Supporting Information

Spatial resolution enhancement in photon-starved STED imaging using deep learning-based fluorescence lifetime analysis

Yuan-I Chen,*[a] Yin-Jui Chang,*[a] Yuansheng Sun,[b] Shih-Chu Liao,[b] Samantha R. Santacruz,[a,c,d] and Hsin-Chih Yeh *[a,e]

[a]Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

[b]ISS, Inc., 1602 Newton Drive, Champaign, IL 61822, USA

[c]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

[d]Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA

[e]Texas Materials Institute, The University of Texas at Austin, Austin, TX 78712, USA

Supplementary Notes

Supplementary Figures

Supplementary Table

Supplementary Notes

Supplementary Note 1. Conversion of decay histogram at each pixel into a phasor point by Fourier transform. Considering the instrument response function (IRF), the noise, $n(t)$, and the measured decay signals, $m(t)$, the response was described as follows,

$$m(t) = IRF \otimes I(t) + n(t)$$

where $I(t)$ represented fluorescence decay before the convolution with the IRF. After applying the fast Fourier transform to the measured data,

$$\vec{M}(\omega) = F(m(t)) = F(IRF \otimes I(t) + n(t)) = F(IRF) \times F(I(t)) + F(n(t))$$

We noticed that in the frequency domain, the convolution operation with the IRF was just a simple multiplication of the IRF vector. Therefore, during the calibration procedure, we would obtain the IRF. Then for the rest of the experiment, we simply subtracted the noise vector and divided the denoised vector by the IRF vector to get the emission fluorescent vector.

Alternatively, the Fourier Transform of the fluorescence decay histogram could be separated into a real number, $g(\omega)$, and an imaginary number, $s(\omega)$.

$$g(\omega) = \frac{\int_0^\infty m(t)\cos \omega t dt}{\int_0^\infty m(t)dt}$$

$$s(\omega) = \frac{\int_0^\infty m(t)\sin \omega t dt}{\int_0^\infty m(t)dt}$$

After calibration, the fluorescent decay histogram (without IRF and noise) at each pixel could be plotted as a single point, termed phasor, on the phasor plot by applying the sine and cosine transforms to the measured decay data, where the modulation frequency ω was the laser repetition angular frequency and was calculated by multiplying the laser repetition rate with 2π.

Supplementary Note 2. VistaVision software for pSTED-SPLIT analysis. VistaVision software is used for instrument control, data acquisition, and data processing (**Fig. S3**). The user can assign the acquisition parameters (e.g., pixel dwell time, image size, image resolution) for the specific purpose. After data acquisition, we conducted the pSTED-SPLIT method in "Multi-image Phasor Analysis," which could be accessed through the top menu of the software. Since the phasor plot can be noisy with limited photon counts, the median filter was applied to denoise and smooth the phasor plot. $\boldsymbol{P}_1$ (G_1, S_1) and $\boldsymbol{P}_2$ (G_2 S_2) phasor components were then assigned manually, followed

by the linear decomposition that separates the long-lifetime photons embedded in each phasor point from the short-lifetime photons (**Fig. S2**). Given the phasor coordinates of the fluorescence decay at an individual pixel is ***P*** (G, S), the fractional contribution of $\boldsymbol{P_1}$ component, f_1, can be obtained by the following equation.

$$\begin{bmatrix} f_1 \\ 1-f_1 \end{bmatrix} = \left(\begin{bmatrix} G_1 & G_2 \\ S_1 & S_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} G \\ S \end{bmatrix}$$

When the $\boldsymbol{P_1}$ component was restored and put back to the original intensity image, the pSTED-SPLIT image was finally obtained (= f_1 × photons per pixel).

Supplementary Note 3. Generative adversarial network structure and training. In this work, we trained a deep neural network using a Wasserstein generative adversarial network (WGAN).[1] While GAN has shown great success with Jensen-Shannon (JS) divergence as the loss function, the difficulty of achieving Nash equilibrium and vanishing gradient makes the GAN training challenging. Instead of JS divergence, Wasserstein distance as the loss function enables the training process to offer strong enough gradients to train the generator as compared with the original GAN. The loss functions of WGAN for the gradient update to the generator and discriminator were as follows,

$$G_{loss} = \frac{1}{n}\sum_{i=1}^{n}\left[-f\left(G(z_i)\right)\right]$$

$$D_{loss} = \frac{1}{n}\sum_{i=1}^{n}\left(-f(x_i) + f\left(G(z_i)\right)\right)$$

where z_i was the normalized low-photon-count fluorescence decay histogram, and x_i represented the normalized ground truth. $G(z)$ was the normalized ground-truth mimicking histogram and $D(x)$ was the probability that x came from the ground-truth decay histogram. Additionally, $f(x)$ was a 1-Lipschitz function that met the following formula: $|f(x_1) - f(x_2)| \leq |x_1 - x_2|$. In the WGAN framework, the function f to calculate the Wasserstein distance was approximated by the discriminator with the weights determined by the clipping parameters. Typically, a high value of the discriminator output was regarded as the ground-truth data, while a lower value meant the low-photon-count histogram.

Another issue in the GAN framework was the mode collapse, where the generator could only produce a single output type. We added two additional loss functions to the G_{loss} to avoid the mode collapse issue. The first one was the mean squared error between the $G(z_i)$ and the x_i in the time domain. The second additional loss was the mean squared error between the $G(z_i)$ and the x_i in the frequency domain. This additional information in the training process forced the generator model to learn the variability in the training dataset, thus stabilizing the training schedule and leading the GAN model to converge.

Generative model (*G*). The model architecture for the generative model consisted of convolutional neural network layers, a multi-task layer with three nodes, a fully-connected decoding layer, and a residual connection (**Fig. S4**). There were two convolutional layers (each had a rectified linear unit activation function, $ReLU(x)$) and pooling

layers (with average operation with the pool size of 2) followed by a flatten layer.

$ReLU(x) = x^+ = \max(0,x)$

The output of the flatten layer was then fed into a multi-task neural network with hard parameter sharing, transforming the high-dimensional output into three tasks. The last layer was the multilayer perceptron with the activation functions as $tanh()$ to force the output range to lie within -1 and 1. Since this last layer mapped the 3 tasks into 256 channels of the fluorescence decay histogram, we termed the layer a decoding layer. Finally, instead of learning a direct mapping from the normalized low-photon-count fluorescence decay histogram to the ground-truth fluorescence decay histogram, we reframed the process with the residual learning framework by introducing a residual connection between the normalized low-photon-count decay histogram and the model's output.

Discriminative model (*D*). The model architecture for the discriminative model consisted of four densely connected neural network layers with 128, 32, 4, and 1 nodes (**Fig. S4**). All the layers except the last one had a sigmoid activation, $S(x)$, forcing the output to lie within the range from 0 to 1, defined as follows,

$$S(x) = \frac{1}{1 + e^{-x}}$$

In contrast, the last layer had the linear activation function to output the scores corresponding to the fluorescence decay histogram fed into *D*.

