Self-contrastive learning enables interference-resilient and generalizable fluorescence microscopy signal detection without interference modeling

Fengdi Zhang[1,2†], Ruqi Huang[1*†], Meiqian Xin[3,4], Haoran Meng[5], Danheng Gao[5], Ying Fu[6], Juntao Gao[3,4*], Xiangyang Ji[7*]

[1]Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China.

[2]Peng Cheng Laboratory, Shenzhen, 518055, China.

[3]Center for Synthetic and Systems Biology, Tsinghua University, Beijing, 100084, China.

[4]Institute of TCM-X, Tsinghua University, Beijing, 100084, China.

[5]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, 130033, China.

[6]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China.

[7]Department of Automation, Tsinghua University, Beijing, 100084, China.

*Corresponding author(s). E-mail(s): ruqihuang@sz.tsinghua.edu.cn; jtgao@mail.tsinghua.edu.cn; xyji@tsinghua.edu.cn;

†These authors contributed equally to this work.

Abstract

Every weak signal in fluorescence microscopy may contain critical biological information. However, the interference resilience required to detect such signals has traditionally relied on task-specific interference modeling, which limits generalizability. Here, we present a self-contrastive learning-based signal detection solution that achieves interference resilience without the need for interference modeling, thereby offering high generalizability. The method, DEPAF (deep pattern fitting), is a module that contrasts asynchronously generated data views from asymmetric model paths to extract signals from interference, while incorporating highly parallel signal recognition and localization in the process. In benchmark tests, we show that DEPAF improves the detection rate of ultra-high-density signals under low signal-to-noise ratio conditions by an order of magnitude. It is also compatible with, and substantially enhances the performance of various imaging techniques, such as super-resolution imaging, spatial transcriptomic imaging, and two-photon calcium imaging. DEPAF is expected to advance the signal-centric fluorescence microscopy techniques and inspire further advancements, especially in the era of image-based multi-omics.

Introduction

In the world of fluorescence microscopy, the interplay between light and molecular signals at the microscopic scale reveals a detailed and dynamic picture of cellular and molecular activities, playing a crucial role in analyzing and understanding biological dynamics and molecular mechanisms. Signal detection—determining the presence of signals and localizing them—is a fundamental step in many advanced fluorescence microscopy techniques. Examples abound, such as localization-based super-resolution imaging [1], omics analysis [2], and neural activity analysis [3]. However, with the growing demand for higher resolution, greater throughput, and broader applicability, signal detection is increasingly challenged by various forms of interference, including decreased signal-to-noise ratio (SNR) due to reduced photon budgets [4–6], strong fluorescence background caused by optical scattering and non-specific staining [7, 8], and severe signal overlap resulting from increased labeling density [4, 9].

Although adjustments in experimental design can partially alleviate the above issues, such as increasing laser excitation intensity, introducing autofluorescence quenching steps, or sparsifying the labels in space and time, these approaches are essentially trade-offs. They often come with side effects, including causing phototoxicity and damaging live cells [4–6, 10], degrading sample quality [7, 8], and compromising the spatiotemporal resolution of observations [2, 4]. Therefore, solutions in data processing are crucial. Detection methods that are resilient to interference and generalizable across various types of fluorescence microscopy signals stand at the fulcrum of this field.

However, challenges persist, as interference-resilience typically depends on interference modeling for specific imaging modalities. This dependency undermines the generalizability of detection methods, leading to the lack of a method that is both generalizable and interference-resilient (Fig. 1a). For instance, self-supervised denoising methods [6, 10–15] enhance signals by leveraging the unpredictability of noise and redundancy of signals. These methods achieve image denoising without modeling the noise but cannot distinguish background interference or detect signals. Conventional statistical methods [16–21] use algorithms like maximum likelihood estimation (MLE) to estimate signals from imaging data without needing interference modeling. Unfortunately, their performance worsens under severe signal overlap, strong noise, or complex backgrounds, showing poor robustness against interference. Supervised learning methods [22, 23] rely heavily on labeled data, limiting their generalizability to tasks requiring high signal density or extremely high accuracy, where labeled data is often unavailable [1, 24]. Simulation learning methods [4, 9, 25–28] use detailed interference models to simulate labeled data for supervised learning, demonstrating high interference resistance. Yet, these methods must be specifically tailored to each imaging modality, limiting their generalizability. Supplementary Table 1 summaries the characteristics of these methods. Therefore, developing an accessible fluorescence microscopy signal detection method that is both generalizable and resistant to interference remains a challenge, both conceptually and practically, for multi-modal biological signal analysis and understanding broader biological phenomena.

Here, we present DEep PAttern Fitting (DEPAF), a self-contrastive learning-based [29, 30] fluorescence microscopy signal detection framework featuring interference robustness *without* interference modeling, high generalizability, and unsupervised training within a compact, user-friendly design (Fig. 1b). Leveraging the inherent structural regularity of fluorescence microscopy signals, such as point spread functions (PSFs) [1, 31], periodic fluctuations [15, 24], and specific frequency distributions [8], DEPAF utilizes repetitive combinations of small data segments containing "patterns of interest" (POIs) to universally represent these diverse signals. By contrasting the consistency between different augmented views of the input data generated asynchronously via asymmetric model paths, DEPAF learns discriminative features determined solely by interference level rather than specific interference models. These features are embedded into a newly introduced subpixel-level dense representation scheme, ensuring strong discriminability between overlapping signals and thereby substantially improving the detectable signal density, even under conditions of strong noise and severe background interference with unknown types and parameters.

We demonstrated the generalizablity and signal detection performance of DEPAF across a range of representative tasks involving diverse imaging modalities (Fig. 1c), requiring only the replacement of POI examples and the adjustment of a single detection sensitivity hyperparameter. In localization-based super-resolution microscopy, DEPAF enabled an order-of-magnitude improvement in 2D PSF detection rate under low-SNR conditions, even at a signal density of 30 PSFs/μm^2. This enhancement improved the temporal resolution of 2D single-molecule localization microscopy (SMLM) from several seconds to 76.8 ms without sacrificing spatial resolution, thereby enabling super-resolution imaging of dynamic minute changes in live cells at the cytoskeletal level with a spatial resolution of $\sim$30 nm. DEPAF also achieved high-robust-accuracy 3D SMLM reconstruction, supporting molecular blinking densities 10 times higher than conventional methods while requiring only 10.8% of the average photon count compared to typical STORM data labeled with Alexa647 [31]. In spatial transcriptomics imaging, DEPAF enables high-sensitivity detection of multiplexed error-robust fluorescence in situ hybridization (MERFISH) spot signals in strong background fluorescence. This results in a 97% increase in detection rate and an improvement of 0.12 in the correlation between detected RNA copy numbers and fragments per kilobase per million reads (FPKM) in a 3D MERFISH sample with strong background fluorescence. In low-SNR two-photon calcium imaging of neurons, DEPAF uses a single pipeline for denoising, background separation, weak spike estimation, and active neuron segmentation. As an unsupervised source extraction-based segmentation method, DEPAF clearly outperforms other methods of the same type, increasing precision by 40% and recall by 43%, and is even competitive with top-performing supervised methods, while reducing the required training data and ground-truth labels by 2 and 1 orders of magnitude, respectively. DEPAF also allows for accurate estimation of arbitrary inhomogeneous fluorescence backgrounds. Across simulated datasets with a wide range of background variation scales and SNR levels, the correlation between DEPAF-estimated and ground-truth backgrounds consistently exceeded 0.95. These organic integrations demonstrate that DEPAF can serve as a practical framework, generally enhancing the signal detection performance under strong interference. With open-source code, flexible POI representation, and task-agnostic modeling, DEPAF is expected to bring broad benefits to related fields in the era of image-based multi-omics [32].

Results

DEPAF architecture. The architecture of DEPAF is illustrated in Fig. 1d. Our method takes an original image, POI examples, and a hyperparameter corresponding to detection sensitivity as inputs, and outputs the POI positions within the image along with a synthesized clean image that contains only the signals. The pipeline of our method comprises two branches. In the first branch, DEPAF treats each pixel as a fitting component and

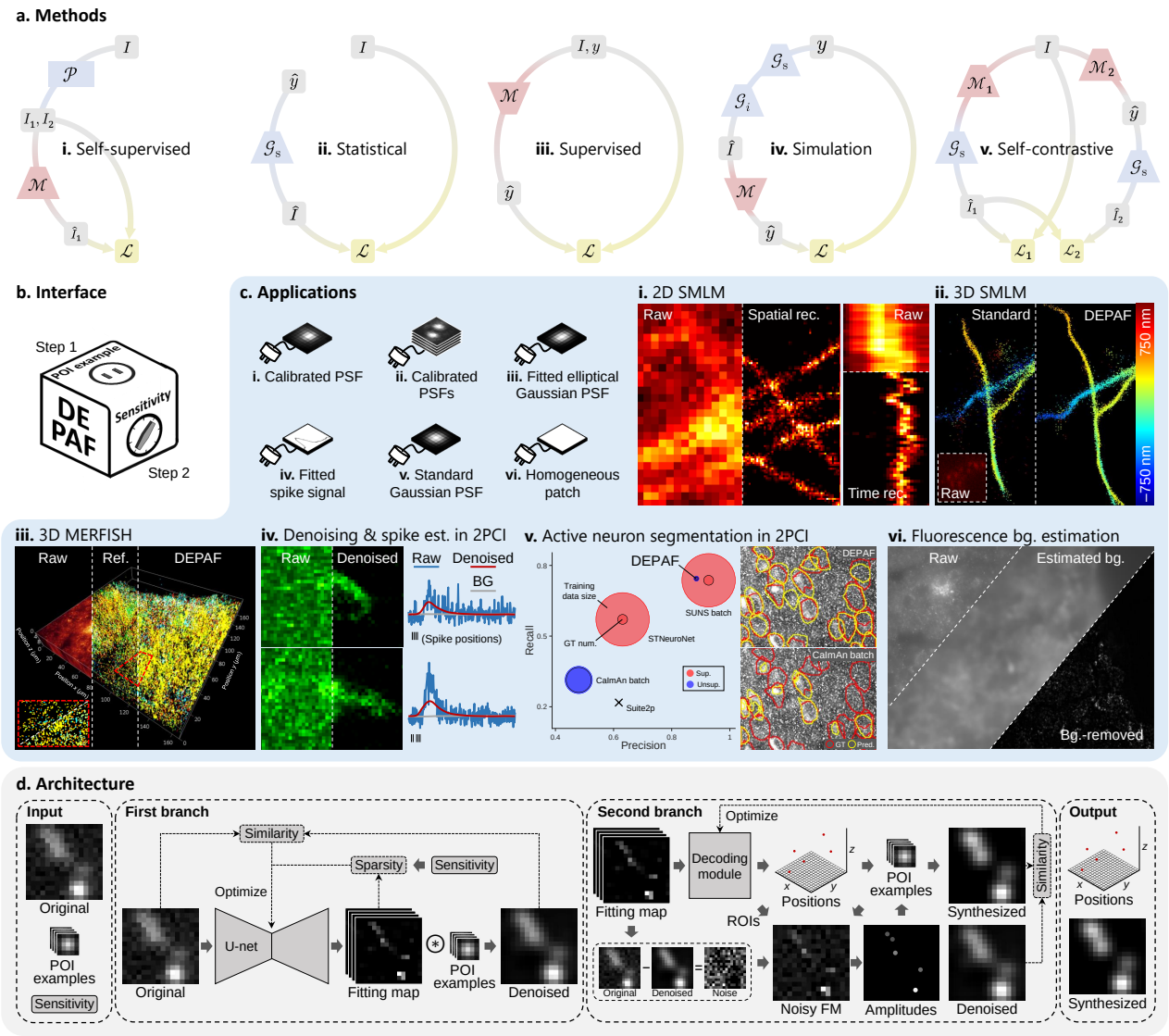


Fig. 1 | DEPAF for interference-resilient and generalizable fluorescence microscopy signal detection without interference modeling. a. Fluorescence microscopy signal processing methods. **i.** Self-supervised denoising methods. Noise-independent but signal-consistent image pairs I_1, I_2 are derived from input images I via pretext tasks $\mathcal{P}$. I_1 are fed into denoising model $\mathcal{M}$ to predict clean images $\hat{I}_1$. $\mathcal{M}$ is optimized by loss $\mathcal{L}$ between $\hat{I}_1$ and I_2. **ii.** Conventional statistical methods. Detection results $\hat{y}$ are passed through signal model $\mathcal{G}_s$ to generate synthesized data $\hat{I}$. $\hat{y}$ are optimized by loss $\mathcal{L}$ between $\hat{I}$ and input images I. **iii.** Supervised learning methods. Input images I are processed by model $\mathcal{M}$ to produce detection results $\hat{y}$. $\mathcal{M}$ is optimized by loss $\mathcal{L}$ between $\hat{y}$ and external ground truths y. **iv.** Simulation learning methods. Starting from generated ground truths y, ideal signals are generated via $\mathcal{G}_s$ and corrupted using interference model $\mathcal{G}_i$ to obtain simulated images $\hat{I}$. $\hat{I}$ are input to model $\mathcal{M}$ to get detection results $\hat{y}$. $\mathcal{M}$ is optimized by loss $\mathcal{L}$ between y and $\hat{y}$. **v.** Self-contrastive learning methods *applied in this work*. Based on input images I, detection results $\hat{y}$ and two views, $\hat{I}_1$ and $\hat{I}_2$, each containing a mixture of signals with noisy signals and signals with noise, respectively, are obtained through asymmetric and asynchronous model paths, involving models $\mathcal{M}_1$, $\mathcal{M}_2$, and signal model $\mathcal{G}_s$ *without* requiring any interference models. $\mathcal{M}_1$ is first optimized by loss $\mathcal{L}_1$ between $\hat{I}_1$ and I, followed by optimizing $\mathcal{M}_2$ using loss $\mathcal{L}_2$ between $\hat{I}_1$ and $\hat{I}_2$. **b,** DEPAF provides a streamlined interface, enabling users to achieve diverse fluorescence microscopy signal detection by simply replacing the POI examples (step 1) and adjusting a single hyperparameter corresponding to detection sensitivity (step 2). **c.** The POI examples used for different applications, along with the corresponding result diagrams. **i.** DEPAF elevates the temporal resolution of 2D SMLM to the millisecond level without compromising the expected spatial resolution. Rec., reconstruction. **ii.** In 3D SMLM, DEPAF maintains stable and high-quality reconstruction performance even under low photon counts and high blinking densities. **iii.** In 3D MERFISH imaging data with strong background, DEPAF notably improves RNA spot signal detection rates and quantitative accuracy. Ref., reference. **iv.** For low-SNR two-photon calcium imaging of neurons, DEPAF achieves efficient denoising, background separation, and detection of weak neural activity. Est., estimation; 2PCI, two-photon calcium imaging; BG, background. **v.** Leveraging the denoised results from **iv**, DEPAF, as an unsupervised source extraction-based method, enables high-performance active neuron segmentation with minimal data requirements, clearly outperforming other such methods, and is even competitive with top-performing supervised methods. GT num., required number of ground-truth labels; Sup., supervised; Unsup., unsupervised; Pred., prediction. **vi.** DEPAF can estimate arbitrary inhomogeneous fluorescence backgrounds, adapting to varying SNR levels and background variation scales. Bg., background. **d,** Schematic depiction of DEPAF architecture. It uses a two-branch pipeline to process an original image, POI examples, and a detection sensitivity hyperparameter to output POI positions and a synthesized clean image with only the signals. FM, fitting map.

utilizes an image-to-image translation deep learning model with a U-net structure [33] to simultaneously (rather than sequentially) resolve all POIs in the original image while also denoising it. The result is represented as a 3D fitting map. The first and second dimensions of this map correspond to the likelihood of a POI's presence

at different pixels, while the third dimension corresponds to the likelihood of each POI's existence at that pixel. The deep learning model in this branch is optimized by regularizing the fitting map to encourage sparsity and maximizing the similarity between the original image and the denoised image, where the denoised image is obtained by group-convolving the fitting map with the POI examples.

