DBlink: Dynamic localization microscopy in super spatiotemporal resolution via deep learning

Alon Saguy[1], Onit Alalouf[1], Nadav Opatovski[2], Soohyen Jang[3,4], Mike Heilemann[3,4], Yoav Shechtman[1,†]

[1] Department of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

[2] Russell Berrie Nanotechnology Institute, Technion – Israel Institute of Technology, Haifa, Israel

[3] Institute of Physical and Theoretical Chemistry, Goethe-University Frankfurt, Germany

[4] Institute of Physical and Theoretical Chemistry, IMPRS on Cellular Biophysics, Goethe-University Frankfurt, Germany

[†] Corresponding Author

Corresponding author mail address: yoavsh@bm.technion.ac.il

Abstract

Single molecule localization microscopy (SMLM) has revolutionized biological imaging, improving the spatial resolution of traditional microscopes by an order of magnitude. However, SMLM techniques require long acquisition times, typically a few minutes, to yield a single super-resolved image, because they depend on accumulation of many localizations over thousands of recorded frames. Hence, the capability of SMLM to observe dynamics at high temporal resolution has always been limited. In this work, we present DBlink, a deep-learning-based method for super spatiotemporal resolution reconstruction from SMLM data. The input to DBlink is a recorded video of SMLM data and the output is a super spatiotemporal resolution video reconstruction. We use a convolutional neural network combined with a bi-directional long short-term memory (CNN-LSTM) network architecture, designed for capturing long-term dependencies between different input frames. We demonstrate DBlink performance on simulated filaments and mitochondria-like structures, on experimental SMLM data under controlled motion conditions, and finally on live cell dynamic SMLM. DBlink's spatiotemporal interpolation constitutes an important advance in super-resolution imaging of dynamic processes in live cells.

Main text

Introduction

The spatial resolution in standard optical microscopes is bounded by the diffraction limit at about half the wavelength of light, corresponding, in the visible range, to ~200-300 nm. Super-resolution microscopy (SRM) methods overcome this limitation and enable higher resolution. Notable methods of this family include stimulated emission depletion (STED)[1], structured-illumination microscopy (SIM)[2], as well as single molecule localization microscopy (SMLM)[3]. Prominent variants of SMLM include photoactivated localization microscopy[4] (PALM), stochastic optical reconstruction microscopy[5] (STORM), points accumulation for imaging in nanoscale topography[6] (PAINT), and DNA-PAINT[7]. The SMLM variants differ in their experimental conditions; however, they share a similar overall pipeline: First, fluorescent molecules are used to label structures in a specimen. Then, a sequence of frames is captured, in which only a sparse, random subset of molecules emit light per-frame. Subsequently, each emission event is detected, and fit to a model of the system point spread function (PSF) allowing highly precise determination of the emitting fluorophore position. Finally, by accumulating the localizations of thousands of emitters, the output of SMLM is a single super-resolved image of the structure, typically with an order of magnitude resolution improvement compared to the diffraction limit.

An inherent limitation in SMLM is its temporal resolution. Accumulating a large enough number (typically millions) of single-molecule emission events to generate a continuous image takes a long time. Moreover, densely labeled structures present a challenge in the detection of single emitters, which limits the maximal number of localizations per frame. Indeed, the typical temporal resolution in SMLM is on the order of minutes, while tens-of-seconds resolutions have also been reported[8,9]. A possible way to increase temporal resolution in SMLM could be to increase the illumination intensity and decrease frame acquisition time[10]. However, high illumination intensity is inherently incompatible with live cell imaging due to phototoxicity[11]. Recent advances in deep learning algorithms have yielded computational algorithms that further improve the capabilities of SMLM. ANNA-PALM[8] significantly reduces the number of frames needed for super-resolution reconstruction. Deep-STORM[12,13] as well as DECODE[14] enable researchers to analyze densely labeled SMLM experiments by training a neural network to perform multi-emitter fitting in super-resolution. Importantly, while these algorithms perform exceptionally well in visualization of nanoscale structures, they are still mostly applicable for the analysis of static data or processes with very slow dynamics.

Existing methods are useful mostly for static data because the typical localization-based reconstruction process does not exploit structural-correlations over long periods of time (longer than the temporal window being reconstructed). Non SMLM based super-resolution methods[15–17], that naturally exhibit high temporal resolution, either compromise on spatial resolution, or do not exploit the abundant super-resolved information of single molecule emission events. A method that combines both spatial and long-range temporal interpolation would be optimal.

