

Published in final edited form as:

Nat Methods. 2020 July 01; 17(7): 734–740. doi:10.1038/s41592-020-0853-5.

DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning

Elias Nehme^{1,2}, Daniel Freedman³, Racheli Gordon², Boris Ferdman^{2,4}, Lucien E. Weiss², Onit Alalouf², Tal Naor², Reut Orange^{2,4}, Tomer Michaeli¹, Yoav Shechtman^{2,4,*}

¹Department of Electrical Engineering, Technion, Haifa, Israel

²Department of Biomedical Engineering & Lorry I. Lokey Center for Life Sciences and Engineering, Technion, Haifa, Israel

³Google Research, Haifa, Israel

⁴Russel Berrie Nanotechnology Intitute, Technion, Haifa, Israel

Abstract

Localization microscopy is an imaging technique in which the positions of individual point emitters (e.g. fluorescent molecules) are precisely determined from their images. This is a key ingredient in single/multiple-particle-tracking and super-resolution microscopy. Localization in three-dimensions (3D) can be performed by modifying the image that a point-source creates on the camera, namely, the point-spread function (PSF). The PSF is engineered to vary distinctively with emitter depth, using additional optical elements. However, localizing multiple adjacent emitters in 3D poses a significant algorithmic challenge, due to the lateral overlap of their PSFs. Here, we train a neural network to localize multiple emitters with densely overlapping PSFs over a large axial range. Furthermore, we then use the network to design the optimal PSF for the multi-emitter case. We demonstrate our approach experimentally with super-resolution reconstructions of mitochondria and volumetric imaging of fluorescently labeled telomeres in cells.

1 Introduction

Determining the nanoscale positions of point emitters forms the basis of localization microscopy techniques such as single particle tracking [1, 2], (fluorescence) photoactivated localization microscopy (f)PALM [3, 4], stochastic optical reconstruction microscopy (STORM) [5], and related single molecule localization microscopy (SMLM) methods. These

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding author: yoavsh@bm.technion.ac.il.

Author contributions

E.N., D.F., T.M. and Y.S. conceived the approach. E.N. performed the simulations and analyzed the data with contributions from all authors. E.N., R.G., B.F., L.E.W., O.A. and T.N. took the data. R.O. fabricated the physical phase mask. T.N. prepared the MEF cells. E.N., D.F., L.E.W., T.M. and Y.S. wrote the paper with contributions from all authors.

Competing interests

The authors declare no competing interests.

Reporting Summary. Further information on research design is available in the Life Sciences Reporting Summary linked to this article.

techniques have revolutionized biological imaging, revealing cellular processes and structures at the nanoscale [6]. Notably, most samples of interest extend in three dimensions, necessitating three-dimensional (3D) localization microscopy [7].

In a standard microscope, the precise z position of an emitter is difficult to ascertain because the change of the point-spread function (PSF) near the focus is approximately symmetric. Furthermore, outside of this focal range ($\approx \pm 350$ nm for a high numerical aperture imaging system), the rapid defocusing of the PSF reduces the signal-to-noise ratio causing the localization precision to quickly degrade. One method to extend the useful z-range and explicitly encode the z position is PSF engineering [8–10]. Here, an additional optical element, e.g. a phase mask, is placed in the emission path of the microscope, modifying the image formed on the detector [11] (Fig. 1a); the axial position can then be recovered via image processing using a theoretical or experimentally-calibrated PSF model [10, 12–15].

In practically all applications, it is desirable to be able to localize nearby emitters simultaneously. For example, in super-resolution SMLM experiments, the number of emitters localized per frame determines the temporal resolution. In tracking applications, PSF overlap from multiple emitters often precludes localization, potentially biasing results in emitter-dense regions. The problem is that localizing overlapping emitters poses a significant algorithmic challenge even in 2D localization, and much more so in 3D. Specifically, encoding the axial position of an emitter over large axial ranges ($>3 \mu\text{m}$) requires the use of laterally large PSFs, e.g. the Tetrapod [10, 16] (Fig. 1b), increasing the possibility of overlap. Consequently, while a variety of methods have been developed to cope with overlapping emitters for the in-focus, standard-PSF [17–19], the performance in high-density 3D localization situations is far from satisfactory [20].

Deep learning has proven to be adept at analyzing microscopic data [21–29], especially for single-molecule localization, handling dense fields of emitters over small axial ranges ($<1.5 \mu\text{m}$) [19, 30–37] or sparse emitters spread over larger ranges [38]. Moreover, an emerging application is to jointly design the optical system alongside the data processing algorithm, enabling end-to-end optimization of both components [36, 39–46]. Here we present DeepSTORM3D, consisting of two fundamental contributions to high-density 3D localization microscopy over large axial ranges. First, we employ a convolutional neural network (CNN) for analyzing dense fields of overlapping emitters with engineered PSFs, demonstrated with the large-axial-range Tetrapod PSF [10, 16]. Second, we design an optimal PSF for 3D localization of dense emitters over a large axial range of $4 \mu\text{m}$. By incorporating a physical-simulation layer in the CNN with an adjustable phase modulation, we jointly learn the optimal PSF (encoding) and associated localization algorithm (decoding). This approach is highly flexible and easily adapted for any 3D SMLM dataset parameters, *i.e.* emitter density, SNRs, and z-range. We quantify the performance of the method by simulation, and demonstrate the applicability to 3D biological samples, *i.e.* mitochondria and telomere.

2 Results

Dense 3D localization with DeepSTORM3D

To solve the high-density localization problem in 3D, we trained a CNN that receives a 2D image of overlapping Tetrapod PSFs spanning an axial range of $4 \mu\text{m}$, and outputs a 3D grid with a voxel-size of $27.5 \times 27.5 \times 33 \text{ nm}^3$ (Fig. 1c). For architecture details and learning hyper-parameters see Supplementary Notes 2.1 and 4. To compile a list of localizations, we apply simple thresholding, local maximum finding, and local averaging on the output 3D grid (Supplementary Note 5).

We compare our method to a fit-and-subtract based Matching Pursuit (MP) approach [47] (see Supplementary Note 7) as we are unaware of any other methods capable of localizing overlapping Tetrapod PSFs. To quantitatively compare our method with MP solely in terms of density, we simulated emitters with high signal-to-noise ratio (30K signal counts, 150 background counts per pixel) at 10 different densities ranging from 1 to 75 emitters per $13 \times 13 \mu\text{m}^2$ field-of-view (for the definition of density see Supplementary Note 1). The results are shown in Fig. 2. As evident in both the Jaccard index [20] (see Supplementary Note 6) and the lateral/axial RMSE (Fig. 2a), the CNN achieves remarkable performance in localizing high-density Tetrapods. In the single-emitter (very low density) case, where the performance of the CNN is bounded by the discretization on the 3D grid, the RMSE of the MP localization is lower (better). This is because for a single-emitter, MP is equivalent to a continuous Maximum Likelihood Estimator (MLE) (Supplementary Note 7), which is asymptotically optimal [48], whereas the CNN's precision is bounded by pixilation of the grid (*i.e.* half voxel of 13.75 nm in xy and 16.5 nm in z). However, quickly beyond the single-emitter case, the CNN drastically outperforms MP at both high and low SNR (see Supplementary Note 7.3). A similar result was obtained when compared to a leading single-emitter fitting method [14] applicable also for the multiple emitter case [20] (see Supplementary Note 8.2). Furthermore, to put our method in context with other existing approaches, we tested DeepSTORM3D on the EPFL Double Helix (DH) high density challenge [20], and got favorable results (see Supplementary Note 8.1).

