Supplementary Materials

High-fidelity reconstruction of structured illumination microscopy by an amplitude-phase channel attention network with multitemporal information

Junchao Fan, [a, †] **Xiaodong Tang,** [a, †] **Xiuli Bi,** [a, †] **Weisheng Li,** [a] **Bin Xiao,** [a, *] **Xiaoshuai Huang,** [b, c, *]

[a]Chongqing Key Laboratory of Image Cognition, College of Computer Science and Technology, Chongqing, Chongqing University of Posts and Telecommunications, China, 400065

[b]Biomedical Engineering Department, Peking University, Beijing, China, 100871

[c]International Cancer Institute, Peking University, Beijing, China, 100191

*Correspondence Author, E-mail: Bin Xiao, xiaobin@cqupt.edu.cn; Xiaoshuai Huang, hxs@hsc.pku.edu.cn

Supplementary Note 1: Explanation of the importance of the phase spectrum.

For image *x*, its frequency domain F_x is composed of amplitude A_x and phase P_x as follows:

$$F_x = A_x \otimes e^{i \bullet P_x} \tag{1}$$

where ⊗ indicates the elementwise multiplication of two matrices. We can obtain the complex frequency values of Image1 and Image2 by DFT. It can be expressed as follows:

$$image_1^F(\mu,\upsilon) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f_1(x,y) \bullet e^{-i2\pi(\frac{\mu x}{M}+\frac{\upsilon y}{N})} \tag{2}$$

$$image_2^F(\mu,\upsilon) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f_2(x,y) \bullet e^{-i2\pi(\frac{\mu x}{M}+\frac{\upsilon y}{N})} \tag{3}$$

where the image size is $M \times N$; (x, y) denotes the coordinate of an image pixel in the spatial domain; $f_1(x, y)$ and $f_2(x, y)$ denote the pixel values of Image1 and Image2 respectively; (μ,υ) represents the coordinate of a spatial frequency in the frequency spectrum; $image_1^F(\mu,\upsilon)$ and $image_2^F(\mu,\upsilon)$ denote the complex frequency values of Image1 and Image2 respectively; and *e* and *i* are Euler's number and the imaginary unit, respectively.

Let $R_1(\mu,\upsilon)$, $I_1(\mu,\upsilon)$, $R_2(\mu,\upsilon)$, and $I_2(\mu,\upsilon)$ be the real and imaginary parts of $image_1^F(\mu,\upsilon)$ and $image_2^F(\mu,\upsilon)$, respectively. $image_1^F(\mu,\upsilon)$ and $image_2^F(\mu,\upsilon)$ can be rewritten as follows:

$$image_1^F(\mu,\upsilon) = R_1(\mu,\upsilon) + I_1(\mu,\upsilon)i \tag{4}$$

$$image_2^F(\mu,\upsilon) = R_2(\mu,\upsilon) + I_2(\mu,\upsilon)i \tag{5}$$

The amplitudes of Image1 and Image2 are defined as follows:

$$A_1 = | image_1^F(\mu,\upsilon) | = \sqrt{R_1(\mu,\upsilon)^2 + I_1(\mu,\upsilon)^2} \tag{6}$$

$$A_2 = | image_2^F(\mu,\upsilon) | = \sqrt{R_2(\mu,\upsilon)^2 + I_2(\mu,\upsilon)^2} \tag{7}$$

The phases of Image1 and Image2 are defined as follows:

$$P_1 = \arctan(\frac{I_1(\mu,\upsilon)}{R_1(\mu,\upsilon)}) \tag{8}$$

$$P_2 = \arctan(\frac{I_2(\mu,\upsilon)}{R_2(\mu,\upsilon)}) \tag{9}$$

According to Eq. (1), by combining the phase spectrum P_1 of Image1 and the amplitude spectrum A_2 of Image2, we can obtain the complex frequency values of Combination1. Similarly, by combining the amplitude spectrum A_1 of Image1 and the phase spectrum P_2 of Image2, we can obtain the complex frequency values of Combination2. It can be formulated as follows:

$$Combination_1^F = A_2 \otimes e^{iP_1} \tag{10}$$

$$Combination_2^F = A_1 \otimes e^{iP_2} \tag{11}$$

To obtain the spatial domain representation of Combination1 and Combination2, we used the inverse discrete cosine transform (IDFT).

$$Combination1 = IDFT(Combination_1^F) \tag{12}$$

$$Combination2 = IDFT(Combination_2^F) \tag{13}$$

When the phase spectrum of the reconstructed image is from Image1, it has the content and texture of Image1. On the other hand, when the phase spectrum of the reconstructed image comes from Image2, it has the content and texture of Image2 in Fig. 1b. This means that the phase information of an image contains the texture details of the image, which is very important for image recovery tasks. Therefore, in addition to the amplitude spectrum, we cannot ignore the role played by the phase spectrum.

Supplementary Note 2: Amplitude-phase channel attention blocks

The input of APCAB first goes through two 3×3 convolutional layers to further extract features. It can be denoted as

$$f = Conv_{3\times3}(ReLU(Conv_{3\times3}(z))) \tag{14}$$

where *z*, *f*, and *ReLU* denote the input of APCAB, the feature map of the output of the two 3×3 convolutional layers, and the activation function of rectified linear unit, respectively.