Recently, methods utilizing both spatial and temporal information have emerged[18–20]. Super-resolution radial fluctuations (SRRF)[19] measures the image intensity gradients in subpixel accuracy and detects the imaged structure by searching gradient convergence points. Since single molecules often blink over multiple frames, additional pixelwise cross-correlation analysis helps distinguish adjacent fluorophores. Recently, enhanced SRRF (eSSRF)[21] has enabled live-cell 3D video reconstruction at spatial resolution of ~70 nm and temporal resolution of ~1 reconstructed volume (20 x 20 x 3.6 μm^3) per second. Although eSRRF achieves a significant improvement in the spatiotemporal resolution, it does not exploit long term temporal correlations, that contain valuable structural information.

In this paper, we present DBlink, a novel method that increases the spatiotemporal resolution in the reconstruction of live-cell dynamic SMLM data. We use a convolutional bi-directional long short-term memory (CNN-LSTM) network, that receives as input a video containing super-resolved localization maps and outputs a video of a dynamic super-resolved structure (Fig. 1). The super-resolved localization maps can be obtained by using existing methods, e.g., Deep-STORM[12] or ThunderSTORM[22]. In order to perform spatiotemporal interpolation, DBlink relies on long-term inter-frame structural-correlation, and on prior information regarding the imaged sample – namely, its type (e.g. microtubules, mitochondria, etc.). We first demonstrate the ability of DBlink to reconstruct super spatiotemporal resolution videos of simulated filament dynamics. Then, we validate our network performance on experimental data in controlled conditions where we possess ground truth information. Next, we present super spatiotemporal resolution reconstructions of microtubule and endoplasmic reticulum (ER) dynamics in live cells, achieving spatial resolution of ~30 nm and temporal resolution of 15 ms. Finally, we demonstrate the reconstruction of mitochondrial dynamics from live-cell PAINT data using a non-covalent, weak affinity fluorophore label.