Next, we validated our method for super-resolution imaging of fluorescently labeled mitochondria in COS7 cells (Fig. 3 and Supplementary Videos 1-3). We acquired 20K diffraction limited frames of a $50 \times 30 \mu\text{m}^2$ FOV and localized them using the CNN in about $\approx 3\text{h } 20\text{m}$, resulting in $\approx 360\text{K}$ localizations. The Tetrapod PSF was implemented using a fabricated fused-silica phase-mask (see Supplementary Note 9.1), and the CNN was trained solely on simulations matching the experimental conditions (see Supplementary Note 4.1). The estimated resolution was $\approx 40 \text{ nm}$ in xy, and $\approx 50 \text{ nm}$ in z (see Supplementary Note 9.3). To visually evaluate localization performance in a single frame (Fig. 3a top), we regenerated the corresponding 2D low-resolution image, and overlaid the recovered image with a uniform photons scaling on top of the experimental frame (Fig. 3a bottom and Supplementary Video 1). As seen in the overlay image, the emitter PSFs (3D positions) are faithfully recovered by the CNN. Moreover, emitters with extremely low number of signal photons were ignored. For further acceleration in acquisition time see Supplementary Note 9.2.

Optimal PSF design for dense 3D imaging

The Tetrapod is a special PSF that has been optimized for the single emitter case by Fisher information maximization [10, 16]. However, when considering the multiple-emitter case, an intriguing question arises: What is the optimal PSF for high density 3D localization over a large axial range? To answer this question we need to rethink the design metric; extending the Fisher information criterion [10] to account for emitter density is not-trivial, and while it is intuitive that a smaller-footprint PSF would be preferable for dense emitters, it is not clear how to mathematically balance this demand with the requirement for high localization precision per emitter.

Our PSF-design logic is based on the following: since we have already established that a CNN yields superior reconstruction for high-density 3D localization, we *are interested in a PSF (encoder) that would be optimally localized by a CNN (decoder)*. Therefore, in contrast to a sequential paradigm where the PSF and the localization algorithm are optimized separately, we adopt a co-design approach (Fig. 4a). To jointly optimize the PSF and the localization CNN, we introduce a differentiable physical simulation layer, which is parametrized by a phase mask that dictates the microscope's PSF. This layer encodes 3D point sources to their respective low-resolution 2D image (see Supplementary Note 3). This image is then fed to the localization CNN which decodes it and recovers the underlying 3D source positions. During training, the net is presented with simulated point sources at random locations (projected on a fine grid) and, using the difference between the CNN recovery and the simulated 3D positions quantified by our loss function (see Supplementary Note 4.2), we optimize both the phase mask and the localization CNN parameters in an end-to-end fashion using the backpropagation algorithm [49] (see Supplementary Note 3.3 and Supplementary Video 4). The learned PSF (Fig. 4b) has a small lateral footprint, which is critical for minimizing overlap at high densities. Moreover, the learned phase mask twists in a spiral trajectory causing the PSF to rapidly rotate throughout the axial range, a trait that was previously shown to be valuable for encoding depth [8].

To quantify the improvement introduced by our new PSF, we first compare it to the Tetrapod PSF in simulations. Specifically, we train a similar reconstruction net for both the Tetrapod and the learned PSF using a matching training set composed of simulated continuous 3D positions along with their corresponding 2D low-resolution images. The learned PSF performs similar to the Tetrapod PSF for low emitter densities (Fig. 4d-f). However, as the density goes up (higher than $\approx 0.2 \left[\frac{\text{emitters}}{\mu\text{m}^2} \right]$) the learned PSF outperforms the Tetrapod PSF in both localization precision (Fig. 4e,f) and in emitter detectability (Jaccard index) (Fig. 4d). This result is not surprising, as the learned PSF has a smaller spatial footprint, and hence it is less likely to overlap than the Tetrapod (Fig. 4c). For further analysis of the learned PSF see Supplementary Note 10.

Volumetric telomere imaging and tracking

Next, we demonstrate the superiority of the new PSF experimentally by imaging fluorescently labeled telomeres (DsRed-hTRF1) in fixed U2OS cells. The cell contains tens of telomeres squeezed in the volume of a nucleus with $\approx 20 \mu\text{m}$ diameter (Fig. 5a, b). From a

single snapshot focused inside the nucleus, the CNN outputs a list of 3D positions of telomeres spanning an axial range of $\approx 3 \mu\text{m}$. Using the Tetrapod PSF snapshot, the Tetrapod-trained CNN was able to recover 49 out of 62 telomeres with a single false positive, yielding a Jaccard index of 0.77 (Fig. 5d). In comparison, using the learned PSF snapshot, the corresponding CNN was able to recover 57 out of the 62 telomeres with only 2 false positives, yielding a Jaccard index of 0.89 (Fig. 5e). The recovered positions were compared to approximated ground-truth 3D positions (Fig. 5c), obtained by axial scanning and 3D fitting (see Supplementary Note 12 and Supplementary Video 5). The precision of both PSFs was calibrated experimentally using a z-scan of a fluorescent micro-sphere (see Supplementary Note 10.4 and Supplementary Video 6).

To qualitatively compare the recovered list of localizations to the acquired snapshot, we fed this list to the physical simulation layer and generated the matching 2D low-resolution image (Fig. 5d,e). As verified by the regenerated images, the 3D positions of the telomeres are faithfully recovered by the CNNs. Moreover, the misses in both snapshots were either due to local aberrations and/or an extremely low number of signal photons (see Supplementary Note 13.1 for more experimental results).

Finally, a great advantage facilitated by our scan-free learned PSF is increased volumetric temporal resolution. To demonstrate the full capability of our method for dense multi-particle localization, we simultaneously track 48 telomeres diffusing within the volume of a live MEF cell, at 10 Hz over 50 seconds (Fig. 6). Such a measurement can provide information on 3D nuclear rotation (Fig. 6a and Supplementary Video 7) and heterogeneity in motion-type (Fig. 6b-d), at time-scales that are typically unexplored by volumetric localization microscopy [50].

3 Discussion

In this work we demonstrated 3D localization of dense emitters over a large axial range both numerically and experimentally. The described network architecture exhibits excellent flexibility in dealing with various experimental challenges, e.g. low signal-to-noise ratios and optical aberrations. This versatility is facilitated in three ways: (1) the net was trained solely on simulated data, thus producing sufficiently large datasets for optimization; (2) the phase mask which governs the PSF was optimized with respect to the implementation in the imaging system, *i.e.* the pixels of the spatial light modulator, rather than over a smaller subspace, e.g. Zernike polynomials [10]; (3) the CNN localization algorithm was designed in coordination with the development of the PSF, thus the system was optimized for the desired output [36] rather than a proxy.

Attaining a sufficiently large training dataset has thus far been a major limitation for most applications of CNNs. With this limitation in mind, the application of CNNs to single-molecule localization would seemingly be an ideal one, since each emitter's behavior should be approximately the same. This uniformity is broken, however, by spatially-varying background, sample density, and variable emitter size in biological samples (Supplementary Note 4.1), all of which diversify datasets and necessitate relevant training data. By implementing an accurate simulator, we have shown that it is possible to build a robust

network entirely in silico, generating arbitrarily large, realistic datasets with a known ground truth to optimize the nets. This aligns with our previous work in 2D SMLM [19].

For super-resolution reconstructions using the Tetrapod PSF, the simulator was particularly important due to the highly variable SNR of emitters in the sample. Here, our net was able to selectively localize the emitters even in very dense regions by focusing on those with a high SNR (Fig. 3). To optimize a PSF while simultaneously training the net, the simulator was also essential, as it would be prohibitively time consuming to experimentally vary the PSF, while recording and analyzing images to train the net.