In the second branch, DEPAF uses an optimizable decoding module that leverages the principles of centroid-based discrete-to-continuous space transformation to decode the POI positions in continuous coordinate space from the fitting map. Then, the fitting map and the noise image, obtained by subtracting the denoised image from the original image, are combined to create a noisy fitting map, from which the amplitudes of the POIs are estimated. With the examples, positions and amplitudes of POIs, a synthesized image is generated. The decoding module is then optimized by maximizing the similarity between the synthesized and denoised image to remove noise-induced POI positions while retaining valid ones. After all the optimizations are complete, new image can be sequentially input into the fixed-parameter deep learning model and the decoding module for fast POI localization and image denoising.

Millisecond-level dynamic 2D SMLM and high-robust-accuracy 3D SMLM via DEPAF. Diffraction causes a lens-based microscope to image a point light source as a PSF with a width of 200–300 nanometers [1], making structures smaller than this scale appear blurred and difficult to distinguish. SMLM techniques, such as STORM [34], dSTORM [35], PAINT [36], and PALM [37], are dedicated to addressing this issue by randomly activating fluorescent molecules within the samples, recording their "blinking" behavior in numerous camera frames, and then using localization algorithms to precisely localize the nanoscale position of each PSF for super-resolution imaging rendering. These techniques have transformed biological imaging, enabling nanoscale observation of subcellular structures.

Temporal resolution is inherently limited in SMLM by the sequential activation of fluorescent molecules and the requirement for capturing numerous frames [5]. Conventional SMLM typically achieve temporal resolution on the order of minutes, with some reports of tens of seconds [38, 39]. Solutions include: (1) allowing a higher density of PSFs per frame, (2) reducing the activation time of fluorescent molecules, and (3) increasing the imaging frame rate. This requires the localization algorithms of SMLM to accurately detect and localize highly overlapping PSFs in strong noise caused by fast frame rates.

To quantitatively test the performance of DEPAF in PSF localization, we set a calibrated PSF image as the POI example and evaluated it on 18 simulated 2D SMLM datasets (Methods), which spanned a wide range of SNR levels and high PSF densities. Compared to SMAP [16], the top-performing conventional statistical method in a public SMLM challenge [31], and DECODE [4], a simulation-based method that outperformed all existing methods across all 12 datasets in the same public challenge, DEPAF showed clear performance improvements in localizing ultra-high-density PSFs and maintained exceptional robustness under low SNR imaging conditions (Fig. 2a and Extended Data Fig. 2). For instance, in the most challenging dataset with a low SNR and an ultra-high density of 30 PSFs/μm^2, DEPAF increases the Jaccard index, which measures the detection rate, from 3.87% with DECODE and 1.64% with SMAP to 36.41% (Extended Data Fig. 2a), achieving an order-of-magnitude improvement. This demonstrates DEPAF's strong ability to address the temporal resolution limitations in SMLM.

Using the dSTORM technique [35], we performed SMLM imaging on a live-cell sample to capture cytoskeletal fluctuations with high blinking density at a frame rate of ∼781 frames per second (FPS) (Fig. 2b) (Methods). Subsequently, using a calibrated 2D PSF as the POI example, DEPAF achieves the rendering of super-resolution videos with a temporal resolution of ∼76.8 ms (∼13 FPS) and a spatial resolution of ∼30 nm (Fig. 2b,c and Supplementary Video 1). The spatial resolution is quantified using decorrelation analysis [40] (Supplementary Fig. 1). This enables the observation of rapid, minute changes in subcellular structures at the cytoskeletal level. In comparison, methods such as SMAP (Supplementary Video 2), which relies on sparse signal conditions, DECODE (Supplementary Video 3), which have been reported to achieve a temporal resolution of a few seconds [4], and eSRRF (Supplementary Video 4), which can generate one super-resolution volume per second but compromises spatial resolution to ∼70 nm [15], as well as DBlink (Supplementary Video 5 and 6), which achieves millisecond-level resolution but is limited by interpolation-induced artifacts [5], are all unable to achieve high-quality super-resolution reconstructions under such challenging imaging conditions.

PSF engineering can employ special optical components or modulation methods to modify the PSF shape produced by emitters at different z-axis positions [41]. This allows for the estimation of the z-axis positions of fluorescent molecules based on their PSF shapes in 2D SMLM frames, enabling 3D SMLM [31, 42]. These engineered PSFs often show non-orthogonality and greater overlap in the frames due to their larger area, which raises the performance demands on localization algorithms.

Using 151 calibrated double helix PSFs [43] as the POI examples, we evaluated DEPAF on a simulated 3D SMLM benchmarking dataset, where the average PSF density per frame was 10 times higher than normal, and the average photon count was only 10.8% of the typical data acquired from Alexa647-labeled STORM samples [31] (Extended Data Fig. 3a–c). The results demonstrate that, compared to conventional statistical methods

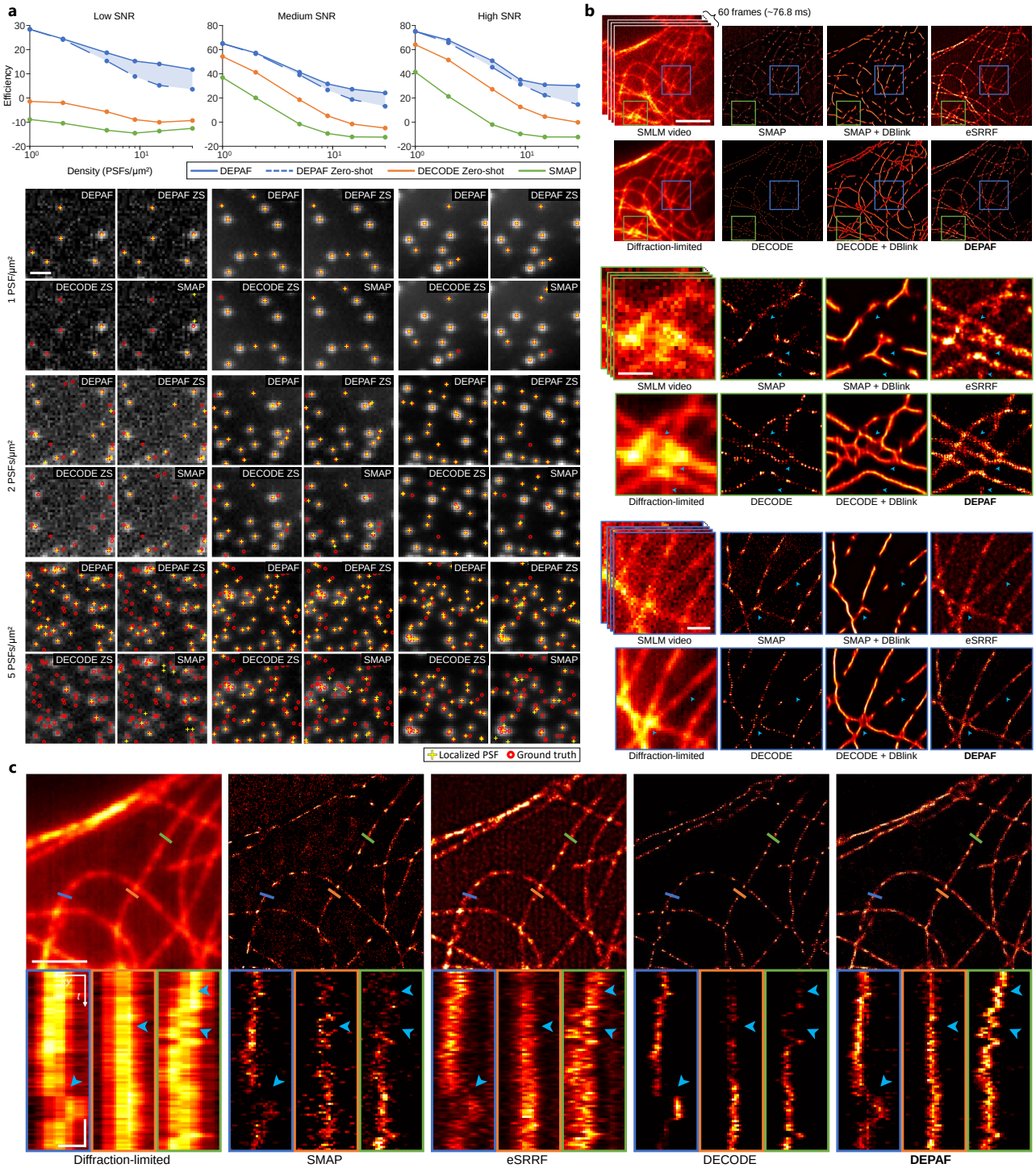


Fig. 2 | DEPAF outperforms other methods in all 18 simulated SMLM datasets with a wide range of SNR levels and high PSF densities and achieves millisecond-level dynamic 2D SMLM. a, The line charts compare the *efficiency* metrics of different methods across different SNR levels and PSF densities (log scale) in 18 simulated datasets (see Extended Data Fig. 2 for further details). Each column of subplots below visualizes the localization results at PSF densities of 1, 2, and 5 PSFs/μm^2 for the corresponding low, medium, and high SNR levels. Scale bar, 1 μm. ZS, zero-shot testing. **b,** Top: super-resolution reconstruction results of different methods based on SMLM videos captured with a high-speed camera at a frame rate of 1.28 ms/frame and the dSTORM technique with high blinking density. Scale bar, 5 μm. Middle and bottom: magnified details of the reconstruction results. Blue arrows indicate specific regions to highlight differences in resolution and structural details among the methods. Scale bar, 1 μm. **c,** The subplots below display the time-lapse images along the line segment in the top images within the reconstructed super-resolution video. Differences in resolution and structural details between the methods are indicated by blue arrows in specific areas. Vertical scale bar, 1 s. Horizontal scale bar, 300 nm.

SMAP, which are effective for SMLM videos with sparse PSFs, DEPAF can effectively handle this challenging condition and achieve high reconstruction quality (Extended Data Fig. 3d).

High-sensitivity 3D MERFISH spot signal detection using DEPAF. Single-molecule fluorescence in situ hybridization (smFISH) utilizes fluorescence labeling, imaging, and algorithm-based spot signal detection to precisely localize specific RNAs with high resolution [44, 45]. This makes smFISH a powerful tool for studying

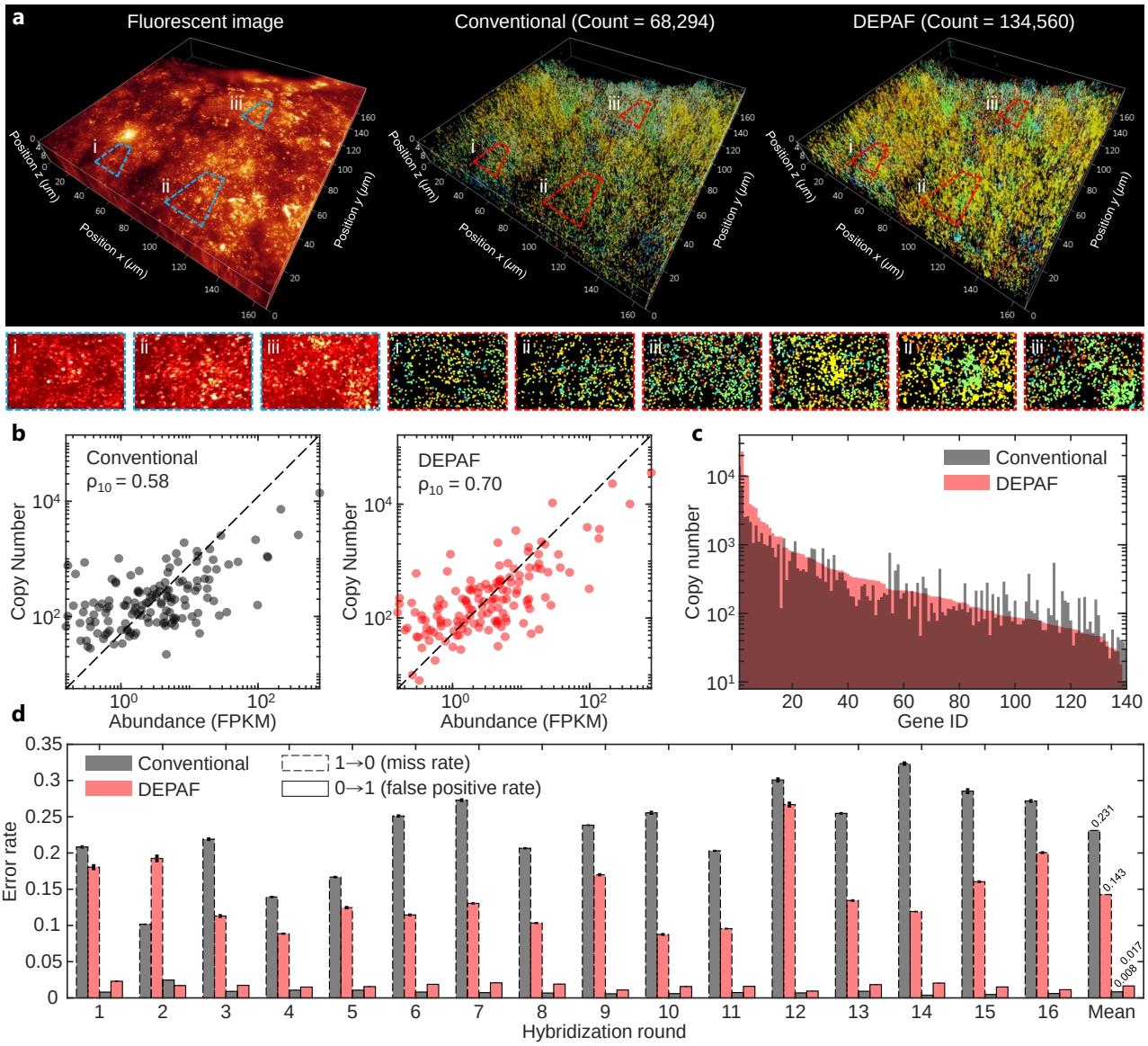


Fig. 3 | DEPAF doubles the detection rate of MERFISH spot signals even under complex imaging conditions with strong backgrounds, high labeling density, and a large FOV. a, In a 3D MERFISH data labeling 140 RNA species acquired in challenging conditions (left), the number of RNA molecules detected by DEPAF (right, 134,560 counts) is considerably higher than that detected by the conventional method (middle, 68,294 counts). The zoomed-in view at the bottom provides detailed visualization of the image and detection results. The 3D fluorescence image is shown as a maximum intensity projection created by combining images from different hybridization rounds after alignment and bleaching correction. The subplots below display maximum intensity projections of different slices derived from this image. Different colors represent different RNA species. **b,** Scatter plots compare the copy numbers detected by the conventional method (left, black) and DEPAF (right, red) with the abundance measured by bulk sequencing. The Pearson correlation coefficients between the $\log_{10}$-transformed copy numbers and abundances (ρ_{10}) are 0.70 and 0.58, with a P value of 5×10^{-22} and 5×10^{-14}, respectively. **c,** The bar plot shows the RNA copy numbers detected by DEPAF (red) and the conventional method (black) across different RNA species ($\log_{10}$ scale). Gene IDs are sorted in descending order based on the RNA copy numbers detected by DEPAF. **d,** The bar plot displays the miss rates (1→0) and false positive rates (0→1) calculated based on error correction with MHD4 coding [2] across multiple hybridization rounds for DEPAF (red) and the conventional method (black). Error bars represent standard error of the mean (SEM, $n = 16$ z-layers).