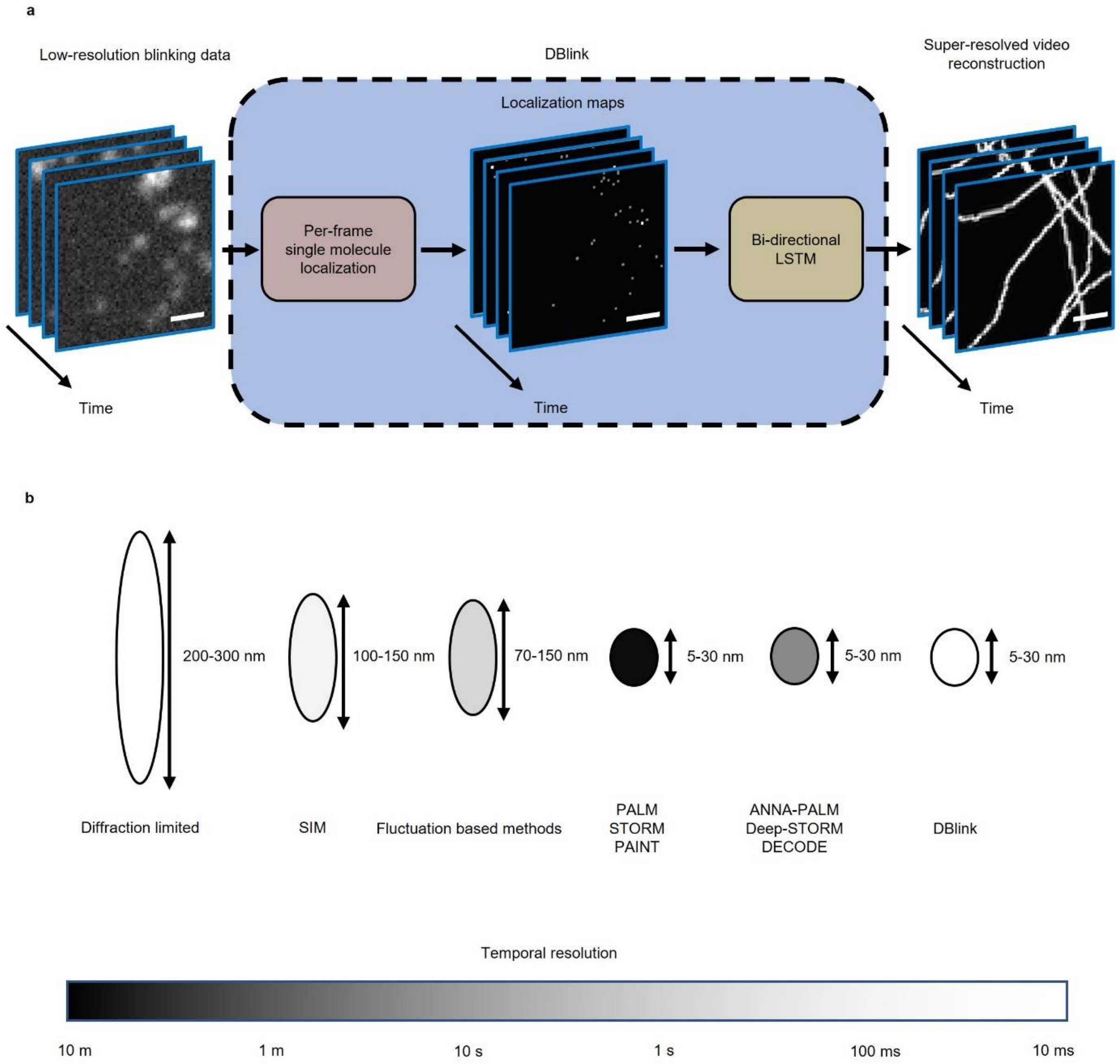

Figure 1: DBlink concept. **a** Low-resolution frames containing stochastic blinking events are analyzed by a localization method, in our case, Deep-STORM[12], which generates super-resolved localization maps for each input frame. The localization maps serve as input to a CNN-LSTM network that provides as output super-resolution video reconstruction of the imaged structure. Scale bar = 2.5 μm. **b** Spatial resolution of different super-resolution microscopy methods[23]. The color represents temporal resolution.

Results

Our goal is to extend the temporal resolution of SMLM beyond its inherent limitation, dictated by the trade-off between emitter density and localization ability, while maintaining high-spatial resolution. Conceptually, the problem at hand is spatiotemporal interpolation of a 3D manifold (2D + time), that describes the continuous movement of a 2D object, from noisy discrete samples in space (2D localizations) and time (due to the camera acquisition rate). Clearly, there is insufficient information per frame, and there are multiple valid solutions, mathematically. The strategy we chose here is to train a neural network with realistic regularization based on prior knowledge on the imaged sample.

Injecting relevant prior knowledge to our network requires thousands of videos containing ground truth information at super spatiotemporal resolution. Experimental acquisition of this amount of data is extremely challenging; furthermore, ground truth information is not available at high spatiotemporal resolution. Therefore, we took the approach of simulating the training videos based on relatively simple static biological models and applying to them time-varying affine transformations. The challenge in simulation-based training is to generate simulations that resemble experimental data; in our case, the biological models facilitated this step, and we only had to tune several experimental parameters (e.g. blinking density, motion velocities, etc.) to generate realistic simulated videos. To capture long term dependencies between different video frames, we employ an LSTM network architecture, which has previously proven itself as a good solution for this task[24,25]. LSTM networks are suitable for sequential analysis tasks since each layer is designed to carry information from previous frames throughout the sequential analysis of the video. Because our method analyzes experiments in retrospect, we also possess information from future frames; therefore, we use a bi-directional LSTM network, and concatenate the forward pass and the backward pass. As expected, this strategy boosts the performance in comparison to a one-directional LSTM network architecture (see one-directional LSTM evaluation section in the SI and Supplementary Video S15). Each LSTM cell consists of a convolutional layer aiming to capture local spatial correlations between adjacent pixels. Each output element of the CNN-LSTM network is analyzed by an additional convolutional neural network (CNN) that provides the final reconstructed frame (Supplementary Fig. S1). In the following example applications, we show that this approach is feasible and produces high quality results. In addition, the recurrent part of the network is not sensitive to different temporal blinking patterns; hence, we obtain clean and uniform structures in the reconstructions.

