Neural-physics adaptive reconstruction reveals 3D subcellular nanostructures over a large depth of field

Shuang Fu[1], Wei Shi[1], Yue Fei[1], Ruixiong Wang[1], Tianlun Zhang[1], Yiming Li[1,*]

[1] Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[*] Correspondence should be addressed to Y.L. (liym2019@sustech.edu.cn)

Abstract

Achieving large depth-of-field super-resolution imaging deep inside samples is often hindered by unknown aberrations and overlapping molecular signals in 3D single-molecule localization microscopy. Here, we present LUNAR, a blind localization method that simultaneously resolves overlapping molecular signals and corrects for aberrations using a neural-physics model. Through self-supervised learning on single-molecule data without requiring prior knowledge or accurate calibration, LUNAR synergistically optimizes a physical model with a neural network to estimate key physical parameters (e.g., 3D positions, photons, aberrations) of molecules. Its hybrid Transformer network effectively handles PSFs of varying sizes, achieving theoretically maximum localization precision of consecutive blinking events. Extensive simulations and experiments demonstrate that LUNAR consistently outperforms current state-of-the-art methods, reducing localization error by up to sixfold in the presence of unknown aberrations and molecular overlaps, enabling high-fidelity whole-cell reconstruction of mitochondria, nucleus, and neuronal cytoskeleton at great depths.

Introduction

As a member of the super-resolution (SR) microscopy family, single-molecule localization microscopy (SMLM) has improved spatial resolution by an order of magnitude or more beyond the diffraction limit, while maintaining outstanding molecular specificity and high contrast[1,2]. Recent advancements have pushed its resolution down to Ångström scales[3–5] and enabled whole-cell 3D SR imaging at high throughput [6,7], unlocking unprecedented opportunities for structural

cell biology research. In SMLM, sparse blinking molecules are localized with a point spread function (PSF) model. The high precision permitted by SMLM requires accurate PSF modeling, which restricts most SMLM applications to sample of interest positioned just above the coverslip, with a limited depth-of-field (DOF).

When imaging deeper within the samples, mismatched refractive index (RI) between the sample medium and objective immersion medium compromises PSF calibration accuracy, especially when combined with sample-induced aberrations[8]. Increasing the DOF using PSF engineering often leads to larger PSF and greater emitter overlaps, affecting reconstruction quality. PSF models are usually derived from isolated point emission patterns either from beads[9] or single-molecule blinking events[10,11]. This poses significant challenges for whole-cell longitudinal imaging at greater depth, where in situ PSF models are very different from bead-calibrated PSFs on coverslip and isolated emitters are difficult to obtain.

In recent years, deep learning (DL) methods like convolutional neural networks (CNNs) have demonstrated excellent performance in SMLM, especially for difficult tasks such as high-density molecule localization[10,11], live-cell imaging[12], aberration estimation[13], and spatially variant PSF analysis[6]. However, these methods heavily rely on labeled training data, either obtained through laborious experimental collection or simulation using deterministic models. Existing DL localization methods require carefully calibrated PSF models to supervise the training process, where any model mismatch can degrade performance. This (PSF-)supervised learning mechanism not only limits the generalizability of DL models[14,15], but also fails to fully leverage the abundant information contained in unlabeled raw data.

A promising approach for blind SMLM is unsupervised learning, which might learn useful representations directly from the data itself, e.g., PSF models and molecule positions. However, this remains challenging as knowledge of both samples and aberrations are absent. Although unsupervised/self-supervised learning has been widely applied to image denoising and deconvolution, these methods were usually image-to-image transformations and designed to capture relationships between image pixels with learning mechanisms based on prior knowledge such as spatial redundancy[16,17],

spatiotemporal dependence[18,19], or structure consistency[20]. In SMLM, it is essential to learn the accurate physical process behind the data and corresponding inverse functions, which should localize single molecules with a precision one order of magnitude smaller than the image pixel size. Previous work is limited to Gaussian PSF model which is only valid for few imaging modalities[21]. Therefore, a deep neural network that can efficiently learn the precise experimental PSF model from the data is still missing in the field. Properly training DL models for single-molecule localization without labels or PSF constraint is quite difficult as the localization network and physical knowledge must be jointly learned in a well-coordinated manner, while minimizing the use of prior information (regularization) to reduce bias.

Different from the popular image-to-image transformations where networks can be readily trained with simple pixel-wise loss functions, deep learning-based single-molecule localization is a structured prediction task which retrieves 3D point cloud from 2D diffraction-limited single molecule images with high precision. Besides, blinking events in SMLM are spatiotemporally stochastic: an emitter can last over multiple frames, and its neighbors also vary during this period. The capability of the DL network to fully extract this information from single-molecule movies with various PSF sizes and complex shapes is key for high-accuracy physical parameters retrieval (e.g. 3D positions, aberrations). Although existing DL methods for SMLM have tried to utilize CNNs to capture this spatiotemporal information[6,11], the inherent locality of convolution kernels restricts that it only works well for small PSFs within few frames. Therefore, a network architecture that can fully exploit the spatiotemporal information of single-molecule patterns with different sizes across multiple frames, which is a typical scenario for large-DOF SR imaging, is also missing in the field.

To tackle these challenges, we present LUNAR, a blind localization method that utilizes a neural-physics model to learn the imaging formation model from the data itself, thereby enabling high-accuracy large-DOF SR imaging at great depth (**Fig. 1**). LUNAR uniquely integrates the training of a localization network with the learning of a physical PSF to maximize the likelihood of the experimental data. It is based on probabilistic prediction method with uncertainty estimation which can be applied for

reliable biological studies. This integration allows LUNAR to fully exploit information from overlapped single-molecule data with unknown aberrations. Additionally, the localization capability of the LUNAR network is bolstered by a hybrid CNN-Transformer architecture that efficiently utilizes spatiotemporal information of the blinking single molecules. Through a synchronized learning strategy designed for this neural-physics model, LUNAR enhances SMLM reconstruction with improved robustness and accuracy. We demonstrate that LUNAR could precisely localize overlapped emitters with unknown aberrations, facilitating high-fidelity whole-cell SR imaging, which managed to resolve nanostructures of neuronal cytoskeleton inside tissue slices.

Results

Design of LUNAR framework

LUNAR is a deep generative model that jointly estimates the 3D positions and optical aberrations of fluorescent molecules directly from the SMLM raw images (**Fig. 1**). Inspired by variational autoencoders (VAEs)[22], LUNAR employs an encoder-decoder architecture with intermediate outputs as latent variables. However, unlike typical VAEs that rely solely on neural networks, LUNAR integrates a neural network-based encoder with a physical model-based decoder (**Methods**), which we term neural-physics model. This design ensures that the generation process strictly follows physical laws, with intermediate outputs having explicit meanings. To briefly illustrate this idea, we start with a probabilistic model describing the SMLM imaging process:

$$p_\theta(d,h) = p_\theta(d|h)p_\theta(h), \tag{1}$$

where $p_\theta(d,h)$ is the joint probability distribution of the observed data d and latent variables h. In this work, h refers to the unknown positions and brightnesses of the emitters that need to be localized. The prior distribution $p_\theta(h)$ is set to a uniform distribution as molecules can appear anywhere. The conditional probability $p_\theta(d|h)$, which is modeled by the physics decoder, describes the stochastic image formation process of deterministic emitters. It incorporates both the optical system response and camera acquisition process. θ represents the generative parameters that need to be learned, and in this work, we set θ as a vector of Zernike coefficients to represent the

aberrations of the PSF model (**Methods**, **Supplementary Fig. 1** and **Supplementary Note 2**). The multiple-emitter localization task can be viewed as the inverse of the imaging process, solving the posterior distribution $p_\theta(h|d)$ of emitter positions h given the data d (**Methods** and **Supplementary Note 1**). However, this posterior is generally intractable due to the complexity involved in the imaging process. Therefore, we use the network encoder (**Fig. 1** and **Supplementary Fig. 2**) to learn an approximate posterior function $q_\phi(h|d)$, where ϕ denotes the neural network weights.

Due to the unknown overlapped emitters and intractable posterior, directly maximizing the marginal likelihood $p_\theta(d)$ is infeasible in this context. Instead, we maximize its variational lower bound, as is done in classic VAEs (**Methods** and **Supplementary Note 1**). To efficiently train our neural-physics model, we developed a synchronized learning strategy inspired by the reweighted wake-sleep[23] and expectation-maximization[24] algorithms (**Fig. 1a** and **Methods**). The synchronized learning process alternates between two phases: physics learning and localization learning. In the physics learning phase, the network encoder receives the raw SMLM frame as the input and outputs the prediction about the emitter posterior. The physics decoder then draws Monte Carlo samples from this posterior to generate synthetic data with importance weights. The reconstruction loss, which measures the probability of the raw data and synthetic data, is computed to update the physics decoder. In the localization learning phase, the physics decoder samples the prior to simulate data with known molecule positions. These data are then used to train the network encoder through a probabilistic localization loss (**Methods** and **Supplementary Note 1**). This helps the network encoder better approximate the posterior given the current imaging model, thereby enabling precise 3D localizations.

The network encoder in LUNAR plays a crucial role by providing both the posterior for PSF learning and the localizations for SR image reconstruction. Unlike previous works[6,11] that relied on 2D U-Nets for two-stage feature extraction (**Supplementary Fig. 3**), we adopted a decoupled network design (**Fig. 1b**) that explicitly accounts for two important characteristics of SMLM data: (1) Each emitter

pattern appears only within a local area corresponding to the PSF size; (2) Emitters may persist or reappear over multiple frames, providing additional positional information. To capture spatial information, we developed a feature extraction module (FEM) based on U-NeXt (**Methods**). This module incorporates designs from ConvNeXt[25], which greatly improves traditional convolution blocks on vision tasks. We believe this design facilitates precise feature extraction for PSFs with complex shape and various sizes. To further extract the temporal information that blinking single molecules appear in multiple frames, we developed a temporal attention module (TAM) based on the Transformer[26] architecture (**Methods**). The self-attention mechanism of Transformer allows spatial features to interact with each other across the time domain, outperforming conventional CNNs in terms of localization across multi-frame single-molecule data. Finally, a convolution-based output module is used to predict the position parameters with uncertainties (**Methods**).