Network training schedule. In this work, we focused on the training of the generative model. In the generative model training stage, the total iteration was 2,000. We randomly selected ~10% of training samples from the dataset pool within each iteration. The discriminative model was updated five times while the generative model was kept untrainable, then the generative model was updated once while keeping the other one untrainable. The Glorot uniform initializer randomly initialized the generative and discriminative models. The generative model was optimized using Adam optimizer with a learning rate of 1 x 10^{-4}. The discriminative model was optimized using RMSprop optimizer with a starting learning rate of 5 x 10^{-5}. The final generative and discriminative models for each application in this work were selected at around the 700th iteration, which took ~2h to train the model. Training without the discriminative loss and predictive cost could result in over-smoothed images, as the generative model optimized only a specific group of statistical metrics. Therefore, it was imperative to incorporate a discriminator to train the generative model.

Software and hardware used for STED-*flimGANE* development. This analysis was developed with TensorFlow framework version 1.12.0 and Python version 3.6.5 in the Microsoft Windows 10 operating systems. The training was performed on the same operating system.

Supplementary Note 4. Automatic assignment of P_1 and P_2 positions in the denoised phasor plot. Previously, the SPLIT method relied on the manual assignment of the locations of two phasors, P_1 and P_2, which corresponded to the center and the periphery of the excitation spot, respectively. With our high-quality phasor plot, we proposed an automatic procedure to assign the locations with our high-quality phasor plot and then retrieved the desired

signals *via* photon separation. First, as the center of the excitation had the largest S coordinate according to the STED trajectory (**Figs. S4-5**), we identified multiple peak pixels on the *G*-generated S image *via* a local maximum filter. Then the exclusion criteria were applied to exclude the identified peaks if any of the following situations has been met: (1) the distance between peaks was less than 40 pixels, (2) the distance to the border of the image was less than 15 pixels. Second, the decay histograms at these selected peaks and the neighboring pixels were extracted to ensure the identified peaks represented the center of excitation. When the distance of the fluorophore to the center of excitation increased, the depletion effect kicked in and increased the decay rate. After summing all the decay histograms at peak pixels, followed by the normalization, we converted the normalized decay into a phasor. The location of this phasor was determined to be $\boldsymbol{P}_1$. Third, according to the STED trajectory, the phasor of the periphery of the excitation spot was located at the other end of the trajectory. Therefore, $\boldsymbol{P}_2$ was assigned to the other end of the STED trajectory with a certain threshold (e.g., $MAX^{0.9}$, where MAX represents the maximum occurrence on the phasor plot; **Fig. S4**). When evaluating the fluorescence decays at the pixels away from the identified peaks, the phasor location with the threshold of $MAX^{0.9}$ matched well with the peripheral region of the excitation spot. Finally, the STED-*flimGANE* image was obtained by performing the linear decomposition algorithm on the *G*-generated phasors smoothed with a Gaussian filter ($\sigma = 1$).

Assume that a pixel (with the total number of collected photons N) had a phasor located on $\boldsymbol{P}$, equal to $\boldsymbol{P} = f_1\boldsymbol{P}_1 + f_2\boldsymbol{P}_2$, where f_1 and f_2 represented the fractional components of the detected photons. Given $\boldsymbol{f} = [f_1; f_2]$ and $\mathrm{M} = [\boldsymbol{P}_1^{\mathrm{T}}, \boldsymbol{P}_2^{\mathrm{T}}]$, we could write the linear system in the matrix form $\boldsymbol{P} = \mathrm{M}\boldsymbol{f}$. The solution $\boldsymbol{f} = \mathrm{M}^{-1}\boldsymbol{P}$ enabled us to separate the photons emitted by molecules in the center of the excitation ($= f_1 N$). Eventually, the resulting image with improved resolution could be obtained by iterating this process for each pixel in the image. To validate the proposed automatically selected method, we applied this approach to the imaging of the fluorescent beads (**Fig. 2**).

Supplementary Note 5. Generation of the simulated dataset (*in silico*) with the Monte Carlo (MC) method. First, multiple sets of ground truth were determined based on the lifetimes of the natural fluorophores ($\tau_{i,fl}$) and the STED power-induced ($\tau_{i,STED}$). For each ground truth, different photon counts (*pcs*) and the number of duplicates (e.g., 100) were assigned to construct the training dataset. For every training sample, it was assigned a value of natural fluorophore lifetime ($\tau_{i,fl}$), shorter STED-induced lifetime ($\tau_{i,STED}$), and photon counts (*pcs*). The IRF was obtained by averaging across all the pixels of the calibration image taken at the beginning of the experiment. These parameters were employed to generate the probability mass function that described the distribution of the photon arrival time *via* the equation:

$$P(t) = N(IRF(t) \otimes I(t))$$

$$I_i(t) \propto \begin{cases} e^{-t/\tau_{i,STED}} & if\ 0 \le t < T_{STED} \\ e^{-T_{STED}/\tau_{i,STED}} e^{-(t-T_{STED})/\tau_{i,fl}} & if\ t \ge T_{STED} \end{cases}$$

$$N(f(t)) = \frac{f(t)}{sum(f(t))}$$

Given the probability mass function, we performed the Monte Carlo simulation method to extract specified numbers (photon counts, *pcs*) of samples. Those extracted samples were then used to generate the simulated (degraded) decay histogram (**Fig. S5**).

Supplementary Note 6. SNR characterization on phasor plot. The phasor elongation along the direction of the STED trajectory, Δ_d, was mostly affected by STED-induced temporal dynamics modulation (**Fig. S7**). The higher depletion power, the larger Δ_d. The spread along the direction perpendicular to the STED trajectory, Δ_N, represented the SNR, where higher SNR led to smaller Δ_N. We first identified the STED trajectory based on the given phasor plot. We then calculated Δ_N for each pixel. Ultimately, the average Δ_N served as a metric to compare the SNR of the phasor plot among different methods.

Supplementary Note 7. Average photon counts per pixel quantification. Since the fluorescence lifetime analysis becomes unreliable when the number of photons acquired is extremely low, the pixels with extremely low photon counts (e.g., 10 photons) were excluded from the analysis. Due to the perfectly circular shape of the fluorescent beads, we were able to clearly identify their positions. We thus only considered the pixels with the distance to the center less than 10 pixels (~200 nm) when calculating the average photons. Due to stronger background and imperfectly circular shape in the imaging of the fixed COS-7 cells, instead of searching for the pixels with a certain distance to the center, we calculated the average photon from the pixels with the intensity value greater than 10 counts per pixel.