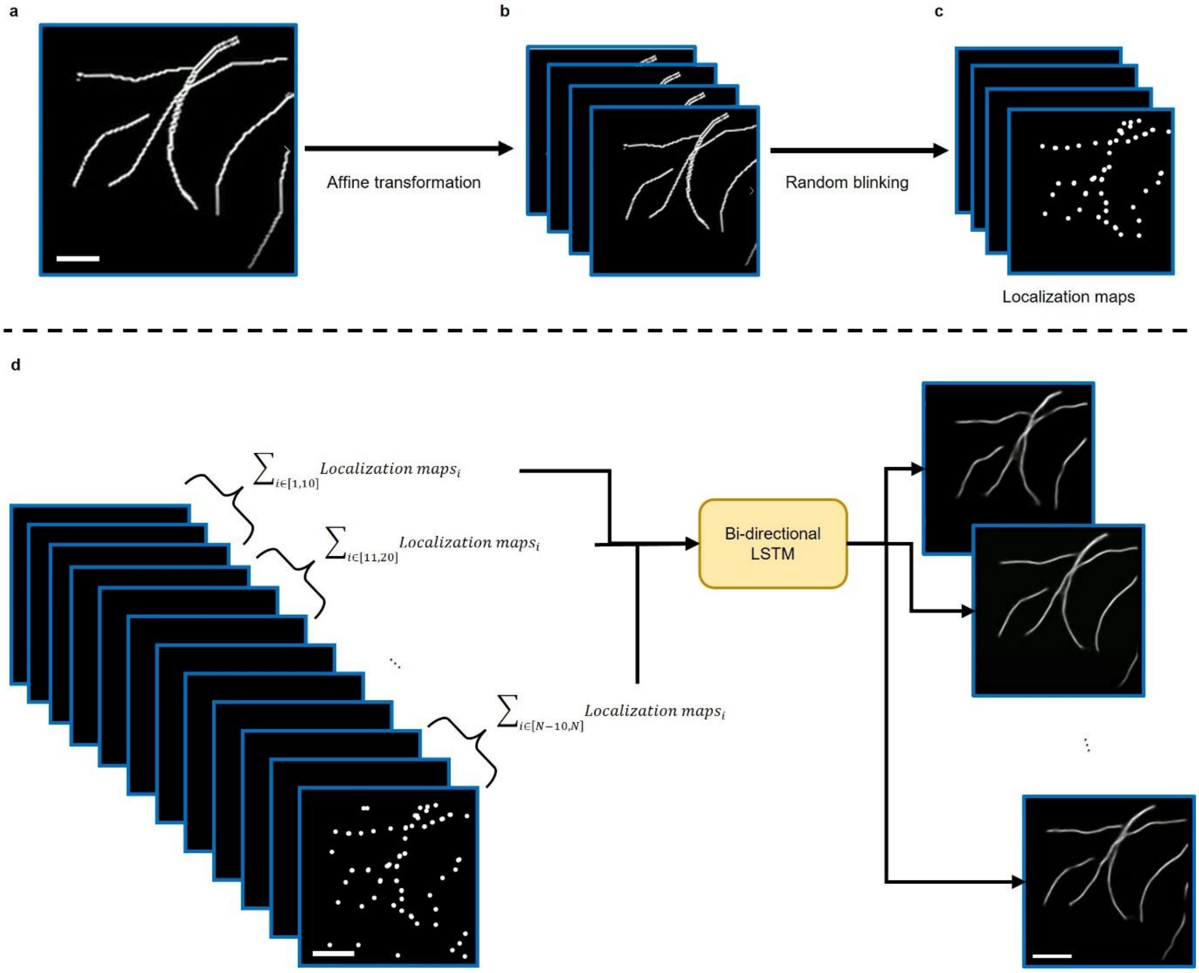
a
b
c
Affine transformation
Random blinking
Localization maps
d
$\sum_{i\in[1,10]} Localization\ maps_i$
$\sum_{i\in[11,20]} Localization\ maps_i$
$\sum_{i\in[N-10,N]} Localization\ maps_i$
Bi-directional LSTM

Figure 2: Generation and analysis of simulated filament data. **a** We simulated a random number of filaments in the field of view (FOV) according to the model of Shariff, et al[26]. **b** Then, we applied gradually increasing affine transformations over a predefined video length of N frames. **c** Next, we generated random blinking localizations

based on the simulated structure. **d** Finally, we summed the simulated localizations every 10 frames and inserted the summed frames (total of N/10 frames) to the LSTM. The output of the LSTM was a super-resolution reconstruction video of length N/10. Scale bar = 2.5 μm.