LUNAR could be used in two modes: localization learning (LL) and synchronized learning (SL). LL is performed in PSF-supervised manner and suitable for experiments with well-calibrated PSFs, typically imaged near coverslips in RI-matched medium. SL is performed by self-supervised learning using experimental data and aims at addressing the more challenging scenarios, where PSF calibration is inaccurate or unavailable. By applying these advanced learning algorithms and network design, LUNAR provides a better understanding of the probability distributions in the localization process, allowing for more accurate and robust reconstruction of super-resolved images (**Fig. 1c**).

LUNAR achieves superior performance for PSFs of different sizes and DOF ranges

To evaluate LUNAR's performance, we first quantified the single-molecule localization error and compared it against theoretical precision limits, the Cramér-Rao Lower Bound (CRLB, **Supplementary Note 3**). The localization error was measured as the 3D root mean square error ($RMSE_{3D}$, **Supplementary Note 4**) at each investigated z-position. $RMSE_{3D,\ avg}$ represents the average $RMSE_{3D}$ across the DOF range. We selected three PSFs of different DOF ranges and sizes: astigmatic PSF, 3 μm Tetrapod

PSF and 6 μm Tetrapod PSF (**Methods**). Single molecules were simulated across multiple frames (1/3/5/7/9 frames) to investigate networks' capability of utilizing temporal information for single-molecule localization.

For comparison, we evaluated the state-of-the-art FD-DeepLoc[6] (**Supplementary Fig. 3**), a network that leverages CNN to capture spatiotemporal information, alongside LUNAR, which integrates CNN with Transformer. As shown in **Supplementary Fig. 4**, FD-DeepLoc could achieve CRLB for astigmatic PSF, whose size is relatively small, in the single-frame condition. However, as the PSF size increases, FD-DeepLoc cannot achieve theoretically limited localization accuracy anymore for 3 μm and 6 μm Tetrapod PSF. For the 6 μm Tetrapod PSF, the $RMSE_{3D,\ avg}$ of LUNAR (28.7 nm) is 19% better than that of FD-DeepLoc (35.4 nm). Moreover, FD-DeepLoc cannot achieve CRLB localization accuracy for single molecules across multiple frames for all three PSFs investigated, while LUNAR approaches the multi-frame CRLB for various PSFs and temporal context lengths (**Supplementary Fig. 5 and Supplementary Fig. 6**). For example, at a 5-frames temporal context length, the $RMSE_{3D,\ avg}$ of LUNAR on the astigmatic PSF is 9.1 nm, compared to 11.0 nm for FD-DeepLoc, representing a ~20% improvement in localization accuracy. For the 6 μm Tetrapod PSF, the $RMSE_{3D,\ avg}$ of LUNAR is 13.5 nm, compared to 21.6 nm for FD-DeepLoc, highlighting a ~40% improvement (**Fig. 2b** and **c**). We attribute these improvements to the advanced design of LUNAR network, which includes the larger receptive filed provided by the U-NeXt and the enhanced ability to capture temporal relationships through the Transformer.

Next, we assessed LUNAR's performance using simulated blinking molecules of different densities and PSFs (**Fig. 2d-f**). We employed the same photophysical model as the SMLM challenge[27], with a practical camera noise model to accurately simulate the imaging and data acquisition process (**Supplementary Note 5**). The three aforementioned PSFs with different DOFs were used. For each PSF type, we applied three levels of molecule densities (low, medium, high, **Supplementary Fig. 7**). Additional coma and spherical aberrations were introduced to better mimic real-world conditions that include unknown aberrations. To minimize the impact of training randomness, each model was trained five times with identical settings, and performance

metrics were reported as mean ± standard deviation. Detection accuracy (Jaccard index, JI), 3D localization error ($RMSE_{3D}$), and 3D efficiency were used for evaluation (**Supplementary Note 4**).

Similar to the previous single-molecule localization results, LUNAR outperforms FD-DeepLoc across all datasets with different temporal context lengths (**Supplementary Fig. 8** and **9, Supplementary Table 1** and **2**), as indicated by both better detection efficiency and localization accuracy. When using only single frame for inference, LUNAR achieves average improvements in $RMSE_{3D}$ of 5%, 8% and 12% compared to FD-DeepLoc for datasets based on astigmatic PSF, 3 μm and 6 μm Tetrapod PSF, respectively. Notably, the improvement is more significant for larger PSFs. When using a 3-frames temporal context, the $RMSE_{3D}$ improvement of LUNAR over FD-DeepLoc for the 3 μm Tetrapod PSF dataset increases to 24%, and further rises to 34% with a 7-frames temporal context, highlighting LUNAR's exceptional ability to leverage temporal information. Additionally, LUNAR exhibits more stable performance for large-DOF PSFs in repetitive tests (**Supplementary Figs. 8**), with about 10 times smaller 3D efficiency variance compared to that of FD-DeepLoc. These results fully demonstrate the outstanding performance and robustness of the LUNAR network.

Finally, we evaluated LUNAR on these datasets in a more realistic scenario: training without knowing the ground truth PSF. We employed the self-supervised synchronized learning strategy to train LUNAR network, referred to as LUNAR SL. For comparison, supervised localization learning approaches using a predefined PSF (i.e. LUNAR LL and FD-DeepLoc) were also evaluated. For each dataset, we trained LUNAR SL five times with randomly initialized PSFs. Amplitude of each aberration was initialized by adding zero mean Gaussian noise with a standard deviation of $\lambda/50$ to the ground truth. Without knowing the ground truth PSF, LUNAR SL achieved performance comparable to that of LUNAR LL, where the PSF models of training data and test data are the same (**Supplementary Fig. 10** and **Supplementary Table 3**). We then introduced additional aberrations to the test data to mimic the unknown aberrations arisen from the sample. LUNAR SL, trained by the test data without knowing the

ground truth PSF, was able to learn PSFs from the data itself and successfully corrected the unknown aberrations. The reconstructed image by LUNAR SL closely resembles the ground truth, as evidenced by clearly resolving the hollow structure of the microtubules (**Fig. 2h-m**) and a sixfold smaller localization error. In contrast, PSF-supervised algorithms like FD-DeepLoc tend to be biased with incorrect PSF models, resulting in dispersed localizations and grid artifacts due to high uncertainty during molecule localization[11].

Synchronized learning allows precise localization with poor PSF calibrations

In typical SMLM laboratories, preparing beads for PSF calibration is a routine task. The frequency of PSF calibration depends on the stability of the optical system. Without regular calibration, it becomes difficult to accurately account for changes in the system, leading to localization errors. Therefore, regular PSF calibration is necessary to ensure optimal results, although this process can be time-consuming and costly.

To illustrate how LUNAR tackles this problem on real biological samples, we used nuclear pore complexes (NPCs) labeled with Nup96 as a 3D reference standard[28]. Nup96 in NPCs form two rings with a diameter of ~107 nm and an axial separation of ~50 nm. We began with a well-calibrated data where supervised learning methods FD-DeepLoc and LUNAR LL could resolve the lateral ring profiles and the axial two-layer profiles of NPCs. The Fourier ring correlation (FRC) resolution of images reconstructed by LUNAR LL (15.4 nm) is better than that of FD-DeepLoc (17.6 nm, **Supplementary Fig. 11**), indicating superior performance of LUNAR LL due to the more accurate and robust molecule localization powered by the advanced network design. However, the upper focal plane region, where is further away from the coverslip, showed blurred NPC images reconstructed by both supervised learning methods (**Supplementary Fig. 11e**). This is probably due to the model mismatch introduced by more sample-induced aberrations with the increase of imaging depth. We then applied LUNAR SL to learn the in situ PSF model from the raw data and obtained clear two-ring structures in the upper focal plane region (**Supplementary Fig. 11f**).

Next, we investigated the performance of the DL-based methods when the calibrated PSF deviated from the single-molecule data acquired. The PSF calibration

was performed several weeks before the imaging experiments, potentially leading to PSF model mismatch between calibration and SMLM data. We first employed supervised learning using the calibrated PSF model with a robust training strategy, which introduced a small perturbation to the PSF model during the training process to compensate for calibration uncertainties[6]. In the top view, we found that more molecules were detected by LUNAR than FD-DeepLoc (**Fig. 3d**). In the side view of the plane near the equator, both LUNAR LL and FD-DeepLoc reconstructed distorted ring structure due to the relatively large model discrepancy between the calibrated PSFs and SMLM data (**Fig. 3f**). We then employed the synchronized learning method, LUNRA SL, to learn the neural-physics model from the single-molecule data itself. During training, LUNAR SL iteratively improved the PSF model constrained by the physics decoder. The reconstructed images by LUNAR SL along the axial direction in the equator of cell nucleus could resolve more ring like NPC structures (**Fig. 3f**). As shown in **Fig. 3g-i,** LUNAR SL successfully restored the two-ring structures of NPCs with higher FRC resolutions compared to that of supervised learning-based networks (i.e., FD-DeepLoc and LUNAR LL).

Whole-cell longitudinal SR imaging with high fidelity

Large-DOF SMLM for whole-cell imaging poses significant challenges for localization analysis software as the data often comprises overlapped single-molecule pattern with complex shapes. Although DL methods have shown great capability for analyzing overlapped single-molecule data, they all relied on the supervised learning with deterministic PSF model. A major obstacle to estimate the precise PSF model from large-DOF SMLM imaging experiments is that it is difficult to extract isolated individual PSF patterns from the overlapped single-molecule data. Furthermore, PSFs with large DOF often extend in a large area. It is still quite challenging to accurately retrieve molecules' 3D positions from the extended 2D patterns using conventional CNNs. To demonstrate LUNAR SL in these challenging scenarios, we imaged mitochondria labeled with TOM20 using a 6 μm Tetrapod PSF. As shown in **Supplementary Fig. 12**, LUNAR SL could reduce the residual error between the raw single-molecule data and the learned model, indicating the superior model learning

capability of LUNAR over FD-DeepLoc. With the help of accurately inferred PSF model and precise large-DOF single-molecule localization by LUNAR's network, we were able to obtain high-quality SR image of mitochondria across the whole cell, as indicated by the protein aggregates on the mitochondria membrane (**Fig. 4a-d**). Punctate protein aggregates on the outer membrane could be clearly visualized in side-view cross-sections of the images reconstructed by LUNAR, whereas FD-DeepLoc, trained with bead-calibrated PSF, produced more blurred side-view reconstructions where protein aggregates were not easily distinguishable (**Fig. 4d**).