An intriguing aspect of our optimization approach is that the optimized PSF is found by continuously varying the pixels of an initialized mask while evaluating the output of the localization net, thus the final result represents a local minimum (Fig. 4). By changing the initialization conditions, we have recognized several patterns that indicate how the optimal PSF varies with the experimental conditions: namely, density, axial range, and SNR (see Supplementary Note 2.2). Some of the recurrent features are intuitive: for example, in dense fields of emitters with limited SNR, the optimized PSFs have a small footprint over the designed axial range, enabling high density and compacting precious signal photons into as few pixels as possible. What distinguishes the net PSFs over predetermined designs is the utilization of multiple types of depth encoding; namely, simultaneously employing astigmatism, rotation, and side lobe movement (Fig. 4), all of which have been conceived of and implemented previously, but never simultaneously!

This work, therefore, triggers many possible questions and research directions regarding its capabilities and limitations. For example, how globally-optimal is the resulting PSF? Similarly, how sensitive is the resulting PSF and its performance to different loss functions, CNN architectures, initializations (e.g. with an existing phase mask), and the sampled training set of locations? Currently, it is unclear how each of these components affects the learning process although we began to partially answer them in simulations (see Supplementary Notes 2.2 and 10). Finally, the co-design approach employed here paves the way to a wide variety of interesting applications in microscopy where imaging systems have traditionally been designed separately from the processing algorithm.

Methods

Sample preparation

COS7 cells were grown for 24 hr on cleaned 22×22 mm, $170 \mu\text{m}$ thick coverslips in 6-well plate in Dulbecco's Modified Eagle Medium (DMEM) With 1g/l D-Glucose (Low Glucose), supplemented with Fetal bovine serum, Penicillin-Streptomycin and glutamine, at 37°C , and 5% CO_2 . The cells were fixed with 4% paraformaldehyde and 0.2% glutaraldehyde in PBS, pH 6.2, for 45 min, washed and incubated in 0.3M glycine/PBS solution for 10 minutes. The coverslips were transferred into a clean 6-well plate and incubated in a blocking solution for 2 hr (10% goat serum, 3% BSA, 2.2% glycine, and 0.1% Triton-X in PBS, filtered with $0.45 \mu\text{m}$ PVDF filter unit, Millex). The cells were then immunostained overnight with anti TOMM20-AF647 (Abcam, ab209606) 1:230 diluted in the blocking buffer, and washed X5 with PBS. Cover glasses (22×22 mm, $170 \mu\text{m}$ thick) were cleaned in an ultrasonic bath

with 5% Decon90 at 60°C for 30 min, then washed with water, incubated in ethanol absolute for 30 min and sterilized with 70% filtered ethanol for 30 min.

U2OS cells were grown on cleaned 0.18 μm coverslips in 12-well plate in Dulbecco's Modified Eagle Medium (DMEM) With 1g/l D-Glucose (Low Glucose), supplemented with Fetal bovine serum, Penicillin-Streptomycin and glutamine, at 37°C, and 5% CO_2 . The day after cells were transfected with a plasmid encoding the fluorescently-tagged telomeric repeat binding factor 1 (DsRed-hTRF1) [50] using Lipofectamine 3000 reagent. 24 hr after transfection cells were fixed with 4% paraformaldehyde for 20 minutes, washed 3 times with PBS and attached to a slide together with mounting medium.

Lamin A double knockout (Imna^{-/-}) Mouse Embryonic Fibroblasts (MEFs) were cultured in phenol red-free DMEM/F-12 media (Gibco, Thermo-fisher Scientific), which was supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin solution (Biological industries, Bet Ha-emek, Israel). Two days prior to imaging, cells were transferred to 15 mm coverglass-bottom culture plates (Nest scientific). After 24 hours, cells were transfected with a plasmid encoding DsRed-hTRF1. The transfection mix was prepared by diluting 8 μg of the plasmid in 50 μl of serum-free DMEM/F-12 and separately diluting 24 μl of transfection reagent Polyjet™ (SignaGen Laboratories, USA) in 26 μl of serum-free DMEM/F-12, then immediately adding the diluted Polyjet solution to the DNA mixture and incubating at room temperature for 20 minutes. The 100 μl DNA-Polyjet mix was then added dropwise to the cell plates. Imaging experiments were conducted about 26 hours after transfection and lasted for 1-2 hours.

Optical setup

Imaging experiments were performed on the experimental system shown schematically in Fig. 1b. The 4f optical processing system was built alongside the side-port of a Nikon Eclipse Ti inverted fluorescence microscope, with a $\times 100/1.45$ numerical aperture (NA) oil-immersion objective lens (Plan Apo $\times 100/1.45$ NA, Nikon).

STORM imaging

For super-resolution imaging, a PDMS chamber was attached to a glass coverslip containing fixed COS7 cells. Blinking buffer (100 mM β -mercaptoethylamine hydrochloride, 20% sodium lactate, and 3% OxyFluor (Sigma, SAE0059), modified from [51], was then added and a glass coverslip was placed on top to prevent evaporation. Low-intensity illumination for recording diffraction-limited images was applied using a Topica laser (640 nm), on the Nikon TI imaging setup described previously, and recorded with an EMCCD (iXon, Andor) in a standard imaging setup. For super-resolution blinking using the Tetrapod PSF, high-intensity (1W at the back of the objective lens) 640 nm light was applied using a 638 nm 2000 mW red dot laser module, whose beam shape was cleaned using a 25 μm pinhole (Thorlabs) in coordination with low-intensity (<5mW) 405 nm light. Emission light was filtered through a 500 nm Long pass dichroic and a 650 nm long pass (Chroma), projected through a 4f system containing the dielectric Tetrapod phase mask (see Supplementary Note 9.1), and imaged on a Prime95b Photometrics camera.

Super-resolution image rendering

Prior to rendering the super-resolved image (Fig. 3b), we first corrected for sample drift using the ThunderSTORM [52] ImageJ [53] plugin. Afterwards, we rendered the 3D localizations as a 2D average shifted histogram, with color encoding the z position.

Telomere imaging

For telomere imaging in fixed cells, the 4f system consisted of two $f=15$ cm lenses (Thorlabs), a linear polarizer (Thorlabs) to filter out the light that is polarized in the unmodulated direction of the LC-SLM, a 1920×1080 pixel LC-SLM (PLUTO-VIS, Holoeye) and a mirror for beam-steering. A sCMOS camera (Prime95B, Photometrics) was used to record the data. The sample was illuminated with 561 nm fiber-coupled laser-light source (iChrome MLE, Toptica). The excitation light was reflected up through the microscope objective by a multibandpass dichroic filter (TRF89902-EM - ET - 405/488/561/647 nm Laser Quad Band Set, Chroma). Emission light was filtered by the same dichroic and also filtered by another 617 nm band pass filter (FF02-617/73 Semrock).

For volumetric telomere tracking in live cells, images were recorded with an EMCCD camera (Andor iXON), exposure time 100 ms and EM-gain 170. The sample was illuminated at ≈ 2 kW/cm² with 561 nm light from a fiber-coupled laser (iChrome MLE, Toptica). All movies were recorded for 50 s (500 frames).

CNN Architecture

In a nutshell, our localization CNN architecture is composed of 3 main modules. First, a multi-scale context aggregation module process the input 2D low-resolution image and extracts features with a growing receptive field using dilated convolutions [54]. Second, an up-sampling module increase the lateral resolution of the predicted volume by a factor of x4. Finally, the last module refines the depth and lateral position of the emitters and outputs the predicted vacancy-grid. For more details regarding the architecture see Supplementary Note 2.