The operation $ASAM_{out}$, which indicates the output of the amplitude spectrum attention module, is denoted as follows:

$$ASAM_{out} = f \times S(Conv_{3\times3}(Cat(Pooling_{avg}(A\_Att_{in}), Pooling_{\max}(A\_Att_{in}))) \tag{15}$$

where

$$A\_Att_{in} = ReLU(Conv_{3\times3}(abs(FFT(f))^{r})) \tag{16}$$

FFT represents the fast Fourier transform, and *r* is applied to enhance the contributions of high-frequency components. $Pooling_{avg}$ and $Pooling_{\max}$ represent the global average pooling and global max pooling, respectively, such that each feature map of A_Att_{in} can be compressed to a representative value. To utilize the results of average and max pooling, the two output vectors obtained are concatenated and fed into a 3×3 convolutional layer, which is similar to the efficient channel attention [1], to reduce the number of parameters of the network. *S* denotes the sigmoid activation function. After a series of operations, they result in a gating mechanism that can self-adaptively calculate the final rescaling factors.

Similar to the operation of $ASAM_{out}$, $PSAM_{out}$, which indicates the output of the phase spectrum attention module, is denoted as follows:

$$PSAM_{out} = f \times S(Conv_{3\times3}(Cat(Pooling_{avg}(P\_Att_{in}), Pooling_{\max}(P\_Att_{in})))) \tag{17}$$

where

$$P\_Att_{in} = ReLU(Conv_{3\times3}(\arctan(imag(FFT(f)+\varepsilon), real(FFT(f)+\varepsilon)) \tag{18}$$

and the *imag* and *real* denote the imaginary and real parts of the *FFT* result, respectively, and ε is applied to prevent division by zero.

To exploit the two results of $ASAM_{out}$ and $PSAM_{out}$, we designed the AWFM to fuse these two outputs in a weighted manner. In contrast to the fixed weighting and concatenated fusion methods, the AWFM allows the network to automatically learn the importance of the two branches. The AWFM is denoted as follows:

$$\alpha, \beta = Softmax(Conv_{1\times1}(\mathrm{Re}LU(Conv_{1\times1}(Pooling_{avg}(f))))) \tag{19}$$

where α and β denote the weight vectors of the ASAM and PSAM, respectively. Full connection is implemented by a 1×1 convolutional layer. Finally, the output of APCAB can be expressed as follows:

$$APCAB_{out} = z + Conv_{1\times1}(\alpha \otimes ASAM_{out} + \beta \otimes PSAM_{out}) \tag{20}$$

where $\otimes$ indicates the multiplication of the corresponding elements.

REFERENCES

1. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11531-11539.

Supplementary Note 3: Calculation of the evaluation metrics NRMSE, MS-SSIM, and PSNR.

Unlike the RMSE, the NRMSE normalizes the RMSE value to between 0 and 1. The NRMSE is defined as follows:

$$NRMSE = \frac{RMSE}{\bar{O}} \tag{21}$$

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}(\hat{Y}_{i,j} - Y_{i,j})^2} \tag{22}$$

where $\bar{O}$ is the average of *Y*.

The MS-SSIM metric assesses the similarity index of the overall structure between inferred SR models and ground truth images. In contrast to SSIM, which is a single-scale approach, MS-SSIM resizes an image with different scales and calculates luminance $L(\hat{Y},Y)$, contrast $C(\hat{Y},Y)$ and structure $S(\hat{Y},Y)$ at each scale. Therefore, it can incorporate the structures of different scales into the assessment. The MS-SSIM metric is defined as follows:

$$MS-SSIM(\hat{Y},Y) = [L_M(\hat{Y},Y)]^{\alpha_M}\prod_{J=1}^{M}[C_J(\hat{Y},Y)]^{\beta_J}[S_J(\hat{Y},Y)]^{\gamma_J} \tag{23}$$

$L_M(\hat{Y},Y)$, $C_J(\hat{Y},Y)$, and $S_J(\hat{Y},Y)$ are formulated as follows:

$$L_M(\hat{Y},Y) = \frac{2\mu_{\hat{Y}}\mu_Y + C_1}{\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1} \tag{24}$$

$$C_J(\hat{Y},Y) = \frac{2\sigma_{\hat{Y}}\sigma_Y + C_2}{\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2} \tag{25}$$

$$S_J(\hat{Y},Y) = \frac{\sigma_{\hat{Y}Y} + C_3}{\sigma_{\hat{Y}} + \sigma_Y + C_3} \tag{26}$$

where C_1, C_2 and C_3 are nonzero constants of small value used to prevent the denominator from being close to zero. They are calculated as follows:

$$C_1 = (K_1L)^2, C_2 = (K_2L)^2, C_3 = C/2 \tag{27}$$

We set $K_1 = 0.01$ and $K_2 = 0.03$ as suggested.

