Nature Communications

https://doi.org/10.1038/s41467-026-71278-2

Article in Press

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

Received: 15 August 2025

Accepted: 18 March 2026

Cite this article as: Hou, Y., Gao, S., Ren, W. *et al.* Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-71278-2

Yiwei Hou, Shu Gao, Wei Ren, Yunzhe Fu, Meiqi Li & Peng Xi

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

Yiwei Hou[1,#], Shu Gao[1,#], Wei Ren[1], Yunzhe Fu[1], 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*

Abstract

Deep learning has demonstrated remarkable success in augmenting fluorescence imaging under photon-limited conditions. However, existing restoration networks are typically devised for training with augmented patches far smaller than the full-view raw data, an overlooked aspect that compromises fidelity and noise-resistance due to the loss of global statistics. To address this limitation, we propose a large-patch network (LargePNet), which synergizes the large effective receptive field provided by shallow ultra-large-kernel convolutions and the nonlinear representation capabilities of deep networks through scale separation. It effectively and efficiently leverages large-view global information for restoration. Directly trained with large-view images, LargePNet shows contrasting advantages over state-of-the-art small-patch networks, with 0.5-2 dB higher peak signal-to-noise ratio across eight representative restoration tasks, involving implementations for single-image, video, and volumetric fluorescence data. For full-view processing, LargePNet generally holds around 4-fold and 20-fold higher computational efficiency compared to advanced convolution-based and Transformer-based networks, respectively. The assistance of LargePNet helps achieve 30-hour-long fluorescence imaging to monitor cytoskeleton dynamics, and hour-long tri-color super-resolution imaging to investigate organelle interaction, showcasing its advancement in live-cell imaging.

Introduction

Deep learning (DL) has emerged as a powerful tool for image restoration in recent years[1-5], outperforming conventional methods when sufficient training data are available. This data-driven approach has been successfully adapted to fluorescence microscopy[6-27] (Supplementary Note 1), where acquired images often suffer from degeneration and require high-fidelity restoration. Notably, in live-cell imaging, the computational compensation techniques can reduce the photon requirement for high-quality recordings, thereby improving several critical imaging parameters such as imaging duration and spatiotemporal resolution[17, 23, 28, 29]. By enabling higher information throughput at lower hardware costs, DL methods facilitate the observation of intricate cellular dynamics with unprecedented clarity.

Among DL architectures, deep convolutional neural networks[30, 31] (CNNs) have proven to be one of the most effective models for image restoration. Their success stems from the convolution operation's ability to exploit local pixel correlations, a universal feature of natural and biological images. For fluorescence image restoration, most state-of-the-art CNNs are derived from the basis of the UNet[32] or RCAN[5] model. Both architectures leverage deep stacking to enhance performance, with skip connections ensuring stable training. However, some recent works suggest that the performance of CNN is inherently restricted by its focus on extracting local information[33-35], originating from the locality of the convolution operation. Transformers[36-38] primarily employ attention operations to integrate global information, achieving superior or competitive performance to CNNs in fluorescence restoration tasks. For example, a recently reported Swin Transformer based network[13] surpassed advanced RCAN-derived models[16] in fluorescence image super-resolution.

Despite these advances, the precision of current models remains to be improved to meet the requirements of scientific research. Under noisy conditions in particular, several studies have reported significant performance degradation[13, 19]. Achieving finer restoration of low-SNR data can further enhance the live-cell imaging speed and duration to observe more intricate biological processes. In current practice, full-size microscopy images (typically ≥512 pixels wide) are routinely cropped into small augmented patches (often 64×64 or 128×128 pixels) for stable network training. While in most application scenarios the fluorescence images are large-view, previous evaluations have largely followed one of two problematic approaches: (1) evaluating model performance solely on small-patches[13, 16], or (2) applying small-patch-trained models directly to full-view images for evaluation[6, 14, 18, 19, 21], justified by the size-invariant nature of convolutional parameters. A critical oversight lies in the inconsistency of patch size across training,

evaluation, and application scenarios. The small patch truncated many contextual structures and forces the network to learn restoration with limited information[39]. However, the application image typically spans large views with much richer contextual structure information, which are important restoration clues but remain largely unexplored. Moreover, the differences of the structure statistics between the large view and small patch can also lead to performance degradation when the small-patch-trained network is applied to large-view images[40, 41]. In a theoretical sense, existing restoration networks and their developmental route pay too much attention to mining the "depth information" of fluorescence images and ignore their "width information".

In this article, we demonstrate that direct large-patch training (≥512 pixels) provides an effective and efficient pathway for superior fluorescence image restoration. To meet this goal, we devise a large-patch network (LargePNet) with scale separation, instance normalization[42, 43], and shallow convolutional layers with ultralarge kernel size that differs from conventional restoration models. Our design holds higher efficiency when processing large-size patches, and builds an effective way to leverage global information from full-view images. Comprehensive evaluations show that LargePNet noticeably outperforms previous state-of-the-art models in a vast range of fluorescence image restoration tasks and turns out to be more robust in noisy conditions, and more efficient for processing full-view images. Empowered by LargePNet, some very challenging imaging tasks can be achieved, such as hour-long tri-color STED imaging and day-long recording of cytoskeleton dynamics in living cells.

Results

Network architecture suitable for aggregating large-view information

Our initiative is to explore the route of effectively leveraging large-view information to improve the performance of current fluorescence image restoration models. A larger patch allows more contextual pixels in the restoration process, inherently providing greater potential for performance improvement. This approach should be more suitable for fluorescence microscopy, where images typically contain abundant similar biological structures that provide rather direct and easy-to-learn semantic information (Supplementary Fig. 1). Moreover, the native resolution of most scientific cameras (≥512 pixels) naturally aligns with and supports our large-view training paradigm, eliminating the need for curating training data through massive croppings.

Effectively leveraging large-patch training requires handling of two key challenges: (1) establishing a sufficiently large effective receptive field[44] (ERF) (Supplementary Note 2), and (2) finely managing the stability and computational cost when processing high-resolution inputs.

While deeper CNNs theoretically provide more chances for long-range pixel interactions, empirical studies show their ERF often remains surprisingly limited, with networks predominantly focusing on local features[34]. Transformer-based architecture builds a relatively explicit long-range information connection within a cropped region. However, when increasing the input size, it suffers from computational inefficiency due to the costly attention operations, thus hard to aggregate large-view global information. Furthermore, for image restoration tasks, such indiscriminate global modeling could be an unnecessary computational burden that may not yield proportional performance benefits. As an echo for our analyses, in practice, the commonly used patch size was 64×64 or 128×128 for fluorescence image restoration CNNs and Transformer models[6-23]. A large patch size often harms the training process of networks[45]. To overcome these constraints, we ought to devise a network architecture that simultaneously achieves effective large-view information aggregation and efficient computation for large-sized inputs.

Our initial trial explored building a network by stacking a few reparameterized ultra-large-kernel (e.g., 25×25) convolutional layers (RepLKConv)[34, 35] as an alternative to deep stacks of 3×3 convolutions or attention operations. The RepLKConv operation was reported to improve the ERF in some high-level vision tasks like natural image classification and target detection[34, 35]. The design initiative of RepLKConv is not specifically for training with very large patches, but as a more efficient way than the costly attention operation to build long-range connections. However, we found that RepLKConv does not perform well in a simple low-level vision task of denoising fluorescence images when training with a 512-size-patch (Supplementary Fig. 2), leaving massive residual noise. This is likely due to that networks with several shallow RepLKConv operations lack sufficient nonlinear representation capability[46] required in the restoration task, especially in noisy conditions. This observation aligns with the notable absence of RepLKConv in state-of-the-art restoration architectures.

Through the above observations, we propose LargePNet for large-patch training to aggregate large-view global information (Fig. 1a, Supplementary Fig. 3, Supplementary Note 3). LargePNet synergizes the efficient long-range modeling of RepLKConv and the nonlinear representation capability of very deep networks. To reduce computational cost and stabilize the gradients during training, LargePNet down-samples the input by multiple rates, which are subsequently processed by different low-frequency feature extractors (LFFE) and a high-frequency feature extractor (HFFE). LFFE is essentially a conventional U-shaped network but modified with instance normalization and average pooling that we termed as InsUNet. LFFEs process the simple down-sampled abstractive feature map matching to stabilize the gradients for training large-sized inputs.

Moreover, the LFFE contains deep architectures to provide a sufficient nonlinear representation. HFFE processes the full-size feature and is composed of four RepLKConv layers stacked with skip connections. It takes the major computational burden that handles the hardest detail recovery and builds a large enough ERF (Fig. 1b) for large-view information aggregation. Especially, like our modification in LFFE, we also employ instance normalization in HFFE to replace the commonly used batch normalization, since we imposed computational burden on the patch-size dimension instead of the batch-size dimension. We found that instance normalization is beneficial for training LargePNet, as shown later.

As indicated in the ERF analysis (Fig. 1c), LargePNet aggregates information from a much larger region, while typical CNN and Transformer architecture focuses on a region in a 128-pixel, or less, window. As evidence to inspect the functioning principle of LargePNet, we found that the feature maps propagated in LFFE are highly abstractive, and HFFE focuses on fine structural details (Fig. 1d). This fundamental difference necessitates our large-patch training paradigm (typically 512×512 pixels) to fully exploit LargePNet's architectural advantages, in contrast to traditional small-patch augmentation approaches (Fig. 1e, Supplementary Table 1). Our optimized design ensures the high efficiency of LargePNet for large-patch training, which is equivalent to a standard RCAN processing small-patches with equivalent volume and is much lower than some more advanced architectures (Fig. 1f). To accomplish a 512-size single-image translation, LargePNet requires ~6 GB GPU memory, which is affordable by most commercial GPUs.

Validation of large-view statistics aggregation through LargePNet

To validate the rationality of the LargePNet design, we conduct an initial evaluation in a simple single-image denoising task by transforming the low-SNR widefield image to the high-SNR one using the BioSR dataset[16]. We first inspect the performance of LargePNet and a standard UNet when training with 128, 256, and 512 patch sizes, to check the favorability of the two networks in terms of patch sizes. The results show that the standard UNet performs optimally with smaller patches, while LargePNet demonstrates superior performance when trained with the largest 512×512 patches (Supplementary Fig. 4, Supplementary Table 2). This characteristic enables LargePNet to surpass UNet considerably, especially in high-noise conditions. We further conduct an ablation investigation on the LargePNet structures by detaching its primary components (Supplementary Fig. 2), showing the importance of combining LFFE and HFFE features and the application of instance normalization for the functionality of LargePNet. Moreover, we also investigate whether further increasing the depth of conventional 3×3 convolution networks can achieve the equivalent outcome. The results show that the ERF does not evidently increase for a

deeper 3×3 convolution network (Supplementary Fig. 5), which is far lower than LargePNet, even when the computational cost is already several folds higher. Another interesting characteristic of LargePNet is that its optimal learning rate is roughly tenfold higher than that usually adopted in conventional networks. Following this initial validation, we progressively evaluate the performance of LargePNet in some high-profile restoration tasks with advanced, tailored competitive models.