Statistics and reproducibility

The STORM experiment was repeated independently for $n = 3$ cells, twice analyzing 20K frames and once analyzing 10K frames all leading to similar performance. The fixed telomere experiment was repeated independently for $n = 10$ U2OS cells all showing similar characteristics and performance. The live telomere experiment was repeated independently for $n = 10$ MEF cells all showing similar characteristics and performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank the Garini lab (Bar-Ilan University, Israel) for the U2OS cells, lamin A double knockout (Imna^{-/-}) Mouse Embryonic Fibroblasts (MEFs), and the plasmid encoding for DsRed-hTRF1. We also thank J. Ries for his help with the application of SMAP-2018 to Tetrapod PSFs. We gratefully acknowledge the support of

NVIDIA Corporation with the donation of the Titan V GPU used for this research. We thank the staff of the Micro-NanoFabrication & Printing Unit (MNF-PU) at the Technion for their assist with the phase mask fabrication. Finally, we thank Google for the research cloud units provided to accelerate this research. E.N., O.A., B.F. and R.O. are supported by H2020 European Research Council Horizon 2020 (802567); T.M. is supported by the Israel Science Foundation (ISF) (grant No. 852/17) and by the Ollendorff Foundation; R.G. and O.A. are supported by Israel Science Foundation (ISF) (grant No. 450/18); Y.S. is supported by the Technion-Israel Institute of Technology Career Advancement Chairship; L.E.W. and Y.S. are supported by the Zuckerman Foundation. D. F. is supported by Google.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code Availability

Code is made publicly available at <https://github.com/EliasNehme/DeepSTORM3D>.

References

- [1]. Katayama Y, et al. Real-time nanomicroscopy via three-dimensional single-particle tracking. *ChemPhysChem*. 2009; 10:2458–2464. [PubMed: 19760694]
- [2]. Manzo C, Garcia-Parajo MF. A review of progress in single particle tracking: from methods to biophysical insights. *Reports on progress in physics*. 2015; 78:124601. [PubMed: 26511974]
- [3]. Betzig E, et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*. 2006; 313:1642–1645. [PubMed: 16902090]
- [4]. Hess ST, Girirajan TP, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*. 2006; 91:4258–4272. [PubMed: 16980368]
- [5]. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature methods*. 2006; 3:793–796. [PubMed: 16896339]
- [6]. Sahl SJ, Moerner W. Super-resolution fluorescence imaging with single molecules. *Current opinion in structural biology*. 2013; 23:778–787. [PubMed: 23932284]
- [7]. von Diezmann A, Shechtman Y, Moerner W. Three-dimensional localization of single molecules for superresolution imaging and single-particle tracking. *Chemical reviews*. 2017; 117:7244–7275. [PubMed: 28151646]
- [8]. Pavani, SRP; , et al. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences*; 2009. 2995–2999.
- [9]. Huang B, Wang W, Bates M, Zhuang X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*. 2008; 319:810–813. [PubMed: 18174397]
- [10]. Shechtman Y, Sahl SJ, Backer AS, Moerner W. Optimal point spread function design for 3d imaging. *Physical review letters*. 2014; 113:133902. [PubMed: 25302889]
- [11]. Backer AS, Moerner W. Extending single-molecule microscopy using optical fourier processing. *The Journal of Physical Chemistry B*. 2014; 118:8313–8329. [PubMed: 24745862]
- [12]. Liu S, Kromann EB, Krueger WD, Bewersdorf J, Lidke KA. Three dimensional single molecule localization using a phase retrieved pupil function. *Optics express*. 2013; 21:29462–29487. [PubMed: 24514501]
- [13]. Babcock HP, Zhuang X. Analyzing single molecule localization microscopy data using cubic splines. *Scientific reports*. 2017; 7:552. [PubMed: 28373678]
- [14]. Li Y, et al. Real-time 3d single-molecule localization using experimental point spread functions. *Nature methods*. 2018; 15:367. [PubMed: 29630062]
- [15]. Aristov A, Lelandais B, Rensen E, Zimmer C. Zola-3d allows flexible 3d localization microscopy over an adjustable axial range. *Nature communications*. 2018; 9:2409.

- [16]. Shechtman Y, Weiss LE, Backer AS, Sahl SJ, Moerner W. Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions. *Nano letters*. 2015; 15:4194–4199. [PubMed: 25939423]
- [17]. Min J, et al. Falcon: fast and unbiased reconstruction of high-density super-resolution microscopy data. *Scientific reports*. 2014; 4:4577. [PubMed: 24694686]
- [18]. Boyd N, Schiebinger G, Recht B. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*. 2017; 27:616–639.
- [19]. Nehme E, Weiss LE, Michaeli T, Shechtman Y. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica*. 2018; 5:458–464.
- [20]. Sage D, et al. Super-resolution fight club: assessment of 2d and 3d single-molecule localization microscopy software. *Nature methods*. 2019; 16:387. [PubMed: 30962624]
- [21]. Rivenson Y, Zhang Y, Günaydin H, Teng D, Ozcan A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*. 2018; 7:17141.
- [22]. Nguyen T, Xue Y, Li Y, Tian L, Nehmetallah G. Deep learning approach for fourier ptychography microscopy. *Optics express*. 2018; 26:26470–26484. [PubMed: 30469733]
- [23]. Weigert M, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*. 2018; 15:1090. [PubMed: 30478326]
- [24]. Ounkomol C, Seshamani S, Maleckar MM, Collman F, Johnson GR. Label-free prediction of threedimensional fluorescence images from transmitted-light microscopy. *Nature methods*. 2018; 15:917–920. [PubMed: 30224672]
- [25]. Krull, A; Buchholz, T-O; Jug, F. Noise2void-learning denoising from single noisy images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2019. 2129–2137.
- [26]. Falk T, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*. 2019; 16:67–70. [PubMed: 30559429]
- [27]. Rivenson Y, et al. Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light: Science & Applications*. 2019; 8:23.
- [28]. Liu T, et al. Deep learning-based super-resolution in coherent imaging systems. *Scientific reports*. 2019; 9:3926. [PubMed: 30850721]
- [29]. Smith JT, et al. Ultra-fast fit-free analysis of complex fluorescence lifetime imaging via deep learning. *bioRxiv*. 2019:523928.
- [30]. Boyd N, Jonas E, Babcock HP, Recht B. Deeploco: Fast 3d localization microscopy using neural networks. *bioRxiv*. 2018:267096.
- [31]. Ouyang W, Aristov A, Lelek M, Hao X, Zimmer C. Deep learning massively accelerates super-resolution localization microscopy. *Nature biotechnology*. 2018; 36:460–468.
- [32]. Diederich B, Then P, Jugler A, Förster R, Heintzmann R. cellstorm—cost-effective super-resolution on a cellphone using dstorm. *PloS one*. 2019; 14
- [33]. Newby, JM; Schaefer, AM; Lee, PT; Forest, MG; Lai, SK. Convolutional neural networks automate detection for tracking of submicron-scale particles in 2d and 3d. *Proceedings of the National Academy of Sciences*; 2018. 9026–9031.
- [34]. Zelger P, et al. Three-dimensional localization microscopy using deep learning. *Optics express*. 2018; 26:33166–33179. [PubMed: 30645473]
- [35]. Liu K, et al. Fast 3d cell tracking with wide-field fluorescence microscopy through deep learning. *arXiv*. 2018
- [36]. Hershko E, Weiss LE, Michaeli T, Shechtman Y. Multicolor localization microscopy and point-spread-function engineering by deep learning. *Optics express*. 2019; 27:6158–6183. [PubMed: 30876208]
- [37]. Speiser A, Turaga SC, Macke JH. Teaching deep neural networks to localize sources in super-resolution microscopy by combining simulation-based learning and unsupervised learning. *arXiv*. 2019
- [38]. Zhang P, et al. Analyzing complex single-molecule emission patterns with deep learning. *Nature methods*. 2018; 15:913. [PubMed: 30377349]