First, we tested our approach on simulated data (Fig. 2). To do so, we generated simulated filaments according to the model of Shariff *et al.*[26] (see methods section); then, we shifted and rotated them randomly, while maintaining structural smoothness and continuity in time (see filament simulation section in the SI). The neural network was able to reconstruct random shifting structures over time with high accuracy (Supplementary Video S1), namely, 90% of the simulated binary map matched the predicted structure, and only 1.3% of the predicted structure was hallucinated (see reconstruction accuracy section in the SI). In this simulation, the temporal resolution corresponds to 1 reconstructed frame per 10 simulated blinking frames.

To quantify the spatial resolution of the reconstruction, we performed Fourier ring correlation (FRC) analysis[27] as well as decorrelation analysis[28] between the network reconstruction and a STORM reconstruction on a static sample. The network result was consistent with standard STORM reconstruction using Deep-STORM[12] up to a resolution of 29 nm according to FRC and up to 30 nm according to decorrelation analysis (see spatial resolution quantification section in the SI). Next, as a first validation of our method on experimental data, we reconstructed a static structure which was shifting laterally over time. For this, we captured a STORM experiment of fixed microtubules exhibiting naturally occurring lateral sample-drift. We estimated the drift using Deep-STORM drift correction mechanism, which is based on cross-correlation, and received a total shift of 240 nm in y direction and 400 nm in x direction (Fig 3a). Next, we used the localization maps provided by Deep-STORM and summed them over windows of 100 frames with 50 ms acquisition time. Finally, we input the summed localization maps into our network and received a super spatiotemporal reconstruction of the shifting data at a temporal resolution of 0.2 frames per second (Supplementary Video S2). We predicted the drift according to the cross-correlation between the first reconstructed frame of our network and every other frame in the reconstructed video. The mean distance between our drift prediction and Deep-STORM prediction over the course of the experiment was 38 nm (Fig. S3). Notably, the network did not have any prior knowledge that the sample is static and drifting, namely, that the only motion was a global shift; rather, the network treated this data the same as general dynamic data. A global-motion prior would improve the performance significantly, at the cost of a less generalized solution.

To demonstrate our method performance on a more complex type of motion than lateral shift, while still possessing knowledge of the sample structure to serve as validation, we captured a STORM video of static microtubules while rotating the camera manually (Fig. 3b). We added to the sample fluorescent beads to serve as fiducial markers reporting on sample rotation. At the end of the experiment, we stopped rotating the camera and let the blinking continue for ~15,000 more frames, from which the ground truth structure was obtained using Deep-STORM. To test reconstruction performance, we computationally rotated back the predicted structure in each frame according to the calculated rotation angle. Then, we compared the computationally-rotated video to the static reconstruction obtained by Deep-STORM (Supplementary Video S3). To quantify the prediction error, we measured the consistency between the reconstructed video frames based on the cross-correlation between every two frames in the reconstructed video, achieving a mean consistency score of 0.91, which indicates a consistent reconstruction (see reconstruction accuracy quantification section in the SI). In this experiment, we used a window size of 40 summed frames to generate the reconstructed video, with acquisition time of 20 ms per frame, resulting in temporal resolution of 1.25 frames per second.