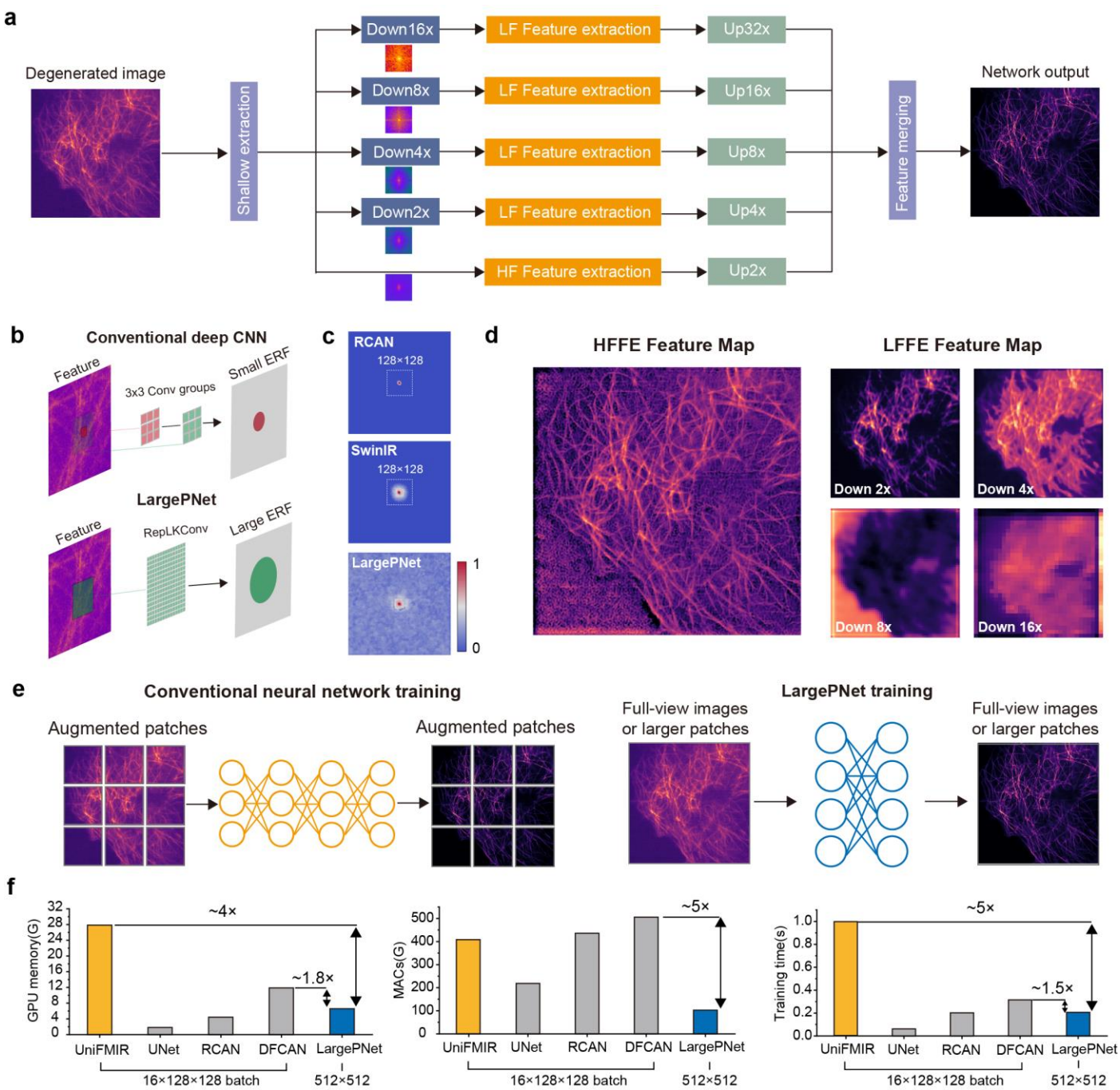


Fig. 1. The architecture of the proposed LargePNet. a. Diagram of the architecture of the proposed

LargePNet. The up-scaling factor can be adjusted to meet the requirements of different restoration tasks. For more details, see Supplementary Fig. 3. **b.** Comparisons of the main feature extraction blocks in the conventional deep convolutional neural network (CNN) and LargePNet. **c.** Analysis of the effective receptive field (ERF) map of RCAN, SwinIR, and LargePNet. **d.** Representative feature maps propagated in the high-frequency feature extractor (HFFE) and low-frequency feature extractor (LFFE) branches in LargePNet. **e.** Difference of the training patches between LargePNet and conventional neural networks. **f.** The peak GPU memory occupation, multiply-accumulate operations (MACs), and training time of LargePNet processing a 512×512-image, and some representative small-patch networks processing an equivalent 16×128×128 batch for single-image translation.

Performance of LargePNet for single-image restoration tasks in widefield systems

We first evaluate the performance of LargePNet on a well-established BioSR single-image super-resolution task[16], which aims to transfer widefield microscopic images of four typical organelles to super-resolution SIM images with 2× upscale. We compare LargePNet with DFCAN[16] and UniFMIR[13], which are the state-of-the-art CNN and Transformer-based networks tailored for this task, respectively. Fine-tuned UniFMIR models are provided for the BioSR super-resolution tasks, serving as an objective performance benchmark. While both comparison models were trained using conventional 128×128 patches, LargePNet leveraged our large-patch training paradigm with 512×512 inputs. We first examine the patch-wise quality of the network results that were previously adopted for the BioSR dataset evaluation (Supplementary Fig. 6). The results show that LargePNet outperforms DFCAN and UniFMIR, with ~1.2 dB higher PSNR than UniFMIR, providing a better resolution of detailed structures and is better at recovering signals from heavy noise.

We progressively investigate the performance on 512×512 full-size widefield images, which could be meaningful for practical application but was less discussed in previous works. When processing an image larger than the training patch size, CNN structures like DFCAN can directly infer the full-size image or stitch patch-wise inferences to full-size, with overlapping to reduce the boundary artifacts (Supplementary Fig. 7). UniFMIR incorporates SwinIR, which successively processes each 8×8-size window, preventing boundary artifacts but only utilizing the information within a 128×128 region. LargePNet, directly trained on large-view images, demonstrates contrasting advantages in the full-view metrics (Fig. 2a, Supplementary Fig. 8, Supplementary Table 3), outperforming UniFMIR by ~2.4 dB higher PSNR. This indicates that the small-patch-trained super-resolution network typically has poor quantitative rigor when transferred to process a full-view image in this super-resolution task. Conventional small-patch CNN architectures like DFCAN could obtain higher fidelity metrics when directly inferring the full-size image in most scenarios, but would lose considerable details (Fig. 2b, c, and Supplementary Figs. 9, 10), due to

the inconsistency of the train-test image size. In the previous BioSR dataset evaluation, only images with average photon count >50 were used for evaluation[16], and we found that lower-intensity conditions would lead to collapsed performance of DFCAN and UniFMIR (Fig. 2d, e). LargePNet demonstrates exceptional noise robustness through global information aggregation. Besides obtaining superior performance, LargePNet holds higher efficiency when inferring full-view images, benefiting from its architecture design, taking only ~1/4 and 1/20 of the time cost of DFCAN and UniFMIR, respectively (Fig. 2f). It also generalizes well when processing images larger than the training patch (Fig. 2g).

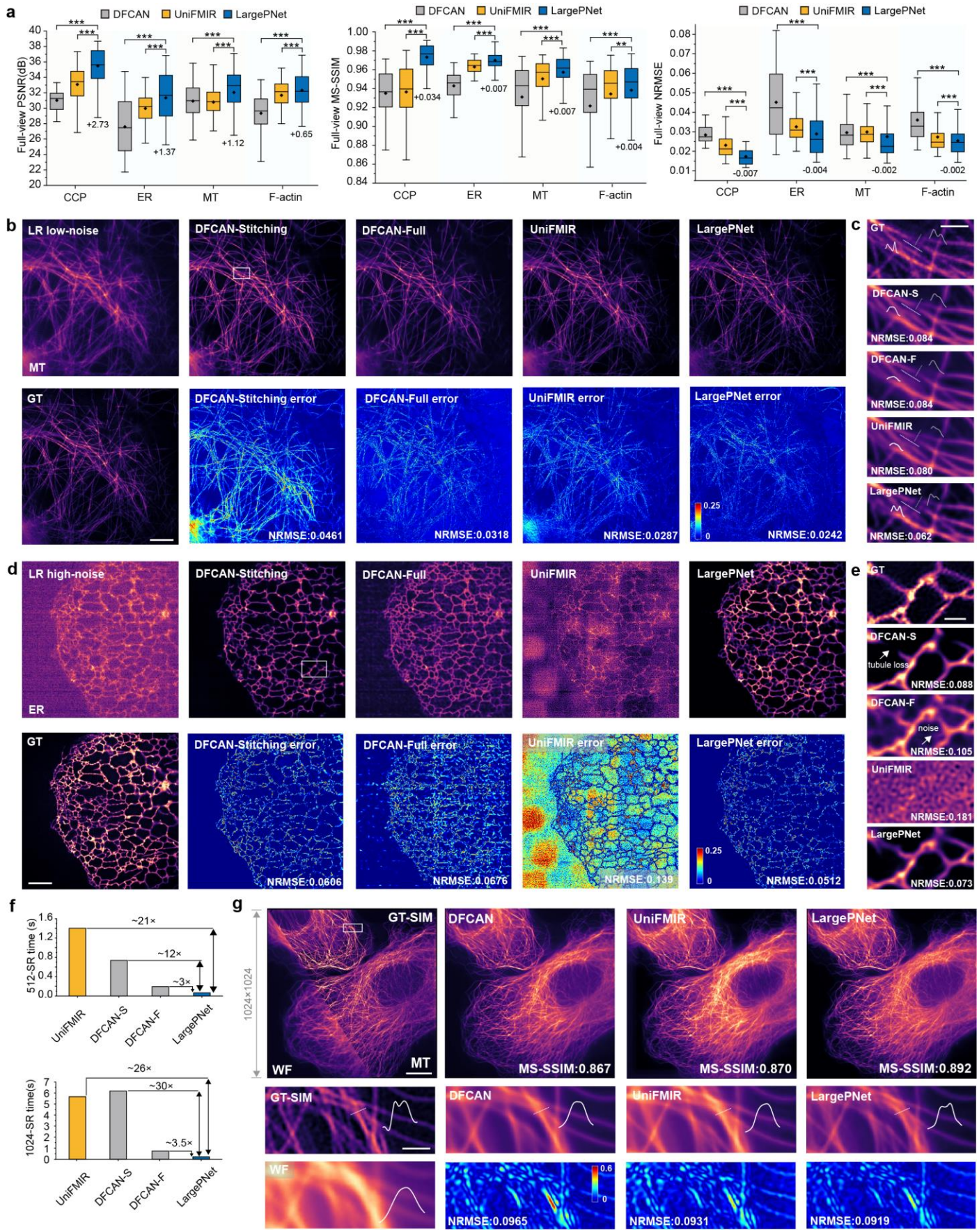

a
DFCAN
UniFMIR
LargePNet
Full-view PSNR(dB)
Full-view MS-SSIM
Full-view NRMSE
CCP
ER
MT
F-actin
b
LR low-noise
DFCAN-Stitching
DFCAN-Full
UniFMIR
LargePNet
GT
DFCAN-Stitching error
DFCAN-Full error
UniFMIR error
LargePNet error
NRMSE:0.0461
NRMSE:0.0318
NRMSE:0.0287
NRMSE:0.0242
c
GT
DFCAN-S
NRMSE:0.084
DFCAN-F
NRMSE:0.084
UniFMIR
NRMSE:0.080
LargePNet
NRMSE:0.062
d
LR high-noise
NRMSE:0.0606
NRMSE:0.0676
NRMSE:0.139
NRMSE:0.0512
e
tubule loss
noise
NRMSE:0.088
NRMSE:0.105
NRMSE:0.181
NRMSE:0.073
f
512-SR time (s)
1024-SR time(s)
~21×
~12×
~3×
~26×
~30×
~3.5×
g
GT-SIM
DFCAN
UniFMIR
LargePNet
WF
MT
1024×1024
MS-SSIM:0.867
MS-SSIM:0.870
MS-SSIM:0.892
NRMSE:0.0965
NRMSE:0.0931
NRMSE:0.0919

Fig. 2. LargePNet shows superior performance on single-image super-resolution tasks in the open-source BioSR dataset. a. Statistical comparisons of the full-view PSNR, MS-SSIM, and NRMSE metrics of the recovered images via different models (n =135, 114, 135, 180 test images in the dataset for CCP, ER, MT, and F-actin). The exact improvement values of LargePNet over the second-best model are marked. The patch-wise metrics are shown in Supplementary Fig. 6. Representative single-image super-resolution results of a 512×512-size microtubule image **b** and ER image **d**, with the magnified white boxed region shown in **c** and **e**. SIM images of the same ROI are shown as the ground truth (GT) for comparison. LR, low resolution. The difference images relative to GT are shown below the corresponding network outputs. **f.** Comparisons of the inference time of the three models. DFCAN-S: stitching method to obtain a full-view image. DFCAN-F: directly infer full-view image. **g.** Performance of the three models on a 1024×1024-size microtubule image. SIM image of the same ROI is shown for comparison. Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: 5 μm (b, d), 1 μm (c, e, g subgraph), 10 μm (g). (***, p<0.001; **, p<0.01; Obtained by two-sided t-test; No adjustment for multiple comparisons was applied; The exact p values in a can be found in Supplementary Table 3).