The PSNR is often used as a metric for image reconstruction quality. It is often defined simply by the mean squared error (MSE). The PSNR is defined as follows:

$$PSNR = 20\log_{10}(\frac{MAX_Y}{MSE}) \tag{28}$$

where

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\|\hat{Y}(i,j) - Y(i,j)\|^2 \tag{29}$$

Supplementary Note 4: Calculation steps of the image resolution estimation algorithm.

For an image $I(r)$, the calculation steps of the previously proposed resolution estimation algorithm, namely, deconvolution analysis [1], are as follows:

1) In the fast Fourier transform (*FFT*) operation, the edge of the input image $I(r)$ is eliminated to eliminate the artifacts caused by the edge effect;
2) *FFT* is performed on the image $I(r)$ after edge elimination, and the result is named $I(k)$;
3) $I(k)$ is normalized using the formula $I_{norm}(k) = I(k)/|I(k)|$;
4) $I_{norm}(k)$ is multiplied by *m* binary masks $M_i(k)$, $i = [1, 2, ..., m]$. The radius of the mask increases uniformly from 0 to 1 in the normalized coordinates of the frequency domain. For example, $r_i = r_{i-1} + \Delta r$; $\Delta r = 1/m$;
5) The Pearson correlation is calculated with $d(r_i)$ between $M_i(k) \cdot I_{norm}(k)$ and $I(k)$. The calculation method is as follows:

$$d(r_i) = \frac{\int \mathrm{Re}\{I(k) M_i(k) I_{norm}(k)\} dk_x dk_y}{\sqrt{\int |I(k)|^2 dk_x dk_y \int |M_i(k) I_{norm}(k)|^2 dk_x dk_y}} \tag{30}$$

   where $\tilde{r}$ represents the radius of the binary mask when $d(r_i)$ reaches the maximum;
6) Using Gaussian high pass filters $H_j(k)$, $j = [1, 2, ..., n]$. The apodization frequency is uniformly increased from 0 to 1 in the frequency domain normalized coordinates to obtain the high-frequency information component left by *I*(*k*) after passing through each high pass filter, $H_1(k) \bullet I(k)$, $H_2(k) \bullet I(k) ...$, $H_n(k) \bullet I(k)$. For each left high-frequency information component, we repeat steps (3) to (5) to calculate $\tilde{r}$;
7) The resolution of the final $I(r)$ can be calculated by the following formula:

$$Resolution = \frac{2 \times d_{xy}}{\max(\tilde{r}_1, \tilde{r}_2, ..., \tilde{r}_n)} \tag{31}$$

In this experiment, we set $m = 40$ and $n = 10$.

REFERENCES

1. A. Descloux, K. S. Grußmayer, and A. Radenovic, "Parameter-free image resolution estimation

based on decorrelation analysis," Nature methods, vol. 16, pp. 918-924, 2019.

Supplementary Figures:

Fig. S1. The amplitude-phase channel attention block (APCAB). An APCAB block, which is included in the residual groups, consists of amplitude-phase dual branch channel attention and an adaptive weight fusion module. Four amplitude-phase channel attention blocks constitute a residual group in APCAN.