Supplementary Note 8. Spatial resolution quantification. In order to quantify the spatial resolution of the image using FWHM, we needed to identify the location of the beads or nuclear pores in the image of the STED probe. We employed the local maximum filter to extract the center of each bead or nuclear pore. Again, certain exclusion criteria was applied to exclude the identified peaks if meeting any of the following situations: (1) the distance between peaks was less than 10 pixels, (2) the distance to the border of the image was less than 5 pixels. We then drew a line profile for each bead or nuclear pore and fitted the data with the Gaussian function, where the FWHM equals 2.35 multiplied by the σ.

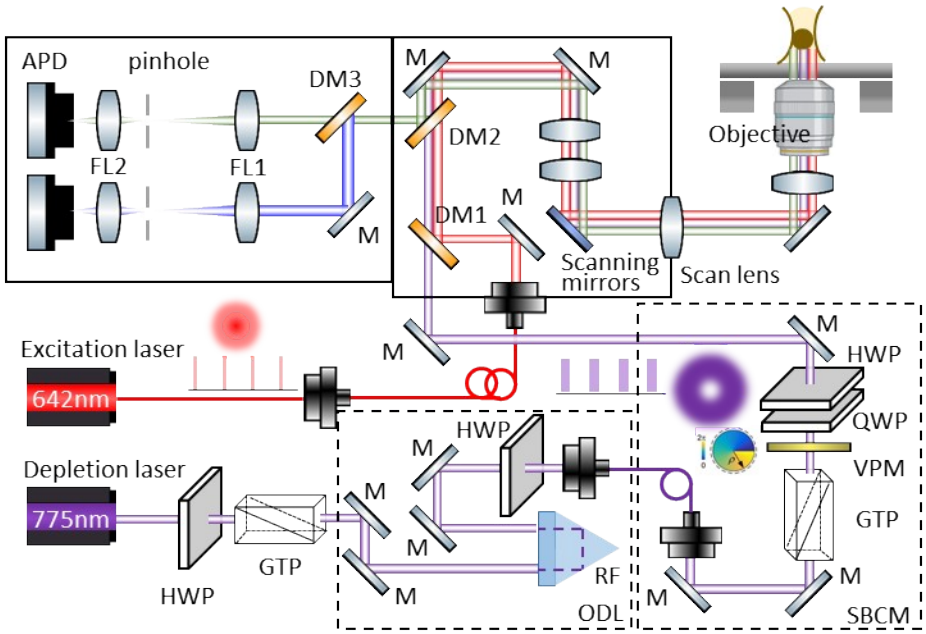

Figure S1 Details of the pulsed STED (pSTED) microscope. HWP: halfwave plate. GTP: Glan-Thompson polarizer. ODL: optical delay line. M: mirror. DM: dichroic mirror. FL: focal lens. RF: retroreflector. SBCM: STED beam conditioning module. QWP: quarter wave plate. VPM: vortex phase mask. APD: avalanche photodiode.

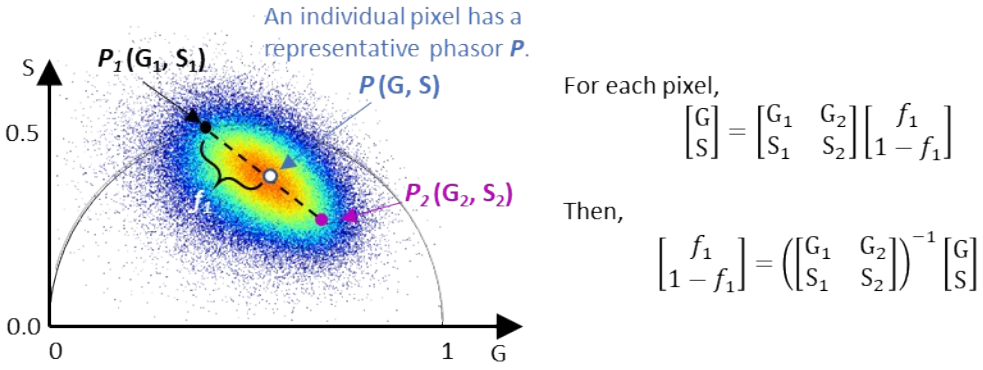

Figure S2 Schematic of linear decomposition algorithm.[2,3] First, $\boldsymbol{P_1}$ (G_1, S_1) and $\boldsymbol{P_2}$ (G_2, S_2) are selected on the phasor plot. Then the following protocol is applied for each measurement at each pixel: (i) force the phasor to the line between $\boldsymbol{P_1P_2}$ and the expected trajectory (black dashed line), leading to the phasor $\boldsymbol{P}$ (G, S); (ii) find the fractional components f_1 by the equations on the right panel. A resulting image is finally obtained by restoring $\boldsymbol{P_1}$ component and putting back to the original intensity image.