Performance of LargePNet for single-image restoration tasks in super-resolution systems

Beyond the restoration tasks in the widefield system, we further evaluate the applicability of LargePNet in super-resolution systems with a higher spatial resolution, such as STED[47] and SMLM[48, 49]. Considering the lack of high-quality open-source single-image STED datasets, we collected STED images of three typical subcellular structures and established a denoising dataset (Methods) including microtubule, mitochondrial inner membrane, and ER, enlightened by the procedure in a previous literature[8]. Practically, DL-assisted STED can reduce the bleaching and improve the acquisition speed. We compare the performance of LargePNet with UNet-RCAN[8] and SwinIR[2], with the former being a state-of-the-art CNN tailored for STED image denosing, and the latter being a strong Swin Transformer baseline for image restoration. In the comparative results, LargePNet outperforms the other two counterparts by providing higher restoration quality metrics (Fig. 3a, Supplementary Figs. 11, 12, Supplementary Table 4) with lower computational cost (Fig. 3f). It provides finer restoration of subtle biological structures in the noisy conditions than SwinIR and UNet-RCAN, such as the inner mitochondrial membrane and closely distanced microtubule filaments (Fig. 3b-e, Supplementary Fig. 13).

Besides denoising the super-resolution STED images, some generative adversarial networks[50] (GANs) can transfer a blurry single image, such as confocal or STED with low depletion power, to a high-resolution STED image[20]. We anticipate that LargePNet can also be utilized to improve the performance of the GAN-based restoration framework by acting as a more powerful generator that aggregates large-view statistics. We devise LargeP-GAN using LargePNet as the generator

and modifying the CNN architecture in ERSGAN[51] to a multiscale version as the discriminator (Supplementary Note 3). We involved UNet[20] and RRDB[7, 51] as the competitor generators of LargePGAN, which were shown to have good performance in the GAN framework for fluorescence image-related tasks. As in visual perception, while conventional regression models such as UNet-RCAN fail to recover plausible high-frequency details, generative models recover most of the blurred details (Supplementary Fig. 14, Supplementary Table 5). Despite GANs achieving finer high-frequency information recovery, it is a challenge to also hold a high pixel-wise fidelity metric value. LargeP-GAN improves this issue by providing higher fidelity metrics while achieving equivalent resolution with UNet-GAN and RRDB-GAN.

We progressively investigated the application of LargeP-GAN in SMLM imaging by validating its performance on transforming undersampled SMLM images to well-sampled counterparts (Methods). Practically, this can improve the acquisition speed and imaging throughput of SMLM[15]. LargeP-GAN also provides the best metrics among all three competitive GAN-based models (Fig. 3g, h, Supplementary Table 6). These evidences validate the efficacy of LargePNet in improving the performance of generative restoration models.

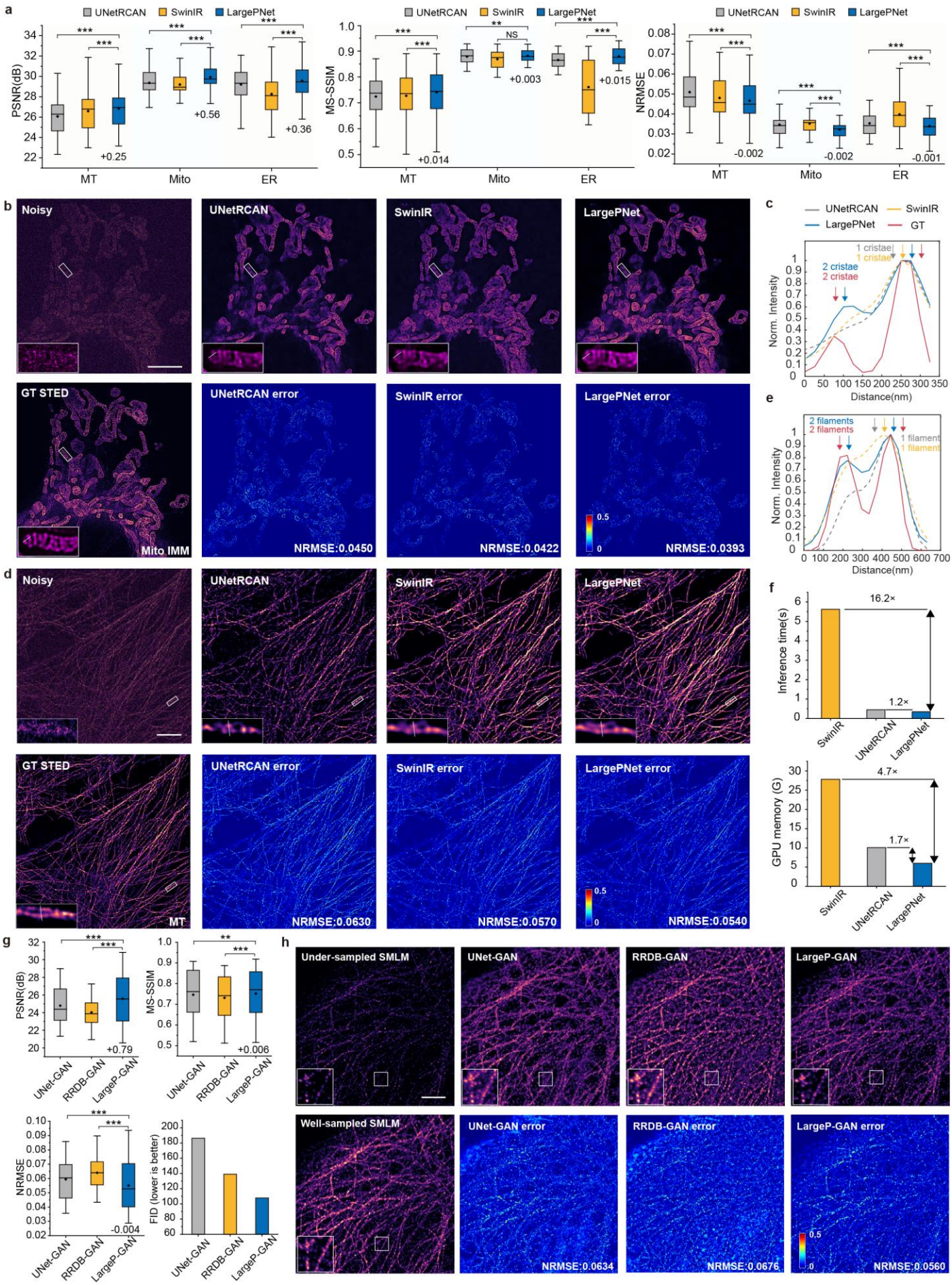

a
UNetRCAN
SwinIR
LargePNet
PSNR(dB)
MS-SSIM
NRMSE
MT
Mito
ER
+0.25
+0.56
+0.36
+0.014
+0.003
+0.015
-0.002
-0.002
-0.001
NS
b
Noisy
UNetRCAN
SwinIR
LargePNet
GT STED
Mito IMM
UNetRCAN error
SwinIR error
LargePNet error
NRMSE:0.0450
NRMSE:0.0422
NRMSE:0.0393
c
GT
1 cristae
2 cristae
Norm. Intensity
Distance(nm)
e
2 filaments
1 filament
d
MT
NRMSE:0.0630
NRMSE:0.0570
NRMSE:0.0540
f
Inference time(s)
16.2×
1.2×
GPU memory (G)
4.7×
1.7×
g
UNet-GAN
RRDB-GAN
LargeP-GAN
+0.79
+0.006
-0.004
FID (lower is better)
h
Under-sampled SMLM
Well-sampled SMLM
NRMSE:0.0634
NRMSE:0.0676
NRMSE:0.0560

Fig. 3. LargePNet shows superior performance in restoring images captured by super-resolution systems. a. Statistical comparisons of the 1024×1024 full-view PSNR, MS-SSIM, and NRMSE metrics of the recovered images via different models (n = 24 test images in the dataset for each structure). The exact improvement values of LargePNet over the second-best model are marked. **b, d.** Representative inference results of a 1024×1024-size inner mitochondrial membrane (IMM) and microtubule image. High-SNR STED images of the same ROI are shown as the ground truth (GT) for comparison. The difference images relative to GT are shown below the corresponding network outputs. **c, e.** Intensity profiles along the two arrowheads in b and d, which mark a mitochondrial cristae and two adjacent microtubule filaments, respectively. **f.** The peak GPU memory occupation in training with equivalent volume and inference time when processing a 1024×1024 image. **g.** Metrics of the recovered 1024×1024-size under-sampled SMLM images via different DL models (n=97 test images in the dataset). The exact improvement values of LargePNet over the second-best model are marked. **h.** Representative inference results via different models. Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: 5 μm (b, d), 2 μm (h). (***, p<0.001; **, p<0.01; Obtained by two-sided t-test; No adjustment for multiple comparisons was applied; The exact p values in a and g can be found in Supplementary Tables 4 and 6, respectively).

Generalization of LargePNet to video and volumetric modalities

The above experiments demonstrate that LargePNet can substantially improve the performance of single-image-based restoration tasks. We also wonder whether its advantages can be generalized to time-lapse and 3D volumetric data by slightly modifying the original architecture.

We devise LargeP-TISR that can leverage temporal information for video super-resolution (Supplementary Note 4). It incorporates an optical flow reconstruction network[52] (OFRNet), which leverages deep-wise convolution and channel shuffle to analyze temporal information. The output from OFRNet is consecutively sent to LargePNet for super-resolution. We test the performance of LargeP-TISR on the open-source BioTISR dataset[18], which contains well-matched pairs of widefield and super-resolution SIM videos. We compared the performance of LargeP-TISR with DPA-TISR[18], which is a state-of-the-art network devised for this dataset and is provided with well-trained models for an objective benchmark. DPA-TISR was trained with augmented 7×128×128-size video patches and was directly applied to infer the full-size 7×512×512 video in the dataset. For LargeP-TISR, we directly trained it with a 7×512×512-size video, which is computationally impractical for DPA-TISR since it contains complex deformable phase-space alignment operations. By aggregating large-view statistics, LargeP-TISR provides clearer super-resolution results (Fig. 4a-c) and shows robustness under different noise levels (Supplementary Figs. 15, 16). Quantitative assessment shows that our LargeP-TISR outperforms DPA-TISR by ~0.6 dB higher PSNR and ~0.03 higher MS-SSIM metrics (Fig. 4e, Supplementary Table 7). Besides, while achieving better fidelity metrics, LargeP-TISR reduces the training memory cost by ~5-fold and inference time by

~4-fold compared to DPA-TISR (Fig. 4f), demonstrating the efficiency of our large-patch training paradigm for this video super-resolution task.

We progressively investigate our method in processing volumetric data, and develop 3D-LargePNet that majorly upgrades each operation in LargePNet to its 3D form. In volumetric microscopic imaging, the axial resolution and sampling rate are usually several times lower than the lateral ones. Considering this feature, we have devised a cuboid ultra-large kernel-size convolution to save memory and computational cost while holding the performance (Supplementary Note 4). Benefitting from this, with equivalent memory and computational cost, 3D-LargePNet can process 3D-stacks with much larger sizes compared with 3D-RCAN[6], a universal state-of-the-art model for restoring volumetric fluorescence data.

We first benchmark 3D-LargePNet on an open-source dataset of deblurring volumetric confocal images[6], a task that 3D-RCAN is specifically good at. By aggregating larger-view information, 3D-LargePNet improves the baseline of 3D-RCAN with ~0.2 higher 3D-PSNR metrics (Supplementary Fig. 17). Besides the tasks similar to 2D scenarios that aim at removing random noise or blurring, we further explore removing scattering noise and background fluorescence in volumetric imaging, in which the global statistics could have a pivotal effect. We established a dataset of four tissue structures by capturing well-matched volumetric pairs of widefield and confocal modality to train 3D-RCAN and 3D-LargePNet (Methods). The results show that 3D-LargePNet provides more precise removal of background and resolves clearer structures (Fig. 4g-i), providing a contrasting ~1.4 dB higher 3D-PSNR in the four tested samples (Fig. 4j, Supplementary Figs. 18, 19, Supplementary Table 8), while maintaining a lower computational cost (Fig. 4k). These evidences demonstrate the advantage of 3D-LargePNet in volumetric image restoration tasks.