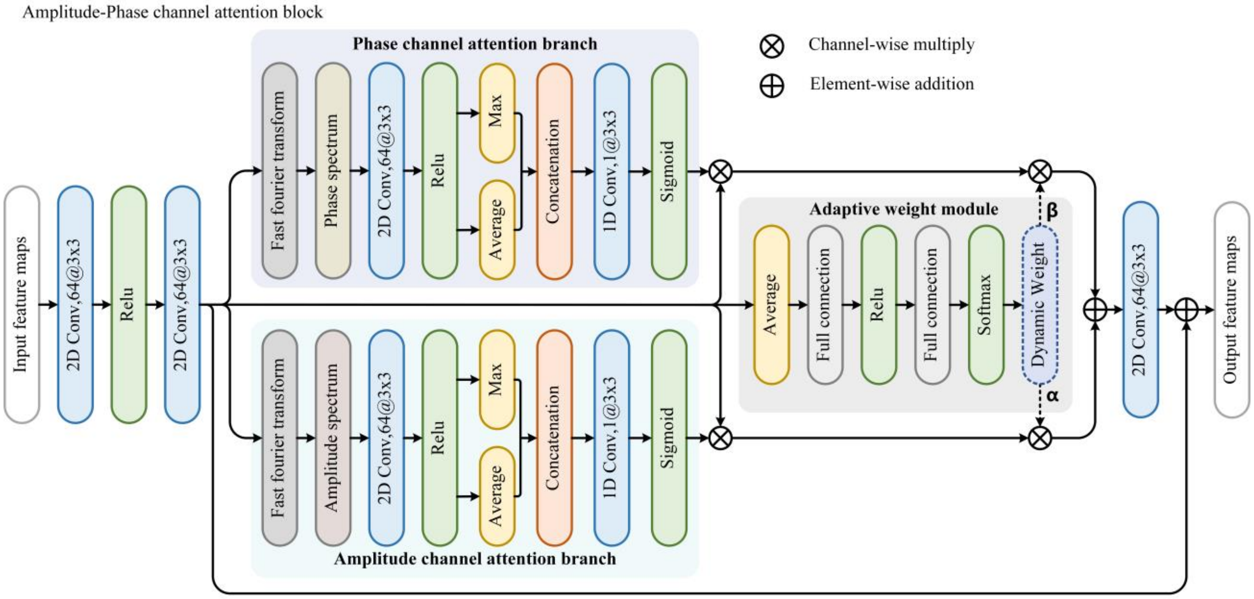

Fig. S2. Comparison of the reconstruction results of Wiener, scUNet, DFCAN and APCAN on the simulated dataset. a, Process of constructing the simulated dataset, which includes the generation of the SIM raw data and the corresponding ground truth. b, Comparison of the Wiener, scUNet, DFCAN and APCAN reconstruction results and the corresponding ground truth on the simulated dataset. c, Evaluation results of the Wiener, scUNet, DFCAN, and APCAN reconstruction images in terms of the three metrics NRMSE, MS-SSIM, and PSNR, with n=50. The simulation experimental results show that noise has a large impact on the reconstruction process, especially for traditional methods, which leads to more artifacts in column 2 of Fig. S2b. For the deep learning methods, since they all have strong noise immunity, they can obtain better reconstruction results. However, the reconstruction quality of APCAN is considerably better than that of DFCAN, which indicates that employing both amplitude and phase information can improve the reconstruction performance of the network in columns 4 and 5 of Fig. S2b.

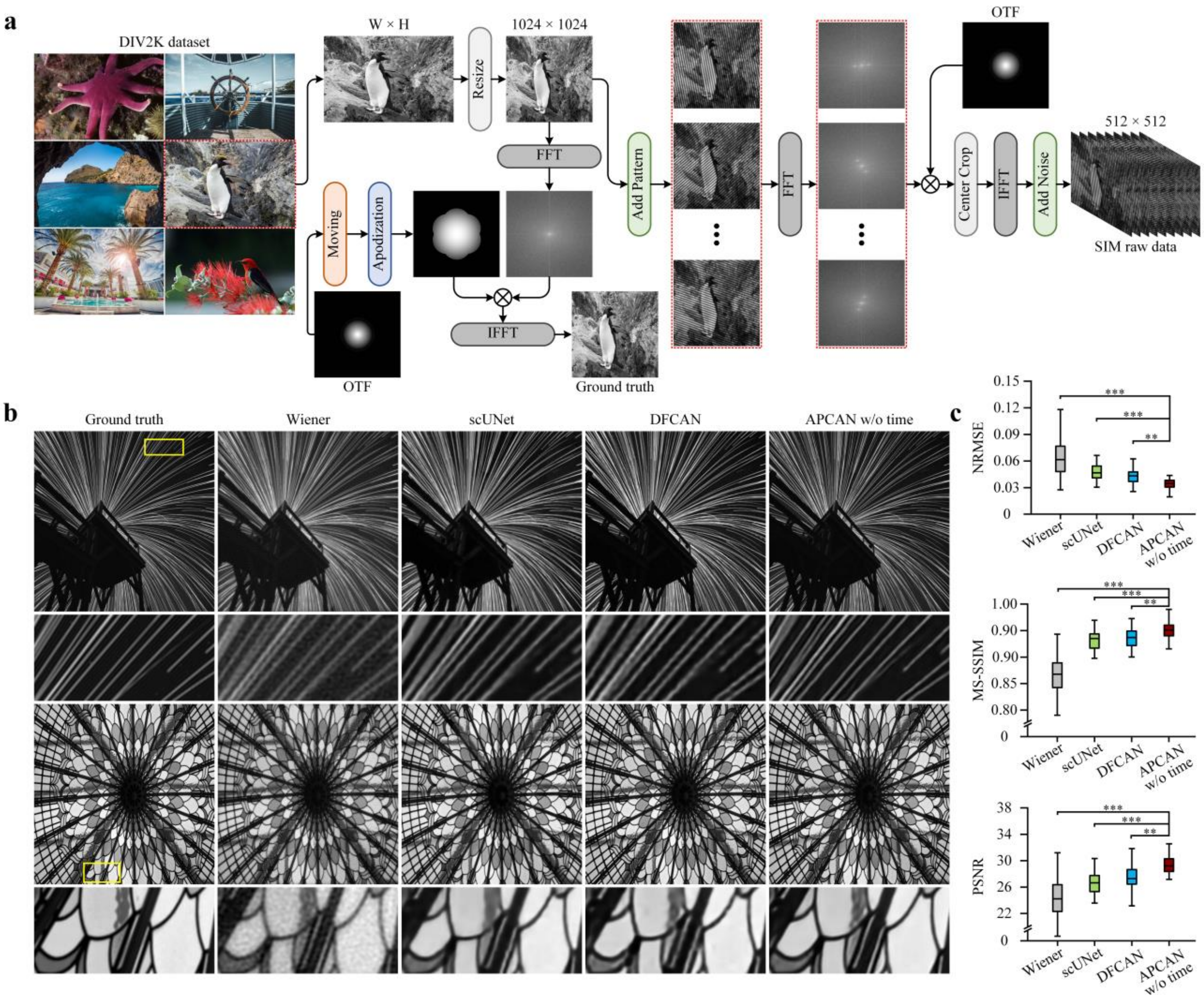

Fig. S3. a, Reconstructions of scUNet, DFCAN, and APCAN (first row) and the difference between these reconstructions and the ground truth image (second row) at high SNR. b, The profile along the line in (a). Scale bar: 2 μm; Axial: 0.6 arbitrary units (a.u.); Lateral: 0.5 μm.