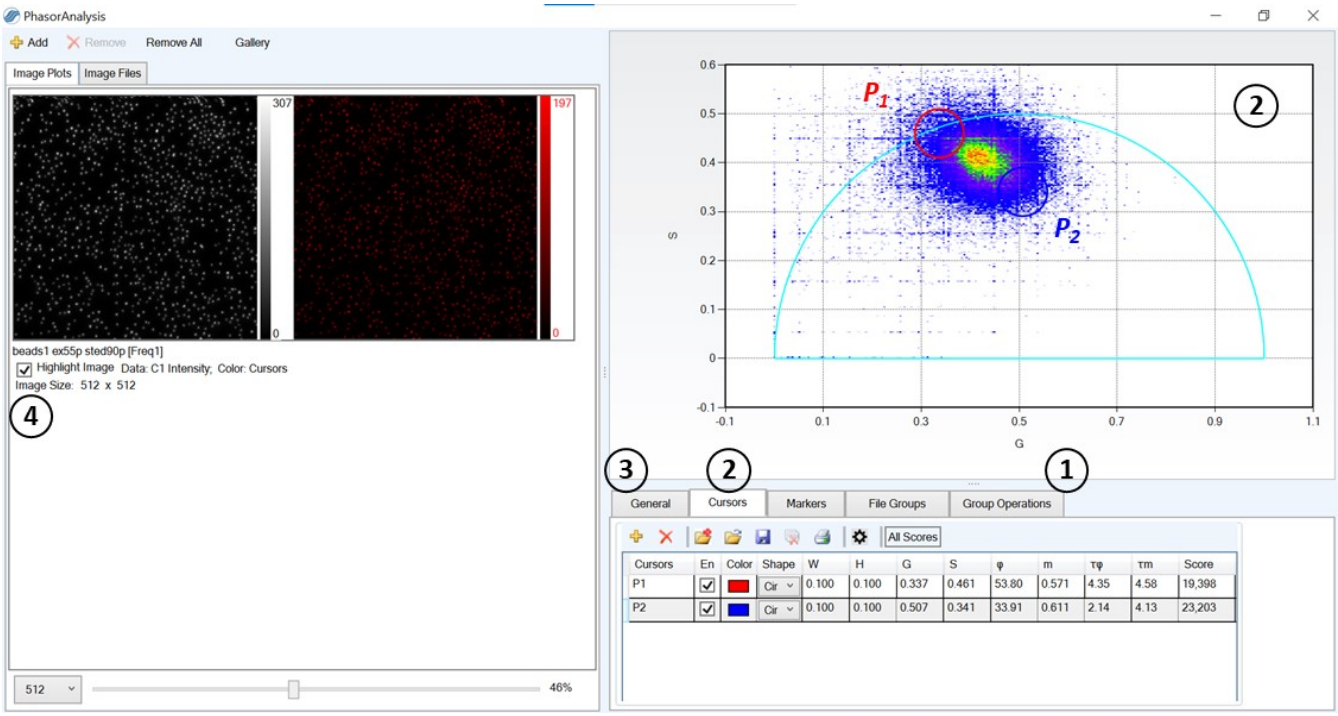

Figure S3 Interface of VistaVision software for pSTED-SPLIT implementation. To obtain the pSTED-SPLIT image, we started with opening the "Multi-image Phasor Analysis" window from "Analysis" at the top menu of the VistaVision software, yielding the "PhasorAnalysis" window. First, we applied the median filter under the "Group Operations" panel. Second, we add two cursors, each representing $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$, under the "Cursors" panel. The $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ were assigned manually based on the phasor plot after smoothing with the median filter. Next, we selected "First two Cs" for "Highlight mode" under the "General" panel. Finally, the pSTED-SPLIT image was obtained and visualized on the right side of the raw STED image by checking the "Highlight Image".

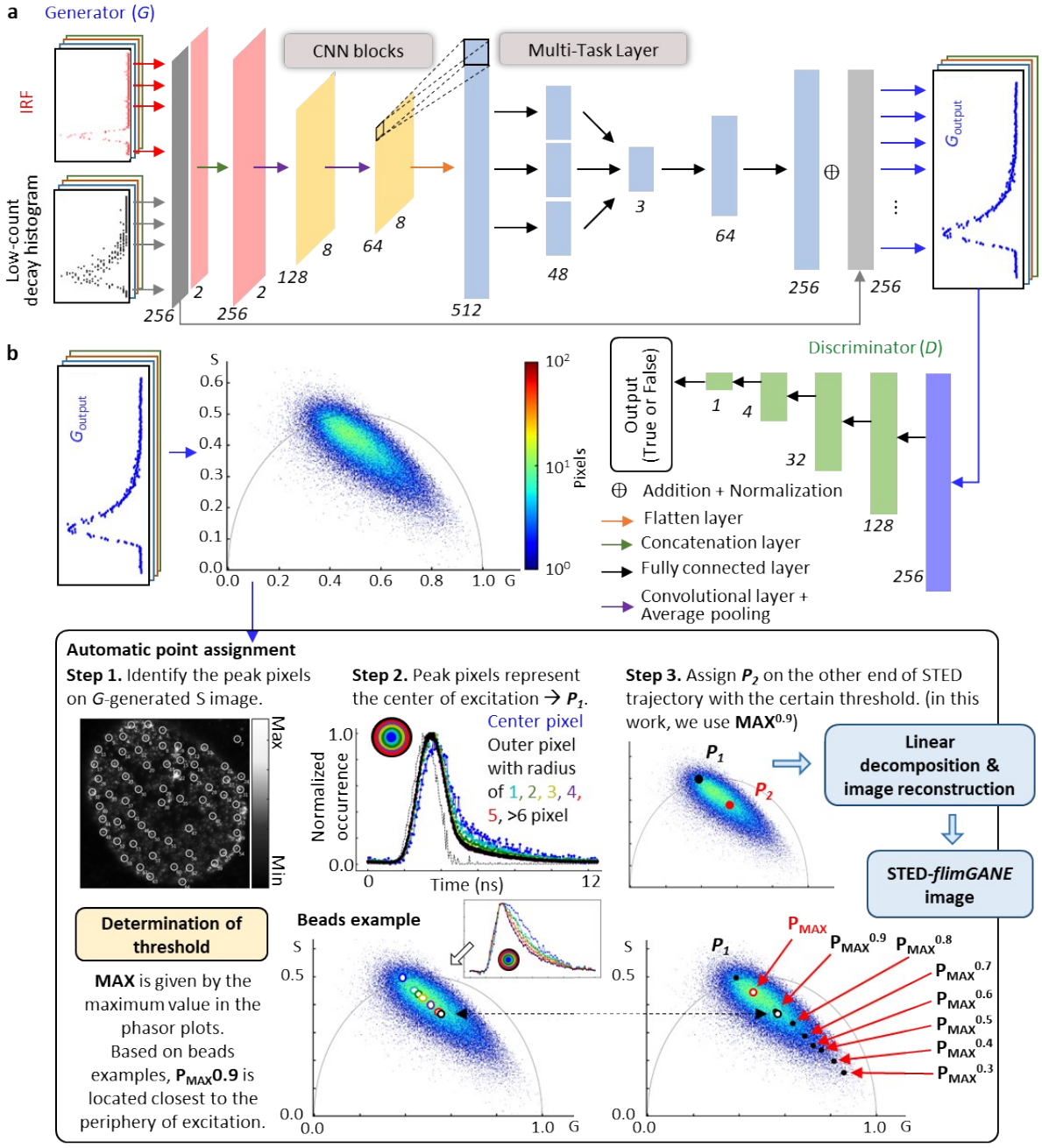

Figure S4 Schematic of the STED-*flimGANE* method. (a) A generator (*G*) transformed the acquired decay curve into a high-photon-count decay. It comprised a concatenation layer, two CNN blocks, each of which comprised one convolutional layer followed by an average pooling layer. The CNN section was followed by a flatten layer. Then a multi-task layer converted data into virtual lifetime parameters, followed by two fully connected layers. Skip connection was used to pass data between layers of the same level. Discriminator (*D*) consisted of four fully connected layers. (b) G_{output} from the well-trained generator was converted into phasor domain for automatic assignment of $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. First, peak pixels on *G*-generated S image were identified. Second, their corresponding decay histograms and the neighboring pixels were extracted to ensure the selected peaks represented the center of excitation. The phasor coordinates of these center pixels were determined to be $\boldsymbol{P_1}$. Third, $\boldsymbol{P_2}$ was

assigned on the other end of the STED trajectory with a certain threshold (e.g., $MAX^{0.9}$, where MAX represented the maximum occurrence on the phasor plot). The representative beads example showed that the location with the threshold of $MAX^{0.9}$ matched well with the fluorescence decays at the periphery of the excitation. Finally, the STED-*flimGANE* image was obtained by photon separation with our automatic assignment of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$.

