}24(***)$
	MS-SSIM	$p=1.25\text{e-}13(***)$	$p=5.36\text{e-}68(***)$	$p=4.30\text{e-}99(***)$	$p=5.02\text{e-}266(***)$
	NRMSE	$p=2.97\text{e-}22(***)$	$p=8.25\text{e-}27(***)$	$p=8.50\text{e-}12(***)$	$p=1.90\text{e-}17(***)$
512×512 Full-view metrics					
Model	Metric	CCPs	ER	Microtubules	F-Actin
DFCAN	PSNR (dB)	31.0205±1.7055	27.6028±4.8866	30.9200±3.2214	29.3412±3.5111
	MS-SSIM	0.9350±0.0313	0.9429±0.0167	0.9311±0.0577	0.9217±0.0488
	NRMSE	0.0284±0.0058	0.0451±0.0280	0.0296±0.0116	0.0360±0.0151
UniFMIR	PSNR (dB)	33.0813±3.6723	29.9711±2.9006	30.7877±2.7212	31.6725±2.3337
	MS-SSIM	0.9365±0.0403	0.9630±0.0118	0.9504±0.0290	0.9344±0.0313
	NRMSE	0.0231±0.0107	0.0326±0.0117	0.0299±0.0089	0.0273±0.0074
LargePNet	PSNR (dB)	35.8077±2.3331	31.3432±4.5586	32.0415±4.8193	32.3224±3.3113
	MS-SSIM	0.9708±0.0181	0.9699±0.0122	0.9574±0.0286	0.9383±0.0290
	NRMSE	0.0166±0.0044	0.0289±0.0169	0.0276±0.0160	0.0254±0.0089
Significance of LargePNet improvement through t-test					
Model	Metric	CCPs	ER	Microtubules	F-Actin
Versus DFCAN	PSNR (dB)	$p=1.50\text{e-}67(***)$	$p=4.22\text{e-}15(***)$	$p=3.82\text{e-}71(***)$	$p=1.01\text{e-}33(***)$
	MS-SSIM	$p=5.38\text{e-}47(***)$	$p=5.79\text{e-}20(***)$	$p<2.2\text{e-}308(***)$	$p=6.74\text{e-}15(***)$
	NRMSE	$p=1.70\text{e-}70(***)$	$p=1.97\text{e-}6(***)$	$p=4.31\text{e-}17(***)$	$p=1.55\text{e-}4(***)$

Versus UniFMIR	PSNR (dB)	$p=1.49\text{e-}26(***)$	$p=6.41\text{e-}5(***)$	$p=8.54\text{e-}85(***)$	$p=2.23\text{e-}5(***)$
	MS-SSIM	$p=2.12\text{e-}33(***)$	$p=1.46\text{e-}10(***)$	$p=1.56\text{e-}289(***)$	$p=0.0044(**)$
	NRMSE	$p=7.94\text{e-}24(***)$	$p=1.49\text{e-}15(***)$	$p=8.49\text{e-}27(***)$	$p=6.24\text{e-}27(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 4. Metrics of UNetRCAN, SwinIR and LargePNet on STED image denoising task