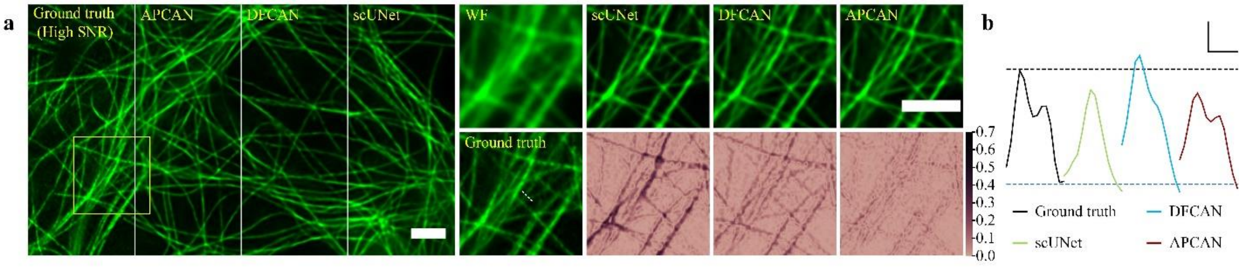

Fig. S4. Super-resolution reconstruction comparison of scUNet, DFCAN, and APCAN for F-actin in BioSR dataset. a, Difference between the ground truth and the reconstructed F-actin of scUNet, DFCAN, and APCAN w/o time under three different signal-to-noise ratios condition. b, Evaluation

results of the reconstructed SR images of scUNet, DFCAN, and APCAN w/o time in terms of NRMSE, MS-SSIM, and PSNR metrics, for SIM raw data with different signal-to-noise ratios, with sample size n=64. Scale bar: 2 μm.