We then imaged whole-cell NPCs using the 6 μm Tetrapod PSF due to their well-defined geometry (**Fig. 4e-l** and **Supplementary Video 1**). Compared to FD-DeepLoc, LUNAR SL reconstructed sharper images with more punctate localizations, as indicated by the white arrows and line profiles (**Fig. 4g-i**). The difference is more profound in the edge areas of the cell nucleus where the ring structure of NPC can be only observed in the side-view images. As show in **Fig. 4j**, LUNAR SL provided superior axial reconstructions with clear ring structure on the side view of nuclear envelope over a large axial range, whereas FD-DeepLoc could only resolve disk-like structures. With improved image quality, we were able to extract more NPCs for quantitative analysis (**Fig. 4k, l**). Under the same rejection criteria of the automatic NPC geometry analysis pipeline in SMAP[29], LUNAR SL retained 490 valid NPCs after the clean-up process for ring radius analysis, compared to only 191 valid NPCs retained for FD-DeepLoc (**Supplementary Note 7**). The radii measured by LUNAR SL aligned closely with previously reported values[30], whereas dispersed localizations in FD-DeepLoc resulted in slightly larger radius. These results underscore the high fidelity of the images reconstructed by LUNAR for large-DOF single-molecule localization, enabling statistical analysis of super-resolved images with more reliable data.

LUNAR enables high-accuracy SR imaging of neuronal cytoskeleton in thick samples

Sample induced aberrations lead to information loss with reduced SNR, which is normally irreversible only by image post-processing[8]. An important method to recover this lost information is to employ adaptive optics (AO) such as a deformable mirror, to

compensate for optical aberrations[31] during data acquisition. AO has also been utilized to engineer the PSF to increase information content of single molecules[32,33], thereby improving the theoretically achievable localization precision (CRLB). As a computational algorithm designed to exploit information from the raw data, LUNAR can perfectly complement with the AO hardware, enhancing the final reconstruction from both data acquisition and post-processing perspectives.

Here, we applied the deformable mirror based optimal PSF engineering (DMO PSF)[32], optimizing the 3D CRLB within a 4 μm axial range, to image the cultured rat neurons labeled with βII-spectrin. We then employed LUNAR SL to simultaneously estimate the in situ PSF model within the sample and localize single molecules. Different from previous observations using astigmatic PSF which is limited to a DOF of a few hundred nanometers[34], the combination of LUNAR SL and DMO PSF allowed us to visualize the full 3D distribution of βII-spectrin across a large axial range without scanning (**Fig. 5a-e**). Through axially separating the reconstruction of densely packed neurites, we were able to clearly resolve the membrane-associated periodic skeletons (MPS) along the whole axon at different depths.

Next, we analyzed a published imaging dataset which was performed on rat brain slices with the help of AO to improve the SNR of the raw data[35] (**Fig. 5f-m** and **Supplementary Video 2**). Neurons were labeled with βIV-spectrin which normally locates at the axon initial segment (AIS). Notably, the imaging depth was about 50 μm inside the tissue slice where the elliptical shape of astigmatic PSF was lost due to large spherical aberrations[36]. The AO method REALM successfully compensated for these aberrations, allowing for proper 3D AIS reconstruction using a classical Gaussian fit. However, Gaussian PSF model cannot fully describe the in situ PSF inside the complex imaging environments, potentially limiting REALM's performance. Without knowing the calibrated PSF model, LUNAR SL successfully learned the precise PSF model from the raw single-molecule data. Leveraging the accurate PSF model inferred and advanced localization network, LUNAR SL reconstructs clearer profiles with fewer distortions, particularly in side-view cross-sections (**Fig. 5 h, i**). The MPS could be observed on both the top and bottom of the axon. From the side-view, we could directly

observe the ring-like structure wrapped around the circumference and evenly spaced along the axon (**Fig. 5j, k**). The line profiles and their autocorrelation analysis indicate that LUNAR SL produces sharper and more informative reconstructions (**Fig. 5l, m**). These results demonstrate that the combination of AO and LUNAR could enable high-quality SR imaging of neuron cytoskeletons inside the tissues, offering new insights into the complex structures of the brain.

Discussion

In this work, we proposed LUNAR, a powerful blind localization method for SMLM. Unlike conventional methods need PSF modeling based on isolated beads data or single-molecule events, LUNAR integrates with a neural-physics model that could learn the precise in situ PSF model directly from the overlapped single-molecule data. Furthermore, LUNAR introduces a hybrid CNN-Transformer localization network that can effectively utilize the spatiotemporal information of 2D single-molecule patterns to retrieve their 3D positions, which is especially important for large-DOF imaging at great imaging depths. These advancements enable LUNAR to overcome significant challenges in whole-cell longitudinal and high-density single-molecule imaging scenarios.

Our results demonstrate that LUNAR not only achieves superior performance in localization accuracy compared to state-of-the-art methods but also adapts seamlessly to SMLM data with uncalibrated or distorted PSFs, enabling SR imaging at greater depth. By leveraging synchronized learning strategies, LUNAR bridges the gap between theoretical models and experimental data, making it highly adaptable to complex and dynamic imaging environments. Synchronized learning strategy is derived purely from probabilistic modeling without requiring structural priors about the sample. This avoids subjective biases and ensures the convergence of both the physical model and the corresponding network in predicting the inverse physical process. These capabilities were validated through extensive simulations and experimental datasets, including whole-cell imaging and neuronal cytoskeleton reconstruction, where LUNAR achieved exceptional levels of resolution and robustness.

With these innovations, LUNAR represents a robust DL localization tool capable

of adaptively correcting mismatched PSF models during training. This capability frees researchers from the need for frequent PSF calibration to maintain unbiased SR imaging. Moreover, as a computational framework, LUNAR could complement with hardware-based AO, guiding AO devices to correct for the aberrations and encode position sensitive aberrations into the data. This synergy effectively enhances the quality and depth of SR imaging, offering promising in situ investigation of cellular structures and functions inside thick tissues.

Despite its advancements, LUNAR's reliance on physical model flexibility highlights opportunities for further development. Future work could explore extending neural-physics frameworks to account for imaging conditions that deviate from traditional vectorial PSF models, paving the way for broader applications in SR microscopy. Additionally, LUNAR is computationally intensive compared to traditional calibration and fitting methods, posing challenges for real-time applications. Looking forward, we envision developing an integrated platform that features user-friendly graphical user interfaces (GUIs) with minimal parameter settings, as well as integrated downstream analysis pipelines, thereby making advanced SMLM analysis more accessible to a broader range of researchers.

References

1. Schermelleh, L. *et al.* Super-resolution microscopy demystified. *Nat Cell Biol* **21**, 72–84 (2019).
2. Sigal, Y. M., Zhou, R. & Zhuang, X. Visualizing and discovering cellular structures with super-resolution microscopy. *Science* **361**, 880–887 (2018).
3. Balzarotti, F. *et al.* Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355**, 606–612 (2017).
4. Bates, M. *et al.* Optimal precision and accuracy in 4Pi-STORM using dynamic spline PSF models. *Nat Methods* **19**, 603–612 (2022).
5. Reinhardt, S. C. M. *et al.* Ångström-resolution fluorescence microscopy. *Nature* **617**, 711–716 (2023).
6. Fu, S. *et al.* Field-dependent deep learning enables high-throughput whole-cell 3D super-resolution imaging. *Nat Methods* **20**, 459–468 (2023).
7. Barentine, A. E. S. *et al.* An integrated platform for high-throughput nanoscopy. *Nat Biotechnol* **41**, 1549–1556 (2023).
8. Liu, S., Huh, H., Lee, S.-H. & Huang, F. Three-Dimensional Single-Molecule Localization Microscopy in Whole-Cell and Tissue Specimens. *Annu. Rev. Biomed.*

Eng. **22**, 155–184 (2020).
9. Li, Y. *et al.* Real-time 3D single-molecule localization using experimental point spread functions. *Nat Methods* **15**, 367–369 (2018).
10. Nehme, E. *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nat Methods* **17**, 734–740 (2020).
11. Speiser, A. *et al.* Deep learning enables fast and dense single-molecule localization with high accuracy. *Nat Methods* **18**, 1082–1090 (2021).
12. Saguy, A. *et al.* DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning. *Nat Methods* **20**, 1939–1948 (2023).
13. Zhang, P. *et al.* Deep learning-driven adaptive optics for single-molecule localization microscopy. *Nat Methods* **20**, 1748–1758 (2023).
14. Möckl, L., Roy, A. R. & Moerner, W. E. Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments [Invited]. *Biomed. Opt. Express* **11**, 1633 (2020).
15. Li, X., Zhang, Y., Wu, J. & Dai, Q. Challenges and opportunities in bioimage analysis. *Nat Methods* **20**, 958–961 (2023).
16. Li, X. *et al.* Spatial redundancy transformer for self-supervised fluorescence image denoising. *Nat Comput Sci* **3**, 1067–1080 (2023).
17. Qu, L. *et al.* Self-inspired learning for denoising live-cell super-resolution microscopy. *Nat Methods* **21**, 1895–1908 (2024).
18. Li, X. *et al.* Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nat Biotechnol* **41**, 282–292 (2023).
19. Eom, M. *et al.* Statistically unbiased prediction enables accurate denoising of voltage imaging data. *Nat Methods* **20**, 1581–1592 (2023).
20. Qiao, C. *et al.* Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy. *Nat Commun* **15**, 4180 (2024).
21. Speiser, A. *et al.* Teaching deep neural networks to localize single molecules for super-resolution microscopy. Preprint at http://arxiv.org/abs/1907.00770 (2020).
22. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. in *Proc. International Conference on Learning Representations (ICLR)* (2014).
23. Bornschein, J. & Bengio, Y. Reweighted Wake-Sleep. in *Proc. International Conference on Learning Representations (ICLR)* (2015).
24. Dempster, A. P., Laird, N. M. & work(s):, D. B. R. R. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1–38 (1977).
25. Liu, Z. *et al.* A ConvNet for the 2020s. in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)* (2022).
26. Vaswani, A. *et al.* Attention Is All You Need. in *Proc. International Conference on Neural Information Processing Systems (NeurIPS)* (2017).
27. Sage, D. *et al.* Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nat Methods* **16**, 387–395 (2019).
28. Thevathasan, J. V. *et al.* Nuclear pores as versatile reference standards for quantitative superresolution microscopy. *Nat Methods* **16**, 1045–1053 (2019).
29. Ries, J. SMAP: a modular super-resolution microscopy analysis platform for