Model	Metric	Microtubules	Mito	ER
UNetRCAN	PSNR (dB)	26.0681±2.7485	29.3609±2.3268	29.2095±2.3496
	MS-SSIM	0.7241±0.1283	0.8785±0.0517	0.8655±0.0443
	NRMSE	0.0509±0.0173	0.0346±0.0099	0.0353±0.0100
SwinIR	PSNR (dB)	26.5923±2.9001	29.1935±2.0527	28.2601±3.4403
	MS-SSIM	0.7269±0.1412	0.8687±0.0511	0.7609±0.1490
	NRMSE	0.0481±0.0173	0.0352±0.0078	0.0398±0.0138
LargePNet	PSNR (dB)	26.8437±2.7108	29.9172±1.6248	29.5671±2.4895
	MS-SSIM	0.7408±0.1307	0.8815±0.0366	0.8800±0.0570
	NRMSE	0.0466±0.0155	0.0322±0.0063	0.0339±0.0090
Significance of LargePNet improvement through t-test				
Model	Metric	Microtubules	Mito	ER
Versus UNetRCAN	PSNR (dB)	$p=1.09\text{e-}17(***)$	$p=7.50\text{e-}6(***)$	$p=1.56\text{e-}4(***)$
	MS-SSIM	$p=2.03\text{e-}18(***)$	$p=0.1754(\text{NS})$	$p=6.19\text{e-}8(***)$
	NRMSE	$p=5.21\text{e-}17(***)$	$p=1.98\text{e-}4(***)$	$p=1.53\text{e-}4(***)$
Versus SwinIR	PSNR (dB)	$p=1.92\text{e-}13(***)$	$p=9.46\text{e-}8(***)$	$p=4.78\text{e-}13(***)$
	MS-SSIM	$p=2.60\text{e-}6(***)$	$p=0.0013(**)$	$p=3.77\text{e-}5(***)$
	NRMSE	$p=1.09\text{e-}12(***)$	$p=6.39\text{e-}10(***)$	$p=1.63\text{e-}6(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 5. Metrics of UNet-GAN, RRDB-GAN and LargeP-GAN on STED image deblurring task

Model	Metric	Microtubules	Mito	ER
UNet-GAN	PSNR (dB)	31.2023±3.4856	31.2392±2.8824	32.2368±2.4157
	Decorr. Resolution (nm)	86.6893	66.6949	99.0295
RRDB-GAN	PSNR (dB)	31.2811±4.2300	31.7111±2.2386	31.2911±3.7552
	Decorr. Resolution (nm)	83.5607	61.3701	80.6322
LargeP-GAN	PSNR (dB)	31.6990±5.8926	32.8693±3.8383	32.7862±2.7386
	Decorr. Resolution (nm)	82.5585	61.7990	83.8082
Significance of LargePNet improvement through t-test				
Model	Metric	Microtubules	Mito	ER
Versus UNet-GAN	PSNR (dB)	$p=3.27\text{e-}4(***)$	$p=1.06\text{e-}12(***)$	$p=8.70\text{e-}4(***)$
Versus RRDB-GAN	PSNR (dB)	$p=3.83\text{e-}4(***)$	$p=1.52\text{e-}6(***)$	$p=2.26\text{e-}4(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 6. Metrics of UNet-GAN, RRDB-GAN and LargeP-GAN on the restoration of undersampled SMLM image

Model	Metric	Microtubules
UNet-GAN	PSNR (dB)	24.7897±3.2976
	MS-SSIM	0.7461±0.1587
	NRMSE	0.0594±0.0200
	FID	186.3
RRDB-GAN	PSNR (dB)	24.0293±2.4831
	MS-SSIM	0.7310±0.1490
	NRMSE	0.0640±0.0167
	FID	139.0
LargeP-GAN	PSNR (dB)	25.5779±3.6021
	MS-SSIM	0.7517±0.1618
	NRMSE	0.0551±0.0246
	FID	107.8
Significance of LargePNet improvement through t-test		
Model	Metric	Microtubules
Versus UNet-GAN	PSNR (dB)	$p=2.38\text{e-}13(***)$
	MS-SSIM	$p=0.0021(**)$
	NRMSE	$p=2.68\text{e-}12(***)$
Versus RRDB-GAN	PSNR (dB)	$p=1.02\text{e-}19(***)$
	MS-SSIM	$p=9.13\text{e-}23(***)$
	NRMSE	$p=2.41\text{e-}22(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 7. Metrics of DPA-TISR and LargeP-TISR on video super-resolution task