- [39]. Chakrabarti A. Learning sensor multiplexing design through back-propagation. *Advances in Neural Information Processing Systems*. 2016:3081–3089.
- [40]. Horstmeyer R, Chen RY, Kappes B, Judkewitz B. Convolutional neural networks that teach microscopes how to image. *arXiv*. 2017
- [41]. Turpin A, Vishniakou I, d Seelig J. Light scattering control in transmission and reflection with neural networks. *Optics express*. 2018; 26:30911–30929. [PubMed: 30469982]
- [42]. Haim H, Elmalem S, Giryes R, Bronstein AM, Marom E. Depth estimation from a single image using deep learned phase coded mask. *IEEE Transactions on Computational Imaging*. 2018; 4:298–310.
- [43]. He L, Wang G, Hu Z. Learning depth from single images with deep neural network embedding focal length. *IEEE Transactions on Image Processing*. 2018; 27:4676–4689.
- [44]. Sitzmann V, et al. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*. 2018; 37:114.
- [45]. Chang, J; Wetzstein, G. Deep optics for monocular depth estimation and 3d object detection. *Proceedings of the IEEE International Conference on Computer Vision*; 2019. 10193–10202.
- [46]. Wu, Y; Boominathan, V; Chen, H; Sankaranarayanan, A; Veeraraghavan, A. Phasecam3d—learning phase masks for passive single view depth estimation. *2019 IEEE International Conference on Computational Photography (ICCP)*; IEEE. 2019. 1–12.
- [47]. Shechtman Y, Weiss LE, Backer AS, Lee MY, Moerner W. Multicolour localization microscopy by point-spread-function engineering. *Nature photonics*. 2016; 10:590. [PubMed: 28413434]
- [48]. Bickel, PJ, Doksum, KA. *Mathematical Statistics: Basic Ideas and Selected Topics, Volumes I-II* Package. Chapman and Hall/CRC; 2015.
- [49]. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature*. 2015; 521:436–444. [PubMed: 26017442]
- [50]. Bronshtein I, et al. Loss of lamin a function increases chromatin dynamics in the nuclear interior. *Nature communications*. 2015; 6:8044.
- [51]. Nahidiazar L, Agronskaia AV, Broertjes J, van den Broek B, Jalink K. Optimizing imaging conditions for demanding multi-color super resolution localization microscopy. *PLoS One*. 2016; 11
- [52]. Ovesný M, Křížek P, Borkovec J, Švindrych Z, Hagen GM. Thunderstorm: a comprehensive imagej plug-in for palm and storm data analysis and super-resolution imaging. *Bioinformatics*. 2014; 30:2389–2390. [PubMed: 24771516]
- [53]. Schindelin J, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*. 2012; 9:676. [PubMed: 22743772]
- [54]. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv*. 2015

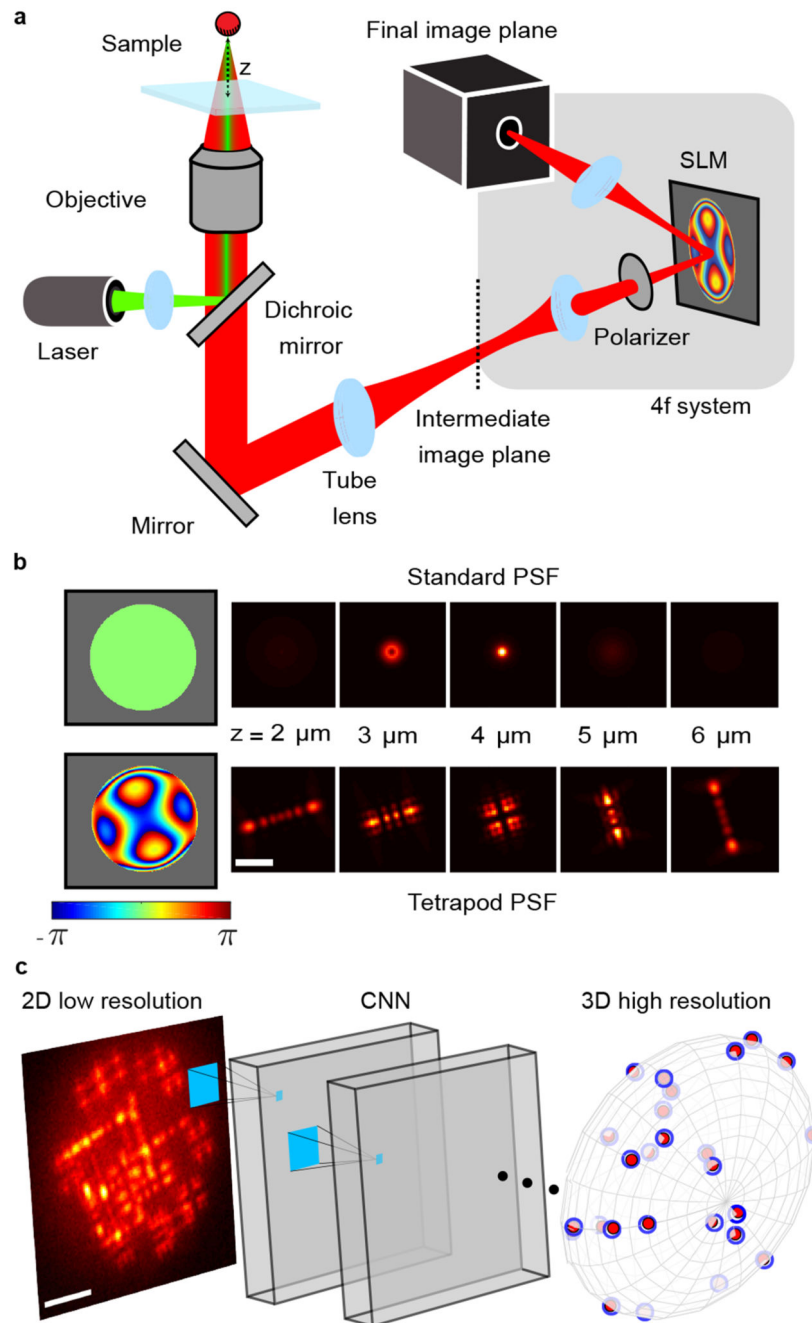


Figure 1. Optical setup and approach overview.

a The light emitted from a fluorescent microscopic particle is collected by the objective and focused through the tube lens into an image at the intermediate image-plane. This plane is extended using a 4f system with a phase mask placed at the Fourier plane in between the two 4f lenses. **b** The implemented phase mask (using either a Liquid Crystal Spatial Light Modulator (LC-SLM) or fabricated fused-silica) dictates the shape of the PSF as function of the emitter's axial position. **c** After training, our CNN receives a 2D low resolution image of overlapping PSFs and outputs a 3D high-resolution volume which is translated to a list of 3D

localizations. Blue empty spheres denote simulated GT positions along the surface of an ellipsoid. Red spheres denote CNN detections. The Tetrapod PSF is depicted here, however the approach is applicable to any PSF, including those optimized by the net itself (Fig. 4). Scale bars are 3 μm .