RNA copy number and spatial organization. With technological advancements, more sophisticated techniques, such as MERFISH, have further expanded the detection capabilities, allowing for the simultaneous analysis of hundreds to thousands of RNA species [2]. However, these advancements also introduce challenges, such as the need for additional rounds of hybridization, increased non-specific hybridization, and higher labeling densities [46]. These factors can result in stronger fluorescence backgrounds [7] and greater signal overlap [47] in the images. Current spot signal localization algorithms still struggle to effectively tackle such complex data.

To demonstrate DEPAF's ability to overcome these challenges, using a fitted Gaussian PSF image as the POI example, we applied it to a 3D z-stack MERFISH data obtained from a fetal liver tissue sample. This data is characterized by strong fluorescence backgrounds, high-density labeling, and a large FOV of $165.7 \times 165.7 \times 9$ μm^3 [47]. In addition to the typical challenges of conventional 2D MERFISH data, z-stack imaging introduces further complexities, such as signal crosstalk between z-layers, PSF defocusing, z-axis drift, and more pronounced fluorescence intensity fluctuations, which pose additional difficulties to the spot signal detection algorithms.

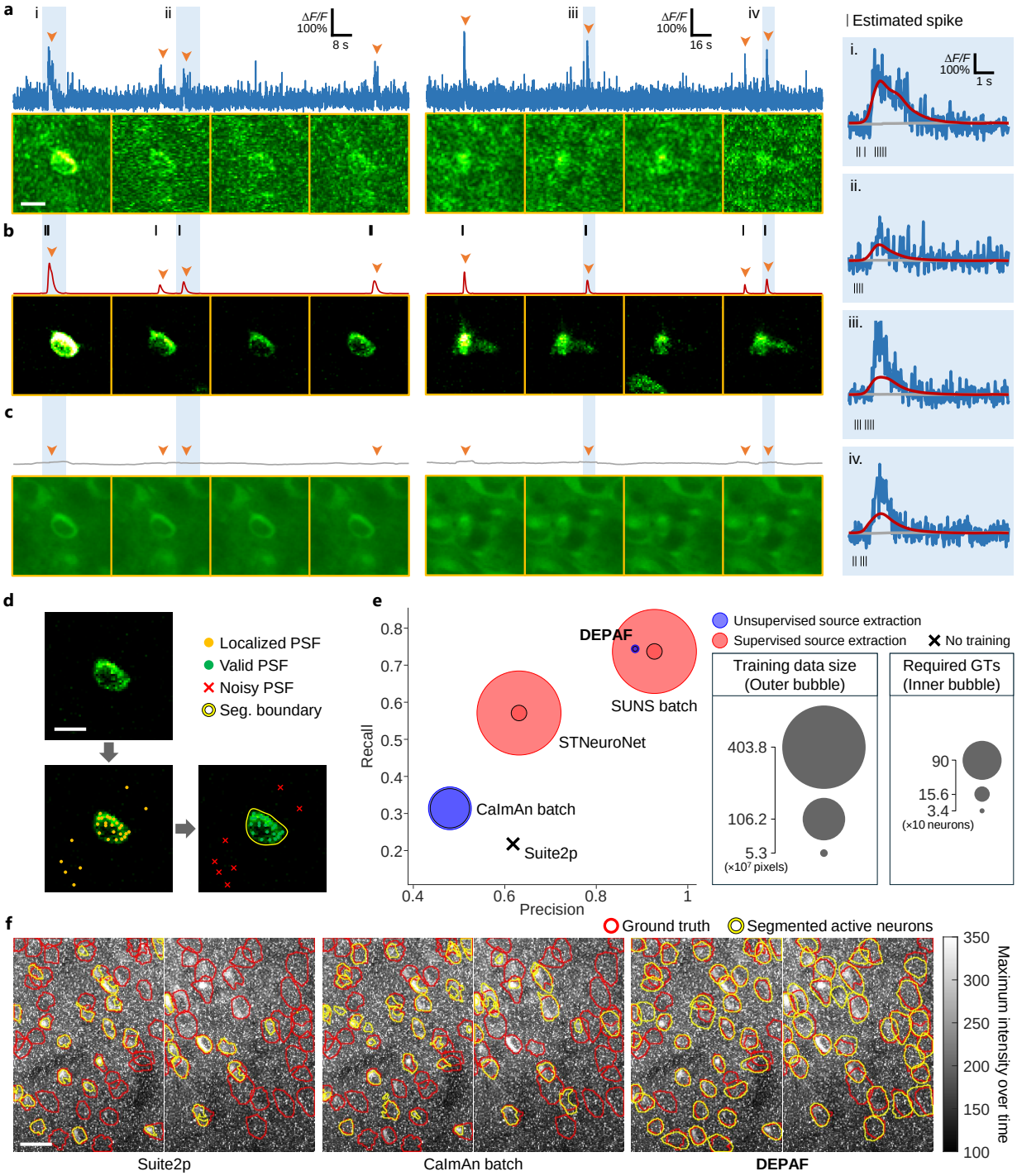


Fig. 4 | DEPAF enables one-pipeline processing for denoising, background separation, weak spike estimation, and data-efficient active neuron segmentation in low-SNR two-photon calcium imaging data. a, Fluorescence traces of two neurons from the low-SNR two-photon calcium imaging data. Frames marked with orange arrows are shown below the traces. ΔF represents the change in fluorescence relative to the baseline fluorescence F_0, which is approximated by the average of the entire trace. Scale bar, 9 μm. **b,** Fluorescence traces and frames after DEPAF denoising, retaining only signals from active neurons. Spike positions are simultaneously estimated with DEPAF and indicated by black vertical lines (variations in thickness are due to overlaps or changes in signal width). **c,** Fluorescence traces and frames of the spatiotemporally inhomogeneous background estimated by DEPAF. **i–iv,** Zoomed-in examples of calcium transients in raw data (blue), denoised data (red), estimated background (gray), and estimated spike positions (black vertical lines). **d,** The process of active neuron segmentation using the denoised data. DEPAF detects potential signals using a standard Gaussian PSF as the POI example, filters valid PSFs via DBSCAN clustering, and generates segmentation boundaries using cubic spline curves based on the valid PSFs. Scale bar, 9 μm. Seg., segmentation. **e,** Segmentation performance of different methods, along with their training data size and ground-truth labels used. GTs, ground-truth labels. **f,** Example segmentation results of all *unsupervised* source extraction-based methods from the second and fourth quadrants of video J123, overlaid on the maximum intensity projection across the time dimension. Scale bar, 20 μm.

Our results show that, compared to the 68,294 RNA molecules detected by the conventional method, DEPAF substantially increases the detection number to 134,560, resulting in a 97% improvement in detection rate (Fig. 3a,c). To verify the reliability of the detection results, we employed a standard evaluation method by comparing the detected RNA copy numbers with the abundance determined by bulk sequencing in FPKM on the same

cell line [2]. The results demonstrate that DEPAF further improves the correlation between RNA copy numbers and FPKM by 0.12 compared to the conventional method, reaching a usable threshold of 0.7 [47] (Fig. 3b). Additionally, compared to the conventional method, DEPAF reduces the average miss rate of bits from 23.1% to 14.3% while slightly increasing the average false positive rate from 0.8% to 1.7% (Fig. 3d).

Low-SNR two-photon calcium imaging data processing with DEPAF. By using genetically encoded calcium indicators and two-photon microscopy, we can transform neuronal action potentials (spikes) into visible fluorescent fluctuations [3, 48]. This allows us to efficiently conduct large-scale optical recordings of neuronal activity in various animal models and associate these activities with neural functions, thereby substantially enhancing our understanding of how local neural networks process information.

Extracting neural activity from two-photon calcium imaging videos involves two key tasks: (1) reconstructing spikes and estimating their positions in the temporal domain [24], and (2) extracting and segmenting active neurons in the spatial domain [19, 20, 22, 23]. However, the imaging data often suffers from significant noise interference, particularly in high temporal resolution analyses [6], and exhibits spatiotemporally inhomogeneous backgrounds [24]. Moreover, a single spike manifests as rapid rises followed by slow decays, which may lead to spike overlap [49]. These factors pose challenges to spike estimation and active neuron segmentation.

We applied DEPAF to the video J123 from the CaImAn datasets, a representative low-SNR two-photon calcium imaging data of hippocampal neurons [19] (Fig. 4a). The results show that, considering the data along the time axis for each pixel in the two-photon calcium imaging videos as a one-dimensional (1D) image and using a fitted spike signal as the POI example, DEPAF can localize overlapping and weak spikes in significant noise (Fig. 4b,(i)–(iv)). Meanwhile, benefiting from DEPAF's accurate estimation of POI positions and amplitudes (Extended Data Fig. 4) and through its image synthesis function (Fig. 1d), noise-free signals separated from the background can be generated (Fig. 4b,c and Extended Data Fig. 5a,b). After reshaping these noise-free signals into 3D videos, denoised two-photon calcium imaging videos that remove the background and retain only active neuron signals can be obtained (Fig. 4b and Extended Data Fig. 5c).

Based on the denoised videos (Supplementary Video 7–10), DEPAF can further detect the ultra-high-density PSFs that constitute the active neuron regions in the spatial domain using a standard Gaussian PSF image as the POI example. After localization, active neurons are extracted and segmented by performing DBSCAN clustering [50] and boundary generation on the localized PSFs (Fig. 4d). Comparative experiments on active neuron segmentation demonstrate that DEPAF, a method based on unsupervised source extraction for identifying the spatial footprints of active neurons, substantially outperforms other methods that likewise do not require ground-truth supervision for source extraction, such as Suite2p [20] and CaImAn batch [19]. In particular, DEPAF achieves a 40% increase in precision and a 43% increase in recall compared to CaImAn batch (Fig. 4e,f). To further benchmark the performance of DEPAF, we also compared it against supervised methods (Supplementary Fig. 2). The results show that DEPAF still outperforms STNeuroNet [23] and even performs competitively with the top-performing supervised method SUNS batch [22], despite requiring two orders of magnitude less training data and one order of magnitude fewer ground-truth labels (Fig. 4e).

Arbitrary inhomogeneous fluorescence background estimation enabled by DEPAF. In fluorescence microscopy images, the term "background" refers to components other than the signals and noise. These components reduce image quality and affect the sensitivity and accuracy of signal detection, necessitating their removal. Assuming the background is homogeneous [16, 17, 21] oversimplifies real imaging environments. Typical specimens, such as cells or tissue slices, contain various components at different spatial scales that may fluoresce spontaneously. Additionally, fluorescent markers might bind non-specifically to other components. Thus, the actual background is an inhomogeneous mixture of various spatial frequencies [4, 8].

Accurately estimating arbitrary inhomogeneous backgrounds by DEPAF hinges on the design of background POI example, specifically how the backgrounds can be broadly represented using the background POI example. Previous studies have shown that background distributions are concentrated in low frequencies [8], meaning the background POI example should capture low-frequency information. Additionally, since the background is distinct from both signals and noise, the background POI example must be uncorrelated with them to prevent DEPAF from mistakenly fitting it to either. Therefore, the background POI example should be designed to be orthogonal to both. To meet these criteria, we designed the background POI example as a homogeneous patch of a certain size, essentially a small image of constant value. This offers the following benefits: (1) Its size limits its ability to fit high-frequency information, making it suitable for low-frequency representation. (2) Due to the periodic nature of signals and the randomness of noise, this homogeneous patch maintains broad orthogonality to both. These features align well with the requirements. Thus, benefiting from DEPAF's capability to fit multiple POI examples simultaneously, by using this background POI example as an additional one alongside signal POI examples, DEPAF can optimize background estimation and signal detection simultaneously, thereby achieving accurate estimation of arbitrary inhomogeneous backgrounds.

Experimental results on 15 simulated datasets (Methods) demonstrate that DEPAF achieves high estimation accuracy in arbitrary inhomogeneous backgrounds across a wide range of background variation scales and SNR

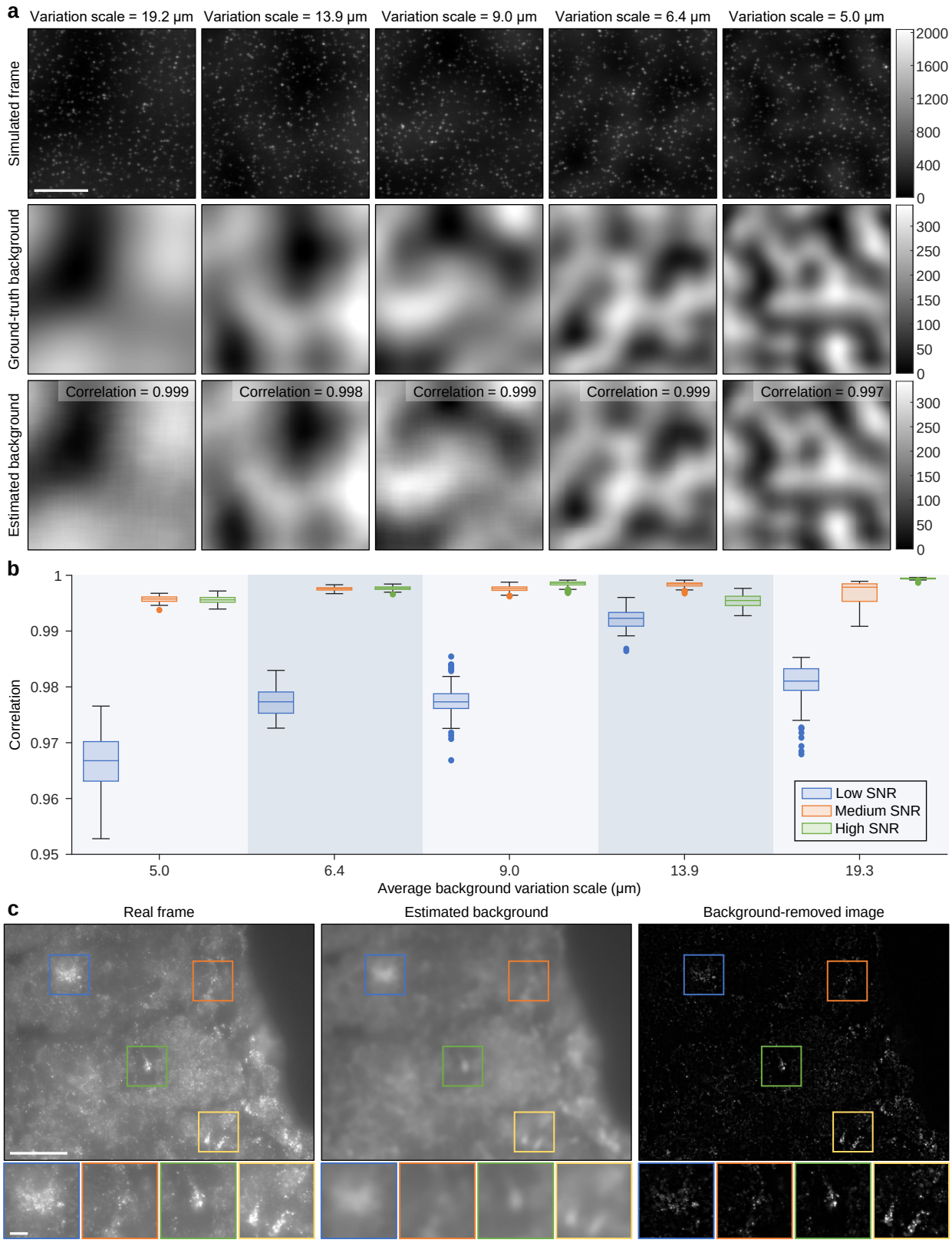


Fig. 5 | DEPAF accurately estimates arbitrary inhomogeneous fluorescence backgrounds. a, The performance of DEPAF in fluorescence background estimation under medium SNR level and different background variation scales. First row: frames containing inhomogeneous backgrounds simulated using Perlin noise [51]. Second row: ground truths of the simulated backgrounds. Third row: backgrounds estimated by DEPAF. The Pearson correlation coefficients between the estimated backgrounds and the ground truths are indicated in the top right corner of each image. Scale bar, 10 μm. **b,** The box plot illustrates the distributions of Pearson correlation coefficients between the backgrounds estimated by DEPAF and the ground truths at different SNR levels and average background variation scales. **c,** The performance of DEPAF in estimating arbitrary inhomogeneous backgrounds in experimental data. Left: a real frame with a complex inhomogeneous fluorescence background. Middle: the background estimated by DEPAF. Right: the result after subtracting the estimated background from the frame. Scale bar, 30 μm. The zoomed-in views at the bottom provides detailed visualization of the images. Scale bar, 5 μm.

levels (Fig. 5a,b and Extended Data Fig. 6), with all correlations to the ground truths consistently exceeding 0.95. It also exhibits excellent performance on experimental data (Fig. 4c, Fig. 5c, and Extended Data Fig. 5).