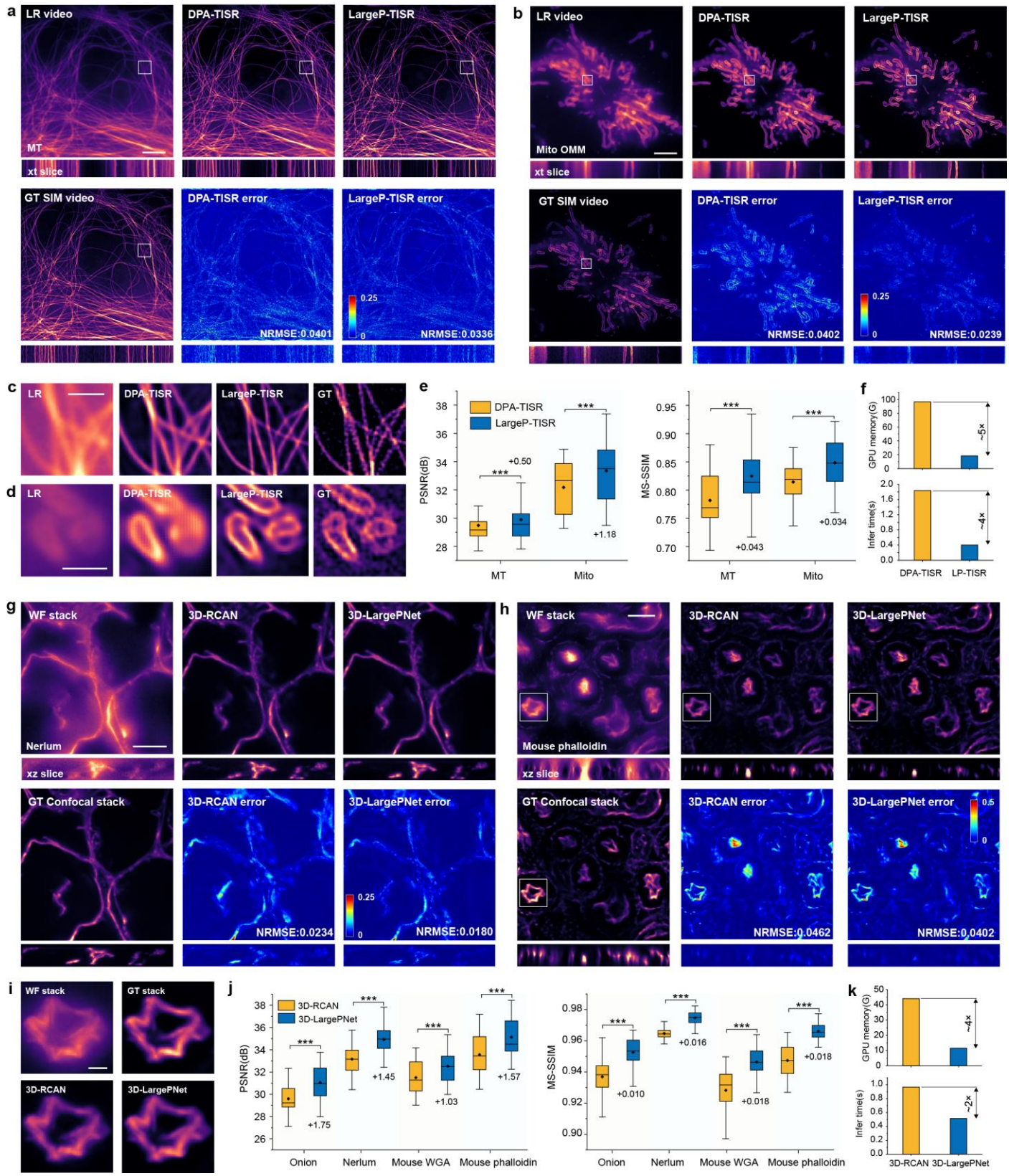


Fig. 4. Extension of LargePNet to restore video and volumetric data. a. b. Representative recovery results of the time-lapse microtubule and mitochondria in the Bio-TISR dataset. LR, low-resolution; GT, ground-truth. **c. d.** Magnified white boxed region in a and b, respectively. **e.** Statistical comparisons of the

full-view PSNR and MS-SSIM metrics of the images recovered by DPA-TISR and LargeP-TISR (n=420 frames for each sample in the dataset). **f.** The peak GPU memory occupation in training (1×7×512×512 input for LargeP-TISR, and 16×7×128×128 input for DPA-TISR) and inference time when processing a 7×512×512 video. **g. h.** Representative recovery results of a widefield stack of Nerlum indicum stem and palloidin-labeled actin in mouse kidney tissue sections, respectively. **i.** Magnified white boxed region in h. **j.** Statistical comparisons of the full-view PSNR(dB) and MS-SSIM metrics (n=32 test images for plant section samples and 24 test images for mouse kidney section samples). **k.** The peak training GPU memory occupation (4×16×256×256 for 3D-RCAN and 1×16×512×512 for 3D-LargePNet) and inference time when processing a 24×512×512 volumetric stack. Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. Scale bars: 10 μm (a, b), 1 μm (c, d), 50 μm (g), 20 μm (h), 2 μm (i). (***, p<0.001; Obtained by two-sided t-test; No adjustment for multiple comparisons was applied; The exact p values in e and j can be found in Supplementary Tables 7 and 8, respectively).

LargePNet improves live-cell fluorescence imaging capabilities

LargePNet shows superior performance in various fluorescence image restoration tasks and improves the noise-robustness of previous networks, thus holding great potential to boost the live-cell imaging performance further. We then investigate how it can enhance the live-cell fluorescence imaging capabilities, allowing for the observation of cellular activities with a higher information flux.

Fluorescence imaging faces the trade-offs between quality and duration, due to the phototoxicity and bleaching caused by laser illumination. We have shown that LargePNet easily improves the imaging duration of microtubule, a light-sensitive organelle, to a day-long level. Our imaging lasts around 30 hours with a half-minute monitoring frequency (Fig. 5a, Supplementary Video 1). At the last frame, the imaging only suffers a moderate bleaching of around 30%, since we employ a low-power laser illumination (Fig. 5b). LargePNet has effectively removed the massive noise in the raw recordings (Supplementary Fig. 20), enabling fine segmentation of the microtubule structures. With the assistance of LargePNet, we can monitor the movement of the microtubule at a day-long level (Fig. 5c, d), maintaining a good state of the cell.

Besides assisting fluorescence imaging in the widefield system, the trained LargePNet models can also be utilized to achieve sustained multicolor super-resolution imaging that can be hard to achieve in the hardware dimension in the point-detecting STED system. With the assistance of LargePNet, we improve the resolution of the confocal recorded tri-color hour-long recordings of ER, mitochondria, and microtubules to STED-level at around 70 nm (Fig. 5e-g, Supplementary Fig. 21, Supplementary Video 2). With the restored images, we observed that one mitochondrial end maintained persistent contact with ER tubules and the microtubule network for most of the 1-hour recording, indicating the close correlation of the three organelles for cellular activities (Fig.

5h). These experiments highlight that LargePNet greatly expands the imaging capabilities of fluorescence microscopy.

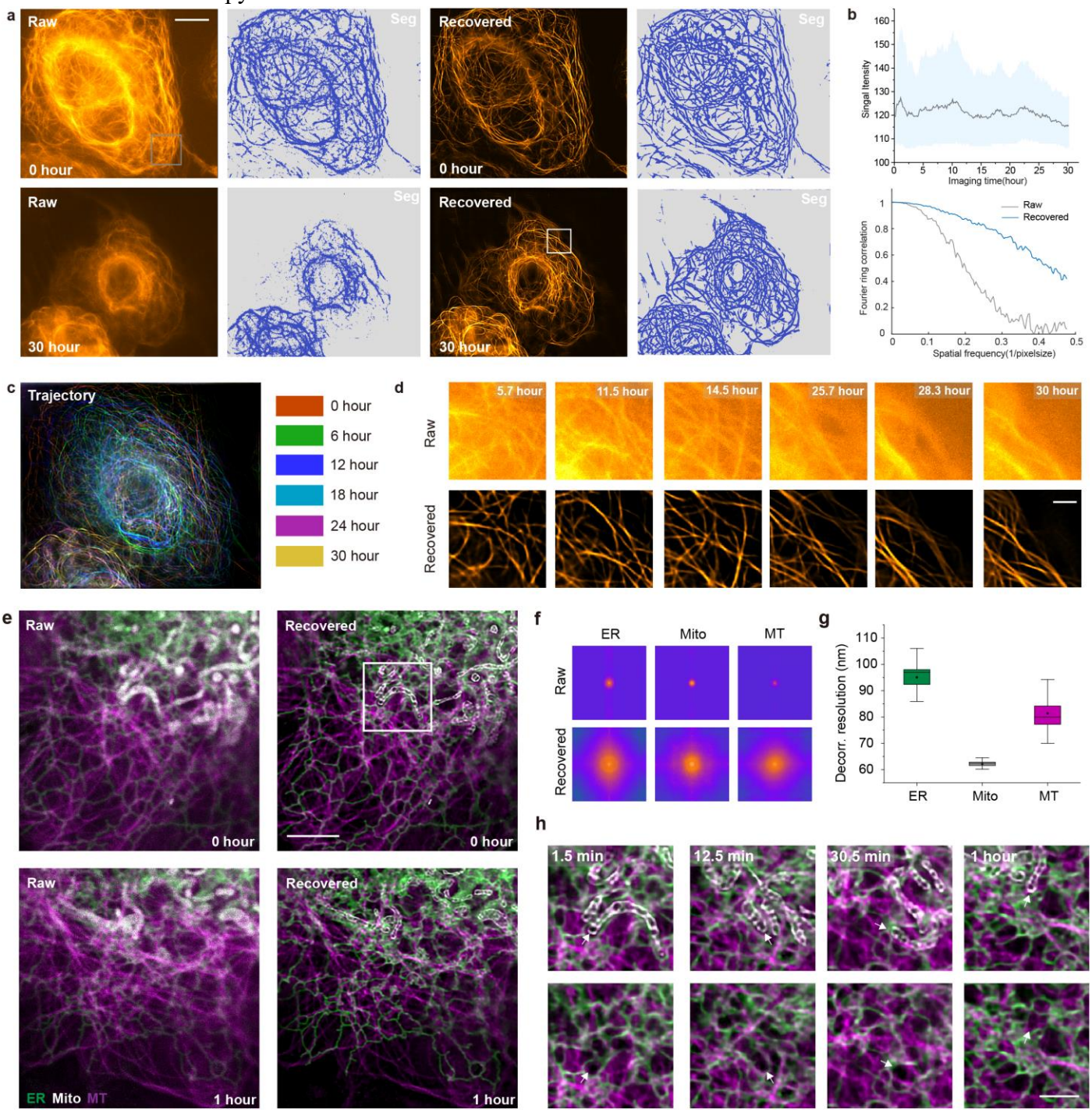


Fig. 5. LargePNet enables finer live-cell imaging. a. Widefield 30-hour-long recordings of the microtubule activities in COS-7 cells with capturing intervals of 30 seconds. Seg: segmentation results, obtained using the TWS toolbox[53]. **b.** Variation of the average, and 10%-90% percentile signal intensity in the gray boxed region in a (top), and a representative FRC of a raw and recovered image (bottom). **c.** Time-coded color image showing movement of the microtubule network. **d.** Magnified boxed region in a at different time points. Top: raw images, bottom: DL recovered. **e.** Representative shots in hour-long tri-color

imaging of ER tubules (green), mitochondria (white), and microtubules (magenta) in a U2OS cell, with capturing intervals of 30 seconds. **f.** The Fourier spectrum before and after recovery. **g.** Resolution of the recovered images is estimated via decorrelation analysis. (n = 120 frames in the video). Box chart: centerline, medians; point, average values; limits, 75% and 25%; whiskers, maxima and minima. **h.** Magnified white boxed region in e. The white arrowhead denotes the position of the mitochondrial cristae end. Scale bars: 10 μm (a), 2 μm (d, h), 5 μm (e).

Discussion

In this work, we highlight the importance of incorporating large-view statistics for fluorescence image restoration. To effectively aggregate such large-view information, we propose LargePNet, which works as the backbone feature extractor in end-to-end training and is versatile to be transplanted to single-image, time-lapse, and volumetric data types. Besides the supervised task presented in the manuscript, we found that LargePNet can also improve some self-supervised learning frameworks, such as the self-inspired noise2noise denoising (Supplementary Fig. 22). LargePNet as the main feature extractor, markedly improves the performance of the previous UNet-SN2N (Supplementary Table 9). Our evaluation results across eight representative fluorescence microscopy restoration tasks demonstrated that LargePNet markedly outperforms previous state-of-the-art small-patch models in various modalities. It should be noted that despite LargePNet grealy improving the restoration performance of previous models, there is still a possibility of hallucinations when the noise is excessively strong (Supplementary Fig. 23). Although we have established a large enough ERF in LargePNet to aggregate large-view information, it should be noted that an even larger ERF is not necessarily better, as shown in Supplementary Table 10. The optimal ERF may be related to factors such as structure size and other relevant parameters[54]. It would also be interesting to discuss why LargePNet shows higher or lower advantages in the restoration of a specific structure. We find that this can be explained by the structural statistics difference in large-view and small-view images (Supplementary Note 6, Supplementary Fig. 24, Supplementary Tables 11-14). The larger the deviation of statistical information in the small patch from the large-view image, the more likely LargePNet is to have a significant restoration advantage.