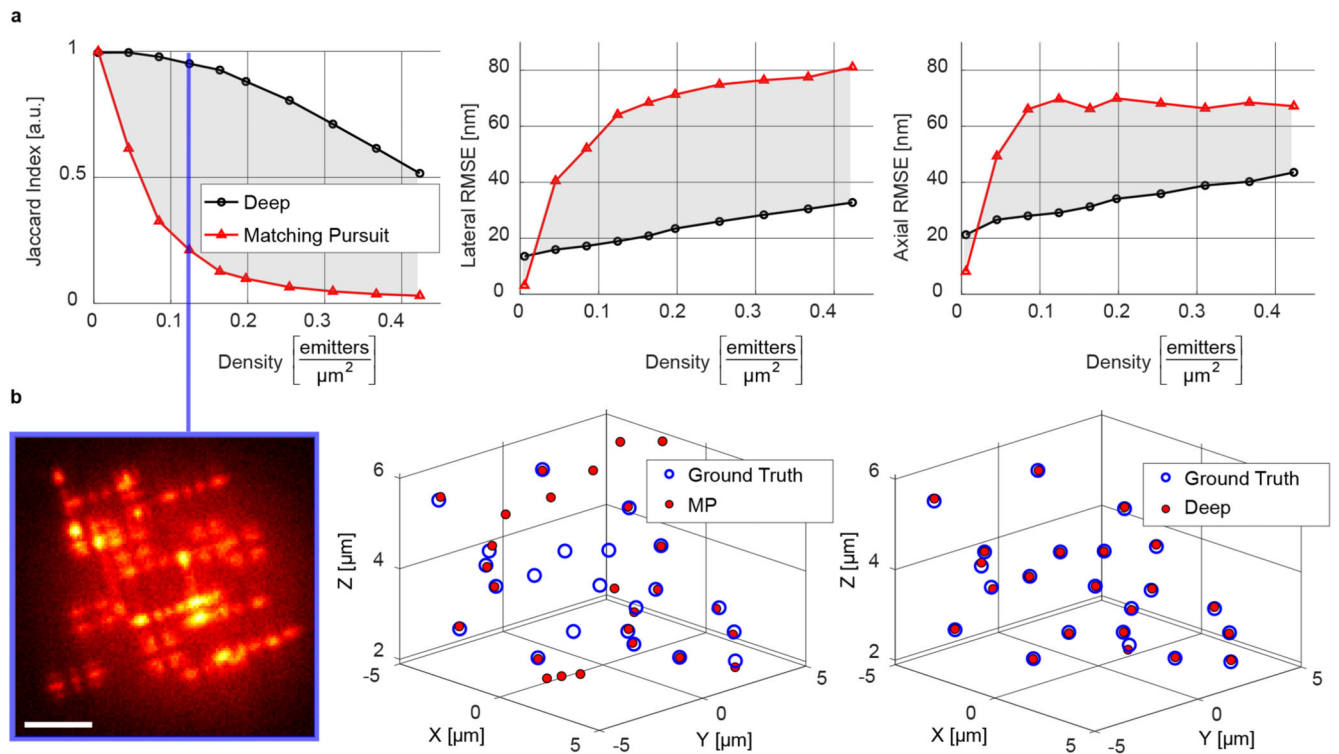


Figure 2. Comparison to MP.

a Performance comparison of a trained CNN (black) and the matching pursuit approach (red) in both detectability (Jaccard index) and in precision (Lateral/Axial RMSE). Matching of points was computed with a threshold distance of 150 nm using the Hungarian algorithm. Each data point is an average of $n = 100$ simulated images. Average standard deviation in Jaccard index was $\approx 6\%$ for both methods, and average standard deviation in precision was ≈ 6 nm for the CNN, and ≈ 15 nm for MP. **b** Example of a simulated frame of density $0.124 \left[\frac{\text{emitters}}{\mu\text{m}^2} \right]$ alongside 3D comparisons of the recovered positions by MP (middle) and by the CNN (right). Scale bar is $2 \mu\text{m}$.

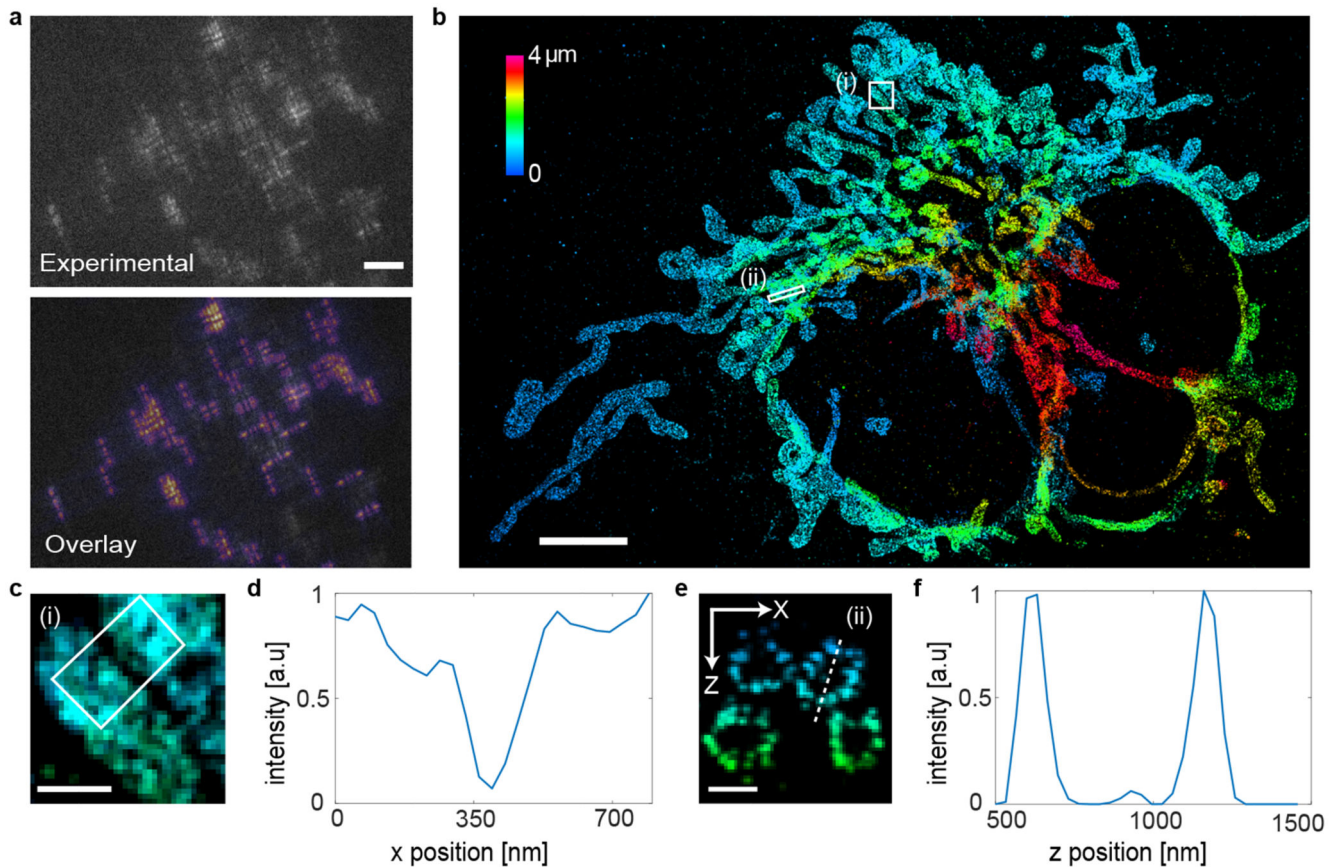


Figure 3. Super-resolution 3D imaging over a $4 \mu\text{m}$ z-range.