SMLM data. *Nat Methods* **17**, 870–872 (2020).
30. Wu, Y.-L. *et al.* Maximum-likelihood model fitting for quantitative analysis of SMLM data. *Nat Methods* **20**, 139–148 (2023).
31. Booth, M., Andrade, D., Burke, D., Patton, B. & Zurauskas, M. Aberrations and adaptive optics in super-resolution microscopy. *Microscopy* **64**, 251–261 (2015).
32. Fu, S. *et al.* Deformable mirror based optimal PSF engineering for 3D super-resolution imaging. *Opt. Lett.* **47**, 3031–3034 (2022).
33. Aristov, A., Lelandais, B., Rensen, E. & Zimmer, C. ZOLA-3D allows flexible 3D localization microscopy over an adjustable axial range. *Nat Commun* **9**, 2409 (2018).
34. Xu, K., Zhong, G. & Zhuang, X. Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal Structure in Axons. *Science* **339**, 452–456 (2013).
35. Siemons, M. E., Hanemaaijer, N. A. K., Kole, M. H. P. & Kapitein, L. C. Robust adaptive optics for localization microscopy deep in complex tissue. *Nat Commun* **12**, 3407 (2021).
36. Mlodzianoski, M. J. *et al.* Active PSF shaping and adaptive optics enable volumetric localization microscopy through brain sections. *Nat Methods* **15**, 583–586 (2018).
37. Siemons, M. E., Kapitein, L. C. & Stallinga, S. Axial accuracy in localization microscopy with 3D point spread function engineering. *Opt. Express* **30**, 28290–28300 (2022).
38. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015).
39. Shi, W. *et al.* Aberration correction for deformable-mirror-based remote focusing enables high-accuracy whole-cell super-resolution imaging. *Photon. Res.* **12**, 821–832 (2024).
40. Zhang, Y. *et al.* Nanoscale subcellular architecture revealed by multicolor three-dimensional salvaged fluorescence imaging. *Nat Methods* **17**, 225–231 (2020).
41. Zhou, R. *et al.* Proteomic and functional analyses of the periodic membrane skeleton in neurons. *Nat Commun* **13**, 3196 (2022).

Methods

LUNAR architecture

LUNAR employs an autoencoder architecture inspired by the VAE[22], leveraging a probabilistic framework to interpret experimental data while accounting for inherent uncertainties. Different from traditional VAEs, which use abstract latent variables and a network-based decoder to generate data, LUNAR incorporates a learnable vectorial PSF model[37] as its decoder. This has two benefits: (1) The data generated strictly obey the light diffraction theory. (2) The latent variables are physically interpretable,

representing emitters with different positions and photons. After proper training, the network encoder can be used separately to make inferences on SMLM data, localizing molecules with uncertainty estimation.

Network encoder. The LUNAR network is comprised of three modules: the feature extraction module (FEM), the temporal attention module (TAM) and the output module (**Supplementary Fig. 2**). The FEM employs a U-NeXt backbone, integrating the design principles of ConvNeXt[25] into the U-Net[38] architecture. We implement several ConvNeXt strategies to enhance the network performance on single-molecule localization: 1) Adjusting the stage compute ratio to 1:1:3:1 (same as Swin Transformer) in the FEM; 2) Replacing the 3×3 convolution with a 5×5 depth wise convolution combined with an inverted bottleneck; 3) Replacing ReLU with GELU as the activation function and using fewer activations.

The FEM outputs spatial features for each input frame in a batch, where each pixel in the spatial dimension corresponds to a batch-length time sequence. These sequences are then fed to the TAM, which is a Transformer block containing a masked self-attention layer and a feed-forward layer[26]. The temporal mask in the self-attention layer enables flexible configuration of the temporal context length that the network can utilize for localization, with a default setting of 7. At the end of the TAM, a temporal dropout layer is applied with a probability of 0.5 to randomly drop this temporal information during training, enhancing the network's generalization ability. A ConvNeXt block follows to aggregate the temporal information with the initial TAM inputs. Notably, if temporal information is not utilized (i.e., using single frame for inference), the TAM will be skipped.

The output module applies several convolution-based prediction heads to output multi-channel maps of equal size to the input images. For each pixel k, it outputs: detection probability $\hat{p}_k$ indicating whether an emitter exists; lateral sub-pixel offsets $\widehat{\Delta x}_k, \widehat{\Delta y}_k$ relative to the pixel center; axial distance $\hat{z}_k$ relative to the focal plane; brightness $\widehat{phot}_k$; associated uncertainties $\sigma^2_{x,k}, \sigma^2_{y,k}, \sigma^2_{z,k}, \sigma^2_{phot,k}$; background $\widehat{bg}_k$. Note that the sub-pixel offsets $\widehat{\Delta x}_k, \widehat{\Delta y}_k$, plus the position of pixel k, form the final

lateral position predictions $\hat{x}_k, \hat{y}_k$.

Physics decoder. To reconstruct the raw SMLM data, we applied a learnable vectorial PSF model as the physics decoder for LUNAR (**Supplementary Fig. 1**). The model is generally formulated as:

$$\mathrm{PSF} \propto \sum_{p=x,y} \sum_{d=x,y,z} \left|E_{p,d}^{\mathrm{img}}\right|^2, E_{p,d}^{\mathrm{img}} = \mathcal{F}_{2D}\left\{A(\rho,\varphi) E_{p,d}^{\mathrm{pupil}} e^{i\psi_{\mathrm{pos}}} e^{i\psi_{\mathrm{aber}}}\right\} \tag{2}$$

where $\mathcal{F}_{2D}$ denotes the 2D Fourier transform. $A(\rho,\varphi)$ is the amplitude function with the apodization factor. $E_{p,d}^{\mathrm{pupil}}$ represents the contribution of the fluorescence dipole components $d = x, y, z$ at the sample plane to the light filed components $p = x, y$ at the pupil plane. ψ_{pos} is the phase shift dependent on the molecule position. ψ_{aber} is the phase term that accounts for both PSF engineering and unknown sample-induced aberrations, expressed as a linear sum of multiple Zernike polynomials:

$$\psi_{\mathrm{aber}} = \sum_{n+|m|\leq 8} C_n^m Z_n^m \tag{3}$$

where C_n^m is the Zernike coefficient corresponding to the Zernike polynomial Z_n^m, n is the radial order, and m is the angular frequency. In this work 21 Zernike coefficients C_n^m that satisfy $2 < n + |m| \leq 8$ are treated as the learnable parameters θ during the physics learning phase. More details are provided in **Supplementary Note 2**.

Synchronized learning strategy

To simultaneously optimize the network encoder and physics decoder, we developed a synchronized learning strategy that updates parameters by alternate iterations between localization learning and physics learning. The objective functions and gradients for these two steps are described below. Implementation details about the LUNAR learning and inference processes can be found in **Supplementary Note 6**.

Physics learning and parameters update. In the physics learning phase, the network encoder $q_\phi(h|d)$, which predicts the posterior distribution of emitters' parameters $h = [x, y, z, photon]$ given data d, with parameters ϕ, is fixed. The physics decoder $p_\theta(d|h)$ accounts for the stochastic image formation process of fluorescence molecules, incorporating the vectorial PSF model with trainable aberration parameters

θ and shot noise. The objective of physics learning is given by:

$$\mathrm{ELBO} = \mathrm{E}_{q_\phi(h|d)}\left[\log\frac{p_\theta(d,h)}{q_\phi(h|d)}\right] \tag{4}$$

where ELBO is the evidence lower bound of the data log likelihood $\log p_\theta(d)$. $\mathrm{E}_q[\cdot]$ represents expectation computation under the distribution q. The joint distribution is defined as $p_\theta(d,h) = p_\theta(d|h)p_\theta(h)$, where $p_\theta(h)$ is the prior distribution of fluorescence molecules. In this work, we make no assumptions about the molecules to be localized and set $p_\theta(h)$ as a uniform distribution.

Since q_ϕ in our framework is represented by a neural network and is not integrable, we apply the importance sampling (IS) estimator[23] to get an optimizable objective (reconstruction loss):

$$\mathrm{ELBO}_{IS}^{K} = \mathrm{E}_{h^{(1)},\ldots,h^{(K)}\sim q_\phi(h|d)}\left[\log\left(\frac{1}{K}\sum_{k=1}^{K} w_k\right)\right], w_k = \frac{p_\theta\left(d,h^{(k)}\right)}{q_\phi\left(h^{(k)}|d\right)} \tag{5}$$

Here, $h^{(1)},\ldots,h^{(K)}$ are sampled independently from the posterior $q_\phi(h|d)$, and we typically set $K = 100$, which works well for all datasets in this work. Finally, maximizing ELBO_{IS}^{K} with respect to the aberration parameters θ yields the estimated gradients:

$$\nabla_\theta \mathrm{ELBO}_{IS}^{K} = \sum_{k=1}^{K} \widetilde{w}_k \nabla_\theta \log p_\theta\left(d,h^{(k)}\right), \widetilde{w}_k = \frac{w_k}{\sum_{k=1}^{K} w_k} \tag{6}$$

where $\widetilde{w}_k$ is the normalized importance weight. These gradients are used to update the aberration parameters θ in the physics decoder, ensuring that the generated synthetic data remains consistent with the physical laws governing the SMLM process.