Moreover, LargePNet achieves superior performance while demonstrating higher efficiency, requiring only 1/4 and 1/20 computational cost of the RCAN or SwinIR-derived advanced models, benefiting its deployment. Besides fluorescence microscopy, we believe that LargePNet could also improve the DL restoration performance in images similar to fluorescence images, such as label-free and electron microscope images. Moreover, LargePNet could be integrated into some task-specific, physics-inspired DL restoration or reconstruction frameworks to further boost the performance[17, 25, 55].

Despite foundation models recently showing great success in some vision areas, the fluorescence imaging restoration fields may still lack a high-quality, uniform, and large-scale dataset to demonstrate their power. In our work, for example, by only training with specific data, our LargePNet surpasses UniFMIR, which was finetuned from a foundational model. To mitigate

the data scarcity in this field, we have open-sourced our self-collected STED, SMLM, and volumetric background removal datasets to strengthen the current available data volume (Supplementary Tables 15, 16). We believe the idea of LargePNet to aggregate large-view information can also benefit the development of foundational models, in cases when sufficient data is given.

Despite LargePNet achieving success in several common fluorescence imaging restoration tasks, it is essential to discuss its deficits for a balanced view. Firstly, LargePNet architecture is devised to aggregate information from a large patch-size (such as 512 pixels). For some cases when only small ROIs are available, LargePNet may perform worse or similarly with state-of-the-art small-patch networks. We showcase this phenomenon in the BioSR dataset (Supplementary Fig. 25). Secondly, while possessing a computational cost lower than some advanced models, LargePNet's processing speed is still evidently lower than the commonly used UNet, despite the total computation amounts being equivalent. This is due to the fact that current off-the-shelf deep-learning tools (such as Pytorch) are more well-optimized for conventional CNNs that stack very deep 3x3 convolutions compared with RepLKConv[34] that occupies major computational cost in LargePNet. Despite this, LargePNet is still much more efficient than some advanced networks (such as DFCAN and UniFMIR) while providing better performance. Optimization of the RepLKConv efficiency can further benefit the deployment of LargePNet.

Methods

Development and training of the LargePNet series model

We implement our large-patch model series using Python 3.10 and the Pytorch 2.0.0 environment. The code is mainly developed using primary neural network components provided by Pytorch, thus, it was insensitive to the CUDA and Pytorch versions through our testing. Our LargePNet series takes large-size patches (usually with a size of 512 pixels or larger) for training, which are generated from the original dataset through augmentation methods including random cropping, flipping, and ±90° rotation. In cases when the original image size is not large enough, we just use the full-size images with only flipping and rotation augmentation without cropping for training. The Adam optimizer is adopted for training LargePNet. It usually takes 800-1,200 512×512 image pairs to well train LargePNet. The typically chosen learning rate is 1e-3 for LargePNet, which is roughly ten times higher than the commonly chosen value for conventional networks. The details for the choice of hyperparameter for each task are listed in Supplementary Note 5. The training was mainly accomplished with a local workstation equipped with an NVIDIA RTX 4080 GPU and a remote server with an NVIDIA RTX 4090 GPU. The Matlab (2022b) software was used for data analysis.

Comparison with other regression models

In this work, we compare the performance of LargePNet with some other advanced restoration models, including UNet[9], RCAN[5], UniFMIR[13], DFCAN[16], SwinIR[2], UNetRCAN[8], RRDB[51], DPA-TISR[18], and 3D-RCAN[6]. In the open-source BioSR[16] and BioTISR[18] datasets, the state-of-the-art UniFMIR and DPA-TISR provide well-trained models, thus could objectively assess the performance of our LargePNet model. For others without well-trained models, we retrained the model using their original code provided in the GitHub repository and following the original hyperparameter selection. Details are listed in Supplementary Note 5.

When comparing GAN-based image-restoration models, we uniformly adopt the relativistic GAN (RaGAN)[51] method. Since the major contribution of this paper is the proposition of a stronger generator, we compare its GAN form performance by equally training UNet-GAN and RRDB-GAN in the RaGAN scheme.

Data splitting

For the self-collected dataset in the manuscript, we adopted random splitting of train, validation, and test data. The data splitting procedure has been chosen to avoid data leakage. For the open-

source dataset, we adopt their original split data. The data splitting mimics anticipated real-world applications. Details can be found through the link in Supplementary Table 15.

Box-chart display

In this paper, we employ the box chart to display the distribution of the restoration metrics. As presented in this paper, the boxes and whiskers are typically relatively long. This is because restored image datasets usually acquire different levels of SNR for a single FOV (e.g., the F-actin dataset in BioSR includes 12 noise levels). The performance of all models degrades significantly as the SNR of the original images decreases, which results in the elongated boxes and whiskers. In addition, although box plots effectively present the fluctuations of restoration metrics across samples with varying SNRs, they may obscure the performance differences between different models. Statistical charts stratified by different noise levels could help to better understand the performance of various models under different initial SNR conditions of original images. Some illustrative examples are provided in Supplementary Figs. 8, 10.

Preparation of STED deblurring and denoising dataset

STED imaging was performed on COS-7 cells grown to ~50%-70% confluency in 35 mm glass bottom plates (Cellvis, #1.5H, D35-20-1.5H). For microtubule labeling, COS-7 cells were first pre-fixed with 0.4% glutaraldehyde (EMS, 16220) and 0.25% Triton X-100 (Sigma, X100-100ML) for 90 s, subsequently fixed with 3% glutaraldehyde for 15 min at room temperature, then 0.1% $NaBH_4$ was used for quenching for 7 min at room temperature. After washing with 1× PBS, 3 times, cells were incubated with 3% BSA and 0.25% Triton X-100 for blocking and permeabilization for 1 h. Afterwards, samples were incubated with primary antibody (anti-alpha-tubulin antibody: Proteintech, 66031-1, 1:200) at 4°C overnight. Afterwards, cells were washed with 1× PBS, 3 times, and incubated with labeled secondary antibody (Abberior, anti-mouse STAR Red, 1:200). After labeling, samples were washed with 1× PBS, 3 times, and used for STED imaging. For mitochondria labeling, COS-7 cells were first labeled with 250 nM PK Mito Orange Fix dye and 10 μM Verapamil (Sigma, 152-11-4) in cell culture medium at 37°C for 15 min. After labeling, cells were washed with 1× PBS and replaced with fresh culture medium in a cell culture incubator for at least 30 min. Then cells were fixed with 2% glutaraldehyde at room temperature for 10 min, and then washed with 1× PBS, 3 times and used for STED imaging. For ER labeling, COS-7 cells stably expressing Sec61β-mEmerald (a kind gift from Dr. Pengli Zheng) were rinsed with 1× PBS, and subsequently fixed with 3% paraformaldehyde and 0.1% glutaraldehyde for 12 min, and 0.1%

$NaBH_4$ was used for quenching for 7 min at room temperature. After washing with 1× PBS, 3 times, cells were used for STED imaging.

We used a commercially available Leica SP8 3× STED microscopy and an Aberrior STEDYCON STED microscope to capture our STED training dataset. The 100× objective lens (1.43 NA, oil immersion) was used to capture the high-resolution STED images. We capture the microtubule structure of 2048×2048 ROI size with 18 independent shots, the mitochondrion and ER structure of 1024×1024 ROI size with 32 independent shots. The pixel size varies at 25-55 nm to let the model comprehensively learn the characteristics at different scales. The data volume is sufficient to train different models.

We investigate two conditions of confined STED images, with one major degeneration originating from random noise due to low dwell time, and the other originating from blurring due to a lower depletion laser intensity. We employed a Gaussian blurry kernel: $g = \exp[-(x^2+y^2)/2\sigma^2]$ to simulate the blurring. In the first condition, we add Poisson-Gaussian mixed noise at different levels to the GT image, after slightly blurring the GT image with a Gaussian kernel with a σ value of 1.0 (pixel), and train DL models to transform the degenerated image to the GT. In the second condition, we blur the GT image with a Gaussian kernel with a σ value of 2.0-4.0 (pixel) with moderate Poisson-Gaussian mixed noise, and then trained DL models to transform the degenerated image to the GT. Previous literature shows that the model trained with images contaminated with degeneration can easily generalize to real conditions, and in our study, we also found that the trained models can be successfully applied to real scenarios (Fig. 5).

Preparation of volumetric background removal dataset

We used a commercially available Airy-NovaSD (Beijing, China) spinning-disk confocal microscope to capture well-matched pairs of volumetric widefield and confocal images. Three commercially available fixed slices were used for imaging, including a mouse kidney section (FluoCells Prepared Slide #3, Invitrogen, F24630), an onion epidermal cell slice (Sagaoptics, China), and a nerlum indicum stem slice (Sagaoptics, China). Considering the characteristics of each structure, we used the 20× objective (NA 0.4, air) to capture the structures in a mouse kidney cell slice, and the 60× objective (NA 1.2, oil immersion) to capture the structures in two plant cell slices. The images were captured with a volumetric size of 1024×1024×24 with at least 30 ROIs for each structure.

Preparation of virtual sampling of the single-molecule localization microscopy dataset

DNA-PAINT imaging was performed on U-2OS cells grown to ~50%-70% confluency in 35 mm glass bottom patri dishes. For microtubule labeling, U-2OS cells were first pre-fixed with 0.4% glutaraldehyde and 0.25% Triton X-100 for 90 s, subsequently fixed with 3% glutaraldehyde for 15 min at room temperature, then 0.1% $NaBH_4$ was used for quenching for 7 min at room temperature. For endoplasmic reticulum labeling, U-2OS cells stably expressing Sec61β-mEmerald were rinsed with 1× PBS, and subsequently fixed with 3% paraformaldehyde and 0.1% glutaraldehyde for 12 min, and 0.1% $NaBH_4$ was used for quenching for 7 min at room temperature. After washing with 1× PBS, 3 times, cells were incubated with 3% BSA and 0.25% Triton X-100 for blocking and permeabilization for 1 h. Afterwards, samples were incubated with 10 μg/ml of primary antibody (anti alpha-tubulin antibody: Proteintech, 66031-1; anti-GFP antibody: Millipore, G1546) at 4°C overnight. Afterwards, cells were washed with 1× PBS, 3 times, and incubated with labeled secondary antibody (Jackson, 715-007-003). Specifically, the secondary antibodies were conjugated to 5′-thiol-modified DNA strands via thiol–maleimide chemistry following a previously established protocol[56]. Briefly, secondary antibodies were first functionalized with maleimide groups using a heterobifunctional NHS-maleimide crosslinker targeting lysine residues. Thiol-modified DNA docking strands (5′-TTGATCTACAT-3′) were reduced and incubated with the maleimide-activated antibodies to form stable thioether bonds via thiol-maleimide coupling. Excess DNA was removed by purification, yielding DNA-conjugated secondary antibodies for imaging. The complementary imager strand（5′-TATCTAGATC-3′) was labeled with Cy3b and synthesized commercially (Thermo Fisher Scientific). Samples were imaged in imaging buffer consisting of 1× PBS (pH 8.0), 500 mM NaCl, and an oxygen-scavenging system containing 1× Trolox, 1× PCA, 1× PCD, and 0.5 nM Cy3b-labeled imager strands. We used a commercially available Airy-NovaSD (Beijing, China) spinning-disk confocal microscope to capture raw images of SMLM, and the results were reconstructed using Huygens Localizer software.

Calculation of the average photon count in the open-source BioSR dataset

To objectively assess the performance of LargePNet, we used the open-source BioSR image super-resolution dataset provided by Qiao et al[16]. It contains the low-high resolution pairs of CCP, MT, ER, and F-actin. Two state-of-the-art models are involved for comparison, including DFCAN and UniFMIR. It also describes a method to calculate the average photon count. In previous assessments, it calculated the average photon count per patch (128×128-size). The patches with an average photon count of 25-600 were selected in the calculation of image metrics, and the calculation method was also described[16]. To assess the performance of LargePNet in this SNR

range, we follow the same procedure and select 128×128 patches to calculate the patch-wise metrics shown in Supplementary Fig. 6.