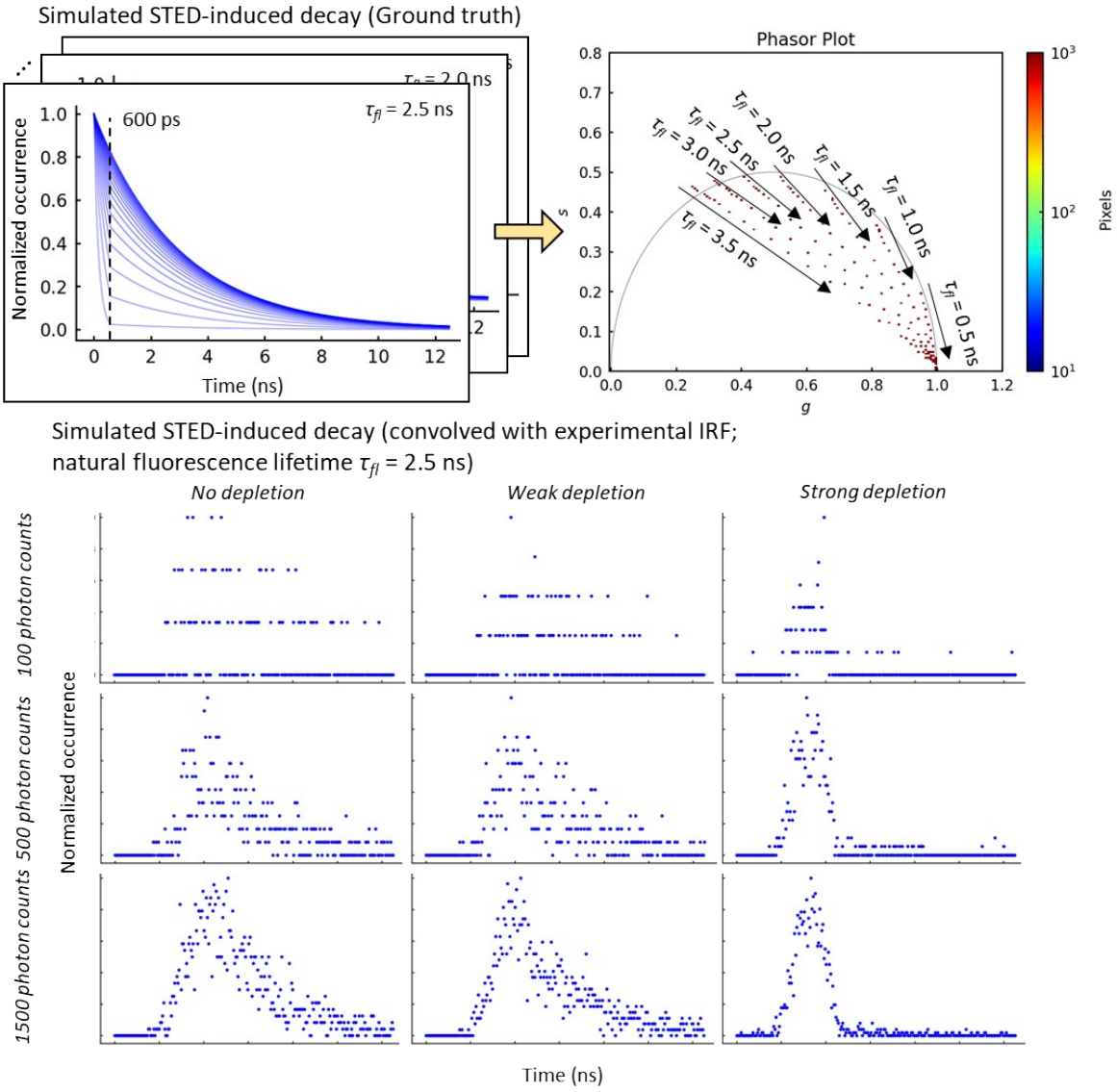

Figure S5 Simulated dataset for STED-*flimGANE* network training. Due to the pulsed STED laser (600 ps), the originally exponential decay was modulated by the STED-induced decay rate during the depletion process. With different simulated natural fluorescence lifetimes and STED-induced decay rates, we observed the corresponding STED trajectory on the phasor plot. Then we performed a Monte Carlo simulation on the instrumental response function-convolved STED-induced decay to generate a large dataset for our model training. A clear and distinct pattern between disparate depletion levels was observed from nine representative simulated STED-induced decays.