Localization learning and parameters update. In the localization learning phase, the physics decoder $p_\theta(d|h)$ is fixed, while the network encoder $q_\phi(h|d)$ learns to approximate the posterior $p_\theta(h|d)$. This is achieved by minimizing the localization loss, which computes the Kullback-Leibler (KL) divergence between the posteriors under the physical model and the network, averaged over the training data generated by the current physical model:

$$\mathrm{E}_{p_\theta(d)}\left[D_{\mathrm{KL}}\left(p_\theta(h|d), q_\phi(h|d)\right)\right] = \mathrm{E}_{p_\theta(d,h)}\left[\log p_\theta(h|d) - \log q_\phi(h|d))\right] \quad (7)$$

Since θ is fixed, the gradient of the above objective is estimated by evaluating $\nabla_\phi \log q_\phi(h|d)$, where $h, d \sim p_\theta(h, d)$. This is equivalent to maximizing the log likelihood $\log q_\phi(h|d)$. In this work, the distribution represented by $q_\phi(h|d)$ is formulated as the joint distribution of three distributions: a Bernoulli distribution for detecting molecule in the pixels, a Gaussian distribution for counting the molecules, and a Gaussian-mixture model (GMM) for molecule parameters regression (**Supplementary Note 1**). This formulation leads to a loss function similar to that of FD-DeepLoc[6]:

$$-\log q_\phi(h|d) = L_{\mathrm{ce}} + L_{\mathrm{count}} + L_{\mathrm{loc}} \quad (8)$$

where L_{ce} computes the cross-entropy between the predicted probability map and the ground truth. L_{count} computes the negative log likelihood of the ground truth molecule number under the predicted Gaussian distribution. L_{loc} computes the negative log likelihood of the ground truth molecule parameters under the predicted GMM. Additionally, a background term is introduced to form the final loss for localization learning:

$$L = L_{\mathrm{ce}} + L_{\mathrm{count}} + L_{\mathrm{loc}} + L_{\mathrm{bg}} \quad (9)$$

where L_{bg} is computed using the simple ℓ_2 norm:

$$L_{\mathrm{bg}} = \left\|\widehat{\mathrm{bg}} - \mathrm{bg}\right\|_2^2 \quad (10)$$

Here, $\widehat{\mathrm{bg}}$ and bg represent the predicted background map and the ground truth background, respectively. This comprehensive loss function ensures that the model not only accurately localizes the molecules but also effectively models the background noise, which is crucial for the physics learning phase.

Training data generation

For physics learning, experimental single-molecule data is used as both training data and targets. For each sampled batch (16 frames in this work), several extra frames are padded at both ends to provide the necessary temporal context for each frame to be reconstructed.

For localization learning, training data and targets are generated using the physics decoder. Each unit in the batch (2 used in this work) consists of a series of frames (8 used in this work) with molecules activated using a simplified photophysical model. Specifically, we use a three-state model with off, on and dark states. Transitions between these states follow simple Bernoulli distributions. This is because only a few consecutive images (typically fewer than 16) are simulated in each unit. The molecules with a predefined density in the off state are first randomly activated. The probabilities for on-state molecules surviving in the next frame or dark-state molecules transitioning back to the on-state are determined by:

$$P_{\text{surv}} = 1 - \left(1 - e^{-1/T_{\text{on}}}\right) \tag{11}$$

$$P_{\text{back}} = 1 - \left(1 - e^{-\frac{1}{T_{\text{dark}}}}\right), \tag{12}$$

where T_{on} and T_{dark} correspond to the average durations of the on and dark states in the photophysical model of the SMLM challenge[27]. In this work, we set $T_{\text{on}} = 1$ frame and $T_{\text{dark}} = 4$ frames to account for the multiple decay paths in the more sophisticated challenge model.

For each molecule in the training data, the position and brightness are drawn from uniform distributions within predefined ranges. These ranges are estimated from the raw data or determined based on experience. Although incorporating structure prior about the sample could enhance performance, we decided not to use it to avoid any potential bias. To account for the spatially variant background in real biological specimens, we employ both a constant background and Perlin noise, which are randomly applied to each unit in the training batch. The background range is estimated from experimental data or manually defined.

For camera noise simulation, both EMCCD and sCMOS noise models can be used (**Supplementary Note 5**), which are the same as in our previous work[6]. For the EMCCD camera, the noise model includes shot noise, electron multiplication noise, and camera readout noise. For the sCMOS camera, only shot noise and readout noise are considered, as there is no electron multiplication process. Notably, the synchronized

learning currently works only for noise models dominated by Poisson shot noise. If the noise involves more complicated distributions, the loss function for the physics learning phase should be rederived to accommodate these additional noise components.

PSF engineering

A deformable mirror (DM140A-35-P01, Boston Micromachines) was used to modulate the PSF for large-DOF experimental imaging. Based on our previous work[32], we optimized the actuator voltages of the deformable mirror to construct a pupil phase that minimizes the average 3D CRLB within a predefined axial range. For simulation tests, we modulated the PSF pupil using Zernike polynomials. The Zernike coefficient for the astigmatic PSF was $C_2^2 = 70$ nm. The optimized Zernike coefficients for the 3 μm Tetrapod PSF were $C_2^{-2} = 120$ nm, $C_4^{-2} = -130$ nm and $C_6^{-2} = -6$ nm. For the 6 μm Tetrapod PSF, the coefficients were $C_2^{-2} = 230$ nm, $C_4^{-2} = -240$ nm and $C_6^{-2} = -17$ nm. All simulations were conducted at a wavelength of 670 nm.

Microscope

In this study, we utilized a custom-built microscope, as described in our previous work[39]. Excitation light from two lasers (iBEAM-SMART-405-S, 150 mW, and iBEAM-SMART-640-S-HP, 200 mW, TOPTICA Photonics) was coupled into a single-mode fiber (P3-405BPM-FC-2, Thorlabs) via a fiber coupler (PAF2-A4A, Thorlabs). By adjusting the translation stage at the fiber output, we can change the illumination angle. The output beam was passed through a laser clean-up filter (ZET405/488/561/640xv2, Chroma) for spectral purification. Beam collimation and reshaping were achieved using a pair of lenses with focal lengths of 75 mm and 400 mm, combined with a slit (SP60, Owis) for beam shaping. The collimated beam was then reflected by a main dichroic mirror (ZT405/488/561/640rpcxt-UF2, Chroma) and directed into the objective for sample illumination. Fluorescence emitted from the sample was collected by a high numerical aperture (NA) objective (NA 1.35, UPLSAPO 100 XS or NA 1.5, UPLSAPO 100XOHR, Olympus) and imaged through a tube lens (TTL-180-A, Thorlabs). To confine the imaging field, a slit (SP40, Owis) was placed immediately after the tube lens. Fluorescence signals were further filtered using two bandpass filters (NF03-405/488/561/635E25 and FF01-676/37-25, Semrock) mounted on a filter wheel,

effectively discriminating the emitted fluorescence from the excitation light. A 4f system was integrated into the imaging path using lenses with focal lengths of 125 mm and 75 mm. A deformable mirror (DM140A-35-P01, Boston Micromachines) positioned at the Fourier plane was employed for PSF engineering and aberration correction. To visualize the back focal plane of the objective, a lens with a focal length of 40 mm was placed before the camera. Image acquisition was carried out using an sCMOS camera (Dhyana 400BSI V3, Tucsen or ORCA-Flash4.0 V3, Hamamatsu), which provided a pixel size of 108 nm in the sample plane. Additionally, a closed-loop focus-lock control system was incorporated into the optical setup to ensure stable focus. This system utilized a 785 nm laser (iBEAM-SMART-785-S, 125 mW, TOPTICA Photonics) that was reflected off the coverslip through a dichroic mirror (FF750-SDi02, Semrock) and detected by a quadrant photodiode (SD197-23-21-041, Advanced Photonix Inc). Feedback from the photodiode was used to control a z-piezo stage (P-726.1CD, Physik Instrumente), maintaining precise focus during imaging.

Sample preparation

Cell Culture and Induced Neuronal Cells Generation. U2OS cells (Nup96-SNAP no. 300444, Cell Line Services) were grown in DMEM (catalog no. 10569, Gibco) containing 10% (v/v) fetal bovine serum (FBS; catalog no. 10099-141C, Gibco), 100 U/ml penicillin and 100 μg/ml streptomycin (PS; catalog no. 15140-122, Gibco) and 1× MEM NEAA (catalog no. 11140-050, Gibco). COS-7 cells (catalog no. 100040, BNCC) were grown in DMEM containing FBS and PS. Cells were cultured in a humidified atmosphere with 5% CO2 at 37 °C and passaged every two or three days. Before cell plating, high-precision 25-mm-round glass coverslips (catalog no. CG15XH, Thorlabs) were cleaned by sequentially sonicating in 1 M potassium hydroxide (KOH), Milli-Q water and ethanol, finally irradiated under ultraviolet light for 30 min. For SR imaging, U2OS and COS-7 cells were cultured on the clean coverslips for 2 days with a confluency of ~80%.

For the generation of induced neuronal cells, mouse embryonic stem cells (mESCs) were transfected with plasmid containing Tet-on promoter driving expression of rtTA, hygromycin resistance, and Ngn2-P2A-puro. Transfected cells were selected with 200

μg/ml hygromycin (Sigma, V900372) for 3-5 days. On day 0, cells were digested and plated (1×10^5 cells/well in 6-well plates) on poly-D-lysine/laminin-coated coverslips in serum/LIF medium. After 12 hours, the medium was replaced with N2B27, a 1:1 mixture of DMEM/F12 (Gibco, 11320033) and Neurobasal (Gibco, 21103049), supplemented with 1% N2 (Gibco, 17502048), 2% B27 (Gibco, 17504044), 1×sodium pyruvate, 1×penicillin-streptomycin, 1×NEAA, 1×GlutaMAX, and 1 mM 2-mercaptoethanol, containing 5 μM retinoic acid (Sigma, R2625) and 2 mg/ml doxycycline (TargetMol, T1687). On day 1, puromycin selection (1 mg/ml) was started for 24 hours. On day 6, Ara-C (2 mM, TargetMol, T1272) was added to inhibit neural stem cell proliferation. From day 2 onward, 50% of the medium was exchanged every 2 days, and induced neuronal cells were assayed on day 21 in most experiments.

SNAP-tag labeling of Nup96. To label Nup96, U2OS-Nup96-SNAP cells were prepared as previously reported[28]. Cells were prefixed in 2.4% (w/v) paraformaldehyde (PFA) for 30s, permeabilized with 0.4% (v/v) Triton X-100 for 3 min, and subsequently fixed in 2.4% PFA for 30 min. All fixation and permeabilization steps were conducted using preheated buffers maintained at 37 °C. Following fixation, cells were quenched in 0.1 M NH_4Cl for 5 min and washed twice with PBS. To minimize nonspecific binding, cells were blocked for 30 min using Image-iT FX Signal Enhancer (catalog no. I36933, Invitrogen). For SNAP-tag labeling, cells were incubated with a dye solution containing 1 μM SNAP-tag ligand BG-AF647 (catalog no. S9136S, New England Biolabs), 1 mM DTT (catalog no. 1111GR005, BioFroxx), and 0.5% (w/v) bovine serum albumin (BSA) in PBS for 2 h. After labeling, cells were washed 3 times with PBS for 5 min each to remove excess dyes. Finally, cells were post-fixed in 4% PFA for 10 min, washed 3 times with PBS for 3 min each, and stored at 4 °C until imaging.