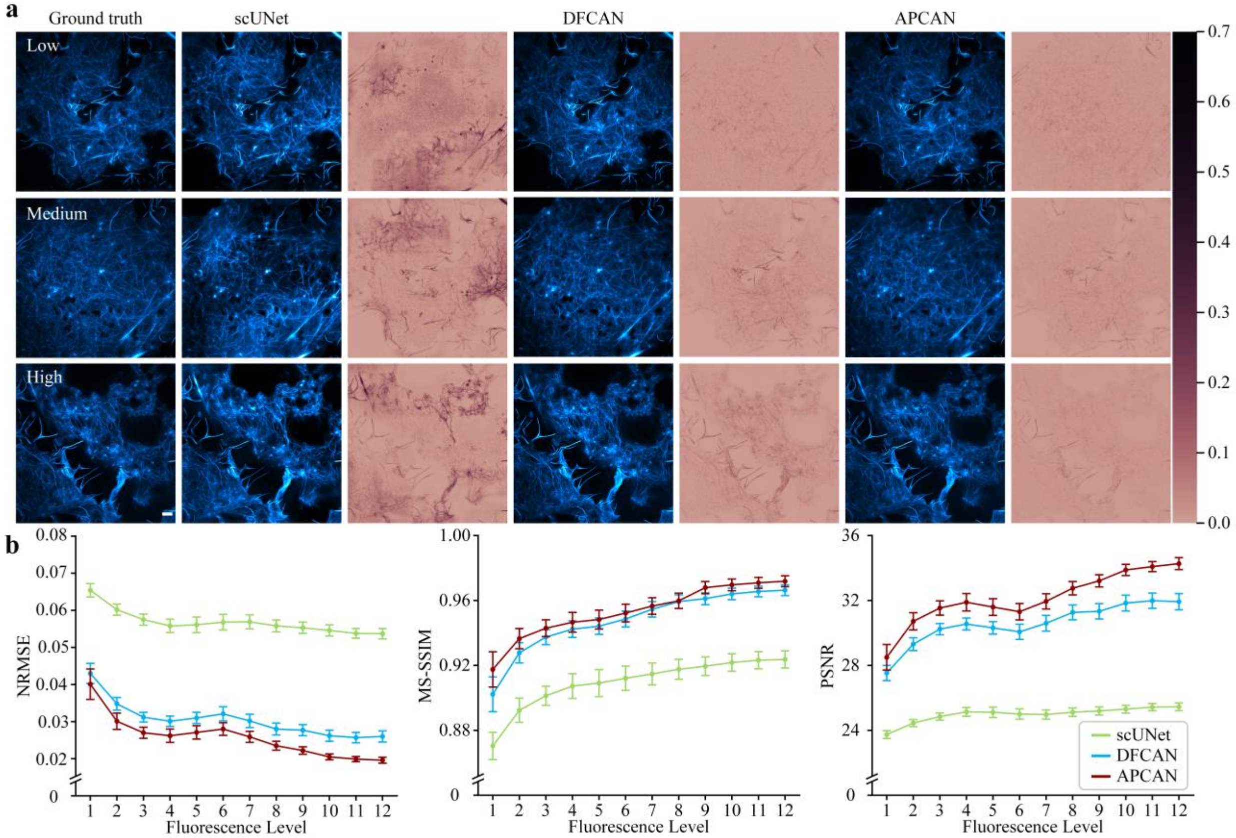

Fig. S5. Super-resolution reconstruction comparison of scUNet, DFCAN, and APCAN for ER data in BioSR dataset. a, Difference between the ground truth and the reconstructed ER of scUNet, DFCAN, and APCAN w/o time under three different signal-to-noise ratios condition. b, Evaluation results of the reconstructed SR images of scUNet, DFCAN, and APCAN w/o time in terms of NRMSE, MS-SSIM, and PSNR metrics, for SIM raw data with different signal-to-noise ratios, with sample size n=33. Scale bar: 2 μm.

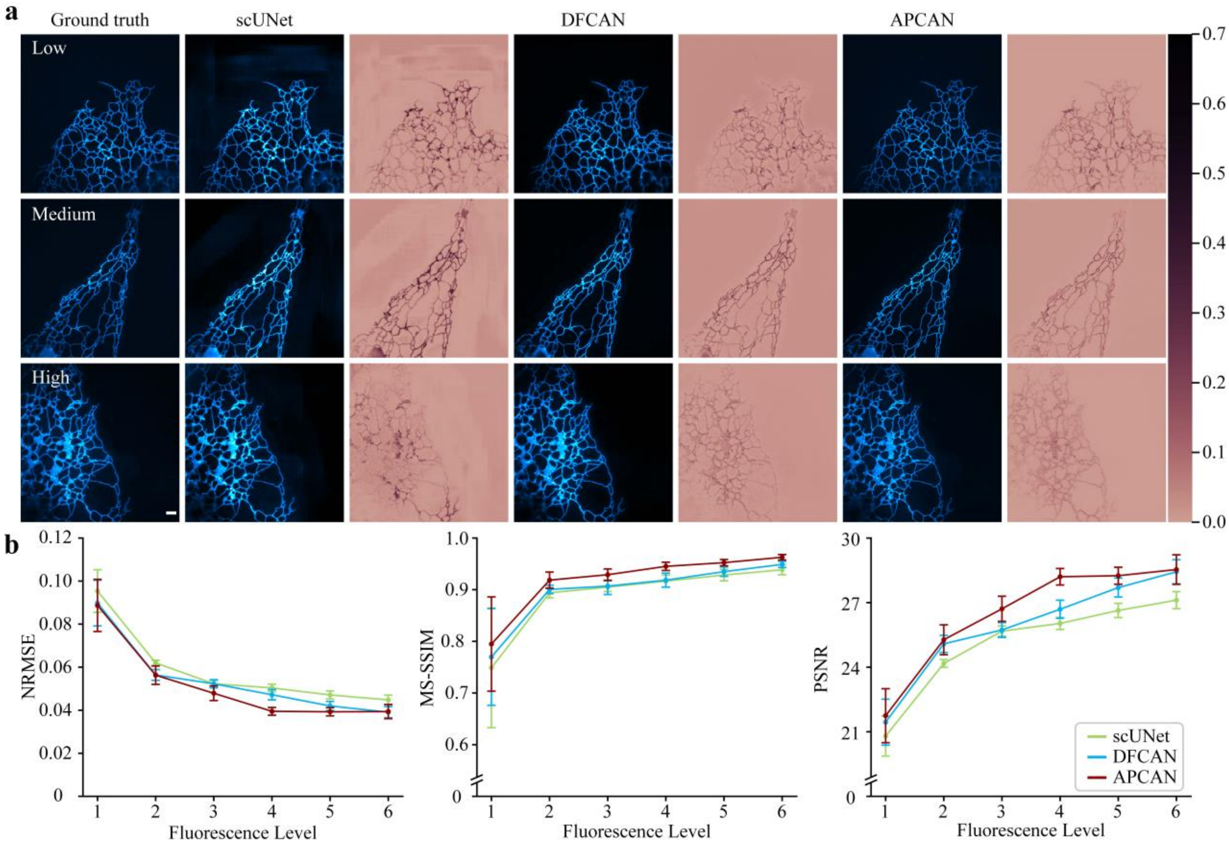

Fig. S6. Super-resolution reconstruction comparison of scUNet, DFCAN, and APCAN for CCPs data in BioSR dataset. a, Difference between the ground truth and the reconstructed CCPs of scUNet, DFCAN, and APCAN w/o time under three different signal-to-noise ratios condition. b, Evaluation results of the reconstructed SR images of scUNet, DFCAN, and APCAN w/o time in terms of NRMSE, MS-SSIM, and PSNR metrics, for SIM raw data with different signal-to-noise ratios, with sample size n=57. Scale bar: 2 μm.