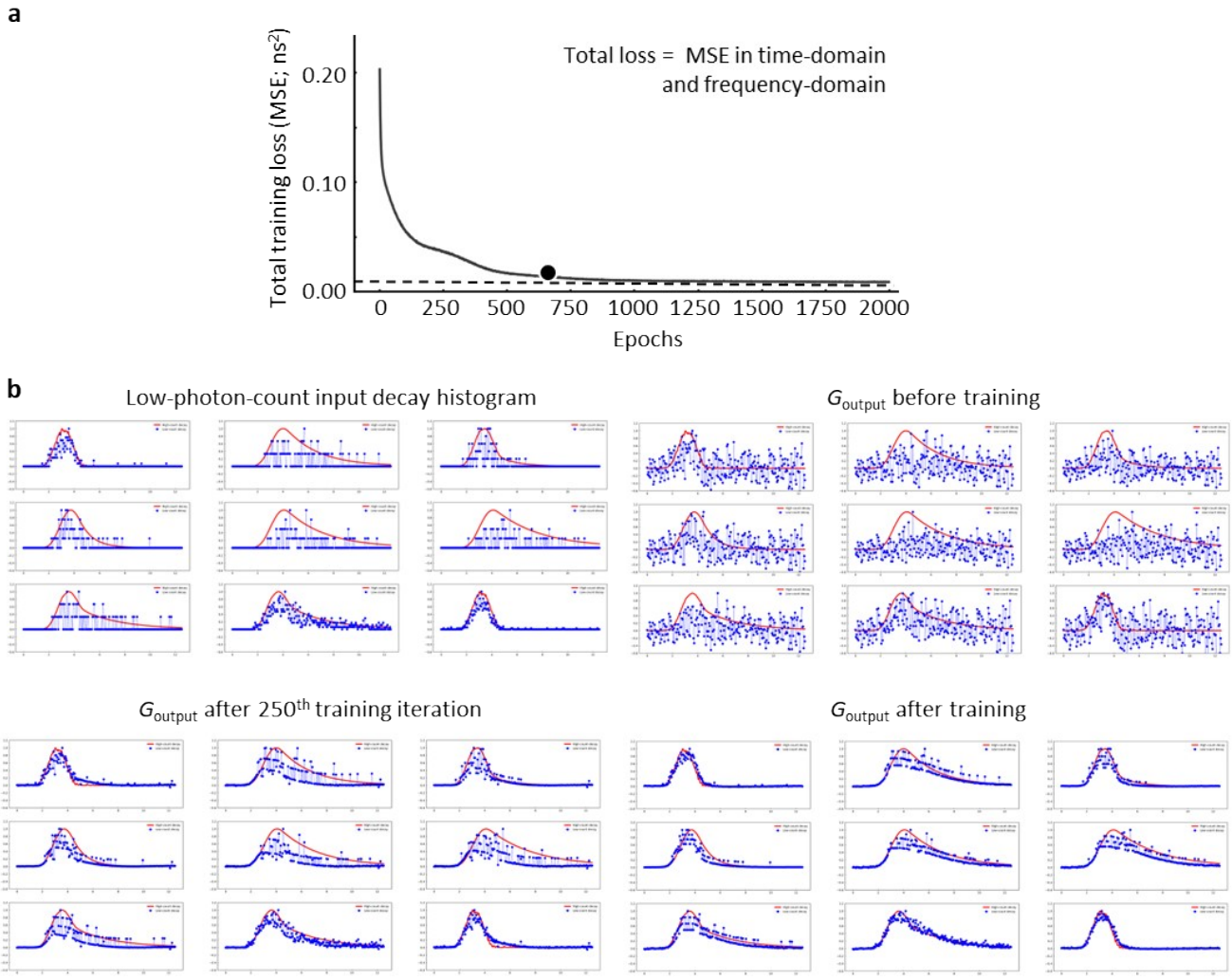

Figure S6 Generator (*G*) can transform a low-photon-count decay histogram into a high-photon-count one. (a) Training loss of STED-*flimGANE* over iterations. The mean squared error between G_{output} and the ground-truth fluorescence decay histograms was visualized over training iterations. We could observe a rapid drop at the beginning of the training, and the loss converged to a certain value (~0.01 ns^2). The model was selected after 700 training iterations (indicated by the back dot). (b) Given the GAN framework, the normalized low-photon-count decay histogram was transformed into the normalized ground-truth mimicking histogram. At the beginning of the training stage, the output from the *G* was chaotic. The generator-inferred fluorescence decay histogram gradually matched the ground truth during the training process.

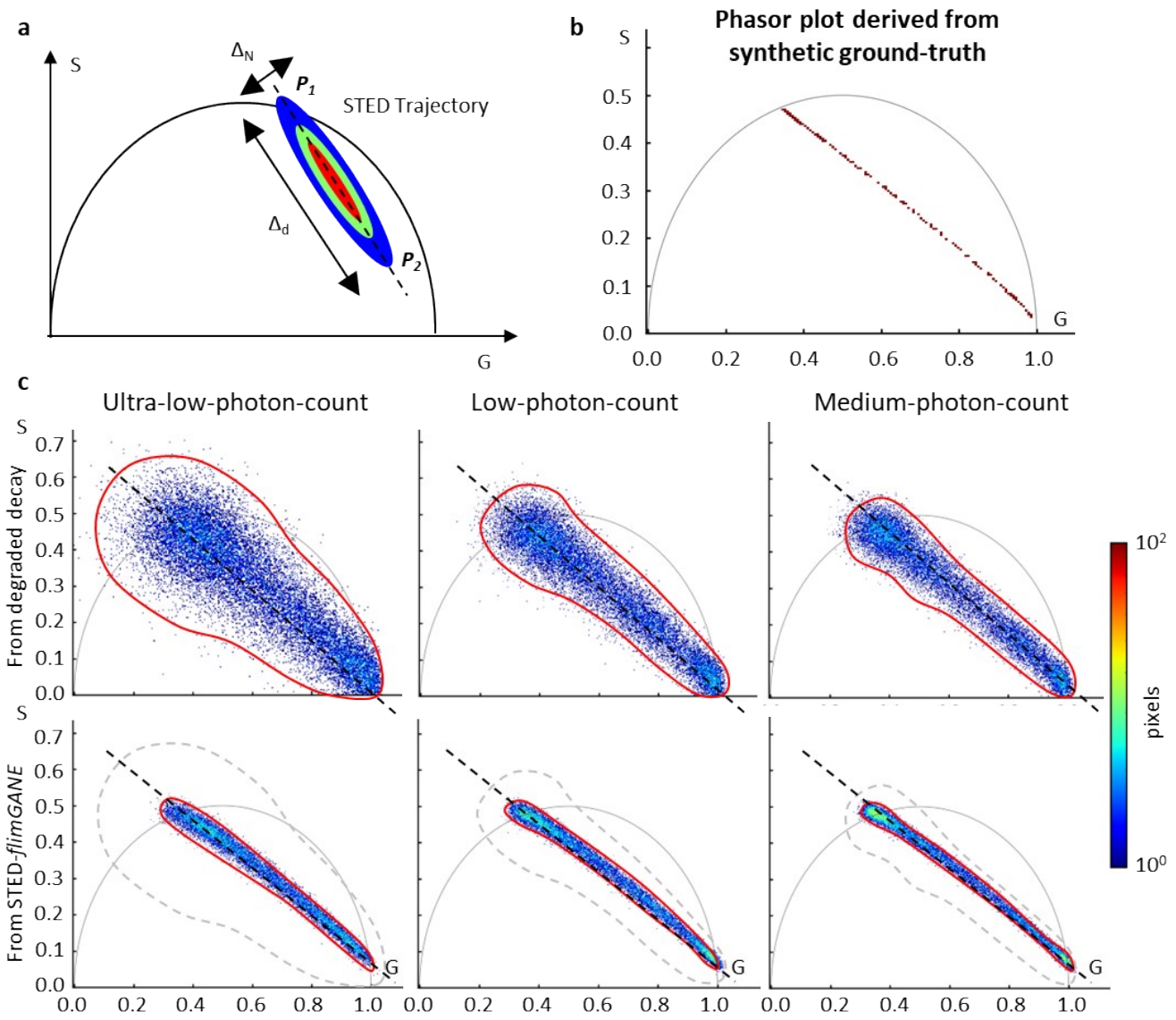

Figure S7 Characterization of SNR improvement on phasor plots. (a) Schematic of SNR characterization in phasor plot. (b) The phasor plot was obtained from the synthetic ground truths for evaluation. (c) The phasor analysis of the synthetic degraded fluorescence decay histogram and the STED-*flimGANE*'s G_{output} under ultra-low- (30-100 photons per pixel), low- (100-200 photons per pixel), and medium-photon-count (200-300 photons per pixel) conditions. Scatter error was significantly reduced after STED-*flimGANE*.