Mitochondria labeling. Mitochondrial samples were prepared as previously described[40]. COS-7 cells were fixed in 4% PFA prepared in PBS and preheated to 37 °C for 12 min. The fixed cells were permeabilized for 3 min in permeabilization buffer containing 0.3% CA-630 (catalog no. I8896, Sigma), 0.05% Triton X-100, 0.1% BSA, and 1× PBS. Following permeabilization, the cells were quenched in 0.1 M NH_4Cl for 5 min. The cells were then washed 3 times with PBS, each wash lasting 5 min, and

subsequently blocked in 3% BSA for 60 min to minimize nonspecific binding. For primary antibody staining, cells were incubated with rabbit anti-Tom20 (catalog no. ab78547, Abcam, 1 mg/ ml) diluted 1:1,000 in 3% BSA for 2 h. Following this step, the cells were washed 3 times with PBS for 5 min each. Secondary staining was performed using AF647-conjugated secondary antibodies (catalog no. A21245, Invitrogen, 2 mg /ml) diluted 1:2,000 in 3% BSA for 2 h. Cells were washed again 3 times with PBS, each wash lasting 5 min. Finally, the cells were post-fixed with 4% PFA for 10 min, washed 3 times with PBS for 5 min each, and stored in PBS at 4 °C.

Neuronal cells labeling. Neuronal samples were labeled following a previously described protocol[41]. Cultured mESC-derived neuronal cells were fixed between days 21 and 25 days in vitro using 4% PFA prepared in PBS and preheated to 37 °C. Fixation was performed for 30 min, followed by 3 washes with PBS. The cells were then permeabilized with 0.15% Triton X-100 in PBS for 10 min. After permeabilization, cells were washed 3 times with PBS for 5 min each and subsequently blocked in 3% BSA for 60 min to minimize nonspecific binding. For primary antibody labeling, the cells were incubated with mouse anti-β-spectrin II antibody (catalog no. 612563, BD Biosiences, 250 μg/ml) diluted 1:100 in 3% BSA. The incubation was performed overnight at 4 °C, followed by 3 washes with PBS, each lasting 5 min. Secondary antibody staining was conducted using AF647-conjugated goat anti-mouse IgG (catalog no. A21235, Invitrogen, 2 mg/ml) diluted 1:800 in 3% BSA. The cells were incubated with the secondary antibody for 2 h at room temperature and then washed 3 times with PBS for 5 min each. Finally, the cells were post-fixed with 4% PFA for 20 min, washed 3 times with PBS for 5 min each, and stored in PBS at 4°C until imaging.

Imaging Buffer. Samples were imaged in a RI-matched buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM NaCl, 10% (w/v) glucose, 0.5 mg/ml glucose oxidase (catalog no. G7141, Sigma), 40 μg/ml catalase (catalog no. C100, Sigma), 35 mM cysteamine, and 28.5% (v/v) 2,2'-thiodiethanol (catalog no. 166782, Sigma). The final RI of the imaging buffer was 1.406.

Beads preparation. Fluorescent beads with a diameter of 100 nm (custom-designed, Invitrogen) were diluted 1:400,000 in Milli-Q water and vortexed for 3-5 min to ensure

uniform dispersion. A 40 μl aliquot of 1 M $MgCl_2$ was pipetted onto the center of a 25-mm-diameter coverslip (cleaned as described in the cell culture section) and mixed with 360 μl of the diluted bead solution. The mixture was incubated at room temperature for 5 min, followed by 3 washes with Milli-Q water to remove unbound beads. The coverslip was then stored in Milli-Q water at 4 °C in the dark until use.

Data availability

The data supporting the findings of this study are publicly available on Zenodo (https://doi.org/10.5281/zenodo.14709467).

Code availability

Source code of LUNAR is provided in Supplementary Software and updated versions are available at: https://github.com/Li-Lab-SUSTech/AI4Loc.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2024YFF0726003 to Y. L.), National Natural Science Foundation of China (62375116 to Y. L., 623B2044 to S. F.), Shenzhen Medical Research Fund (B2302038 to Y. L.), Key Technology Research and Development Program of Shandong Province (2021CXGC010212 to Y. L.), Shenzhen Science and Technology Innovation Program (JCYJ20220818100416036 and KQTD20200820113012029 to Y. L.), Basic and Applied Basic Research Fund of Guangdong Province (2024A1515011565 to Y. L.), SUSTech Presidential Postdoctoral Fellowship (S. F.), Postdoctoral Fellowship Program of CPSF (GZC20240651 to W. S.), Guangdong Provincial Key Laboratory of Advanced Biomaterials (2022B1212010003), and a startup grant from Southern University of Science and Technology.

Author contributions

S.F. and Y.L. conceived the project and designed the experiments. Y.L. supervised the entire project. S.F. developed the methods and wrote the software. W.S. and Y.F. tested the software. W.S. and S.F. performed the imaging experiments and analyzed the data. S.F. and Y. L. wrote the manuscript with the input from all other authors.

Competing interests

The authors declare no competing interests.

Figures

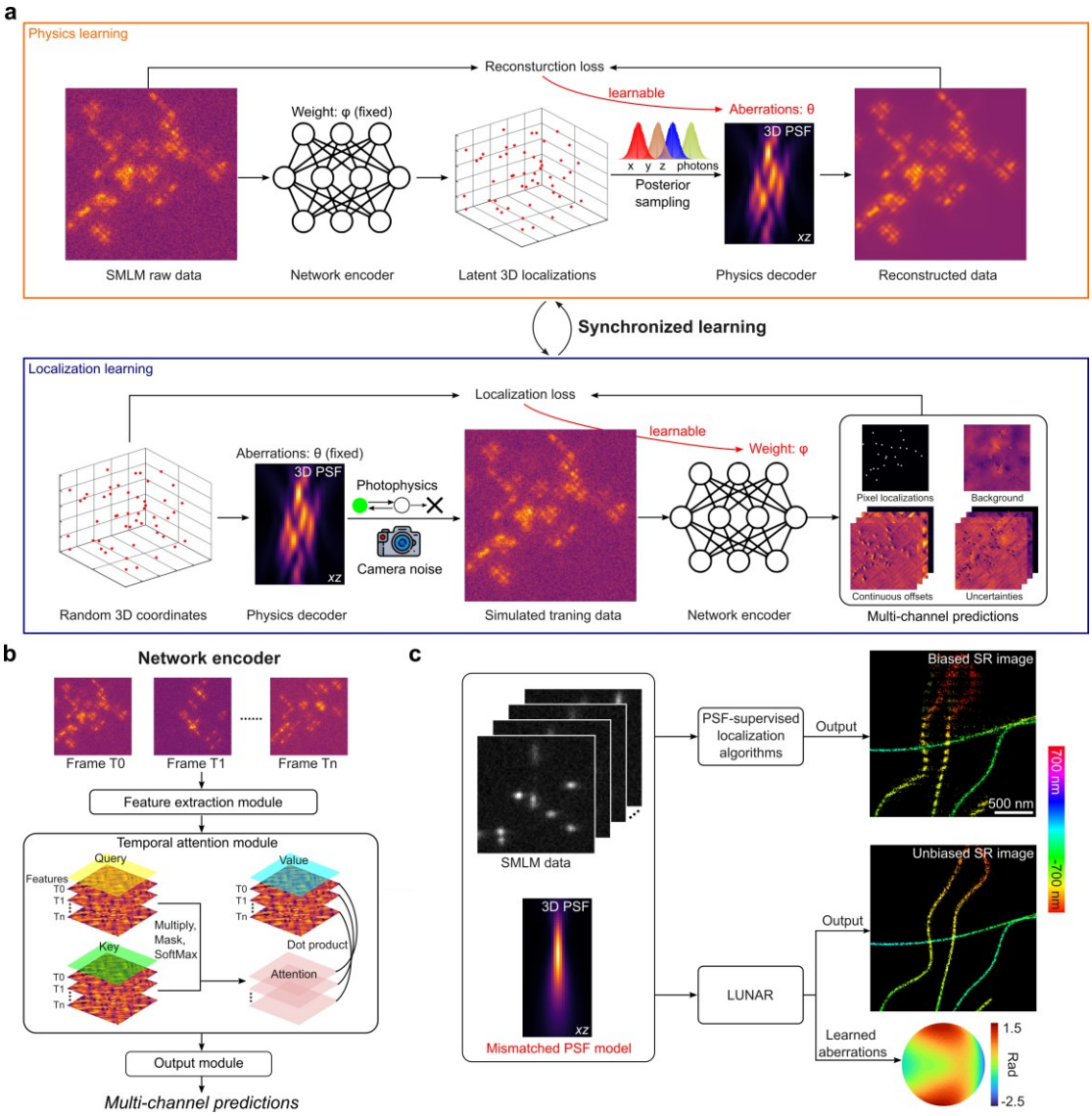

Fig. 1 | Overview of LUNAR framework. a, The synchronized learning strategy for LUNAR alternates between physics learning and localization learning phases. During physics learning, the network is fixed while the physics decoder is updated to minimize the reconstruction loss. During localization learning, the physics decoder is fixed while the network is updated to minimize the localization loss. **b**, The network encoder utilizes a feature extraction module to process each frame independently, while the temporal attention module captures temporal relationships between successive frames. The output module predicts 3D localizations along with their uncertainties for single

molecules in each frame. **c**, For SMLM data with a mismatched PSF model, traditional PSF-supervised localization algorithms produce biased SR images with artifacts. LUNAR overcomes this limitation by simultaneously learning the network encoder for localization and unknown aberrations in the physics decoder, enabling reconstruction of high-fidelity SR images. Scale bar, 500 nm (**c**).