Discussion

In this work, we present DEPAF, an interference-resilient framework for fluorescence microscopy signal detection that eliminates the need for interference modeling. This feature further imparts generalizability to the method, which is tested across multiple tasks that essentially encompass different representative signal types: To demonstrate its effectiveness in detecting single-shape signals, we applied it to 2D SMLM, achieving millisecond-level dynamic super-resolution imaging; To showcase its performance in detecting multi-shape signals, we employed it in 3D SMLM, achieving high-robust-accuracy 3D super-resolution imaging; To test its ability to handle signals without precise shape calibration, we used it in MERFISH analysis, supporting high-sensitivity 3D MERFISH spot signal detection in strong background fluorescence; To further demonstrate its capability in processing cross-dimensional data and various potential applications, such as 1D signal detection, blind denoising, and geometric segmentation through point localization, we applied it to low-SNR two-photon calcium imaging video processing, achieving denoising, background separation, weak spike estimation, and data-efficient active neuron segmentation in a single pipeline; Lastly, to demonstrate its ability to detect signals under broader priors unrelated to signal shape, we applied it to fluorescence background estimation, enabling accurate estimation of arbitrary inhomogeneous backgrounds. These cases suggest that DEPAF holds considerable potential for further development.

DEPAF is user-friendly, as the adaption for different tasks can be achieved by simply replacing the POI examples and adjusting a hyperparameter related to detection sensitivity. This is possible due to the structural and regular features of fluorescence microscopy signals can be broadly and uniformly represented through POI, and the model architecture used strikes a good balance between specificity and generality. For instance, it focuses only on interference level rather than the exact type and parameters of interference; It utilizes general priors aligned with physical intuition to decode fitting results from discrete pixel space to continuous space; It utilizes contrast relationships for data processing without introducing additional manually set parameters or external ground truths. These make DEPAF lower the technical barriers to accessing high-performance fluorescence microscopy signal detection. By simplifying operations, it requires only minimal and intuitive input, eliminating the need for users to collect labor-intensive labeled datasets or parameterize physics-based imaging models that rely on specialized expertise.

The framework of DEPAF is scalable. As a model-agnostic solution, the U-net used in DEPAF serves as a simple end-to-end baseline, which can be easily extended to more advanced network architectures in the future. DEPAF also introduces an important perspective: various fluorescence microscopy signals can be represented through different POI examples. This means DEPAF can expand into a wider range of applications by incorporating more well-designed POI examples. Building on this concept, as demonstrated in tasks such as 3D SMLM and fluorescence background estimation, DEPAF supports co-optimization using multiple POI examples to simultaneously detect multiple types of signals. Therefore, its scalable and integrable architecture positions it as a potential unified model to address the lack of foundational frameworks in the field.

In conclusion, DEPAF has transformed the performance benchmarks of current fluorescence microscopy signal detection methods while maintaining its generalizability and scalability. We envision it as a standardized, high-performance signal detection framework applicable across cellular, subcellular, and molecular scales, facilitating the advancement of emerging techniques such as image-based multi-omics [32], localization-based live-cell super-resolution imaging, and their integration.

Methods

DEPAF pipeline. Our pipeline consists of two branches. In the first branch (Extended Data Fig. 1a), a deep learning model $\mathcal{M}_{\text{net}}$ is initialized and used to transform the original image I_{ori} into a fitting map I_{fit}, which encodes the POI positions and amplitudes. This map is further group-convolved with the POI examples to produce a denoised image I_{dns}. The $\mathcal{M}_{\text{net}}$ is optimized by maximizing the L_2 similarity between I_{ori} and I_{dns} with the L_1 regularization of I_{fit}. Once $\mathcal{M}_{\text{net}}$ is optimized, in the second branch (Extended Data Fig. 1b), an optimizable decoding module $\mathcal{M}_{\text{decode}}$ is initialized and utilized to decode the POI positions from I_{fit}. By subtracting I_{dns} from I_{ori}, a noise image I_{ns} is obtained. This noise image is then merged with $\hat{I}_{\text{fit}}$, the sum of I_{fit} in the channel dimension, to produce a noisy fitting map $\tilde{I}_{\text{fit}}$. The POI amplitudes are estimated based on $\tilde{I}_{\text{fit}}$. Subsequently, the estimated POI amplitudes are combined with the known POI examples and the obtained POI positions to generate a synthesized image I_{syn}. The $\mathcal{M}_{\text{decode}}$ is optimized by maximizing the L_1 similarity between I_{syn} and I_{dns}. After the optimizations are complete, the parameter-fixed $\mathcal{M}_{\text{net}}$ and $\mathcal{M}_{\text{decode}}$ can be utilized for POI detection on new images. The detailed process is presented below.

POI example construction. DEPAF applies min-max normalization to each of the C pre-acquired POI images $\{I_c^{\text{POI}}\}_{c=1}^{C}$, scaling their pixel intensities to $[0, 1]$. The normalized images are then stacked along the channel dimension to form the POI examples $P \in \mathbb{R}^{W \times H \times C}$, where $P(x, y, c) = I_c^{\text{POI}}(x, y)$.

First branch. In this branch, the $\mathcal{M}_{\text{net}}$, a full convolutional U-net neural network [33] (Extended Data Fig. 1c), is used to simultaneously fit all the POIs in I_{ori} and denoise the I_{ori}. The loss function for optimizing $\mathcal{M}_{\text{net}}$ is:

$$\mathcal{L}_{\text{learn}} = \|I_{\text{ori}} - I_{\text{dns}}\|_2^2 + \lambda \|I_{\text{fit}}\|_1, \tag{1}$$

where I_{ori} is the original image input to the model, I_{fit} is the fitting map output from the model, λ is the L_1 regularization strength, and I_{dns} is the denoised image calculated by:

$$I_{\text{dns}} = \sum_{c=1}^{C} I_{\text{fit}}(:, :, c) \circledast \bar{P}(:, :, c), \tag{2}$$

where "$\circledast$" denotes a convolution operation, ":" denotes all indices along a dimension, $\bar{P}$ is obtained by dividing the POI examples by the mean of the total pixel sums across all channels. The loss function $\mathcal{L}_{\text{learn}}$ comprises of an L_2 similarity term and an L_1 regularization term. The L_2 similarity term enforces similarity between I_{dns} and I_{ori} for reconstruction, while the L_1 regularization term promotes the sparsity of I_{fit} for denoising. The regularization strength λ, as a adjustable hyperparameter, can therefore control the denoising strength and sensitivity of POI detection.

Second branch. In this branch, the $\mathcal{M}_{\text{decode}}$, an optimizable decoding module, is used to decode the POI positions from I_{fit}. Since POI examples also share similarities with the noise in I_{ori} (not just with the POIs), I_{fit} may also contain fitting values produced by noise rather than by POIs. These noise fitting values typically do not exceed the noise level. Therefore, we can use a threshold θ to filter them out. However, since the noise level is difficult to estimate [52], an iterative optimization method is employed here to robustly address this issue. Specifically, We first sum I_{fit} along the channel dimension to obtain $\hat{I}_{\text{fit}}$. Then we initialize a θ to be optimized within the pixel value range of $\hat{I}_{\text{fit}}$. The fitting map after filtering out the noise fitting values is obtained by:

$$\hat{I}_{\text{fit}}^{\theta}(x, y) = \begin{cases} \hat{I}_{\text{fit}}(x, y), & \text{if } \hat{I}_{\text{fit}}(x, y) \geq \theta \\ 0, & \text{otherwise} \end{cases}. \tag{3}$$

To further partition the regions of interest (ROIs) containing an individual POI in $\hat{I}_{\text{fit}}^{\theta}$, the following two priors (Extended Data Fig. 1d) are considered.

- **Shape prior:** Due to the discrete nature and inherent size of pixels, a point in continuous coordinate space is represented as 1 to 4 pixels within a 2×2 pixel block in discrete pixel space under a sparsity constraint, which corresponds to the shape of each ROI.
- **Priority prior:** For ROIs containing 1 to 4 non-zero pixels, the range of points in continuous coordinate space they can represent increases sequentially. Following classical probability theory, the likelihood of their occurrence also increases accordingly. Their priority for partitioning is thus in ascending order.

Therefore, the pseudo-code of an algorithm (Extended Data Fig. 1e) for using the above priors to partition ROIs is as Algorithm 1.

Algorithm 1 ROI Partitioning

```
R ← ∅                                                          // Initialize empty ROI set
while IsNonZero(Î^θ_fit) do
    {(x_i, y_i)}_{i=1}^N ← FindNonZeroPixels(Î^θ_fit)          // Get non-zero pixels
    {n_i}_{i=1}^N ← Count8Neighbors({(x_i, y_i)}_{i=1}^N)      // Count 8-connected neighbors
    {(x_j, y_j)}_{j=1}^{N'} ← SelectMinNeighborPixels({(x_i, y_i), n_i}_{i=1}^N)   // Select pixels with fewest neighbors
    for each (x_j, y_j) ∈ {(x_j, y_j)}_{j=1}^{N'} do
        {R̂_k}_{k=1}^4 ← Get2x2BlocksAround(x_j, y_j)           // Four candidate 2 × 2 blocks
        R̂ ← SelectMaxSumBlock({R̂_k}_{k=1}^4, Î^θ_fit)          // Select block with maximum sum
        R ← R ∪ R̂                                              // Add block to ROI set
        ZeroOutBlock(Î^θ_fit, R̂)                               // Zeroing selected region
return R                                                       // Final ROI set
```

After ROI partitioning, the POI position $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ of i-th ROI $\hat{R}_i$ is estimated by calculating the center of mass:

$$\begin{aligned}
\hat{x}_i &= \frac{\sum_{(x,y)\in\hat{R}_i} x\hat{I}_{\text{fit}}(x,y)}{\sum_{(x,y)\in\hat{R}_i} \hat{I}_{\text{fit}}(x,y)}, \\
\hat{y}_i &= \frac{\sum_{(x,y)\in\hat{R}_i} y\hat{I}_{\text{fit}}(x,y)}{\sum_{(x,y)\in\hat{R}_i} \hat{I}_{\text{fit}}(x,y)}, \\
\hat{z}_i &= \frac{1}{C}\sum_{c=1}^{C}\sum_{(x,y)\in\hat{R}_i} cI_{\text{fit}}(x,y,c).
\end{aligned} \tag{4}$$

To further obtain the POI amplitudes, we first derive the noise image I_{ns} by subtracting the denoised image I_{dns} from the original image I_{ori}. We then normalize I_{ns} using the maximum value of $\bar{P}$ to ensure it is on the same scale as $\hat{I}_{\text{fit}}$. These are then synthesized into a noisy fitting map by:

$$\tilde{I}_{\text{fit}}(x,y) = \max\left(\frac{I_{\text{ns}}(x,y)}{\max \bar{P}}, \hat{I}_{\text{fit}}(x,y)\right). \tag{5}$$

The purpose of this step is to use noise to contaminate the fitting values below the noise level while preserving those above the noise level. Afterward, the POI amplitude A_i of i-th ROI $\hat{R}_i$ is estimated by calculating the maximum value after convolving each ROI with the POI example corresponding to its z-axis position:

$$A_i = \max\left\{\left[\tilde{I}_{\text{fit}} \odot M(\hat{R}_i)\right] \circledast \bar{P}(:,:,\hat{z}_i)\right\}, \tag{6}$$

where "$\odot$" represents element-wise multiplication, $M(\cdot)$ is a function that outputs a mask where the pixels within the input set are 1 and the rest are 0. The sub-pixel values of the POI example at $\hat{z}_i$ are obtained using interpolation. Based on the known POI examples and the obtained POI positions and amplitudes, a synthesized image I_{syn} can be generated by:

$$I_{\text{syn}}(x,y) = \sum_i A_i P\left(x - \hat{x}_i + \left\lfloor\frac{H}{2}\right\rfloor, y - \hat{y}_i + \left\lfloor\frac{W}{2}\right\rfloor, \hat{z}_i\right), \tag{7}$$

where H and W are the height and width of the POI examples, respectively. The sub-pixel values of the POI example at $\hat{x}_i$ and $\hat{y}_i$ are also obtained through interpolation.

Building upon the above construction, a critical trade-off emerges: choosing a too small θ that is below the noise level will result in the noise fitting values in $\hat{I}_{\text{fit}}$ being retained. This leads to the generation of multiple POIs in I_{syn} with amplitudes derived from the noise in $\tilde{I}_{\text{fit}}$. Due to the randomness and instability of noise, these noise-induced POIs cannot match the illusionary POIs in I_{dns}, which are obtained by convolving the noise fitting values with the POI examples. Therefore, I_{syn} will deviate from I_{dns}. Conversely, choosing a too large θ that is above the noise level will lead to the erroneous filtering of non-noise fitting values in $\hat{I}_{\text{fit}}$. This results in fewer non-noise-induced POIs being generated in I_{syn} compared to those in I_{dns}, similarly causing I_{syn} to deviate from I_{dns}. Therefore, the θ can be optimized by maximizing the similarity between I_{dns} and I_{syn}, i.e., minimizing the loss function:

$$\mathcal{L}_{\text{decode}} = \|I_{\text{dns}} - I_{\text{syn}}\|_1. \tag{8}$$

Due to the non-differentiability of $\mathcal{L}_{\text{decode}}$ with respect to θ, a gradient-free optimization method is used here. Specifically, we start with Bayesian optimization [53, 54] to quickly estimate the shape of the $\mathcal{L}_{\text{decode}}$-$\theta$ curve and find a suboptimal θ. After that, we perform a grid search method [55] centered around this suboptimal θ to find the optimal θ.

Enhancement techniques. Three enhancement techniques are employed in DEPAF: (1) center-aligned upsampling to improve POI localization accuracy (Supplementary Fig. 3), (2) inhomogeneous background learning to enhance interference resistance (Supplementary Fig. 4), and (3) POI diffusion to promote model convergence in multi-POI detection scenarios (Supplementary Fig. 5).

Center-aligned upsampling. The inherent pixel size limits DEPAF's ability to distinguish overlapping POIs and accurately estimate their positions using pixel-weighted centroids. To address this limitation, we reduce the pixel size by applying the same upsampling factor to both the POI examples and the original images at the start of our pipeline. To prevent localization bias, this process must ensure that the center position of the POI example remains consistent before and after upsampling. A detailed derivation and additional explanations of this method are presented in Supplementary Note 1.

Inhomogeneous background learning. We add a homogeneous patch as the background POI example P_{bg} and an output channel of $\mathcal{M}_{\text{net}}$ to generate a background fitting map I_{fitBG}. This method enables the model to learn arbitrary inhomogeneous backgrounds while detecting signals. When this technique is applied, the loss function to optimize $\mathcal{M}_{\text{net}}$ becomes:

$$\hat{\mathcal{L}}_{\text{learn}} = \|I_{\text{ori}} - (I_{\text{dns}} + I_{\text{fitBG}} \circledast P_{\text{BG}})\|_2^2 + \lambda \|I_{\text{fit}}\|_1 . \tag{9}$$

Note that since the background does not exhibit natural sparsity, I_{fitBG} is not constrained by L_1 regularization here.