Cell culture and maintenance

Human osteosarcoma U-2OS cell line (HTB-96), human adenocarcinoma HeLa cell line (CRM-CCL-2), and cercopithecus aethiops COS-7 cell line (CRL-1651) were purchased from ATCC. We directly purchased the cell line from ATCC, which has been authenticated. We confirm that the cell line we used was tested negative for mycoplasma contamination. COS-7 cells (ATCC, USA) and U-2OS cells (ATCC, USA) used in this study were cultured in Dulbecco's Modified Eagle's medium (Gibco, 11965-092) supplemented with 10% (v/v) heat-inactivated Fetal Bovine Serum (Thermofisher Scientific, A5670701) and 1% (v/v) Pencillin-streptomycin (Thermofisher Scientific, 15140122). Cells were incubated in a sterile and humid incubator with 5% CO_2 at 37°C.

Collection of the three-channel long-term time-lapse confocal imaging dataset

COS-7 cells grown to ~50%-70% confluency in 35 mm glass bottom patri dishes were used for the collection of the three-channel long-term time-lapse confocal imaging dataset. Cells were transfected with Sec61β-mEmerald (a kind gift from Dr. Pengli Zheng) and EMTB-mYongHong (a kind gift from Dr. Zhifei Fu) plasmids using the Lipofectamine 3000 kit. After transfecting for 30 hrs, cells were labeled with 500 nM HBmito Crimson dye in cell culture medium at 37°C for 15 min. Then cells were washed with fresh cell culture medium twice and subsequently used for long-term time-lapse confocal imaging. We collected this dataset using the Leica SP8 3× STED microscopy.

Collection of long-term time-lapse widefield imaging dataset

U-2OS cells grown to ~50%-70% confluency in 35 mm glass bottom patri dishes were used for the collection of a long-term time-lapse widefield imaging dataset. Cells were transfected with EMTB-mYongHong plasmids using the Lipofectamine 3000 kit. After transfecting for 30 hrs, cells were used for long-term time-lapse widefield imaging. We collected this dataset using the Airy polar SIM microscope.

Statistics and Reproducibility

For each restoration task, in addition to the well-trained comparative models that have been previously published, each deep learning model was trained independently three times, demonstrating similar restoration performance. The model with the highest restoration metrics was

used for subsequent analysis. For long-term live-cell imaging experiments, the same scene was repeated three times under identical experimental conditions, and LargePNet achieved satisfactory restoration with consistent observations.

Code availability

The source Python code of LargePNet series, including LargePNet (for single-image restoration), LargeP-GAN (for generative single-image restoration), LargeP-TISR (for time-lapse video restoration), and 3D-LargePNet (for volumetric data restoration) are all publicly available at the GitHub repository[57]: https://github.com/YiweiHou/LargePNet-for-fluorescence-image-restoration. Trained LargePNet models that can reproduce the results in the paper are available at: https://figshare.com/s/05f576c96b08add7eee0.

Data availability

The open-source data used in this study are all publicly available, as listed in Supplementary Table 10. The self-established dataset of STED denoising/deblurring, virtual sampling of SMLM, volumetric background removal datasets, and the source training data for the open-source BioSR and BioTISR datasets are available: https://zenodo.org/records/15694668.

Author contributions

Y. Hou and P. Xi conceived the idea of LargePNet for fluorescence image restoration tasks. P. Xi and M. Li supervised this project. Y. Hou developed all the codes and performed all the training and validation experiments. S. Gao captured the training dataset for STED and SMLM and performed all the live-cell imaging experiments. W. Ren provided part of the STED mitochondria imaging data. Y. Fu provided insightful discussions. Y. Hou, S. Gao, M. Li, and P. Xi wrote the paper after discussions with all the authors.

Competing interests

The authors declare no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (62025501 to P. X.; 62335008, 62405010 to M. L.), National Key R&D Program of China (2022YFC3401100 to P. X.), and Major Basic Research Project of the Natural Science Foundation of Shandong Province (ZR2024ZD27 to P. X.). We thank National Center for Protein Sciences at Peking University in Beijing, China, for assistance with STED super-resolution imaging.

References

1. Ledig, C. et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *In IEEE Conference on Computer Vision and Pattern Recognition*, 105-114 (2017).
2. Liang, J. et al. SwinIR: Image Restoration Using Swin Transformer. *In IEEE/CVF International Conference on Computer Vision Workshops* 1833-1844 (2021).
3. Saharia, C. et al. Image Super-Resolution via Iterative Refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 4713-4726 (2023).
4. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing* **26**, 3142-3155 (2016).
5. Zhang, Y. et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *In European Conference on Computer Vision*, 294-310 (2018).
6. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nature Methods* **18**, 678-687 (2021).
7. Chen, R. et al. Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. *Nature Communications* **14** (2023).
8. Ebrahimi, V. et al. Deep learning enables fast, gentle STED microscopy. *Communications Biology* **6** (2023).
9. Jin, L. et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nature Communications* **11** (2020).
10. 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).
11. Li, Y. et al. Incorporating the image formation process into deep learning improves network performance. *Nature Methods* **19**, 1427-1437 (2022).
12. Lu, C. et al. Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy. *Nature Communications* **15** (2024).
13. 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).
14. 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).
15. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* **36**, 460-468 (2018).
16. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* **18**, 194-202 (2021).
17. Qiao, C. et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nature Biotechnology* **41**, 367-377 (2022).
18. Qiao, C. et al. A neural network for long-term super-resolution imaging of live cells with reliable confidence quantification. *Nature Biotechnology* (2025).
19. Qu, L. et al. Self-inspired learning for denoising live-cell super-resolution microscopy. *Nature Methods* **21**, 1895-1908 (2024).
20. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103-110 (2018).
21. Wang, Z., Xie, Y. & Ji, S. Global voxel transformer networks for augmented microscopy. *Nature Machine Intelligence* **3**, 161-171 (2021).

22. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090-1097 (2018).
23. Zhang, G. et al. Bio-friendly long-term subcellular dynamic recording by self-supervised image enhancement microscopy. *Nature Methods* **20**, 1957-1970 (2023).
24. Chen, X. et al. Self-supervised denoising for multimodal structured illumination microscopy enables long-term super-resolution live-cell imaging. *PhotoniX* **5**, 4 (2024).
25. Bilodeau, A. et al. Development of AI-assisted microscopy frameworks through realistic simulation with pySTED. *Nature Machine Intelligence* **6**, 1197-1215 (2024).
26. Bouchard, C. et al. Resolution enhancement with a task-assisted GAN to guide optical nanoscopy image analysis and acquisition. *Nature Machine Intelligence* **5**, 830-844 (2023).
27. Guo, M. et al. Deep learning-based aberration compensation improves contrast and resolution in fluorescence microscopy. *Nature Communications* **16**, 313 (2025).
28. Hou, Y. et al. Multi-resolution analysis enables fidelity-ensured deconvolution for fluorescence microscopy. *eLight* **4**, 14 (2024).
29. Huang, X. et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nature Biotechnology* **36**, 451-459 (2018).
30. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *In IEEE Conference on Computer Vision and Pattern Recognition* (2016).
31. Lecun, Y. & Bottou, L. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278-2324 (1998).
32. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015).
33. Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C. & Dosovitskiy, A. Do Vision Transformers See Like Convolutional Neural Networks? *In Conference and Workshop on Neural Information Processing Systems* (2021).
34. 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).
35. 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).
36. Dosovitskiy, A., Beyer, L., Kolesnikov, A. & Weissenborn, D. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *In International Conference on Learning Representations* (2021).
37. Liu, Z., Lin, Y., Cao, Y. & HanHu Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *In IEEE/CVF International Conference on Computer Vision* (2021).
38. Vaswani, A., Shazeer, N., Parmar, N. & Uszkoreit, J. Attention is all you need. *In Advances in Neural Information Processing Systems* (2017).
39. Levin, A., Nadler, B., Durand, F. & Freeman, W.T. in Computer Vision – ECCV 2012. (eds. A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato & C. Schmid) 73-86 (Springer Berlin Heidelberg, Berlin, Heidelberg; 2012).
40. Lu, X.C.a.L.C.a.C.C.a.X. Improving Image Restoration by Revisiting Global Information Aggregation. *ECCV* (2022).
41. Luo, Z.Y.a.A.Z.a.Y.M.a.Y.L.a.X.Z.a.P. Scale Calibrated Training: Improving Generalization of Deep Networks via Scale-Specific Normalization. *CVPR* (2020).
42. Chen, L., Lu, X., Zhang, J., Chu, X. & Chen, C. HINet: Half Instance Normalization Network for Image

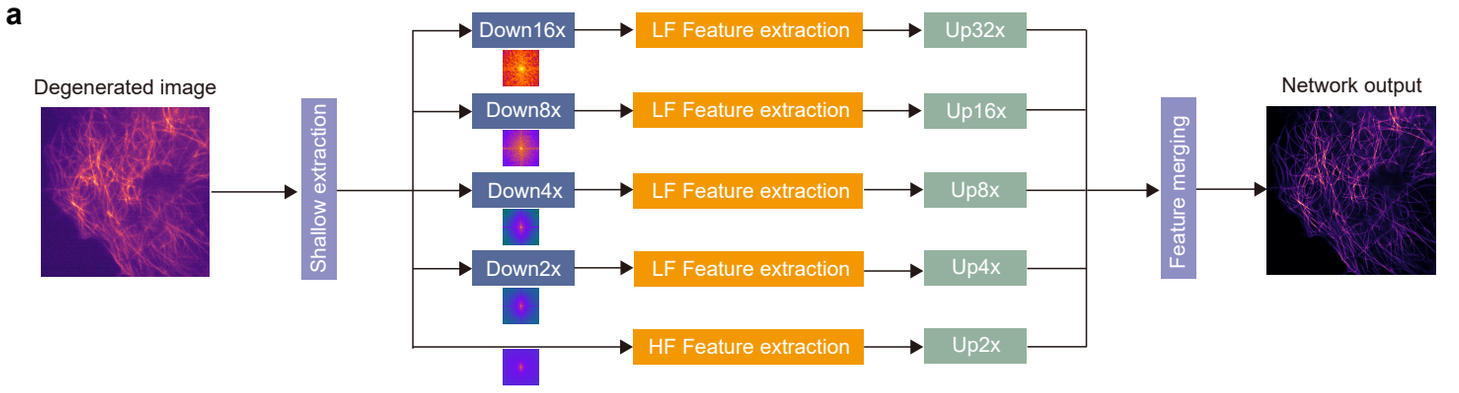
Restoration. *In IEEE Conference on Computer Vision and Pattern Recognition* (2021).

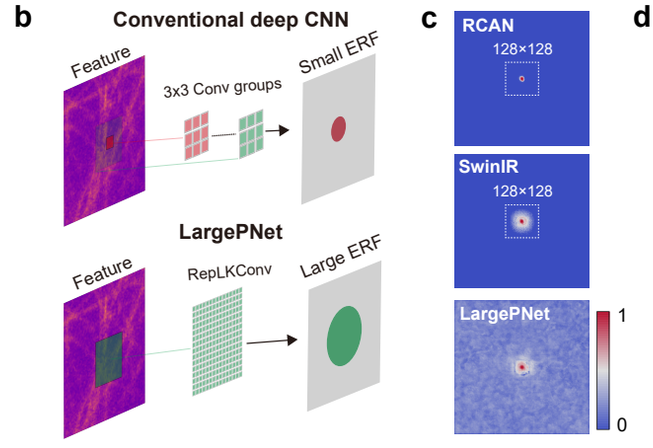
43. Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv preprint arXiv:1607.08022.* (2016).
44. Luo, W., Li, Y., Urtasun, R. & Zemel, R. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *In Advances in neural information processing systems* (2017).
45. Liu, B., Zhu, Y., Song, K. & Elgammal, A. Towards Faster and Stabilized GAN Training for High-fidelity Few-shot Image Synthesis. *In International Conference on Learning Representations* (2021).
46. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *In International Conference on Learning Representations* (2014).
47. Hell, S.W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters* **19** (1994).
48. Betzig, E. et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642-1645 (2006).
49. Rust, M.J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793-796 (2006).
50. Goodfellow, I. et al. Generative Adversarial Nets. *In Conference on Neural Information Processing Systems* (2014).
51. Wang, X. et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *In European Conference on Computer Vision* (2018).
52. 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).
53. Descloux, A., Grußmayer, K.S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods* **16**, 918-924 (2019).
54. Loos, V., Pardasani, R. & Awasthi, N. Demystifying the effect of receptive field size in U-Net models for medical image segmentation. *Journal of Medical Imaging* **11** (2024).
55. Wang, F. et al. Phase imaging with an untrained neural network. *Light: Science & Applications* **9**, 77 (2020).
56. Schnitzbauer, J., Strauss, M.T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nature Protocols* **12**, 1198-1228 (2017).
57. Hou, Y. et al. Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics. *Zenodo* https://doi.org/10.5281/zenodo.18820335 (2026).