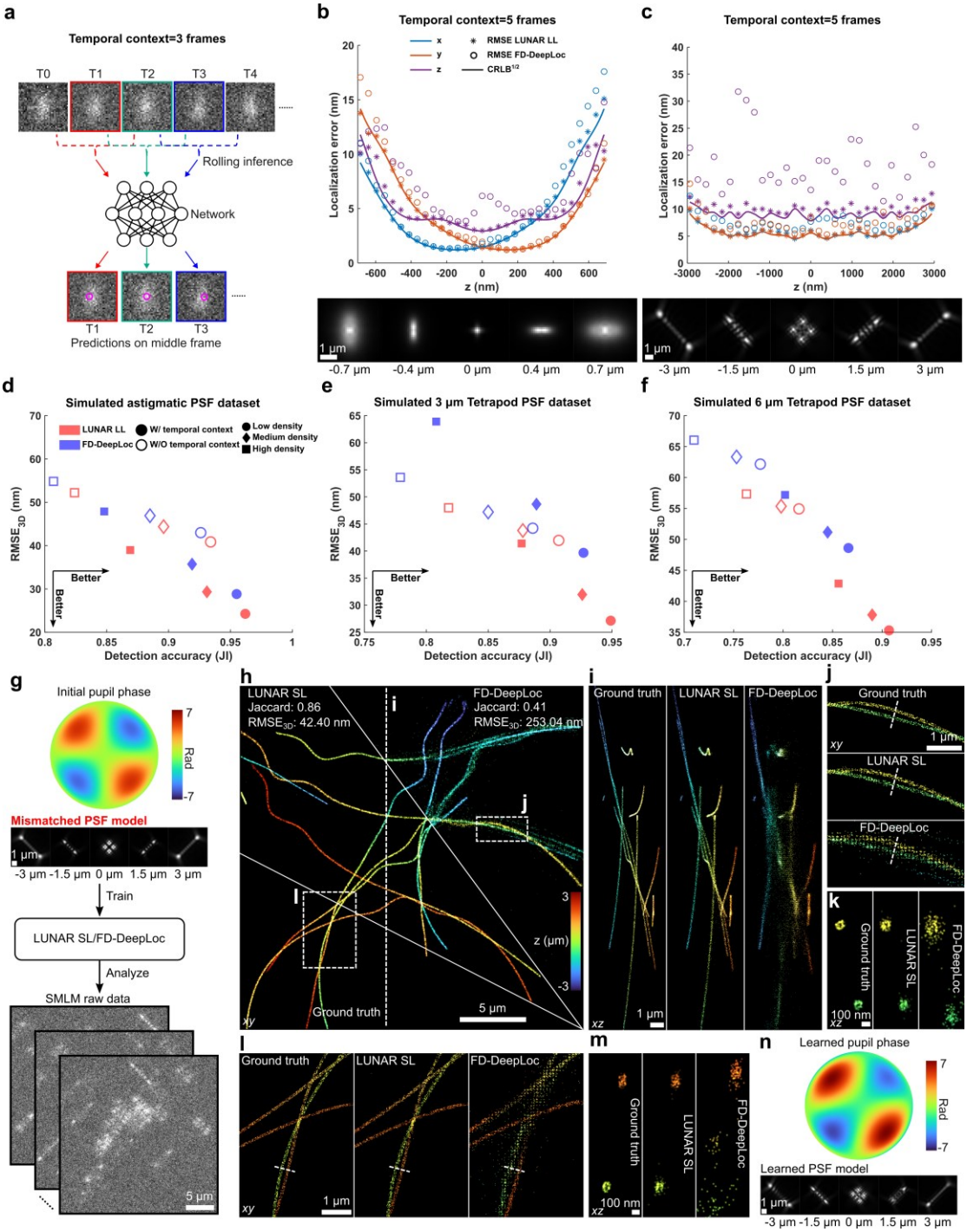

Fig. 2 | Evaluation of LUNAR on simulated datasets. a, Schematic of network localization using a 3-frames temporal context, where successive frames in a batch provide mutual context for localization. **b** and **c,** Comparison of single-molecule localization accuracy between LUNAR LL and FD-DeepLoc using a 5-frames temporal context, evaluated on astigmatic PSF (**b**) and 6 µm Tetrapod PSF (**c**), respectively. Single molecules were simulated with 5,500 photons and 50 background per frame. 1,000 frames were simulated at each of 31 investigated z-positions. **d-f**, Performance

evaluation on simulated datasets of astigmatic PSF (**d**), 3 μm Tetrapod PSF (**e**) and 6 μm Tetrapod PSF (**f**). Both LUNAR LL and FD-DeepLoc were trained using the same training data generated by the ground truth PSF model. **g**, Robustness test on data with unknown aberrations. LUNAR SL and FD-DeepLoc were trained with a mismatched PSF model, which differs from the PSF model used to simulate the test dataset. **h-m**, Comparison of SR images of microtubules reconstructed by LUNAR SL, FD-DeepLoc, and ground truth. **i**, 10 μm-width side-view reconstructions along the dashed line in **h**. **j** and **l**, Magnified views of the region denoted by the dashed boxes in **h**. **k** and **m**, 800 nm-width side-view reconstructions along the dashed lines in **j** and **l**, respectively. **n**, The learned pupil phase and PSF model of LUNAR SL. Scale bars, 5 μm (**g** bottom and **h**), 1 μm (**b**, **c**, **g** top, **i**, **j**, **l** and **n**), and 100 nm (**k** and **m**).

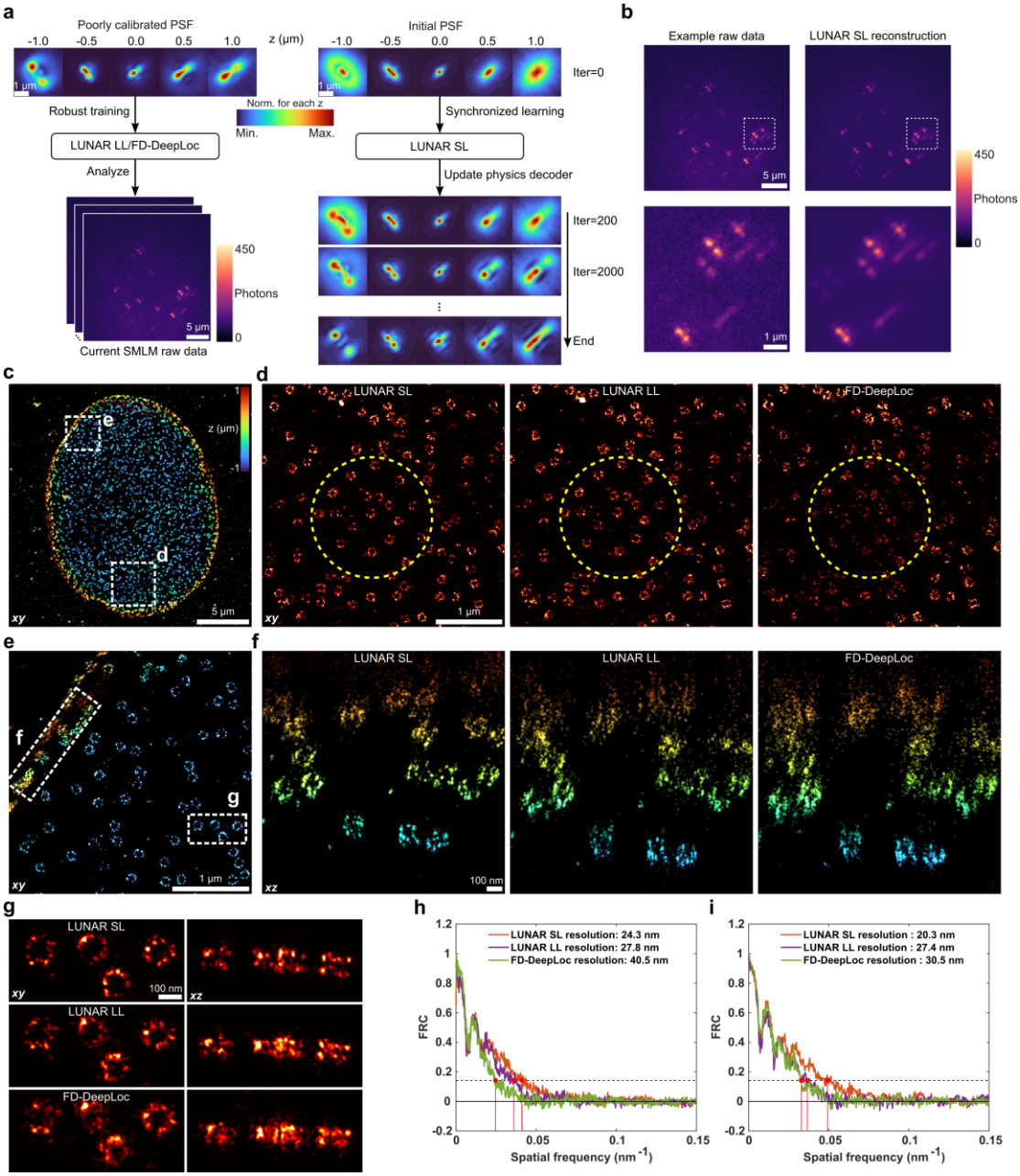

Fig. 3 | Synchronized learning enables high-fidelity SR imaging with poor PSF calibrations. a, PSF-supervised learning algorithms (LUNAR LL, FD-DeepLoc) were trained using a PSF model calibrated several weeks before single-molecule data collection. LUNAR SL was trained with a more inaccurate initial PSF model. **b**, Comparison of experimental and learned single-molecule data reconstructed by LUNAR SL. **c**, Top view of the 3D SR image of NPCs (Nup96) reconstructed by LUNAR SL. **d**, Magnified views of the region denoted by the dashed box d in **c**, reconstructed by LUNAR SL, LUNAR LL and FD-DeepLoc, respectively. Yellow dashed circles highlight area with fewer detections by FD-DeepLoc than LUNAR. **e**,

Magnified view of the region denoted by the dashed box e in **c**. **f** and **g**, Side-view reconstructions of the regions denoted by the dashed boxes in **e,** reconstructed by different algorithms, respectively. **h** and **i**, FRC resolution analysis of the regions in **d** and **e**, respectively. Scale bars, 5 μm (**a** bottom, **b** top and **c**), 1 μm (**a** top, **b** bottom, **d** and **e**), and 100 nm (**f** and **g**).