POI diffusion. When our method is applied to multi-POI detection tasks, the number of POI examples determines the channel number of the output I_{fit} in $\mathcal{M}_{\text{net}}$ (Equation 2). Therefore, as the number of POI examples increases, the search space of I_{fit} also expands, potentially affecting the convergence stability of $\mathcal{M}_{\text{net}}$. To address this, we introduce the POI diffusion technique. This technique first employs bisection to divide all the POI examples into two groups. Within each group, mean pooling is used to merge the POIs into the central channel of that group. In the final convolutional layer of $\mathcal{M}_{\text{net}}$, only the kernels corresponding to the central channels are activated for training. During training, if the validation loss does not decrease after a predefined number of iterations, POI diffusion is triggered. At this point, bisection is recursively applied to determine new active channels, and POI examples are re-merged into their nearest active channels using mean pooling. The kernels corresponding to these newly activated channels then inherit their initial weights from previously activated kernels through nearest copy and are activated for training. This process repeats until all channels are activated.

Implementation details of DEPAF models. In tasks requiring high localization accuracy and the ability to distinguish overlapping POIs, we use an upsampling factor of 1.5 to upsample both the original images and POI examples with cubic spline interpolation [17]. In general cases, a factor of 1 (i.e., no upsampling) is sufficient. Following this, the training data is normalized to a 0-1 scale using the 99th percentile and the minimum value of the entire training set. It is then randomly sampled into overlapping patches for training. If inhomogeneous background learning is used, the background POI example is set to be three times the size of the central 95% pixel-value distribution area of the signal POI example.

The neural network $\mathcal{M}_{\text{net}}$ is trained using the Adam optimizer [56] with an initial learning rate of 0.001. Validation loss is measured after every 20 iterations. If the loss does not fall below the historical minimum over 50 consecutive measurements, POI diffusion is triggered and the historical minimum validation loss is reset. If all the channels of the POI examples are activated, the learning rate is halved until it reaches 0.0001. If the loss still does not decrease after another 50 consecutive measurements, the early stopping method [57] is executed, and the model with the lowest historical validation loss is saved as the optimal model.

In the second branch of our method, Bayesian optimization uses the expected improvement (EI) criterion [53, 54] to select the next evaluation point, with a maximum of 80 evaluations. The grid search comprises 100 steps with a step size set to 1/3000 of the pixel value range of $\hat{I}_{\text{fit}}$. The POI example values in sub-pixel coordinates are obtained using the same interpolation method as in the center-aligned upsampling.

For inferring new images, normalization uses the pre-saved 99th percentile and minimum value from the training set. Since $\mathcal{M}_{\text{net}}$ uses a fully convolutional architecture, our method can handle images of any size. To prevent memory overflow, images are processed using overlapping sliding windows, with the window size equal to the training patch size and a stride of half the window size. To suppress stitching artifacts caused by edge effects, a Gaussian importance weighting method [58] is used.

Comprehensive testing on multiple datasets demonstrated the high effectiveness of the provided settings, rendering additional user adjustments unnecessary. Users only need to adjust one hyperparameter: the regularization strength λ in the loss function for optimizing $\mathcal{M}_{\text{net}}$. The λ determines the model's noise robustness and POI detection sensitivity. A higher λ enhances noise robustness but may miss low-amplitude POIs, and vice versa.

Acquisition of POI examples. For SMLM, the POI examples are obtained by imaging calibrated fluorescent beads under precisely the same conditions as the sample imaging. In MERFISH analysis, an elliptical Gaussian distribution with different sigma values along the x and y axes, fitted from the hybridization image, is used as the POI example. For spike estimation, the POI example is determined by averaging the normalized spikes from the original image that fit a single exponential decay function [59]. For active neuron segmentation, a 2D standard Gaussian distribution with a sigma value of 1 is used as the POI example. The POI example used for inhomogeneous background estimation is a patch of constant value, with a size three times that of the central 95% pixel-value distribution area of the signal POI example.

Simulating data for performance evaluation. To evaluate the performance of DEPAF, we generate multiple simulated datasets for PSF localization benchmarking (Fig. 2a and Extended Data Fig. 2), denoising evaluation (Extended Data Fig. 4), and background estimation evaluation (Fig. 5 and Extended Data Fig. 6). This dataset collection spans a wide range of PSF densities, SNRs, and background variation scales. Details on the simulation process and parameters are provided in Supplementary Note 2 and Supplementary Table 2.

Data processing. *PSF localization benchmarking.* Due to computational constraints, DECODE limits the maximum PSF density for simulated data generation in each training epoch. This makes it impractical to use the same high PSF densities as in most of the 18 simulated datasets for training. Therefore, we adopted a zero-shot testing approach to evaluate the performance of DECODE models. Specifically, the models were trained on datasets with a PSF density of 1 PSF/μm^2 and subsequently tested on datasets with higher PSF densities. This approach reflects a common practical scenario where the PSF density in simulated data may not match the experimental data. For fair comparison, we also conducted the zero-shot testing on DEPAF. To ensure there were no systematic errors in the coordinate system definition, such as offsets, xy-axis flips, or clear scaling errors, the localization results of each comparison method were manually checked and corrected.

Dynamic 2D SMLM. Due to photobleaching, the brightness of fluorescent molecules gradually diminishes over time, hindering the ability of localization algorithms to use consistent sensitivity in detecting single molecules across all SMLM video frames. To address this issue, we employed a histogram-matching-based bleaching correction method [60] to preprocess the collected SMLM videos. This method first analyzes the grayscale histogram of the first frame, then adjusts the grayscale histograms of subsequent frames to match that of the first one, effectively correcting the brightness decay caused by photobleaching and reduces noise increase.

Using single-molecule positions obtained from 60 SMLM video frames, we rendered super-resolution frames. To enhance image smoothness, bilinear interpolation was applied when mapping single-molecule positions to pixel values.

3D SMLM benchmarking. Both the SMLM videos and bead images were directly obtained from a public SMLM challenge [31] without any preprocessing. The localization results were directly rendered as scatter plots (Extended Data Fig. 3d) without applying any visual enhancement techniques. The localization results of each comparison method were manually checked and corrected to ensure that there were no systematic errors in the coordinate system definition.

MERFISH analysis. We applied the histogram-matching-based photobleaching correction method [60] to address brightness fluctuations in MERFISH images across different hybridization rounds and z-layers caused by photobleaching. Specifically, using the first z-layer of the first hybridization round as a template, we corrected photobleaching across all rounds and z-layers, then applied a uniform sensitivity for MERFISH spot signal detection in all images. Compared to the original MERFISH technique, which iteratively adjusted the thresholds (corresponding to sensitivity) in localization algorithms for each round's images through trial and error [2, 46], this method is more concise.

Each round of hybridization image needs to be registered for barcode decoding. Unlike the original MERFISH technique, which calculated the transformation matrix by fitting and matching the fiducial bead centroids, we found that the entire bead images, including fluorescence background, can be used for registration. Specifically, we first computed the KAZE features points [61] for the bead image corresponding to each round of hybridization image and select the bead image with the most feature points as the template. Then, we calculated the transformation matrix between the feature points of each bead image and those of the template bead image. This method offers greater automation and robustness because the KAZE feature points are more numerous and include descriptors beyond mere positional information.

To accommodate the ultra-high-density MERFISH spot signal detection capability provided by our method, we improved the barcode decoding algorithm to address the more frequent matching conflict issue (i.e., the

distance between one spot signal and multiple spot signals is less than the matching threshold). Specifically, we first matched spot signals directly where no conflicts existed and retained all spot signals with matching conflicts. Next, we used hierarchical clustering [62], with the matching threshold as the shortest distance between clusters, to group all conflicting spot signals into clusters. Within each cluster, we applied the Munkres assignment algorithm [63] to obtain the optimal matching that minimizes the overall distance. This method allows ultra-high-density MERFISH spot signal detection results to be decoded into barcodes more accurately while maintaining reasonable time and memory costs and makes the decoding performance less sensitive to the matching threshold.

Two-photon calcium imaging data processing. All two-photon calcium imaging videos were aligned non-rigidly using the NoRMCorre algorithm [64] to eliminate motion artifacts before processing. Each pixel's time-series values in the video were treated as a 1D image and subjected to median filtering to suppress spike noise. These 1D images were then normalized by dividing them by the difference between their median and first quartile values to address spike amplitude variations between different neurons and cortical regions.

Given the low amount of training data required by our method, we selected and cropped neighborhoods of a few pixels with larger values from these 1D images as potential spike data for training the spike estimation model. After training, we obtained noise-free 1D images using our method's synthesis function and reshaped them back to their original video format, yielding denoised two-photon calcium imaging videos.

Similarly, we selected a few frames with high pixel value sums from the denoised videos as potential active neuron signal data. These frames were used to train the source extraction model based on ultra-high-density PSF localization, using a standard Gaussian PSF as the POI example. After training, to generate the final neuron segmentation masks, we first performed spatial clustering of the localized PSF positions for each frame using the DBSCAN algorithm [50]. For each cluster, we calculated the boundary points, generated the boundary curve using cubic spline interpolation, and extended the boundary curve along its normal direction to compensate for boundary shrinkage caused by the standard Gaussian PSF radius. Finally, we converted each boundary curve into a segmentation mask and merged them using the method employed in SUNS [22], based on criteria such as the mask's area, center of mass, intersection-over-union (IoU), consume ratio, and the number of consecutive frames.

All parameters, including those for mask generation and merging, were optimized within a pre-set range using Bayesian optimization [53, 54] based on the ground-truth segmentation masks of the training set, aiming to minimize the loss function:

$$\mathcal{L}_{\text{paramOpt}} = 1 - S_{\text{F1}} + \frac{1 - S_{\text{mIoU}}}{1 + 1000(1 - S_{\text{F1}})}, \tag{10}$$

where S_{F1} is the F1 score and S_{mIoU} is the mean IoU score.

Evaluation metrics. *SMLM.* The evaluation of SMLM encompasses two aspects: the ability to detect all PSFs and the accuracy of the detected PSF positions. The former is evaluated using the Jaccard index (JI), which is defined as:

$$S_{\text{JI}} = \frac{100 N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}, \tag{11}$$

where N_{TP} is the true positive rate, N_{FP} is the false positive rate, and N_{FN} is the false negative rate. The latter is evaluated using the lateral and axial root mean square error (RMSE), which are the distance in nanometers between the localized and the ground-truth positions in lateral and axial direction, respectively. They are defined as:

$$S_{\text{latRMSE}} = \sqrt{\frac{1}{N_{\text{TP}}} \sum_{M} \left[(\hat{x}_i - x_i^{\text{GT}})^2 + (\hat{y}_i - y_i^{\text{GT}})^2 \right]}, \tag{12}$$

$$S_{\text{axRMSE}} = \sqrt{\frac{1}{N_{\text{TP}}} \sum_{M} (\hat{z}_i - z_i^{\text{GT}})^2}, \tag{13}$$

where $M = \{(\hat{x}_i, \hat{y}_i, \hat{z}_i), (x_i^{\text{GT}}, y_i^{\text{GT}}, z_i^{\text{GT}})\}_{i=1}^{N_{\text{TP}}}$ are the matched position pairs within a distance threshold of 1 pixel. The JI and RMSE exhibit a trade-off relationship. To provide a balanced overall assessment of localization performance, the "efficiency" is employed, defined as:

$$S_{\text{eff}} = 100 - \sqrt{(100 - S_{\text{JI}})^2 + \alpha_1^2 S_{\text{latRMSE}}^2 + \alpha_2^2 S_{\text{axRMSE}}^2}, \tag{14}$$

where α_1 and α_2 are weighting parameters set to $1\,\text{nm}^{-1}$ and $0.5\,\text{nm}^{-1}$, respectively [31].

MERFISH analysis. Evaluating MERFISH analysis involves assessing the accuracy of RNA expression quantification and barcode decoding. The accuracy of RNA expression quantification is determined by the Pearson correlation coefficient between the logarithmic scale of the copy number of each RNA species detected through MERFISH and the abundance determined by FPKM. It is defined as:

$$S_{\text{corr}} = r(\log_{10}(\boldsymbol{g}_{\text{detect}}), \log_{10}(\boldsymbol{g}_{\text{FPKM}})), \tag{15}$$

where $\boldsymbol{g}_{\text{detect}}$ is the vector of the copy number of each RNA species detected through MERFISH, $\boldsymbol{g}_{\text{FPKM}}$ is the vector of the abundance of each RNA species determined by bulk sequencing in FPKM, and $r(\cdot,\cdot)$ is the Pearson correlation coefficient function. Since the codebook uses MHD4 code [2], a match is determined if the Hamming distance between the decoded barcode and any barcode in the codebook is less than or equal to 1. The accuracy of barcode decoding is evaluated by the average per-bit $0 \to 1$ and $1 \to 0$ error rates in the matched barcodes. The average per-bit $0 \to 1$ error rate is defined as:

$$S_{0\to 1} = \frac{1}{N_{\text{bit}}} \sum_{j=1}^{N_{\text{bit}}} \frac{\sum_{i=1}^{N_{\text{match}}} \mathbf{1}(b_{(i,j)}^{\text{decode}} = 1 \wedge b_{(i,j)}^{\text{GT}} = 0)}{\sum_{i=1}^{N_{\text{total}}} 1(b_{(i,j)}^{\text{GT}} = 0)}, \tag{16}$$

and the average per-bit $1 \to 0$ error rate is defined as:

$$S_{1\to 0} = \frac{1}{N_{\text{bit}}} \sum_{j=1}^{N_{\text{bit}}} \frac{\sum_{i=1}^{N_{\text{match}}} \mathbf{1}(b_{(i,j)}^{\text{decode}} = 0 \wedge b_{(i,j)}^{\text{GT}} = 1)}{\sum_{i=1}^{N_{\text{total}}} 1(b_{(i,j)}^{\text{GT}} = 1)}, \tag{17}$$

where N_{match} is the total number of matched barcodes, N_{bit} is the number of bits in the codebook, $b_{(i,j)}^{\text{decode}}$ is the j-th bit of the i-th decoded barcode, $b_{(i,j)}^{\text{GT}}$ is the j-th bit of the ground-truth barcode matched with the i-th decoded barcode, and $\mathbf{1}(\cdot)$ is the indicator function, which returns 1 if the input is true and 0 otherwise.

Active neuron segmentation. The neuron segmentation results are evaluated by comparing them with ground-truth labels. To match the segmentation masks with the ground-truth labels, the distance between a ground-truth mask M_i^{GT} and a detected mask M_j is determined based on their IoU score as follows:

$$\text{dist}(M_i^{\text{GT}}, M_j) = \begin{cases} 1 - \text{IoU}(M_i^{\text{GT}}, M_j), & \text{IoU}(M_i^{\text{GT}}, M_j) \geq 0.5 \\ 2, & \text{IoU}(M_i^{\text{GT}}, M_j) < 0.5 \end{cases}. \tag{18}$$

Here,

$$\text{IoU}(M_i^{\text{GT}}, M_j) = \frac{|M_i^{\text{GT}} \cap M_j|}{|M_i^{\text{GT}} \cup M_j|}. \tag{19}$$

In this formula, a distance of 2 is assigned to mask pairs with an IoU score below 0.5, indicating that these masks are too dissimilar to be considered matches. The Munkres assignment algorithm [63] is then applied to this distance matrix to solve the linear assignment problem, identifying matched mask pairs where the distance is less than 2 as true positives. Next, the performance is quantified using F1 score defined as:

$$S_{\text{F1}} = \frac{2}{S_{\text{R}}^{-1} + S_{\text{P}}^{-1}}, \tag{20}$$

and the mean IoU score defined as:

$$S_{\text{mIoU}} = \frac{\sum_{(i,j)\in R_{\text{TP}}} \text{IoU}(M_i^{\text{GT}}, M_j)}{N_{\text{TP}}}, \tag{21}$$

where $S_{\text{R}} = N_{\text{TP}}/N_{\text{GT}}$ denotes recall, $S_{\text{P}} = N_{\text{TP}}/N_{\text{detect}}$ denotes precision, R_{TP} is the set of true positive pairs, N_{TP} is the number of true positives, N_{GT} is the total number of ground-truth labels, and N_{detect} is the number of neurons detected by the evaluated methods.