a Representative experimental frame (top), and rendered frame from the 3D recovered positions by the CNN overlaid on top (bottom). Scale bar is $5 \mu\text{m}$. **b** Superresolved image of mitochondria spanning a $\approx 4 \mu\text{m}$ z-range rendered as a 2D histogram where z is encoded by color. Scale bar is $5 \mu\text{m}$. **c** Zoom-in on white rectangle (i) in **b**. Scale bar is $0.5 \mu\text{m}$. **d** Relative intensity averaged along the shorter side of the white rectangle in **c**. **e** XZ cross-section of white rectangle (ii) in **b**. Scale bar is $0.5 \mu\text{m}$. **f** Relative intensity along the dashed white line in **e**. The experiment was repeated independently for $n = 3$ cells, twice analyzing 20K frames and once analyzing 10K frames all leading to similar performance.

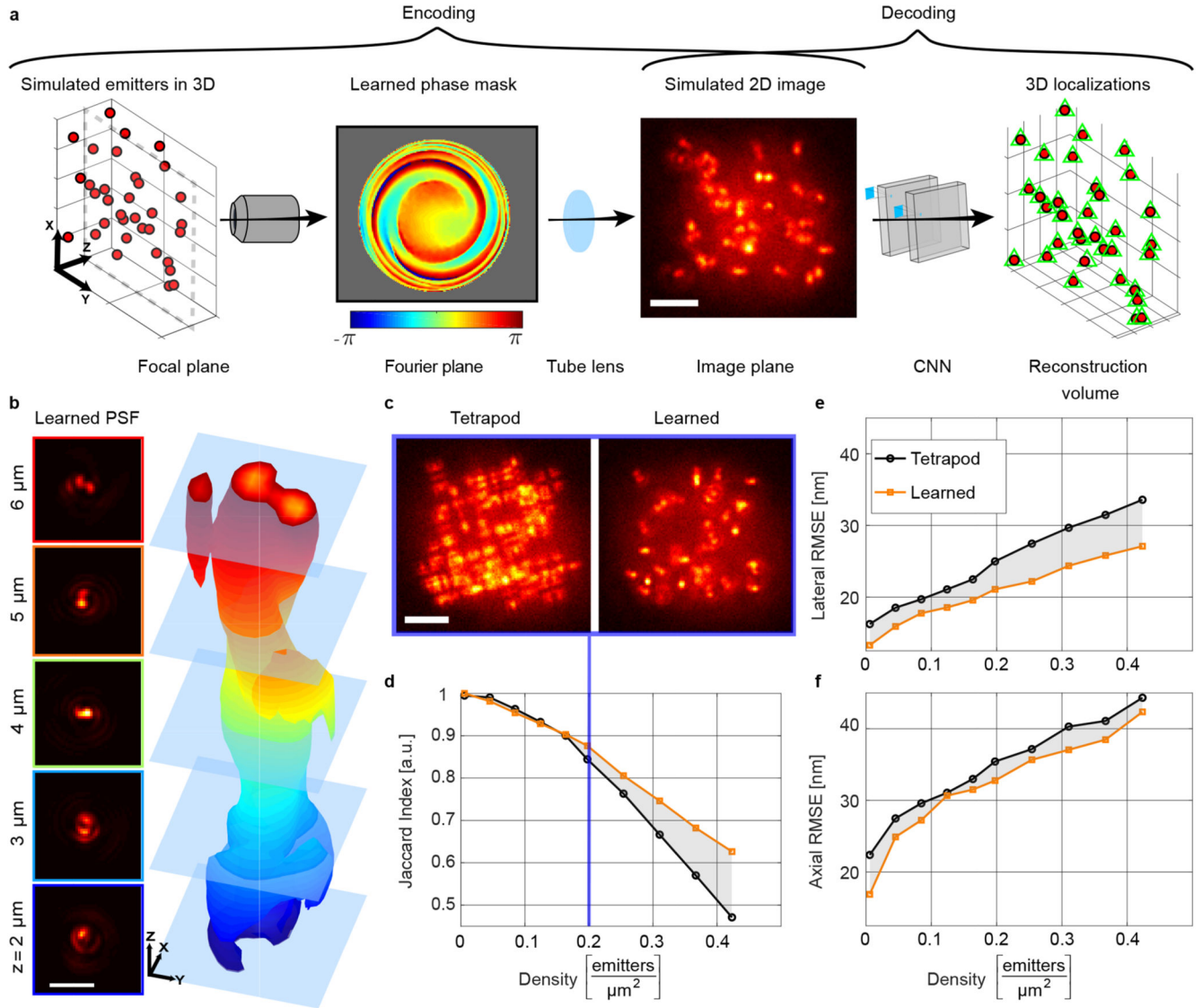


Figure 4. PSF learning for high density 3D imaging.

a Simulated 3D emitter positions are fed to the image formation model to simulate their low resolution CCD image (Encoding). Next, this image is fed to a CNN that tries to recover the simulated emitter positions (Decoding). The difference between the simulated positions and the positions recovered by the CNN is used to jointly optimize the phase mask at the Fourier plane, and the recovery CNN parameters. **b** Simulation of the learned PSF as function of the emitter axial position (left). 3D isosurface rendering of the learned PSF (right). **c** Example frame of density $0.197 \left[\frac{\text{emitters}}{\mu\text{m}^2}\right]$ with the same simulated emitter positions, using the Tetrapod (left) and the learned PSF (right). **d** Jaccard index comparison between two CNNs with the same architecture, one trained to recover 3D positions from 2D images of Tetrapod PSF (black), and the second trained to recover 3D positions from 2D images of the learned PSF (orange). Each data point is an average of $n = 100$ simulated images. Average standard deviation was $\approx 6\%$ for both PSFs. **e-f** Lateral and axial RMSE comparison between the

same two CNNs from **d**. Average standard deviation was ≈ 6 nm for both PSFs. Scale bars are $3 \mu\text{m}$.