Model	Metric	Microtubule	Mitochondrion
DPA-TISR	PSNR (dB)	29.4858±2.1480	32.1757±2.3037
	MS-SSIM	0.7817±0.0700	0.8144±0.0505
LargeP-TISR	PSNR (dB)	29.8913±2.6880	33.3537±3.2645
	MS-SSIM	0.8247±0.0808	0.8486±0.0589
Significance of LargePNet improvement through t-test			
Model	Metric	Microtubule	Mitochondrion
Versus DPA-TISR	PSNR (dB)	$p=5.91\text{e-}72(***)$	$p=5.18\text{e-}51(***)$
	MS-SSIM	$p=1.25\text{e-}73(***)$	$p=4.30\text{e-}58(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 8. Metrics of 3D-LargePNet and 3D-RCAN on background removal task

Model	Metric	Nerlum indicum stem	Onion epidermal cell	Mouse kidney WGA	Mouse kidney phalloidin
3D-RCAN	PSNR (dB)	33.1620±2.5921	29.5969±3.2693	31.4888±2.6814	33.5530±3.6315
	MS-SSIM	0.9647±0.0097	0.9369±0.0258	0.9282±0.0311	0.9474±0.0205
	NRMSE	0.0222±0.0079	0.0336±0.0147	0.0271±0.0083	0.0215±0.0086
3D-LargePNet	PSNR (dB)	34.9152±2.4816	31.0489±3.0687	32.5185±2.8361	35.1223±3.3280
	MS-SSIM	0.9747±0.0077	0.9526±0.0218	0.9464±0.0198	0.9661±0.0112
	NRMSE	0.0182±0.0057	0.0284±0.0115	0.0240±0.0076	0.0179±0.0065
Significance of LargePNet improvement through t-test					
Model	Metric	Nerlum indicum stem	Onion epidermal cell	Mouse kidney WGA	Mouse kidney phalloidin
Versus 3D-RCAN	PSNR (dB)	$p=8.94\text{e-}19(***)$	$p=1.11\text{e-}14(***)$	$p=1.35\text{e-}8(***)$	$p=5.87\text{e-}9(***)$
	MS-SSIM	$p=3.21\text{e-}20(***)$	$p=3.77\text{e-}18(***)$	$p=8.36\text{e-}8(***)$	$p=1.64\text{e-}9(***)$
	NRMSE	$p=2.79\text{e-}18(***)$	$p=2.03\text{e-}14(***)$	$p=1.19\text{e-}7(***)$	$p=1.02\text{e-}7(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 9. Metrics of UNet and LargePNet on self-supervised image denoising task

Model	Metric	CCPs	ER	Microtubules	F-Actin
UNet-SN2N	PSNR (dB)	33.1821±4.9036	21.9195±13.6181	28.6238±2.6499	28.5272±3.1646
	MS-SSIM	0.9665±0.0218	0.8134±0.4345	0.9641±0.0141	0.9432±0.0066
	NRMSE	0.0320±0.0195	0.1327±0.2519	0.0382±0.0121	0.0438±0.0304
LargePNet-S2N2	PSNR (dB)	34.2519±3.8296	25.1455±13.1846	31.5304±2.5705	31.2897±4.6347
	MS-SSIM	0.9685±0.0194	0.8410±0.3974	0.9733±0.0158	0.9553±0.0139
	NRMSE	0.0246±0.0115	0.0826±0.1697	0.0275±0.0081	0.0177±0.0144
Significance of LargePNet improvement through t-test					
Model	Metric	CCPs	ER	Microtubules	F-Actin
Versus UNet-SN2N	PSNR (dB)	$p=1.63\text{e-}6(***)$	$p=6.53\text{e-}21(***)$	$p=4.02\text{e-}29(***)$	$p=5.11\text{e-}13(***)$
	MS-SSIM	$p=0.0237(*)$	$p=3.65\text{e-}10(***)$	$p=3.36\text{e-}30(***)$	$p=2.50\text{e-}12(***)$
	NRMSE	$p=0.0025(**)$	$p=8.28\text{e-}9(***)$	$p=1.06\text{e-}21(***)$	$p=1.67\text{e-}5(***)$

Note: The significance test results are obtained by two-sided t-test. No adjustment for multiple comparisons was applied.