Sample preparation and imaging of dynamic 2D SMLM. See Supplementary Note 3 for details on the sample preparation and imaging of dynamic 2D SMLM.

Data available

The data and trained models for PSF localization benchmarking (Fig. 2a and Extended Data Fig. 2), dynamic 2D SMLM (Fig. 2b,c), denoising evaluation (Extended Data Fig. 4), and background estimation evaluation (Fig. 5 and Extended Data Fig. 6) are available at https://zenodo.org/records/14073540. The full-length supplementary videos 7–10 are available at https://zenodo.org/records/14145974. The 3D SMLM benchmarking dataset (Extended Data Fig. 3) is publicly available at https://srm.epfl.ch/srm/dataset/challenge-3D-simulation/index.html. The video J123 (Fig. 4 and Extended Data Fig. 5) is publicly available at https://zenodo.org/records/1659149. Raw data for MERFISH analysis (Fig. 3) are available upon request from the authors of ref. [47]. All other data supporting this study are available from the corresponding authors upon reasonable request.

Code available

DEPAF is accessible as Supplementary Software. The latest version can be found at https://github.com/zhang-fengdi/DEPAF. The code to reproduce the results in this paper can be accessed at https://github.com/zhang-fengdi/DEPAF/tree/main/paper_reproduction.

References

[1] Lelek, M., Gyparaki, M.T., Beliu, G., Schueder, F., Griffie, J., Manley, S., Jungmann, R., Sauer, M., Lakadamyali, M., Zimmer, C.: Single-molecule localization microscopy. Nat Rev Methods Primers **1** (2021) https://doi.org/10.1038/s43586-021-00038-x

[2] Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., Zhuang, X.: Spatially resolved, highly multiplexed RNA profiling in single cells. Science **348**(6233), 6090 (2015) https://doi.org/10.1126/science.aaa6090

[3] Grienberger, C., Giovannucci, A., Zeiger, W., Portera-Cailliau, C.: Two-photon calcium imaging of neuronal activity. Nat Rev Methods Primers **2**(1), 67 (2022) https://doi.org/10.1038/s43586-022-00147-1

[4] Speiser, A., Muller, L.R., Hoess, P., Matti, U., Obara, C.J., Legant, W.R., Kreshuk, A., Macke, J.H., Ries, J., Turaga, S.C.: Deep learning enables fast and dense single-molecule localization with high accuracy. Nat Methods **18**(9), 1082–1090 (2021) https://doi.org/10.1038/s41592-021-01236-x

[5] Saguy, A., Alalouf, O., Opatovski, N., Jang, S., Heilemann, M., Shechtman, Y.: DBlink: Dynamic localization microscopy in super spatiotemporal resolution via deep learning. Nat Methods **20**(12), 1939–1948 (2023) https://doi.org/10.1038/s41592-023-01966-0

[6] Li, X., Zhang, G., Wu, J., Zhang, Y., Zhao, Z., Lin, X., Qiao, H., Xie, H., Wang, H., Fang, L., Dai, Q.: Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. Nat Methods **18**(11), 1395–1400 (2021) https://doi.org/10.1038/s41592-021-01225-0

[7] Moffitt, J.R., Hao, J., Bambah-Mukku, D., Lu, T., Dulac, C., Zhuang, X.: High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc Natl Acad Sci USA **113**(50), 14456–14461 (2016) https://doi.org/10.1073/pnas.1617699113

[8] Mockl, L., Roy, A.R., Petrov, P.N., Moerner, W.E.: Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network bgnet. Proc Natl Acad Sci USA **117**(1), 60–67 (2020) https://doi.org/10.1073/pnas.1916219117

[9] Nehme, E., Freedman, D., Gordon, R., Ferdman, B., Weiss, L.E., Alalouf, O., Naor, T., Orange, R., Michaeli, T., Shechtman, Y.: DeepSTORM3D: Dense 3D localization microscopy and PSF design by deep learning. Nat Methods **17**(7), 734–740 (2020) https://doi.org/10.1038/s41592-020-0853-5

[10] Li, X., Li, Y., Zhou, Y., Wu, J., Zhao, Z., Fan, J., Deng, F., Wu, Z., Xiao, G., He, J., Zhang, Y., Zhang, G., Hu, X., Chen, X., Zhang, Y., Qiao, H., Xie, H., Li, Y., Wang, H., Fang, L., Dai, Q.: Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. Nat Biotechnol **41**(2), 282–292 (2023) https://doi.org/10.1038/s41587-022-01450-8

[11] Li, X., Hu, X., Chen, X., Fan, J., Zhao, Z., Wu, J., Wang, H., Dai, Q.: Spatial redundancy transformer for self-supervised fluorescence image denoising. Nat Comput Sci **3**(12), 1067–1080 (2023) https://doi.org/10.1038/s43588-023-00568-2

[12] Zhang, G., Li, X., Zhang, Y., Han, X., Li, X., Yu, J., Liu, B., Wu, J., Yu, L., Dai, Q.: Bio-friendly long-term subcellular dynamic recording by self-supervised image enhancement microscopy. Nat Methods **20**(12), 1957–1970 (2023) https://doi.org/10.1038/s41592-023-02058-9

[13] Qiao, C., Zeng, Y., Meng, Q., Chen, X., Chen, H., Jiang, T., Wei, R., Guo, J., Fu, W., Lu, H., Li, D., Wang, Y., Qiao, H., Wu, J., Li, D., Dai, Q.: Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy. Nat Commun **15**(1), 4180 (2024) https://doi.org/10.1038/s41467-024-48575-9

[14] Qu, L., Zhao, S., Huang, Y., Ye, X., Wang, K., Liu, Y., Liu, X., Mao, H., Hu, G., Chen, W., Guo, C., He, J., Tan, J., Li, H., Chen, L., Zhao, W.: Self-inspired learning for denoising live-cell super-resolution microscopy. Nat Methods (2024) https://doi.org/10.1038/s41592-024-02400-9

[15] Laine, R.F., Heil, H.S., Coelho, S., Nixon-Abell, J., Jimenez, A., Wiesner, T., Martinez, D., Galgani, T., Regnier, L., Stubb, A., Follain, G., Webster, S., Goyette, J., Dauphin, A., Salles, A., Culley, S., Jacquemet, G., Hajj, B., Leterrier, C., Henriques, R.: High-fidelity 3D live-cell nanoscopy through data-driven enhanced super-resolution radial fluctuation. Nat Methods **20**(12), 1949–1956 (2023) https://doi.org/10.1038/s41592-023-02057-w

[16] Li, Y., Mund, M., Hoess, P., Deschamps, J., Matti, U., Nijmeijer, B., Sabinina, V.J., Ellenberg, J., Schoen, I., Ries, J.: Real-time 3D single-molecule localization using experimental point spread functions. Nat Methods **15**(5), 367–369 (2018) https://doi.org/10.1038/nmeth.4661

[17] Babcock, H.P., Zhuang, X.: Analyzing single molecule localization microscopy data using cubic splines. Sci Rep **7**(1), 552 (2017) https://doi.org/10.1038/s41598-017-00622-w

[18] Babcock, H., Sigal, Y.M., Zhuang, X.: A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. Opt Nanoscopy **1**(6) (2012) https://doi.org/10.1186/2192-2853-1-6

[19] Giovannucci, A., Friedrich, J., Gunn, P., Kalfon, J., Brown, B.L., Koay, S.A., Taxidis, J., Najafi, F., Gauthier, J.L., Zhou, P., Khakh, B.S., Tank, D.W., Chklovskii, D.B., Pnevmatikakis, E.A.: CaImAn an open source tool for scalable calcium imaging data analysis. Elife **8** (2019) https://doi.org/10.7554/eLife.38173

[20] Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L.F., Dalgleish, H., Carandini, M., Harris, K.D.: Suite2p: Beyond 10,000 neurons with standard two-photon microscopy. Preprint at https://www.biorxiv.org/content/10.1101/061507v2 (2017)

[21] Fazel, M., Wester, M.J., Mazloom-Farsibaf, H., Meddens, M.B.M., Eklund, A.S., Schlichthaerle, T., Schueder, F., Jungmann, R., Lidke, K.A.: Bayesian multiple emitter fitting using reversible jump Markov chain Monte Carlo. Sci Rep **9**(1), 13791 (2019) https://doi.org/10.1038/s41598-019-50232-x

[22] Bao, Y., Soltanian-Zadeh, S., Farsiu, S., Gong, Y.: Segmentation of neurons from fluorescence calcium recordings beyond real-time. Nat Mach Intell **3**(7), 590–600 (2021) https://doi.org/10.1038/s42256-021-00342-x

[23] Soltanian-Zadeh, S., Sahingur, K., Blau, S., Gong, Y., Farsiu, S.: Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. Proc Natl Acad Sci USA **116**(17), 8554–8563 (2019) https://doi.org/10.1073/pnas.1812995116

[24] Deneux, T., Kaszas, A., Szalay, G., Katona, G., Lakner, T., Grinvald, A., Rozsa, B., Vanzetta, I.: Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. Nat Commun **7**, 12190 (2016) https://doi.org/10.1038/ncomms12190

[25] Boyd, N., Jonas, E., Babcock, H., Recht, B.: DeepLoco: Fast 3D localization microscopy using neural networks. Preprint at https://www.biorxiv.org/content/10.1101/267096v1 (2018)

[26] Fu, S., Shi, W., Luo, T., He, Y., Zhou, L., Yang, J., Yang, Z., Liu, J., Liu, X., Guo, Z., Yang, C., Liu, C., Huang, Z.L., Ries, J., Zhang, M., Xi, P., Jin, D., Li, Y.: Field-dependent deep learning enables high-throughput whole-cell 3D super-resolution imaging. Nat Methods **20**(3), 459–468 (2023) https://doi.org/10.1038/s41592-023-01775-5

[27] Zhang, W., Zhang, Z., Bian, L., Wang, H., Suo, J., Dai, Q.: High-axial-resolution single-molecule localization under dense excitation with a multi-channel deep U-net. Opt Lett **46**(21), 5477–5480 (2021) https://doi.org/10.1364/OL.441536

[28] Nehme, E., Weiss, L.E., Michaeli, T., Shechtman, Y.: Deep-STORM: Super-resolution single-molecule microscopy by deep learning. Optica **5**(4), 458–464 (2018) https://doi.org/10.1364/OPTICA.5.000458

[29] Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., *et al.*: Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems **33**, 21271–21284 (2020)

[30] Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15750–15758 (2021)

[31] Sage, D., Pham, T.A., Babcock, H., Lukes, T., Pengo, T., Chao, J., Velmurugan, R., Herbert, A., Agrawal,

A., Colabrese, S., Wheeler, A., Archetti, A., Rieger, B., Ober, R., Hagen, G.M., Sibarita, J.B., Ries, J., Henriques, R., Unser, M., Holden, S.: Super-resolution fight club: Assessment of 2D and 3D single-molecule localization microscopy software. Nat Methods **16**(5), 387–395 (2019) https://doi.org/10.1038/s41592-019-0364-4

[32] Vandereyken, K., Sifrim, A., Thienpont, B., Voet, T.: Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet **24**(8), 494–515 (2023)

[33] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241 (2015)

[34] Rust, M.J., Bates, M., Zhuang, X.: Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods **3**(10), 793–5 (2006) https://doi.org/10.1038/nmeth929

[35] Linde, S., Loschberger, A., Klein, T., Heidbreder, M., Wolter, S., Heilemann, M., Sauer, M.: Direct stochastic optical reconstruction microscopy with standard fluorescent probes. Nat Protoc **6**(7), 991–1009 (2011) https://doi.org/10.1038/nprot.2011.336

[36] Sharonov, A., Hochstrasser, R.M.: Wide-field subdiffraction imaging by accumulated binding of diffusing probes. Proc Natl Acad Sci USA **103**(50), 18911–6 (2006) https://doi.org/10.1073/pnas.0609643104

[37] Boehr, D.D., McElheny, D., Dyson, H.J., Wright, P.E.: The dynamic energy landscape of dihydrofolate reductase catalysis. Science **313**(5793), 1638–42 (2006) https://doi.org/10.1126/science.1130258

[38] Ouyang, W., Aristov, A., Lelek, M., Hao, X., Zimmer, C.: Deep learning massively accelerates super-resolution localization microscopy. Nat Biotechnol **36**(5), 460–468 (2018) https://doi.org/10.1038/nbt.4106

[39] Wang, Y., Jia, S., Zhang, H.F., Kim, D., Babcock, H., Zhuang, X., Ying, L.: Blind sparse inpainting reveals cytoskeletal filaments with sub-nyquist localization. Optica **4**(10), 1277–1284 (2017) https://doi.org/10.1364/OPTICA.4.001277

[40] Descloux, A., Grussmayer, K.S., Radenovic, A.: Parameter-free image resolution estimation based on decorrelation analysis. Nat Methods **16**(9), 918–924 (2019) https://doi.org/10.1038/s41592-019-0515-7

[41] Shechtman, Y., Sahl, S.J., Backer, A.S., Moerner, W.E.: Optimal point spread function design for 3D imaging. Phys Rev Lett **113**(13), 133902 (2014) https://doi.org/10.1103/PhysRevLett.113.133902

[42] Donelan, J.M., Li, Q., Naing, V., Hoffer, J.A., Weber, D.J., Kuo, A.D.: Biomechanical energy harvesting: Generating electricity during walking with minimal user effort. Science **319**(5864), 807–10 (2008) https://doi.org/10.1126/science.1149860

[43] Pavani, S.R.P., Thompson, M.A., Biteen, J.S., Lord, S.J., Liu, N., Twieg, R.J., Piestun, R., Moerner, W.E.: Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. Proc Natl Acad Sci USA **106**(9), 2995–2999 (2009) https://doi.org/10.1073/pnas.0900245106

[44] Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H.: Visualization of single RNA transcripts in situ. Science **280**(5363), 585–90 (1998) https://doi.org/10.1126/science.280.5363.585

[45] Raj, A., Bogaard, P., Rifkin, S.A., Oudenaarden, A., Tyagi, S.: Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods **5**(10), 877–9 (2008) https://doi.org/10.1038/nmeth.1253

[46] Moffitt, J.R., Zhuang, X.: RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). Method Enzymol **572**, 1–49 (2016) https://doi.org/10.1016/bs.mie.2016.03.020

[47] Lu, Y., Liu, M., Yang, J., Weissman, S.M., Pan, X., Katz, S.G., Wang, S.: Spatial transcriptome profiling by MERFISH reveals fetal liver hematopoietic stem cell niche architecture. Cell Discov **7**(1), 47 (2021) https://doi.org/10.1038/s41421-021-00266-1

[48] Zhang, Y., Rózsa, M., Liang, Y., Bushey, D., Wei, Z., Zheng, J., Reep, D., Broussard, G.J., Tsang, A., Tsegaye, G., *et al.*: Fast and sensitive gcamp calcium indicators for imaging neural populations. Nature **615**(7954), 884–891 (2023)

[49] Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., Looger, L.L., Svoboda, K., Kim, D.S.: Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature **499**(7458), 295–300 (2013) https://doi.org/10.1038/nature12354

[50] Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, vol. 96, pp. 226–231 (1996)

[51] Perlin, K.: An image synthesizer. ACM SIGGRAPH Comput Graph **19**(3), 287–296 (1985) https://doi.org/10.1145/325165.325247

[52] Liu, X., Tanaka, M., Okutomi, M.: Single-image noise level estimation for blind denoising. IEEE Trans Image Process **22**(12), 5226–37 (2013) https://doi.org/10.1109/TIP.2013.2283400