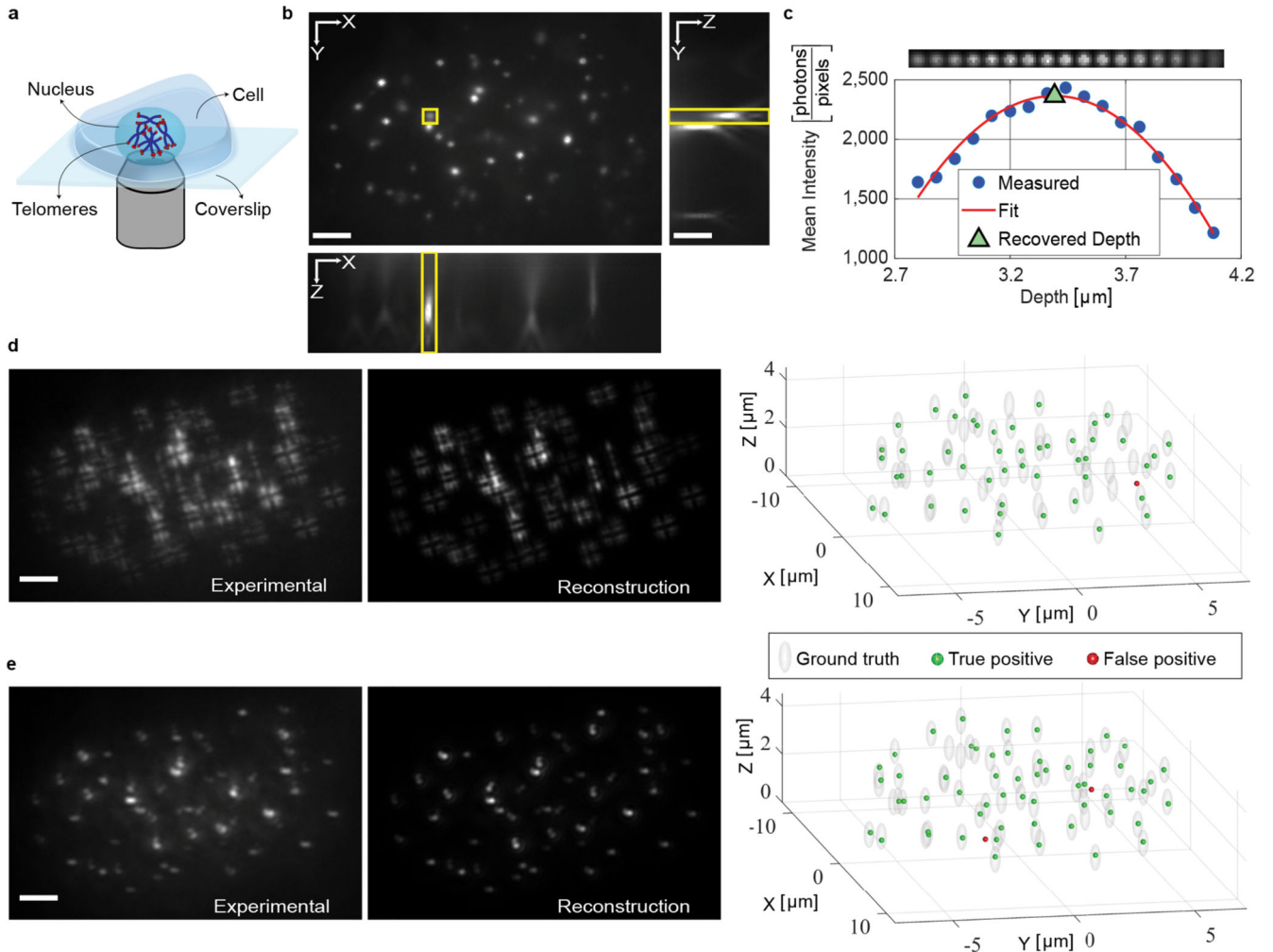


Figure 5. Three dimensional imaging of telomeres in a single-snapshot.

a Schematic of imaging fixed U2OS cells with fluorescently labeled telomeres inside their nucleus. **b** Focus slice with the standard PSF inside a U2OS cell nucleus, obtained via a z-scan. The yellow rectangles mark the same emitter in all three orthogonal planes. **c** Example fit of the mean intensity in sequential axial slices used to estimate the approximate emitter axial position. **d** Experimental snapshot with the Tetrapod PSF (left), rendered image from the 3D recovered positions by the Tetrapod CNN (middle), and a 3D comparison of the recovered positions and the approximate experimental ground truth (right). **e** Experimental snapshot with the learned PSF (left), rendered image from the 3D recovered positions by the learned PSF CNN (middle), and a 3D comparison of the recovered positions and the approximate experimental ground truth (right). Scale bars are $3 \mu\text{m}$. The experiment was repeated independently for $n = 10$ U2OS cells all showing similar characteristics and performance.

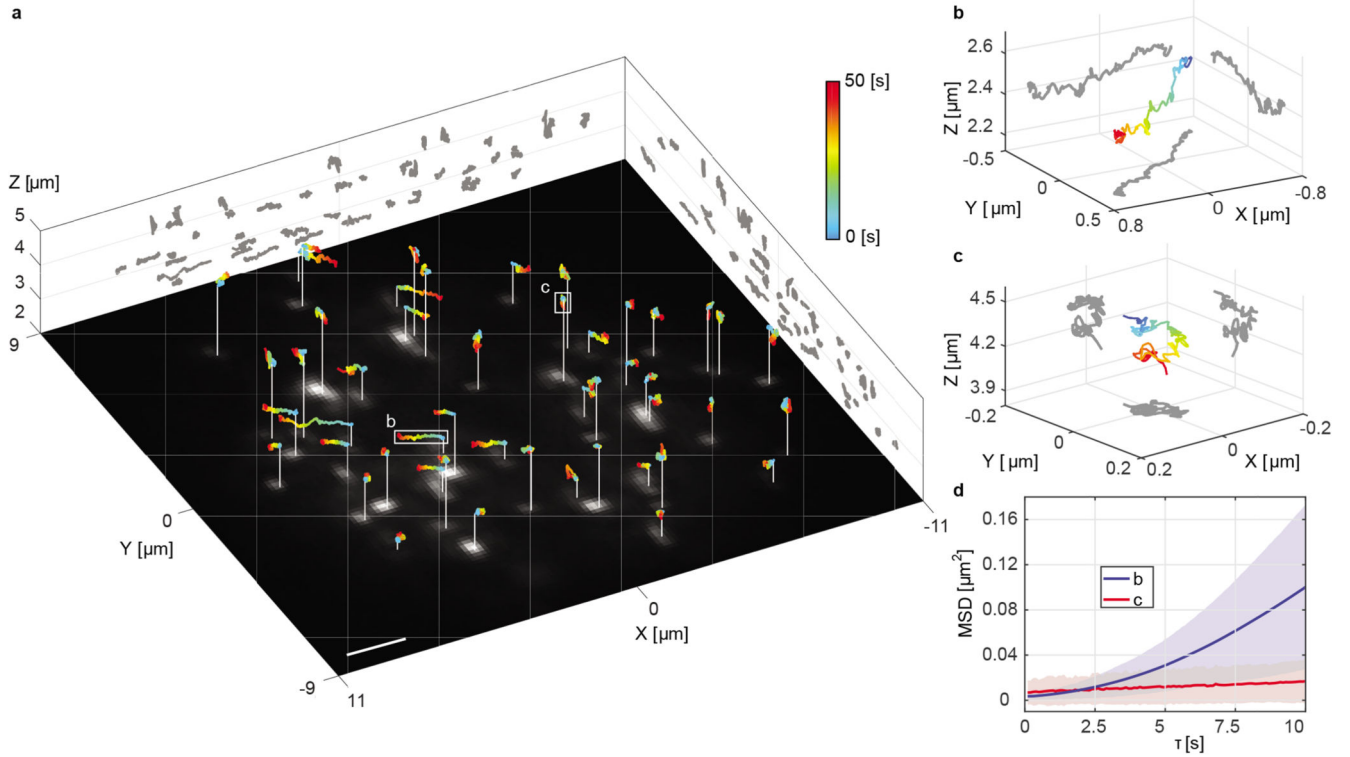


Figure 6. Volumetric tracking of telomeres in live MEF cells.

a 3D trajectories of tracked telomeres on top of the first experimental snapshot. White sticks mark the starting point, and color encodes time. XZ and YZ projections are plotted in gray. Scale bar is $2 \mu\text{m}$. **b** Example trajectory of a telomere (white box in **a**) with a large mean displacement. **c** Example trajectory of a telomere (white box in **a**) with a small mean displacement. **d** Mean Squared Displacement (MSD) of the trajectories in **b** and **c**. Shaded area marks one standard deviation. The experiment was repeated independently for $n=10$ MEF cells all showing similar characteristics and performance.