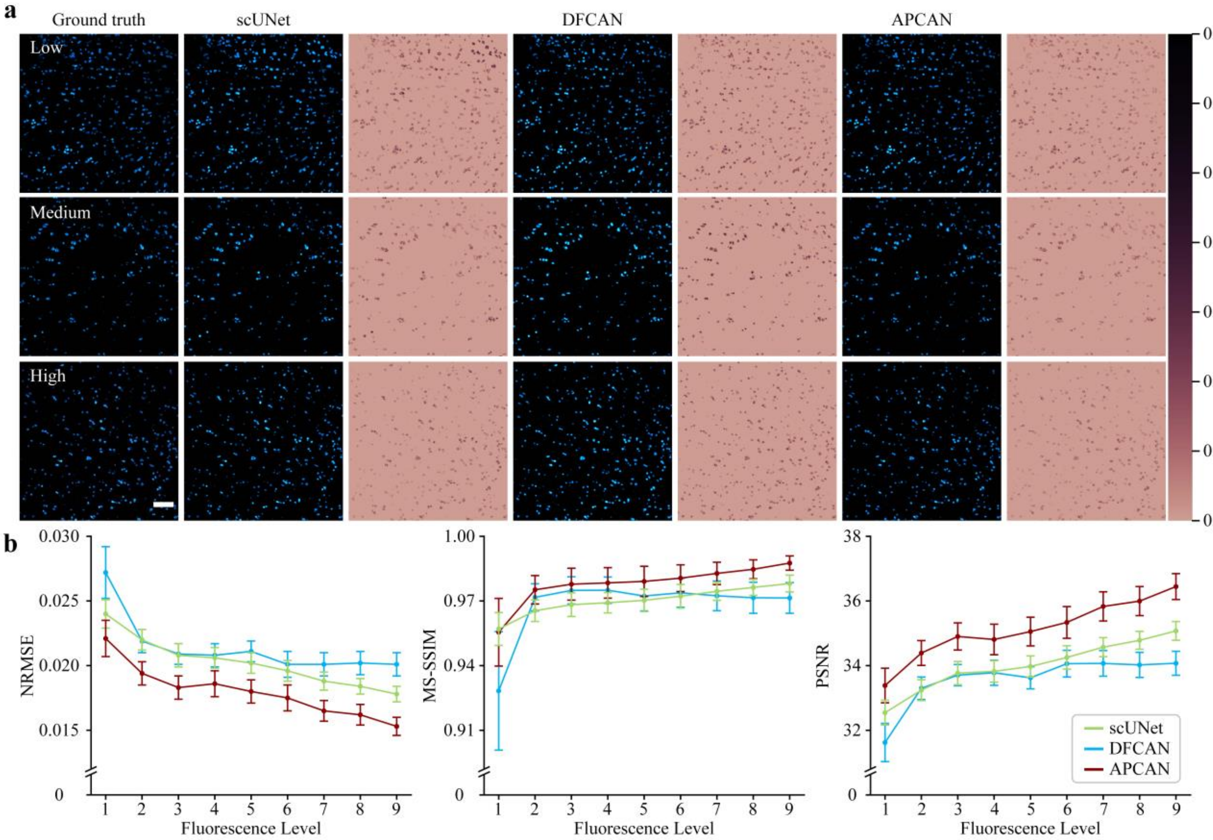

Fig. S7. Mean intensity of SIM raw data over time for ER data. The horizontal coordinate represents the time in seconds, and the vertical coordinate represents the average intensity of 9 frames of SIM raw data. The average intensity of the ER data gradually decreases with time, which indicates that the ER data is affected by photo-bleaching.

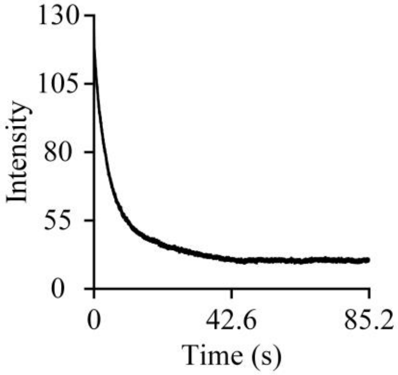

Table S1. Quantitative evaluation results of Wiener, scUNet, DFCAN, and APCAN on the simulated dataset.

	NRMSE	MS-SSIM	PSNR
Wiener	0.0649±0.0052	0.8646±0.0187	24.1786±0.6897
scUNet	0.0505±0.0044	0.9271±0.0166	26.3110±0.5958
DFCAN	0.0439±0.0031	0.9355±0.0099	27.4369±0.5669
APCAN	**0.0354±0.0028**	**0.9502±0.0089**	**29.4527±0.7053**

Table S2. Divided number of low, medium, and high fluorescence levels for the BioSR test data. For BioSR, since each dataset contains SIM raw data with different fluorescence levels, to facilitate testing, we divided the fluorescence levels into three types: low fluorescence level, medium fluorescence level, and high fluorescence level. For the F-actin dataset, there were 12 fluorescence levels in each group of data. Therefore, we classified 1 to 4 fluorescence levels as low fluorescence levels, 5 to 8 fluorescence levels as medium fluorescence levels, and 9 to 12 as high fluorescence levels. For microtubules and CCPs, because each group of data contains 9 fluorescence levels, we classified 1 to 3 fluorescence levels as low fluorescence levels, 4 to 6 as medium fluorescence levels, and finally 7 to 9 as high fluorescence levels. For ER, because each group of data only contains 6 fluorescence levels, we classified 1 to 2 fluorescence levels as low fluorescence levels, 3 to 4 fluorescence levels as medium fluorescence levels, and 5 to 6 fluorescence levels as high fluorescence levels. Table S4 shows the number of test datasets we split from the four datasets.