[53] Snoek, J., Larochelle, H., Adams, R.P.: Practical bayesian optimization of machine learning algorithms. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 2, pp. 2951–2959 (2012)

[54] Gelbart, M.A., Snoek, J., Adams, R.P.: Bayesian optimization with unknown constraints. Preprint at https://arxiv.org/abs/1403.5607 (2014)

[55] Syarif, I., Prugel-Bennett, A., Wills, G.: SVM parameter optimization using grid search and genetic algorithm to improve classification performance. TELKOMNIKA (Telecommunication Computing Electronics and Control) **14**(4) (2016) https://doi.org/10.12928/telkomnika.v14i4.3956

[56] Kingma, D.P.: Adam: A method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014)

[57] Prechelt, L.: Early stopping — but when? In: Neural Networks: Tricks of the Trade, pp. 53–67. Springer, Berlin (2012)

[58] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nat Methods **18**(2), 203–211 (2021) https://doi.org/10.1038/s41592-020-01008-z

[59] Akerboom, J., Chen, T.W., Wardill, T.J., Tian, L., Marvin, J.S., Mutlu, S., Calderon, N.C., Esposti, F., Borghuis, B.G., Sun, X.R., Gordus, A., Orger, M.B., Portugues, R., Engert, F., Macklin, J.J., Filosa, A., Aggarwal, A., Kerr, R.A., Takagi, R., Kracun, S., Shigetomi, E., Khakh, B.S., Baier, H., Lagnado, L., Wang, S.S., Bargmann, C.I., Kimmel, B.E., Jayaraman, V., Svoboda, K., Kim, D.S., Schreiter, E.R., Looger, L.L.: Optimization of a GCaMP calcium indicator for neural activity imaging. J Neurosci **32**(40), 13819–40 (2012) https://doi.org/10.1523/JNEUROSCI.2601-12.2012

[60] Miura, K.: Bleach correction imagej plugin for compensating the photobleaching of time-lapse sequences. F1000Res **9**, 1494 (2020) https://doi.org/10.12688/f1000research.27171.1

[61] Alcantarilla, P.F., Bartoli, A., Davison, A.J.: KAZE features. In: Proceedings of the 12th European Conference on Computer Vision, pp. 214–227 (2012)

[62] Murtagh, F., Contreras, P.: Algorithms for hierarchical clustering: An overview. WIREs Data Min Knowl **2**(1), 86–97 (2011) https://doi.org/10.1002/widm.53

[63] Munkres, J.: Algorithms for the assignment and transportation problems. J Soc Indust Appl Math **5**(1), 32–38 (1957) https://doi.org/10.1137/0105003

[64] Pnevmatikakis, E.A., Giovannucci, A.: NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. J Neurosci Methods **291**, 83–94 (2017) https://doi.org/10.1016/j.jneumeth.2017.07.031

Acknowledgements

J.G. was supported by the National Key Research and Development Program of China (Grant No. 2022YFA1103401). R.H. was supported by the National Key Project (Grant No. 62331006). Y.F. was supported

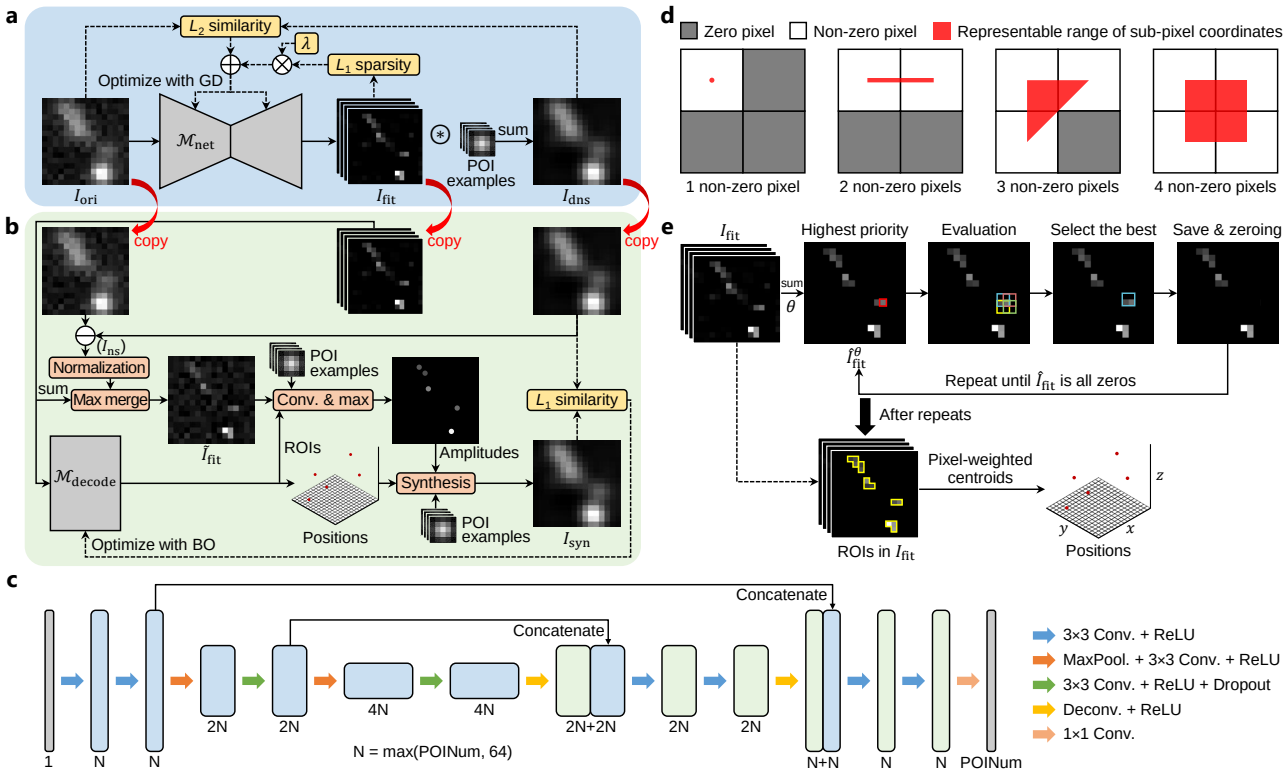
by the National Key R&D Program of China (Grant No. 2022YFC3300704) and the National Key Project (Grant No. 62331006). We thank Prof. Minping Qian from Peking University for her helpful discussions and suggestions.

Author contributions

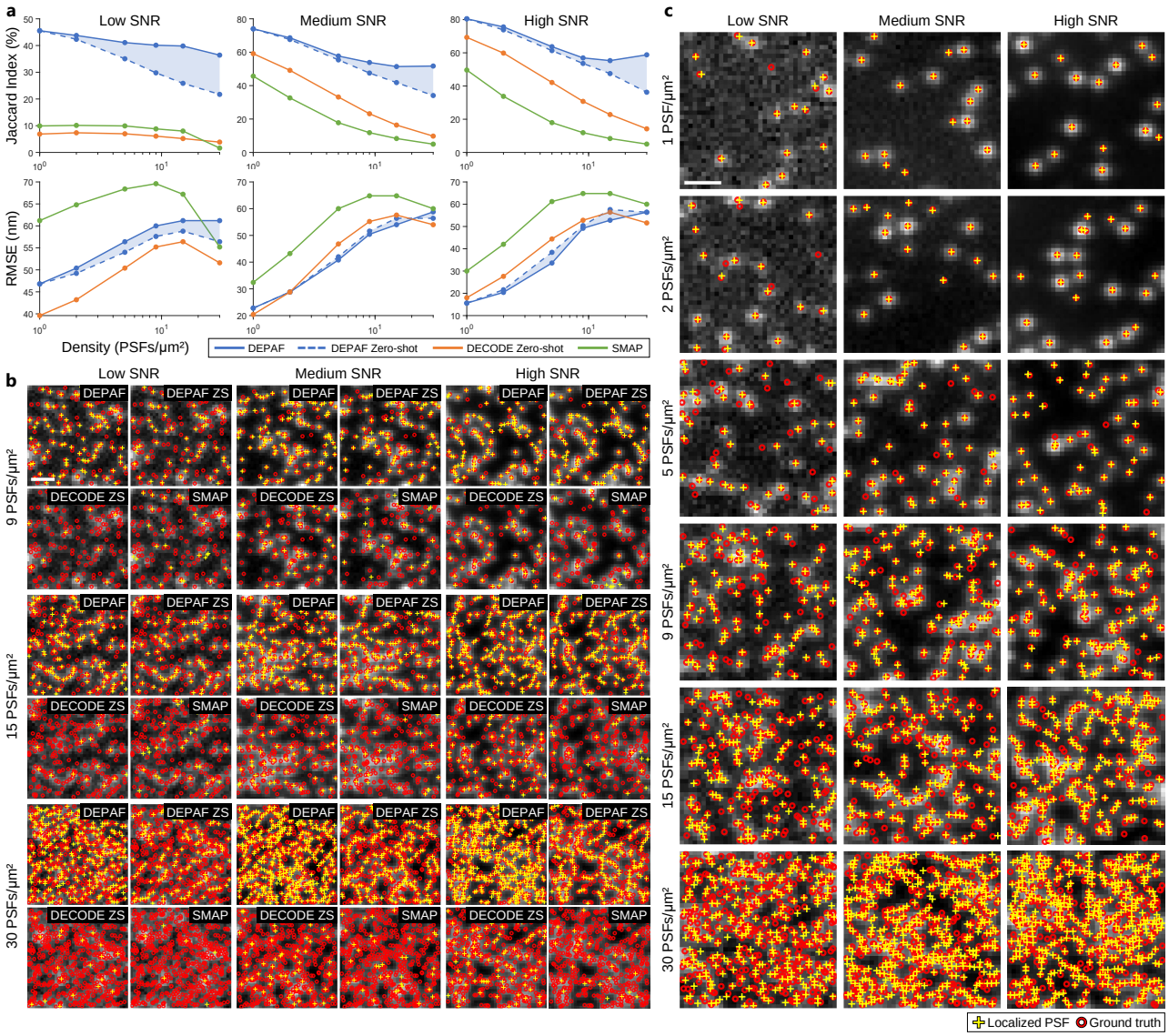
R.H. and J.G. initiated the project. F.Z. developed the software, analyzed the data, and drafted the manuscript with contributions from all authors. R.H., J.G., H.M and D.G. contributed to result interpretation. J.G. and M.X. provided data support. Y.F. assisted with analysis and result discussions. R.H., J.G. and X.J. supervised the project and the final manuscript.

Competing interests

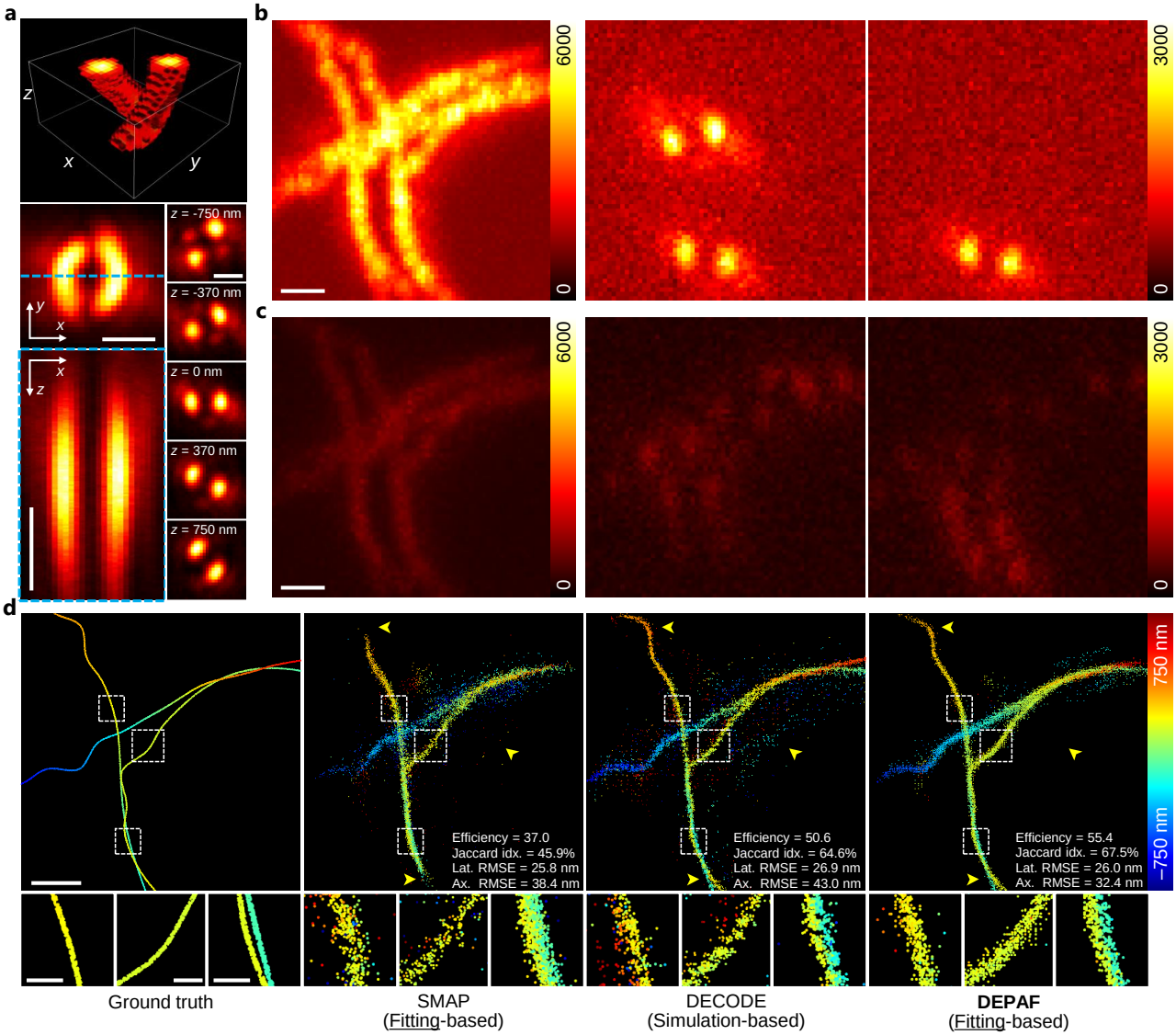
F.Z., J.G., R.H. and M.X. have a pending patent application on the presented method. The other authors declare no competing interests.