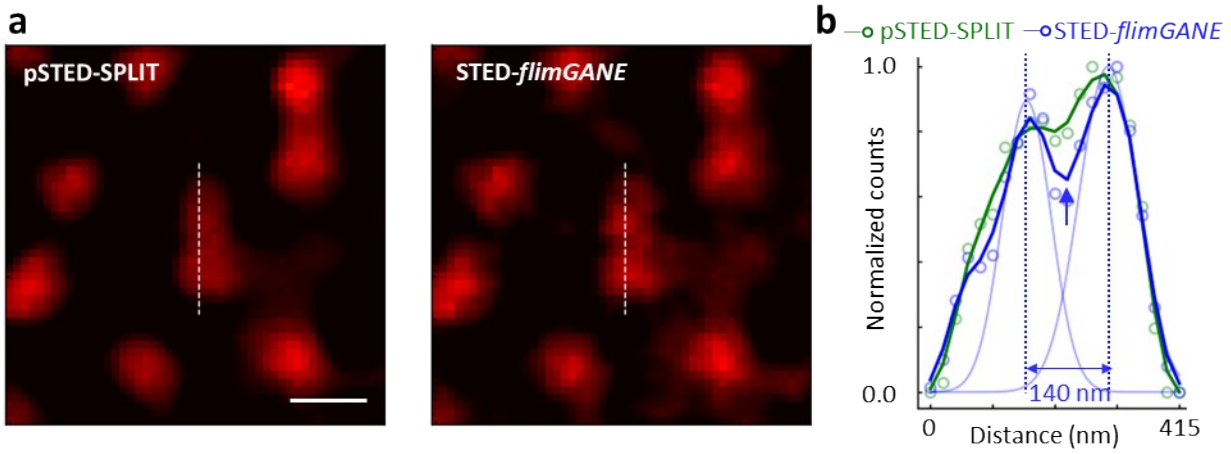

Figure S8 The STED-*flimGANE* enables differentiation of the nuclear pore complex. (**a**) The zoom-in views of the intensity images of NPCs using the pSTED-SPLIT and STED-*flimGANE* microscopy at P_{STED} = 20 mW. Scale bar, 200 nm. (**b**) Line profiles of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images in (a). Blue arrow indicated a clear trough between the adjacent nuclear pore complexes.

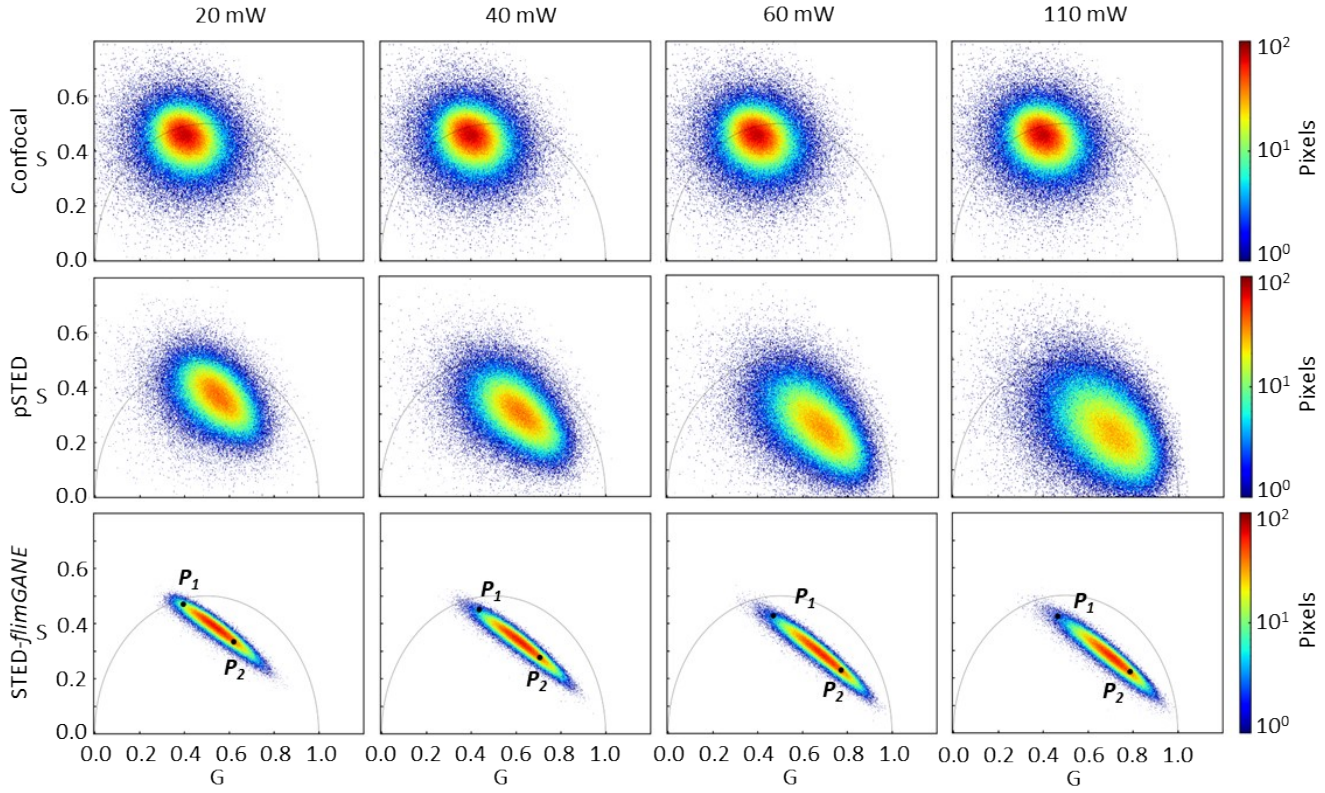

Figure S9 Phasor plots for confocal, pSTED, and STED-*flimGANE* in Fig 3. The *G*-generated phasor plots, P_1 and P_2 assigned by our automatic approach allowed us to determine the photons emitted from the natural fluorophore in the STED experiments with disparate levels of STED power (20, 40, 60, and 110 mW) and the excitation power of 2.2 μW.