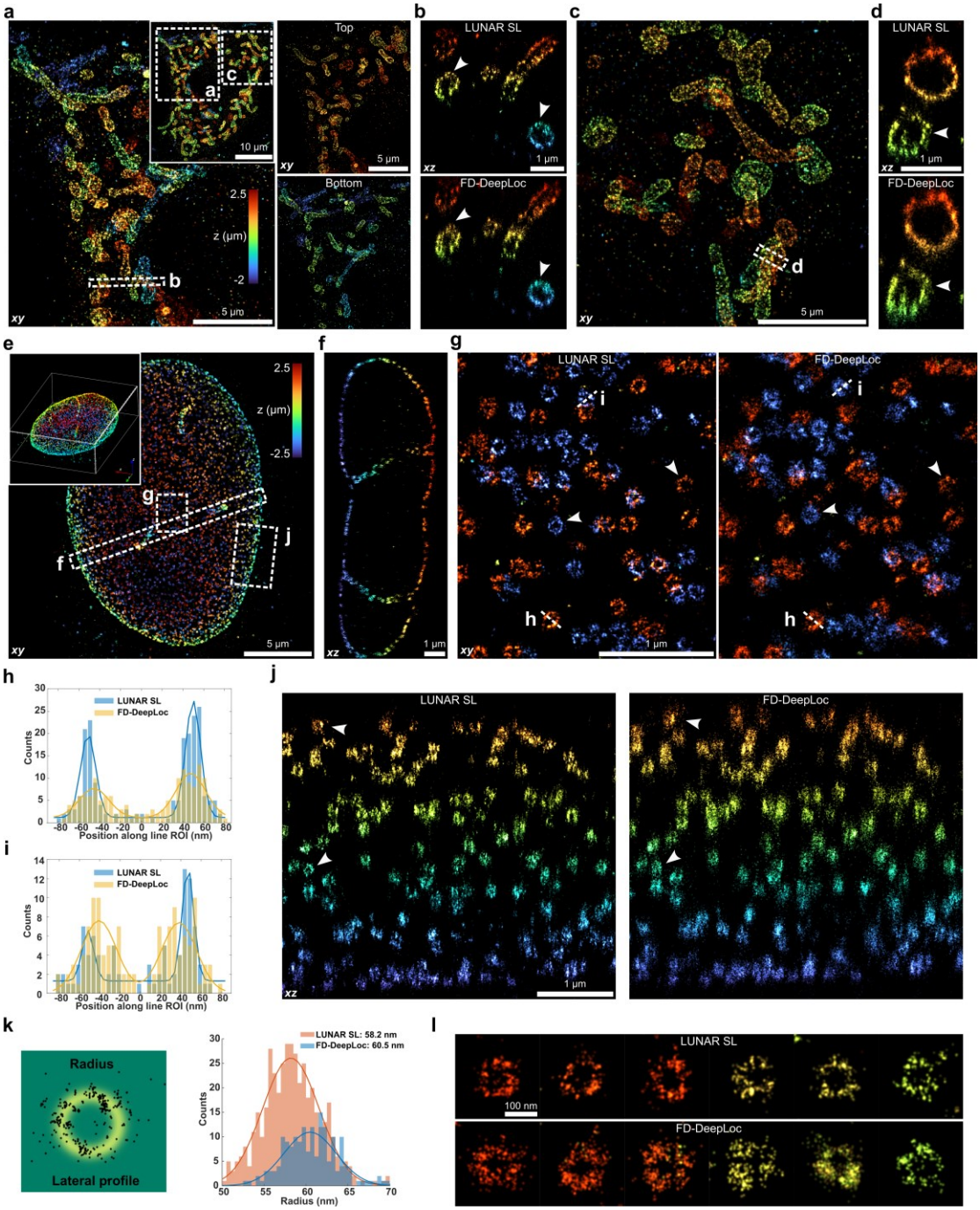

Fig. 4 | Whole-cell longitudinal SR imaging of mitochondria and nucleus. a, SR images of mitochondria (TOM20). Inset is the overview SR image of the whole cell. **b**, Side-view reconstructions of the region denoted by dashed rectangles b in **a. c**, Magnified views of the region denoted by the dashed boxes c in **a**. **d**, Side-view reconstructions of the region denoted by dashed rectangles d in **c**. **e**, Top view and 3D perspective (inset) of the whole-cell SR image of NPCs. **f**, Side view of the region denoted by the dashed rectangle f in **e**. **g**, Comparison of different algorithms by magnified views of the region denoted by the dashed box g in **e**. **h** and **i**, Localization

profiles along the dashed lines in **g**. **j**, Side views of the region denoted by the dashed box j in **e**, reconstructed by LUNAR SL and FD-DeepLoc, respectively. **k**, NPC geometry analysis and distributions of the fitted radii on the top half of the nuclear envelope, with 490 and 191 valid measurements for LUNAR SL and FD-DeepLoc, respectively. **l**, Examples of valid NPC candidates used for radius fitting. Scale bars, 10 μm (**a** inset), 5 μm (**a**, **c**, **g** and **e**), 1 μm (**b**, **d**, **f**, **g** and **j**), and 100 nm (**l**).

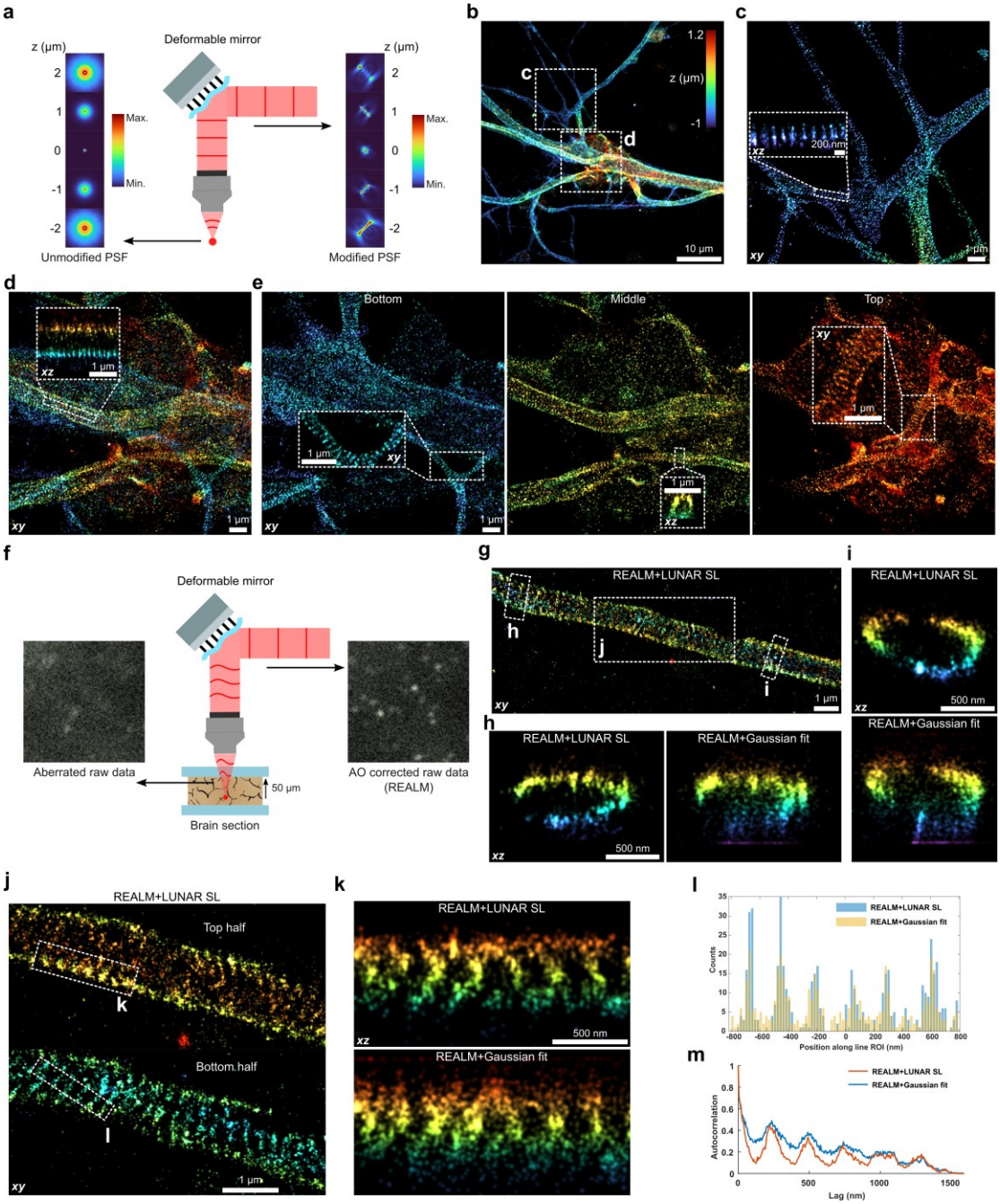

Fig. 5 | LUNAR combined with AO facilitates SR imaging of neuronal cytoskeleton across large depths. a, The use of AO for large-DOF PSF engineering. **b**, Overview SR image of neurons (βII-spectrin) reconstructed by LUNAR SL. **c** and **d**, Magnified views of the regions denoted by the dashed boxes in **b**. **e**, SR images reconstructed using bottom (-1 to -0.2 μm), middle (-0.2 to 0.6 μm), and top (0.6 to 1.2 μm) axial sections of the localization results in **d**. **f**, The use of AO for aberration compensation at 50 μm depth in brain slice. **g**, Overview SR image of an axon (βIV-spectrin) reconstructed by LUNAR SL. **h** and **i**, Side views of the regions denoted by dashed

boxes in **g**, reconstructed by LUNAR SL and Gaussian fit, respectively. **j**, Magnified views of the top and bottom halves of the axon in the region indicated by dashed box j in **g**. **k**, Comparison of LUNAR SL and Gaussian fit for the side views of the region denoted by dashed rectangle k in **j**. **l** and **m**, Localization profiles (**l**) of the region denoted by dashed rectangle l in **j**, along with their autocorrelations (**m**). Scale bars, 10 μm (**b**), 1 μm (**c**, **d**, **e**, **g** and **j**), and 500 nm (**h**, **i** and **k**).