	Test			Total
	low fluorescence level	medium fluorescence level	high fluorescence level	
F-actin	64	64	64	192
Microtubules	60	60	60	180
CCPs	57	57	57	171
ER	33	33	33	99

Table S3. Quantitative evaluation results of scUNet, DFCAN, and APCAN on BioSR at low fluorescence levels.

Models	Metrics	F-actin	Microtubules	ER	CCPs
scUNet	NRMSE	0.0594±0.0074	0.0483±0.0077	0.0724±0.0230	0.0222±0.0039
	MS-SSIM	0.8674±0.0123	0.9104±0.0140	0.8477±0.0588	0.9638±0.0064
	PSNR	24.5817±0.2666	26.4198±0.3431	23.0973±0.5654	33.2082±0.3697
DFCAN	NRMSE	0.0344±0.0086	0.0385±0.0117	0.0679±0.0245	0.0225±0.0044
	MS-SSIM	0.9108±0.0125	0.9125±0.0173	0.8586±0.0495	0.9639±0.0130
	PSNR	29.4978±0.4753	28.6068±0.5745	23.7635±0.6762	33.0900±0.4035
APCAN	NRMSE	0.0304±0.0111	0.0354±0.0142	0.0642±0.0279	0.0198±0.0045
	MS-SSIM	0.9209±0.0111	0.9310±0.0159	0.8768±0.0478	0.9700±0.0114
	PSNR	30.7648±0.6448	29.5800±0.7520	24.4008±0.7940	34.2558±0.4663

Table S4. Quantitative evaluation results of scUNet, DFCAN, and APCAN on BioSR at the medium fluorescence level.

Models	Metrics	F-actin	Microtubules	ER	CCPs
scUNet	NRMSE	0.0564±0.0076	0.0450±0.0064	0.0533±0.0081	0.0202±0.0033
	MS-SSIM	0.8917±0.0090	0.9275±0.0105	0.9083±0.0142	0.9706±0.0051
	PSNR	25.0490±0.2905	27.0213±0.3155	25.5681±0.3200	34.0163±0.3387
DFCAN	NRMSE	0.0303±0.0068	0.0312±0.0082	0.0533±0.0131	0.0207±0.0036
	MS-SSIM	0.9396±0.0068	0.9465±0.0106	0.9060±0.0178	0.9737±0.0065
	PSNR	30.5592±0.4527	30.3657±0.4927	25.7092±0.5143	33.8211±0.3782
APCAN	NRMSE	0.0261±0.0063	0.0244±0.0073	0.0456±0.0135	0.0180±0.0038
	MS-SSIM	0.9428±0.0071	0.9624±0.0085	0.9327±0.0138	0.9793±0.0066
	PSNR	31.8989±0.4859	32.5396±0.5483	27.1348±0.5779	35.0668±0.4639

Table S5. Quantitative evaluation results of scUNet, DFCAN, and APCAN on the BioSR at the high fluorescence level.

Models	Metrics	F-actin	Microtubules	ER	CCPs
scUNet	NRMSE	0.0544±0.0057	0.0445±0.0063	0.0466±0.0087	0.0183±0.0025
	MS-SSIM	0.9026±0.0067	0.9337±0.0084	0.9337±0.0104	0.9763±0.0039
	PSNR	25.3441±0.2339	27.118±0.3102	26.7817±0.3888	34.8096±0.2869
DFCAN	NRMSE	0.0264±0.0059	0.0271±0.0058	0.0413±0.0103	0.0201±0.0035
	MS-SSIM	0.9553±0.0044	0.9631±0.0064	0.9422±0.0082	0.9718±0.0069
	PSNR	31.7704±0.4734	31.5237±0.4410	27.9480±0.5333	34.0550±0.3792
APCAN	NRMSE	0.0205±0.0034	0.0216±0.0049	0.0402±0.0117	0.0160±0.0031
	MS-SSIM	0.9626±0.0044	0.9722±0.0042	0.9574±0.0063	0.9850±0.0044
	PSNR	33.8593±0.3554	33.5157±0.4648	28.2485±0.5899	36.0911±0.4312

Table S6. Quantitative evaluation results of HiFi-SIM, DFCAN, APCAN without time (APCAN w/o time), and APCAN using temporal information (APCAN w time) on the Actin dataset in T-SIM.

	NRMSE	MS-SSIM	PSNR
HiFi-SIM	0.0302±0.0011	0.9219±0.0080	30.4877±0.3126
DFCAN	0.0412±0.0089	0.8856±0.0123	27.8975±0.4573
APCAN w/o time	0.0344±0.0043	0.9006±0.0068	29.3383±0.2649
APCAN w time	**0.0254±0.0008**	**0.9357±0.0064**	**31.9805±0.2605**