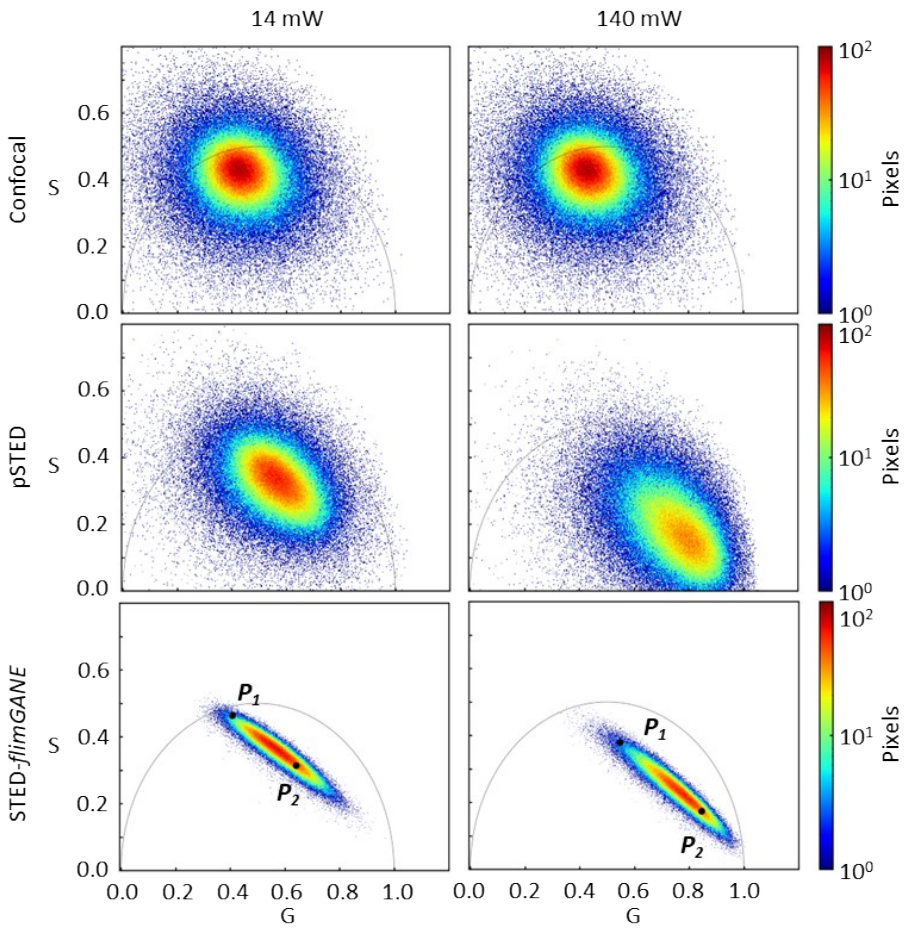

Figure S10 Phasor plots for confocal, pSTED, and STED-*flimGANE* in Fig. 4. The *G*-generated phasor plots, P_1 and P_2 assigned by our automatic approach allowed us to determine the photons emitted from the natural fluorophore in the STED experiments with disparate levels of STED power (14, 140 mW) and the excitation power of 2.2 μW.

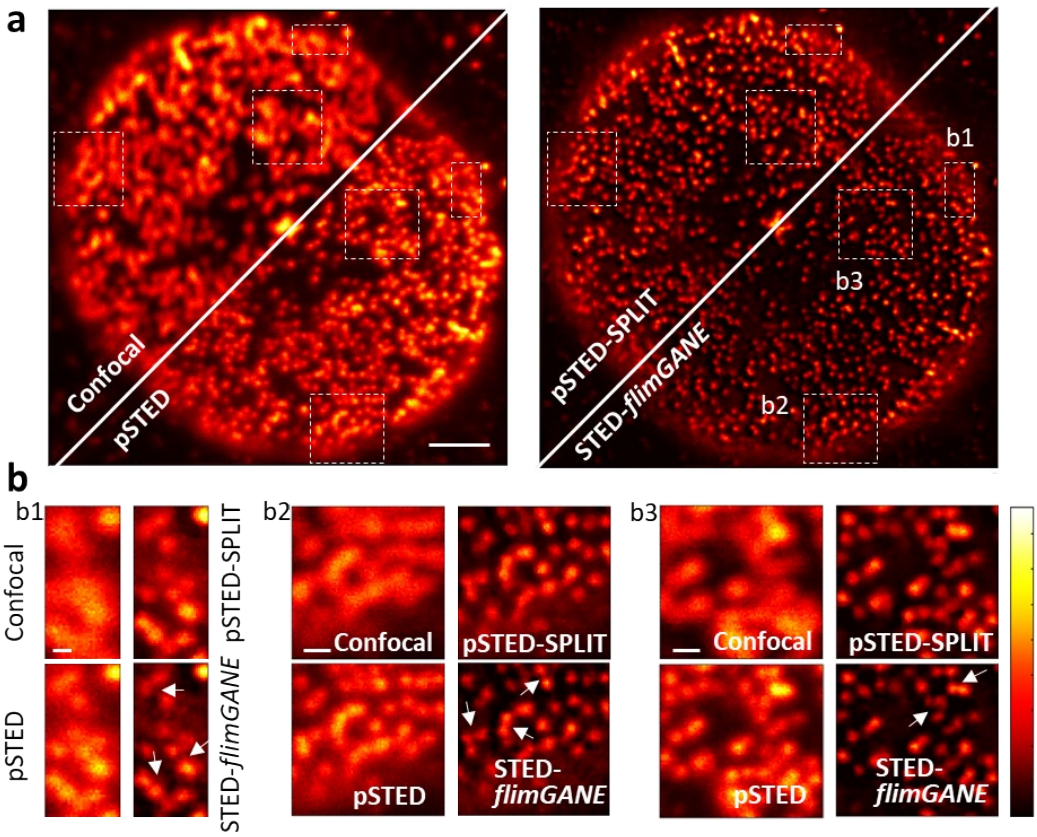

Figure S11 Nuclear pore complex (NPC) imaging results on COS-7 cells at extremely-low depletion power. (a) The intensity images of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* for the STED probe under extremely-low depletion conditions (14 mW). Scale bars, 1 μm. (b) Zoom-in views from the dashed box regions in (a) demonstrated that only the STED-*flimGANE* could discriminate adjacent nuclear pores. Scale bars, 200 nm for b1, 400 nm for b2 and b3.

Table S1 Colormap range for visualization

Colormap range	Fig. 3 20 mW	Fig. 3 40 mW	Fig. 3 60 mW	Fig. 3 110 mW	Fig. 4 & Fig. S10 14 mW	Fig. 4 140 mW
Confocal	[0, 656]	[0, 656]	[0, 656]	[0, 656]	[0, 533]	[0, 533]
pSTED	[0, 781]	[0, 624]	[0, 489]	[0, 266]	[0, 600]	[0, 286]
pSTED-SPLIT	[0, 606]	[0, 537]	[0, 371]	[0, 177]	[0, 567]	[0, 247]
STED-*flimGANE*	[0, 626]	[0, 596]	[0, 346]	[0, 151]	[0, 443]	[0, 219]

Supplementary References

1. M. Arjovsky, S. Chintala and L. Bottou, *International Conference on Machine Learning*, 2017.
2. L. Lanzanò, I. Coto Hernández, M. Castello, E. Gratton, A. Diaspro and G. Vicidomini, *Nature communications*, 2015, **6**, 1-9.
3. G. Tortarolo, Y. Sun, K. W. Teng, Y. Ishitsuka, L. Lanzanó, P. R. Selvin, B. Barbieri, A. Diaspro and G. Vicidomini, *Nanoscale*, 2019, **11**, 1754-1761.