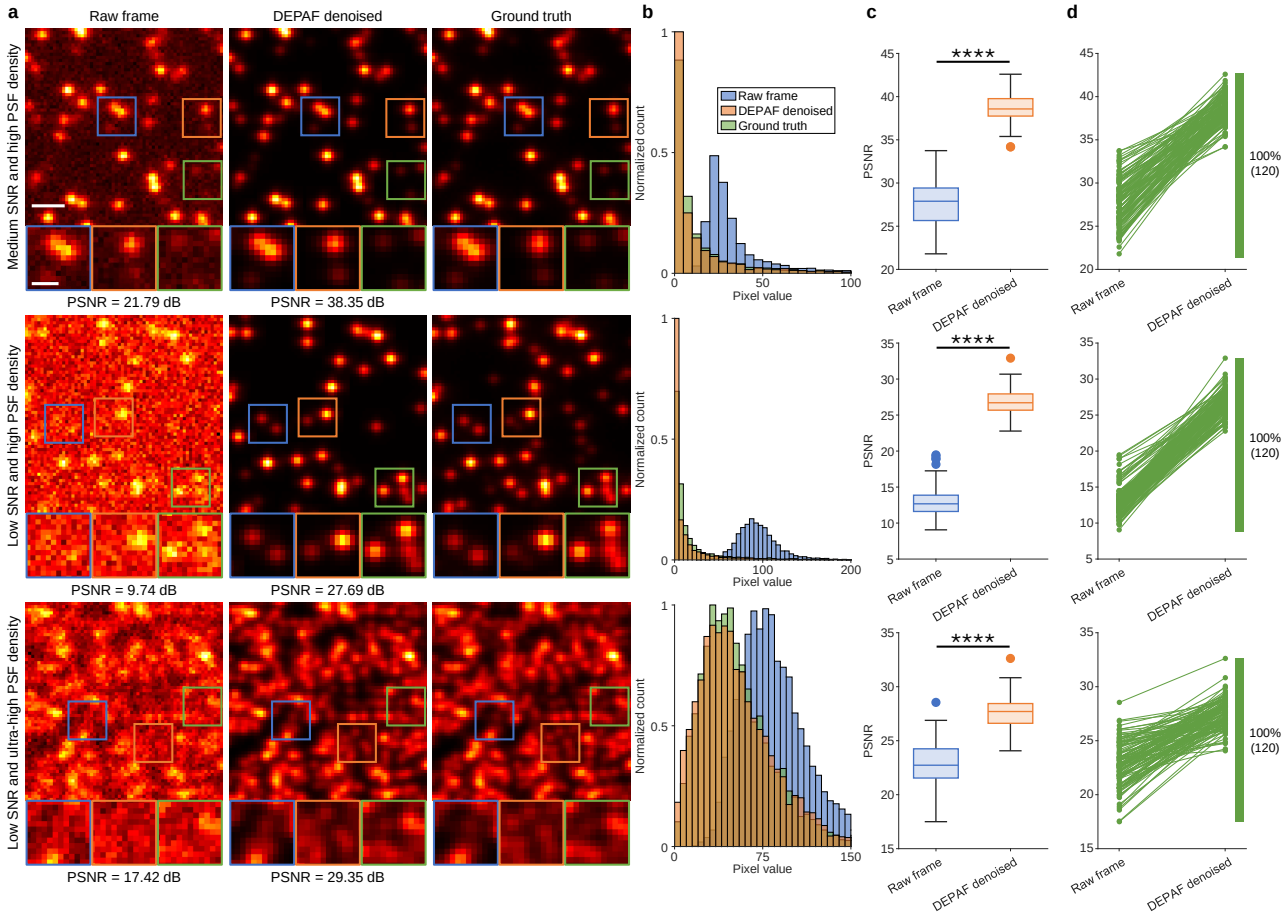
Extended Data Fig. 1 | Method detailed overview. a, The first branch of our method. GD, gradient descent. **b,** The second branch of our method. Conv., convolution; Max, maximum; BO, Bayesian optimization. **c,** The architecture of $\mathcal{M}_{\text{net}}$. This neural network comprises a down-sampling stage followed by an up-sampling stage. Each stage consists of several (de)convolutional layers with 3×3 filters, along with non-trainable operations including max-pooling (stride of 2), dropout (probability of 50%), and rectified linear unit (ReLU) activation. The first convolutional layer in the down-sampling stage has N filters, where N is the maximum of 64 and the number of POI examples. In each down-sampling phase, the resolution is halved while the number of filters is doubled; the reverse is true for each up-sampling phase. Black arrows denote skip connections. POINum, the number of POI examples; Conv., convolutional layer; MaxPool., max-pooling layer; Deconv., deconvolutional layer. **d,** Schematic depiction of the priors used in $\mathcal{M}_{\text{decode}}$. For ROIs containing 1 to 4 non-zero pixels, their representable range of sub-pixel coordinates in continuous coordinate space expands progressively. According to classical probability theory, their occurrence probability increases correspondingly, resulting in ascending partitioning priority ordered by the number of non-zero pixels. **e,** Schematic depiction of the processing procedure for $\mathcal{M}_{\text{decode}}$. Starting from the image $\hat{I}^{\theta}_{\text{fit}}$ obtained by summing the channels of I_{fit} and filtering it with threshold θ, the algorithm iteratively selects pixels with the minimal number of neighbors, computes pixel sums in surrounding 2×2 blocks, selects and saves the block with the largest pixel sum, and updates the image until all ROIs are partitioned.



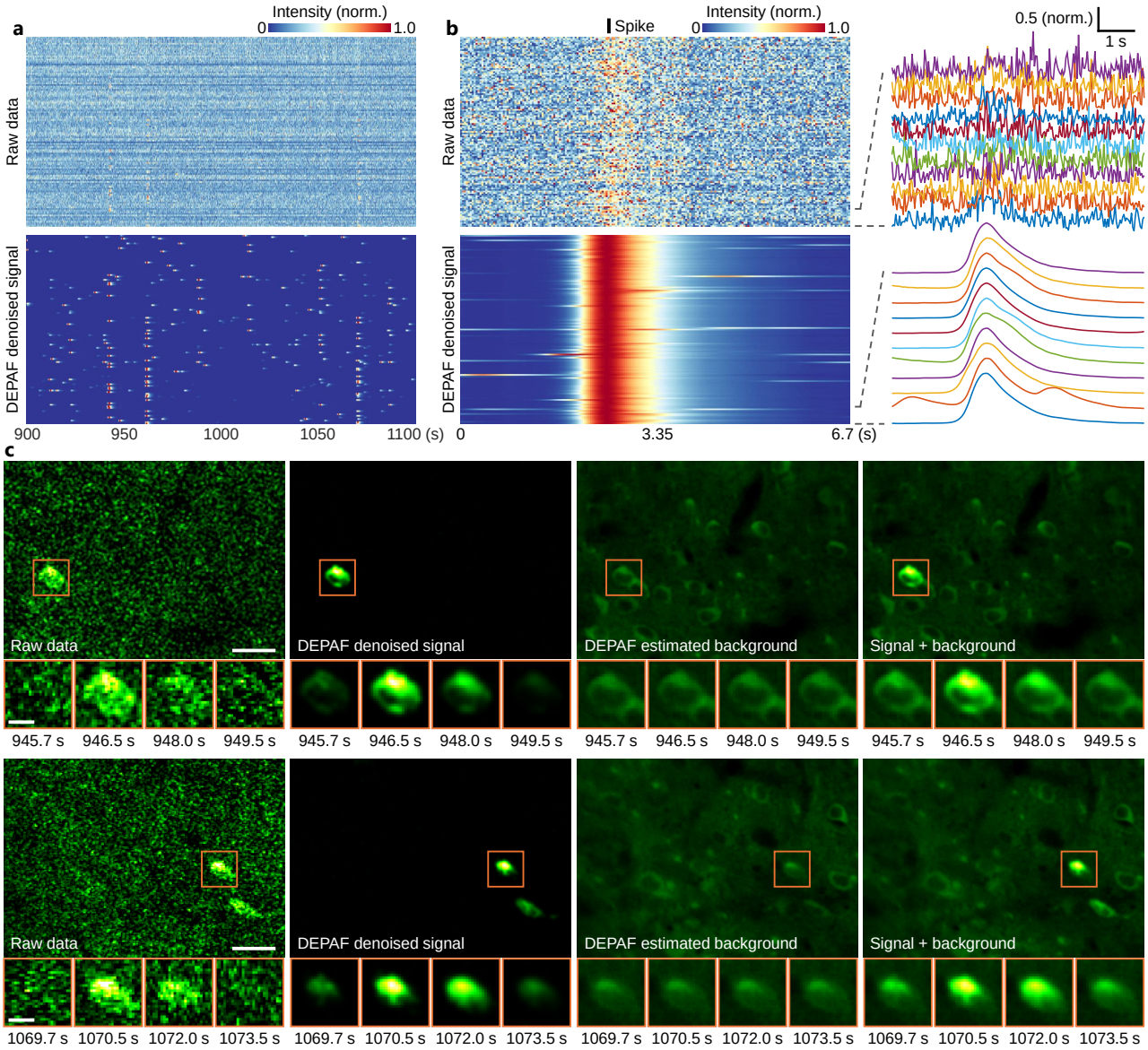
Extended Data Fig. 2 | Detailed comparison results of localization performance across different methods on 18 simulated SMLM datasets. a, The line charts compares the Jaccard index and RMSE of different methods across different SNR levels and PSF densities (log scale) in 18 simulated datasets. **b,** Each column of the images visualizes the localization results of various methods at PSF densities of 9, 15, and 30 PSFs/μm^2 for low, medium, and high SNR levels, respectively. Scale bar, 1 μm. ZS, zero-shot testing. **c,** Each column of images provides a more detailed visualization of DEPAF's localization results at PSF densities of 1, 2, 5, 9, 15, and 30 PSFs/μm^2, corresponding to low, medium, and high SNR levels. Scale bar, 1 μm.



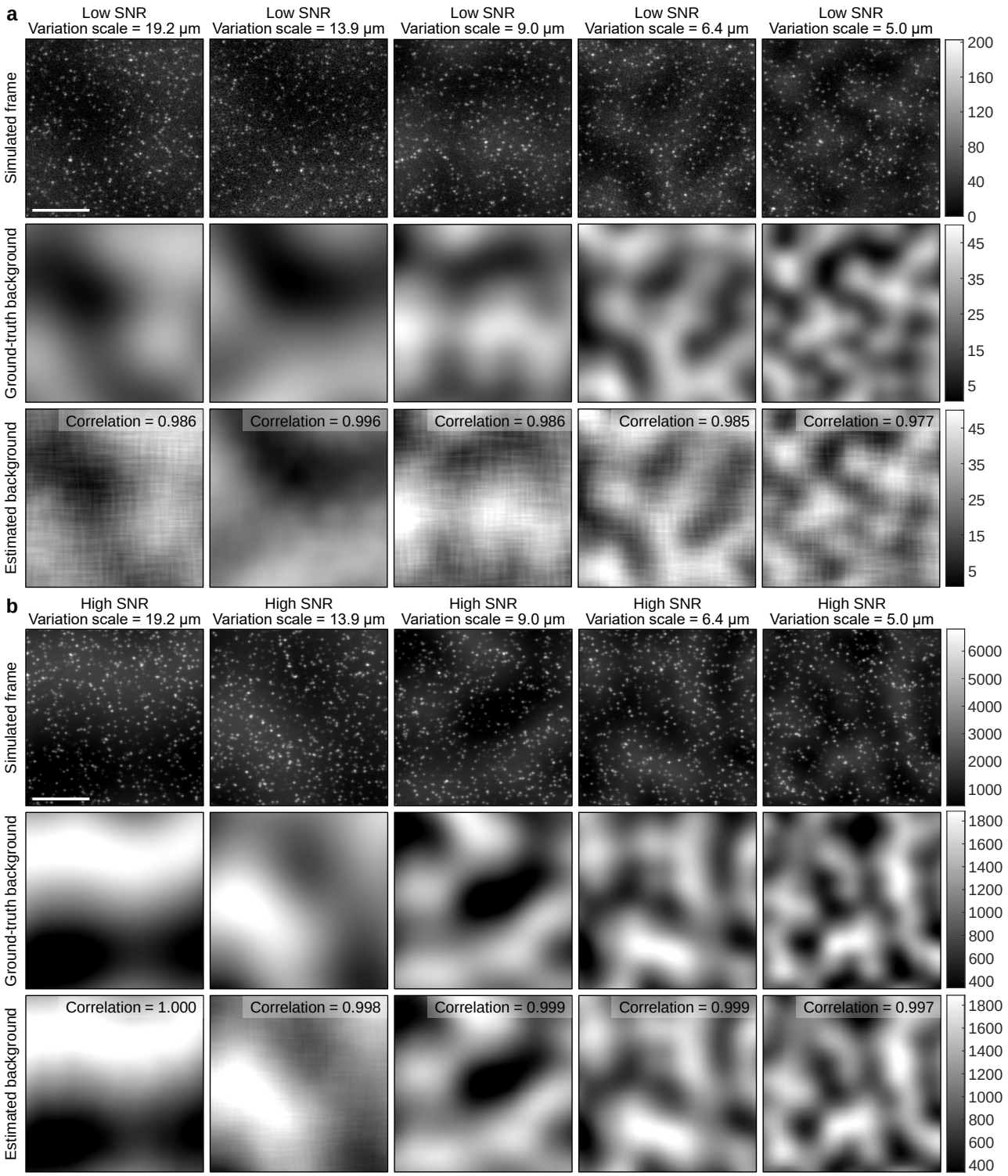
Extended Data Fig. 3 | DEPAF enables high-robust-accuracy 3D SMLM in imaging conditions with low SNR and high PSF density. **a,** 3D view of real double-helix PSFs acquired using calibrated beads at 151 different z-axis positions (ranging from -750 nm to 750 nm, with a 10 nm interval) for generating the simulated dataset (top). Maximum intensity projection of the double-helix PSFs in the xy view (middle left). Scale bar, 1 μm. Intensity profile in the xz view along the blue dashed line in the maximum intensity projection (bottom left). Scale bar, 500 nm. Representative PSF images at different z-axis positions (bottom right). Scale bar, 1 μm. **b,** Temporal maximum intensity projection (left) and example frames (middle and right) from 2500 simulated frames showing typical Alexa647-labeled STORM data [31]. Scale bar, 1 μm. **c,** Data used for reconstruction, with 10 times the PSF density and an average photon count reduced to 10.8% of that in **b**. Scale bar, 1 μm. **d,** Ground-truth 3D super-resolution image for generating the simulated dataset, and the reconstructed image using SMAP, DECODE and DEPAF. Yellow arrows indicate specific regions to highlight differences in resolution and structural details among the methods. Scale bar, 1 μm. Idx., index. Lat., lateral. Ax., axial. Zoomed-in views of the boxed regions present detailed visualization of the reconstructions. Scale bar, 200 nm.



Extended Data Fig. 4 | DEPAF achieves strong denoising performance across different SNRs and PSF densities through its image synthesis function, benefiting from its accurate estimation of POI positions and amplitudes. **a,** Comparison of simulated data (Methods) in raw frames, DEPAF denoised frames, and ground truths. Each row shows images under different SNR and PSF densities. Scale bar, 1 μm. The subplots below display magnified details of the denoising results. Scale bar, 500 nm. The peak signal-to-noise ratio (PSNR) is indicated below each raw frame and DEPAF denoised frame. **b,** Histograms of pixel values under different conditions, comparing the distributions of input frames (blue), denoised frames (orange), and ground truths (green). **c,** Box plots of PSNR values for each condition, showing the performance difference between original images and DEPAF denoised images. P values calculated by one-sided paired t-test are specified with asterisks; $****P < 0.0001$ for all comparisons. **d,** Comparison of PSNR between original images and denoised images under different conditions, with PSNR markedly improved in each condition.



Extended Data Fig. 5 | Extended results of DEPAF in denoising and background separation for two-photon calcium imaging videos. a, Fluorescence traces from 120 pixels. Top, raw recording. Bottom, background-separated signals denoised by DEPAF. Norm., normalized. **b,** Calcium fluctuations induced by 120 action potentials. All spikes were normalized and temporally aligned to the black bar. Zoomed-in traces are displayed in the right panel. Norm., normalized. **c,** First column: raw somatic signal recording. Second column: background-separated signals denoised by DEPAF. Images were spatially smoothed using a standard Gaussian kernel. Third column: backgrounds estimated by DEPAF. Fourth column: sum of backgrounds and signals. Scale bar, 20 μm. Zoomed-in views of the boxed regions present calcium transients within a 3.8-second time window. Scale bar, 6 μm.



Extended Data Fig. 6 | Extended results for DEPAF in estimating arbitrary inhomogeneous backgrounds under low and high SNR levels and different background variation scales. The results for low and high SNR levels are shown in **a** and **b**, respectively. In both panels, the layout follows the order below. First row: frames with inhomogeneous backgrounds simulated using Perlin noise [51]. Second row: ground truths of the simulated inhomogeneous backgrounds. Third row: backgrounds estimated by DEPAF. The Pearson correlation coefficients between the estimated and ground-truth backgrounds are displayed in the top-right corner of each estimated background image. Scale bar, 10 μm.