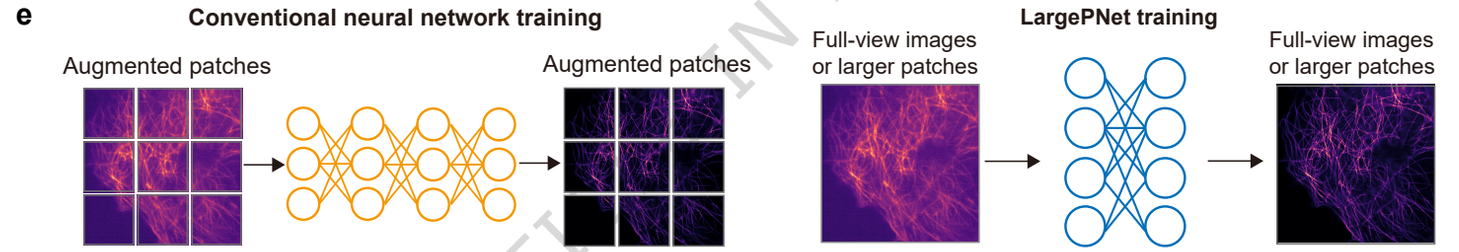
Editor's Summary

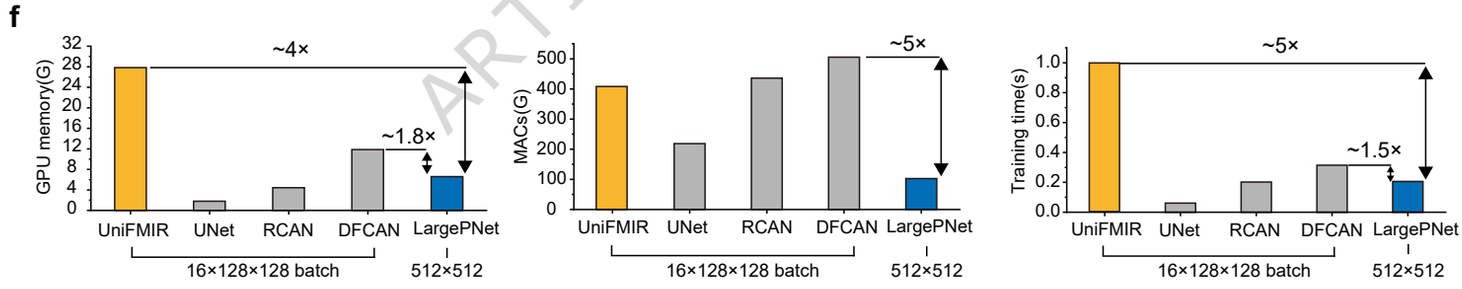
Existing small patch design in CNN and Transformer restoration network deviates from characteristics of fluorescence microscopy. LargePNet overcomes this context-deficiency, significantly improving state-of-the-art application scenarios.

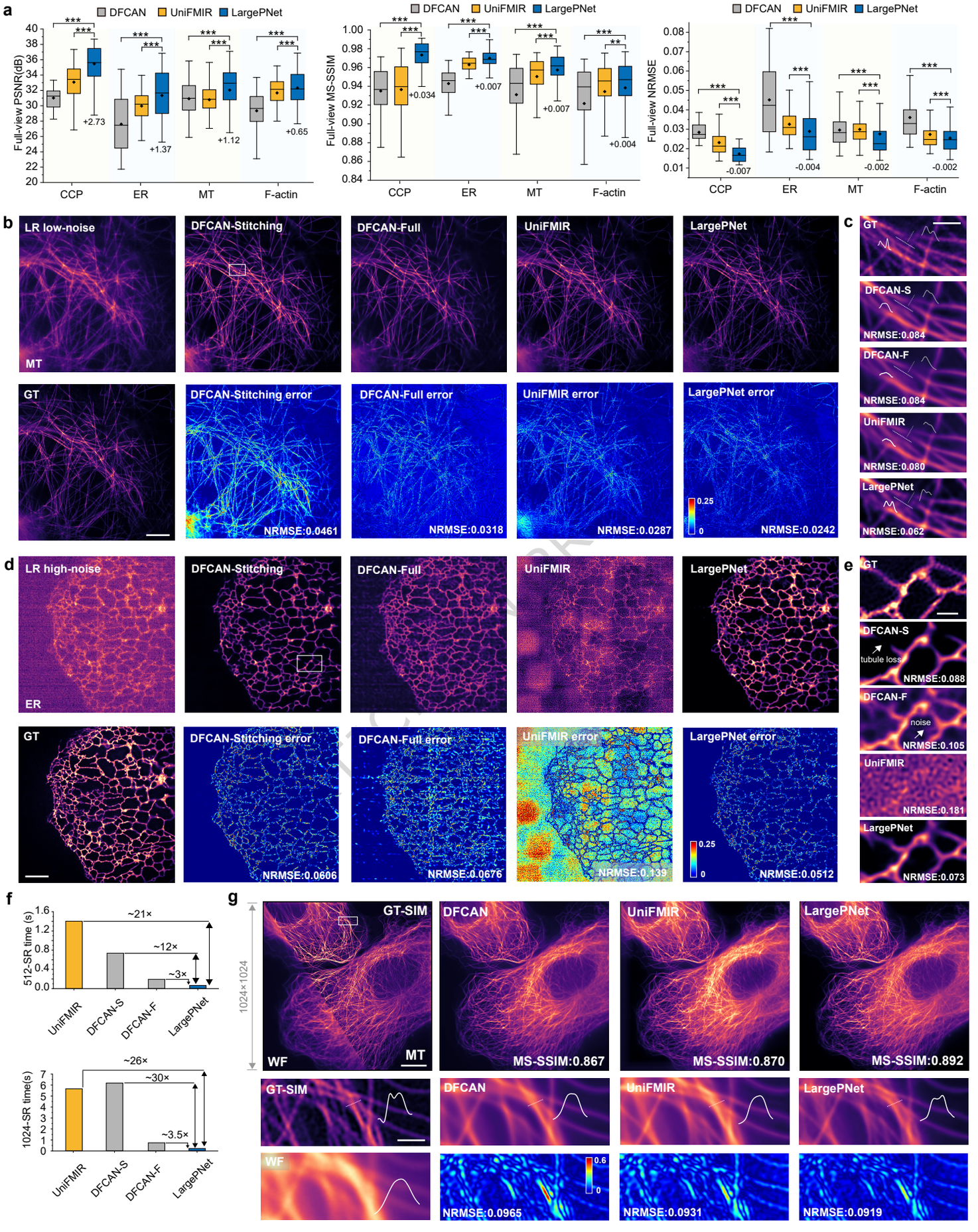
Peer Review Information: *Nature Communications* thanks Ian Dobbie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

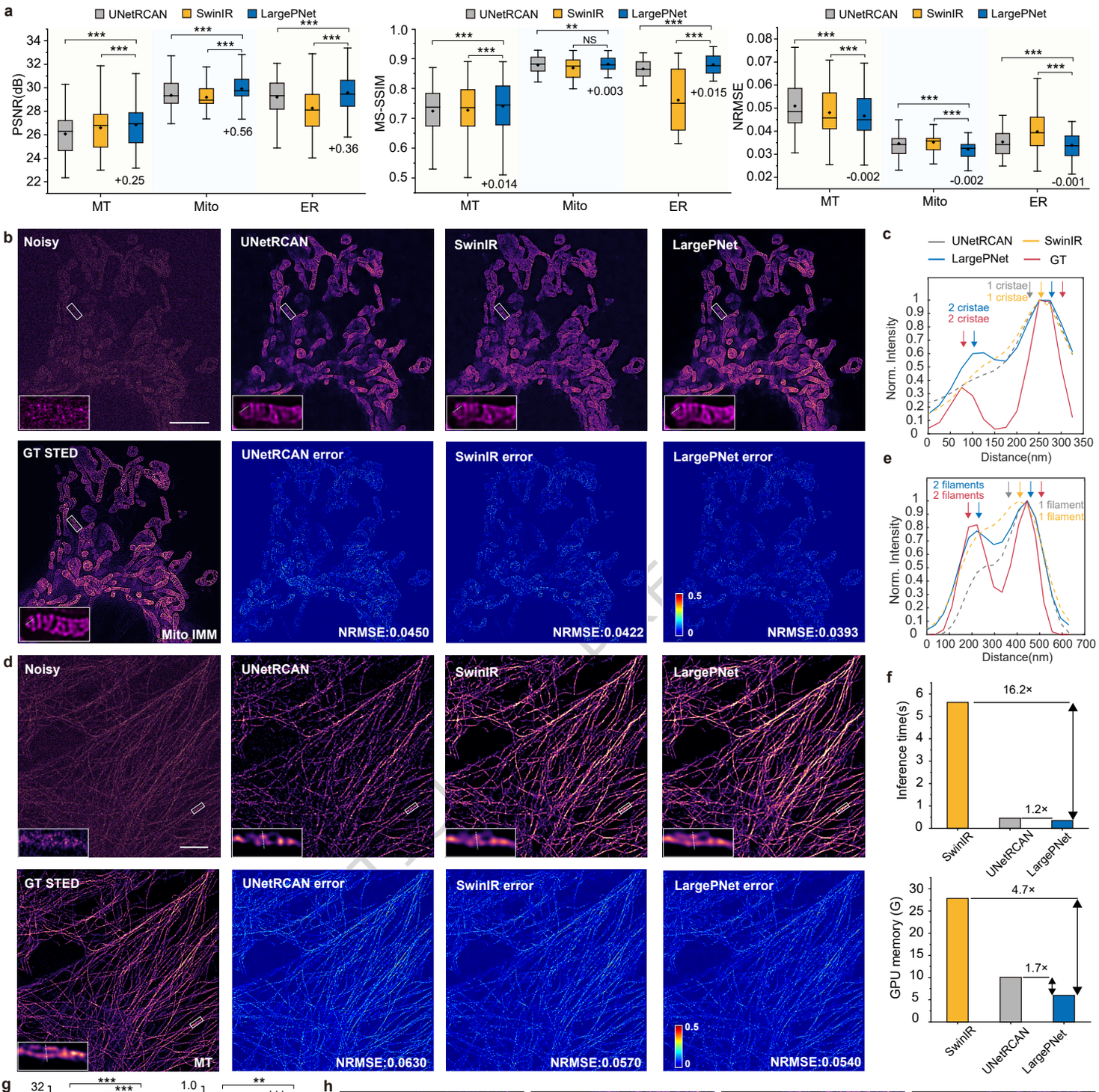

a
Degenerated image
Shallow extraction
Down16x
Down8x
Down4x
Down2x
LF Feature extraction
LF Feature extraction
LF Feature extraction
LF Feature extraction
HF Feature extraction
Up32x
Up16x
Up8x
Up4x
Up2x
Feature merging
Network output