Supplementary Table 10. PSNR(dB) metrics of LargePNet with different ERFs on the BioSR dataset.

Structure	LargePNet-11	LargePNet-13	LargePNet-15	LargePNet-17	LargePNet-19
CCP	35.5593	35.6468	35.8077	35.7443	35.781
ER	31.4821	30.6697	31.3432	31.7895	31.4765
MT	31.7426	31.8876	32.0415	32.2077	31.1861
Actin	31.7963	31.4313	32.3224	31.5853	30.6305

Note: LargePNet-k represents LargePNet with k×k-size RepLKConv. A larger k value leads to a larger ERF.

Supplementary Table 11. Explaining structure-dependent advantage of LargePNet in the BioSR single-image-super-resolution task using the GLCM analysis method.

LargePNet PSNR gain compared with the best small-patch model				
Structure	CCP	ER	MT	F-actin
PSNR(dB) gain	+2.73dB	+1.37dB	+1.12dB	+0.65dB
GLCM Contrast statistics difference analysis				
Large-view	1.54-fold	3.28-fold	1.76-fold	2.08-fold
Small-view	1.55-fold	3.29-fold	2.17-fold	2.47-fold
Large-small Diff	1.0-fold	1.0-fold	1.2-fold	1.2-fold
GLCM Energy statistics difference analysis				
Large-view	0.185-fold	0.168-fold	0.280-fold	0.352-fold
Small-view	0.065-fold	0.135-fold	0.189-fold	0.319-fold
Large-small Diff	2.8-fold	1.2-fold	1.5-fold	1.1-fold
GLCM Correlation statistics difference analysis				
Large-view	0.0061	0.2351	0.0031	0.0334
Small-view	0.0379	0.2710	0.0363	0.0555
Large-small Diff	0.0318	0.0359	0.0332	0.0221

Supplementary Table 12. Explaining structure-dependent advantage of LargePNet in the STED denoising task using the GLCM analysis method.

LargeP-TISR PSNR gain compared with the best small-patch model			
Structure	MT	Mito IMM	ER
PSNR(dB) gain	+0.25dB	+0.56dB	+0.36dB
GLCM Contrast statistics difference analysis			
Large-view	7.15-fold	5.79-fold	12.97-fold
Small-view	6.28-fold	5.82-fold	8.07-fold
Large-small Diff	1.14-fold	1.00-fold	1.61-fold
GLCM Energy statistics difference analysis			
Large-view	5.23-fold	7.58-fold	1.84-fold
Small-view	4.76-fold	2.39-fold	1.17-fold
Large-small Diff	1.10-fold	3.17-fold	1.57-fold
GLCM Correlation statistics difference analysis			
Large-view	0.1018	0.0497	0.8172
Small-view	0.1530	0.1195	0.8134
Large-small Diff	0.0512	0.0698	0.0038

Supplementary Table 13. Explaining structure-dependent advantage of LargePNet in the BioTISR video super-resolution task using the GLCM analysis method.

LargeP-TISR PSNR gain compared with the best small-patch model		
Structure	MT	Mito OMM
PSNR(dB) gain	+0.50dB	+1.18dB
GLCM Contrast statistics difference analysis		
Large-view	4.29-fold	2.22-fold
Small-view	4.37-fold	3.09-fold
Large-small Diff	1.02-fold	6.85-fold
GLCM Energy statistics difference analysis		
Large-view	2.16-fold	3.28-fold
Small-view	3.10-fold	1.27-fold
Large-small Diff	1.43-fold	4.16-fold
GLCM Correlation statistics difference analysis		
Large-view	0.1652	0.0062
Small-view	0.1286	0.0225
Large-small Diff	0.0366	0.0395

Supplementary Table 14. Explaining structure-dependent advantage of LargePNet in the STED deblurring task using the GLCM analysis method.

LargeP-GAN PSNR gain compared with the best small-patch model		
Structure	MT	Mito IMM
PSNR(dB) gain	+0.42dB	+1.16dB
GLCM Contrast statistics difference analysis		
Large-view	4.80-fold	1.02-fold
Small-view	3.27-fold	1.41-fold
Large-small Diff	1.47-fold	1.38-fold
GLCM Energy statistics difference analysis		
Large-view	0.085-fold	0.155-fold
Small-view	0.058-fold	0.313-fold
Large-small Diff	1.47-fold	2.02-fold
GLCM Correlation statistics difference analysis		
Large-view	0.0177	0.0961
Small-view	0.0361	0.0309
Large-small Diff	0.0184	0.0652

Supplementary Table 15. A summary of the dataset used in this study

Dataset Name	Type	Structures	Link	Figure
BioSR single-image super-resolution	Open-source	Clathrin-coated pit, endoplasmic reticulum, microtubule and F-actin filament	https://doi.org/10.6084/m9.figshare.13264793	Figure 2 Supplementary Figure 22
BioTISR video super-resolution	Open-source	Microtubule, mitochondrial outer membrane	https://doi.org/10.5281/zenodo.13924483	Figure 4
Volumetric confocal to STED	Open-source	Nuclear pore complex, microtubule	https://zenodo.org/record/4624364#.YF3jsa9Kibg.	Supplementary Figure 17
Volumetric background removal	Self-established	Nerlun indicum stem, onion epidermal cells, WGA and pallocidin labeled mouse kidney tissue sections	https://doi.org/10.5281/zenodo.15694668	Figure 4
Single-image STED denoising and deblurring	Self-established	Microtubule, endoplasmic reticulum, mitochondrial inner membrane	https://doi.org/10.5281/zenodo.15694668	Figure 3
SMLM virtual sampling	Self-established	Microtubule	https://doi.org/10.5281/zenodo.15694668	Figure 3

Supplementary Table 16. The microscopy methods in the collected datasets

Self-collected dataset					
Dataset	Structure	Sampling	Excitation wavelength	Objective	Detectors
STED (Fig. 3, Ex Fig. 5)	Microtubule	25-55nm	610 nm	100x(Oil), 1.4NA	Leica HyD hybrid detector
STED (Fig. 3, Ex Fig. 5)	Mitochondria	25-55nm	610 nm	100x(Oil), 1.4NA	Leica HyD hybrid detector
STED (Fig. 3, Ex Fig. 5)	ER	25-55nm	610 nm	100x(Oil), 1.4NA	Leica HyD hybrid detector
SMLM (Fig.3)	Microtubule	15 nm	488 nm	100x(Oil), 1.4NA	sCMOS
SD-confocal (Fig. 4)	Nerlum	108 nm lateral 800 nm z step	610 nm	20x(Air), NA	sCMOS
SD-confocal (Fig. 4)	Onion	108 nm lateral 800 nm z step	610 nm	20x(Air), NA	sCMOS
SD-confocal (Fig. 4)	Mouse WGA	325 nm lateral 800 nm z step	488 nm	10x(Air), NA	sCMOS
SD-confocal (Fig. 4)	Mouse Phalloidin	325 nm lateral 800 nm z step	561 nm	10x(Air), NA	sCMOS
Open-source dataset					
Dataset	Structure	Sampling	Excitation wavelength	Objective	Detectors
BioSR SIM (Fig. 2, Ex. Fig.6)	CCP	31.3 nm	488 nm	100x(Oil) 1.41NA	sCMOS
BioSR SIM (Fig. 2, Ex. Fig.6)	ER	31.3 nm	488 nm	100x(Oil) 1.35NA	sCMOS
BioSR SIM (Fig. 2, Ex. Fig.6)	Microtubule	31.3 nm	488 nm	100x(Oil) 1.35NA	sCMOS
BioSR SIM (Fig. 2, Ex. Fig.6)	F-actin	31.3 nm	488 nm	100x(Oil) 1.41NA	sCMOS
BioTISR SIM (Fig. 4)	Microtubule	30.6 nm spatial 0.25 s temporal	488 nm	100x(Oil) 1.35NA	sCMOS
BioTISR SIM (Fig. 4)	Mitochondrion	30.6 nm spatial 0.5 s temporal	488 nm	100x(Oil) 1.35NA	sCMOS

Note: the self-collected SMLM data require reconstruction, which is performed using the commercial Huygens Localizer software. The open-source BioSR and BioTISR SIM data already contained the reconstructed SIM data. The reconstruction is performed through the home-written SIM reconstruction algorithms by the dataset establishers. Please find details in the data link and their papers.