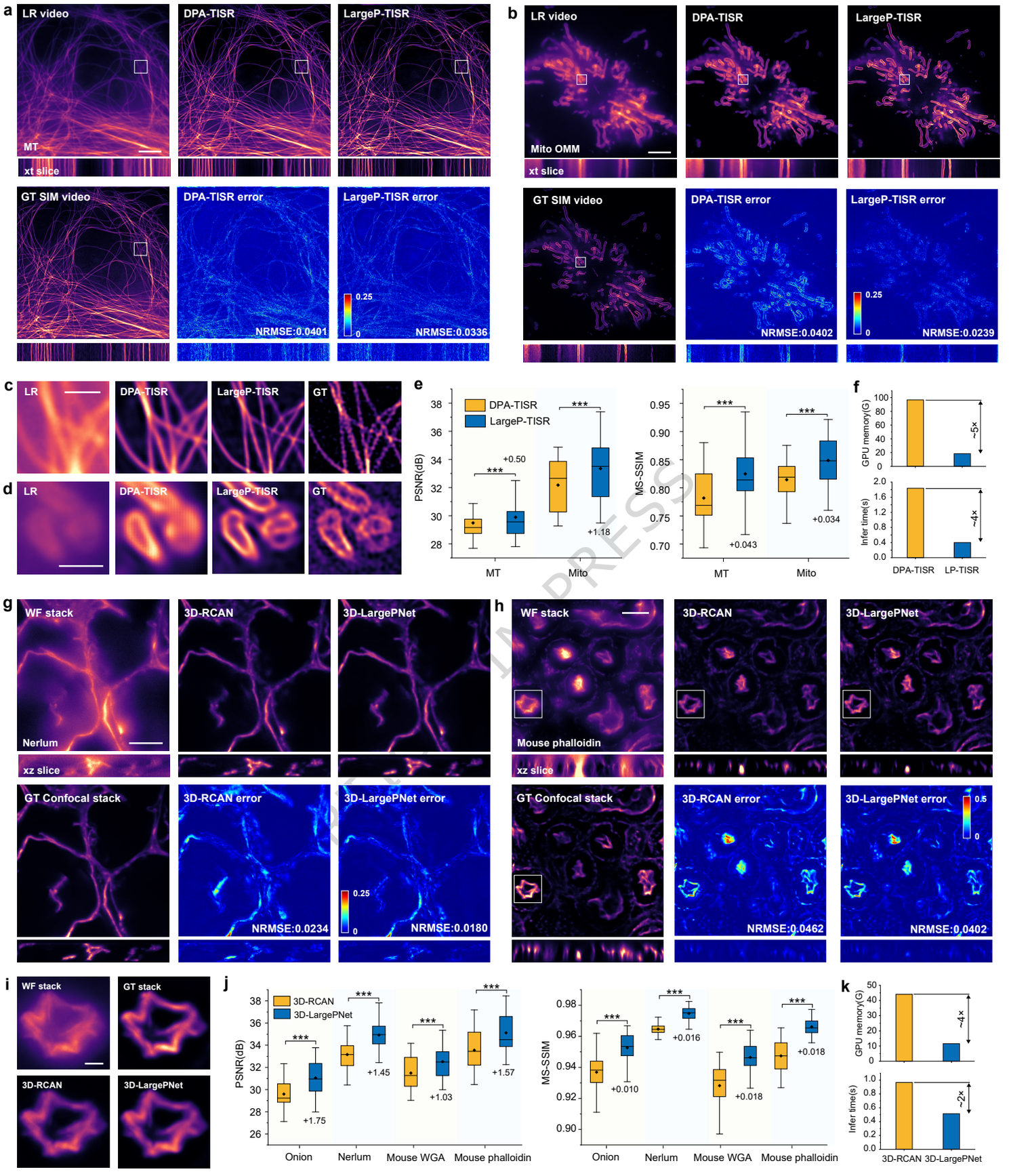

b
Conventional deep CNN
Feature
3x3 Conv groups
Small ERF
LargePNet
Feature
RepLKConv
Large ERF
c
RCAN
128×128
SwinIR
128×128
LargePNet
1
0
d
HFFE Feature Map
LFFE Feature Map
Down 2x
Down 4x
Down 8x
Down 16x

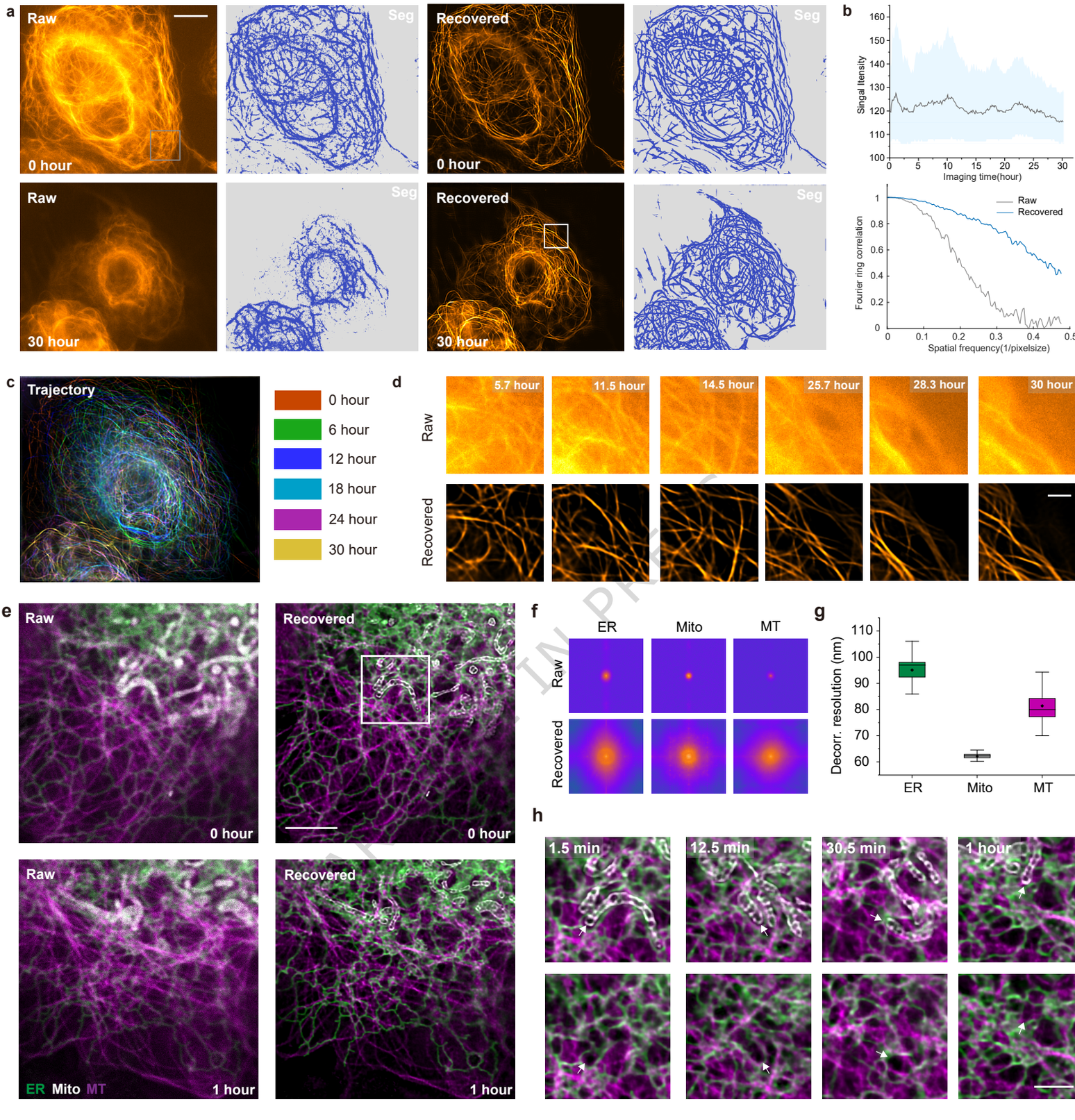

e
Conventional neural network training
Augmented patches
Augmented patches
LargePNet training
Full-view images
or larger patches
Full-view images
or larger patches


f
GPU memory(G)
~4×
~1.8×
UniFMIR
UNet
RCAN
DFCAN
LargePNet
16×128×128 batch
512×512
MACs(G)
~5×
Training time(s)
~5×
~1.5×


a
DFCAN
UniFMIR
LargePNet
Full-view PSNR(dB)
Full-view MS-SSIM
Full-view NRMSE
CCP
ER
MT
F-actin
+2.73
+1.37
+1.12
+0.65
+0.034
+0.007
+0.007
+0.004
-0.007
-0.004
-0.002
-0.002
b
LR low-noise
MT
DFCAN-Stitching
DFCAN-Full
UniFMIR
LargePNet
GT
DFCAN-Stitching error
DFCAN-Full error
UniFMIR error
LargePNet error
NRMSE:0.0461
NRMSE:0.0318
NRMSE:0.0287
NRMSE:0.0242
c
GT
DFCAN-S
NRMSE:0.084
DFCAN-F
NRMSE:0.084
UniFMIR
NRMSE:0.080
LargePNet
NRMSE:0.062
d
LR high-noise
ER
DFCAN-Stitching
DFCAN-Full
UniFMIR
LargePNet
GT
DFCAN-Stitching error
DFCAN-Full error
UniFMIR error
LargePNet error
NRMSE:0.0606
NRMSE:0.0676
NRMSE:0.139
NRMSE:0.0512
e
GT
DFCAN-S
tubule loss
NRMSE:0.088
DFCAN-F
noise
NRMSE:0.105
UniFMIR
NRMSE:0.181
LargePNet
NRMSE:0.073
f
512-SR time (s)
~21×
~12×
~3×
UniFMIR
DFCAN-S
DFCAN-F
LargePNet
1024-SR time(s)
~26×
~30×
~3.5×
g
GT-SIM
WF
MT
DFCAN
UniFMIR
LargePNet
1024×1024
MS-SSIM:0.867
MS-SSIM:0.870
MS-SSIM:0.892
NRMSE:0.0965
NRMSE:0.0931
NRMSE:0.0919


a
UNetRCAN
SwinIR
LargePNet
PSNR(dB)
MS-SSIM
NRMSE
MT
Mito
ER
+0.25
+0.56
+0.36
+0.014
+0.003
+0.015
-0.002
-0.002
-0.001
NS
b
Noisy
UNetRCAN
SwinIR
LargePNet
GT STED
Mito IMM
UNetRCAN error
SwinIR error
LargePNet error
NRMSE:0.0450
NRMSE:0.0422
NRMSE:0.0393
c
GT
1 cristae
2 cristae
Norm. Intensity
Distance(nm)
d
MT
NRMSE:0.0630
NRMSE:0.0570
NRMSE:0.0540
e
2 filaments
1 filament
f
Inference time(s)
GPU memory (G)
16.2×
1.2×
4.7×
1.7×
g
UNet-GAN
RRDB-GAN
LargeP-GAN
+0.79
+0.006
-0.004
FID (lower is better)
h
Under-sampled SMLM
Well-sampled SMLM
UNet-GAN
RRDB-GAN
LargeP-GAN
UNet-GAN error
RRDB-GAN error
LargeP-GAN error
NRMSE:0.0634
NRMSE:0.0676
NRMSE:0.0560


a
LR video
DPA-TISR
LargeP-TISR
MT
xt slice
GT SIM video
DPA-TISR error
LargeP-TISR error
NRMSE:0.0401
NRMSE:0.0336
b
LR video
DPA-TISR
LargeP-TISR
Mito OMM
xt slice
GT SIM video
DPA-TISR error
LargeP-TISR error
NRMSE:0.0402
NRMSE:0.0239
c
LR
DPA-TISR
LargeP-TISR
GT
d
LR
DPA-TISR
LargeP-TISR
GT
e
DPA-TISR
LargeP-TISR
PSNR(dB)
MS-SSIM
MT
Mito
+0.50
+1.18
+0.043
+0.034
f
GPU memory(G)
Infer time(s)
~5×
~4×
DPA-TISR
LP-TISR
g
WF stack
3D-RCAN
3D-LargePNet
Nerlum
xz slice
GT Confocal stack
3D-RCAN error
3D-LargePNet error
NRMSE:0.0234
NRMSE:0.0180
h
WF stack
3D-RCAN
3D-LargePNet
Mouse phalloidin
xz slice
GT Confocal stack
3D-RCAN error
3D-LargePNet error
NRMSE:0.0462
NRMSE:0.0402
i
WF stack
GT stack
3D-RCAN
3D-LargePNet
j
3D-RCAN
3D-LargePNet
PSNR(dB)
MS-SSIM
+1.75
+1.45
+1.03
+1.57
+0.010
+0.016
+0.018
+0.018
Onion
Nerlum
Mouse WGA
Mouse phalloidin
k
GPU memory(G)
Infer time(s)
~4×
~2×
3D-RCAN
3D-LargePNet


a
Raw
Seg
Recovered
Seg
0 hour
0 hour
Raw
Seg
Recovered
Seg
30 hour
30 hour
b
Singal Itensity
Imaging time(hour)
Fourier ring correlation
Raw
Recovered
Spatial frequency(1/pixelsize)
c
Trajectory
0 hour
6 hour
12 hour
18 hour
24 hour
30 hour
d
5.7 hour
11.5 hour
14.5 hour
25.7 hour
28.3 hour
30 hour
Raw
Recovered
e
Raw
Recovered
0 hour
0 hour
Raw
Recovered
ER Mito MT
1 hour
1 hour
f
ER
Mito
MT
Raw
Recovered
g
Decorr. resolution (nm)
ER
Mito
MT
h
1.5 min
12.5 min
30.5 min
1 hour