References

1. Lecun, Y. & Bottou, L. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278-2324 (1998).
2. Glorot, X., Bordes, A. & Bengio, Y. Deep Sparse Rectifier Neural Networks. *Journal of Machine Learning Research* **15**, 315-323 (2011).
3. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *In IEEE Conference on Computer Vision and Pattern Recognition* (2016).
4. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nature Methods* **18**, 678-687 (2021).
5. Ebrahimi, V. et al. Deep learning enables fast, gentle STED microscopy. *Communications Biology* **6** (2023).
6. Li, X. et al. Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nature Biotechnology* **41**, 282-292 (2022).
7. Ma, C., Tan, W., He, R. & Yan, B. Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration. *Nature Methods* **21**, 1558-1567 (2024).
8. Qu, L. et al. Self-inspired learning for denoising live-cell super-resolution microscopy. *Nature Methods* **21**, 1895-1908 (2024).
9. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090-1097 (2018).
10. Zhang, G. et al. Bio-friendly long-term subcellular dynamic recording by self-supervised image enhancement microscopy. *Nature Methods* **20**, 1957-1970 (2023).
11. Chen, R. et al. Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. *Nature Communications* **14** (2023).
12. Jin, L. et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nature Communications* **11** (2020).
13. Li, Y. et al. Incorporating the image formation process into deep learning improves network performance. *Nature Methods* **19**, 1427-1437 (2022).
14. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* **36**, 460-468 (2018).
15. Qiao, C. et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nature Biotechnology* **41**, 367-377 (2022).
16. Qiao, C. et al. A neural network for long-term super-resolution imaging of live cells with reliable confidence quantification. *Nature Biotechnology* (2025).
17. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103-110 (2018).
18. Wang, Z., Xie, Y. & Ji, S. Global voxel transformer networks for augmented microscopy. *Nature Machine Intelligence* **3**, 161-171 (2021).
19. Ounkomol, C., Seshamani, S., Maleckar, M.M., Collman, F. & Johnson, G.R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature Methods* **15**, 917-920 (2018).
20. Bilodeau, A. et al. Development of AI-assisted microscopy frameworks through realistic simulation with pySTED. *Nature Machine Intelligence* **6**, 1197-1215 (2024).
21. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* **18**, 194-202 (2021).
22. Chen, X. et al. Self-supervised denoising for multimodal structured illumination microscopy enables long-term super-resolution live-cell imaging. *Photonix* **5**, 4 (2024).

23. Luo, W., Li, Y., Urtasun, R. & Zemel, R. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *In Advances in neural information processing systems* (2017).
24. Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv preprint arXiv:1607.08022*. (2016).
25. Ding, X. et al. Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs. *In IEEE Conference on Computer Vision and Pattern Recognition* (2022).
26. Ding, X. et al. UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio, Video, Point Cloud, Time-Series and Image Recognition. *In IEEE Conference on Computer Vision and Pattern Recognition* (2023).
27. Wang, X. et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *In European Conference on Computer Vision* (2018).
28. Wang, L., Guo, Y., Liu, L., Lin, Z. & An, W. Deep Video Super-Resolution using HR Optical Flow Estimation. *IEEE Transactions on Image Processing* **PP**, 1-1 (2020).
29. Liang, J. et al. SwinIR: Image Restoration Using Swin Transformer. *In IEEE/CVF International Conference on Computer Vision Workshops* 1833-1844 (2021).