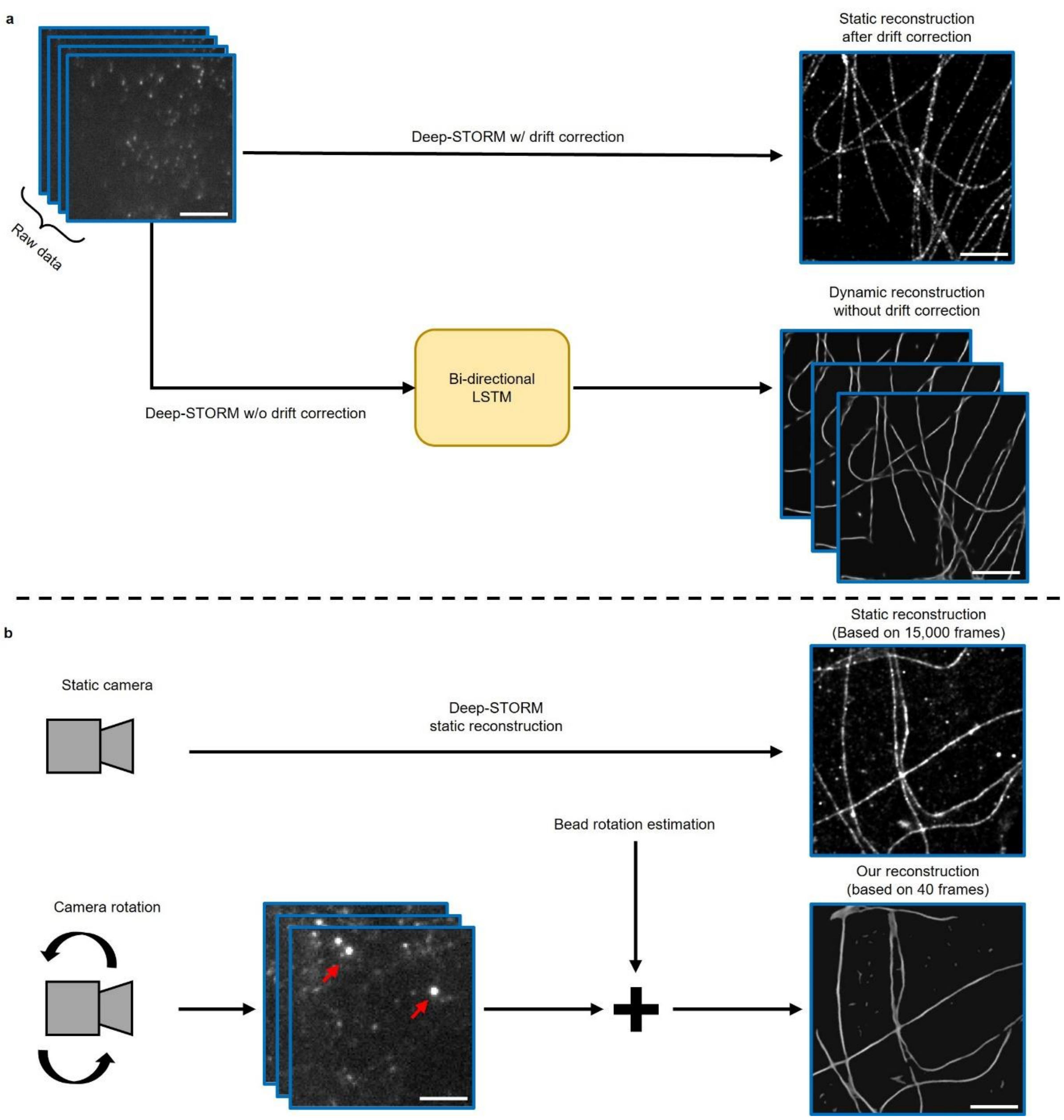

Figure 3: Static structure reconstruction during global motion. **a** STORM experiment containing undesirable drift was captured over 10,000 frames. We used Deep-STORM[12] to obtain a super-resolution reconstruction of the microtubule structure. Then, we used Deep-STORM's drift correction tool to predict the drift over the course of the experiment and acquire a single reconstructed frame in super resolution. The same localizations, without drift correction, are analyzed using DBlink. Scale bar = 2.5 μm. **b** A STORM movie of static microtubules is captured while manually rotating the camera. Fluorescent beads (red arrows) serve as fiducial markers reporting on the rotation. At some point, rotation is stopped, and a static STORM video is captured for 15,000 more frames. This video is used to produce a ground truth static structure via Deep-STORM. The static structure is then compared to each frame in the dynamic reconstructed video, rotated appropriately. Scale bar = 2.5 μm.

For the validation of DBlink on experimental data that contained dynamic morphological changes, and for which we could also possess ground-truth information at super-resolution, we reconstructed dynamic motion of dynein motors moving on static microtubules[29] (Supplementary Video S4). The static microtubule reconstruction served as the structural ground truth in super-resolution for the traces that the dynein motors could track. To simulate moving filaments, instead of single molecules, we summed the localization of dynein motors over short spatiotemporal windows. For independent reconstruction validation, we used a different algorithm for the static localization (ThunderSTORM) and for the dynamic DBlink localization-inputs (Deep-STORM). DBlink reconstruction of the dynein dynamics agreed with the static reconstruction of ThunderSTORM spatially (Fig. 4 and Supplementary Video S5). Quantitatively the hallucination percentage of DBlink reconstruction relative to ThunderSTORM's reconstruction was 0.1 % (see reconstruction accuracy quantification section in the SI). To validate the temporal aspect of the reconstruction, we marked the edges of the temporal window containing the input localizations for DBlink.

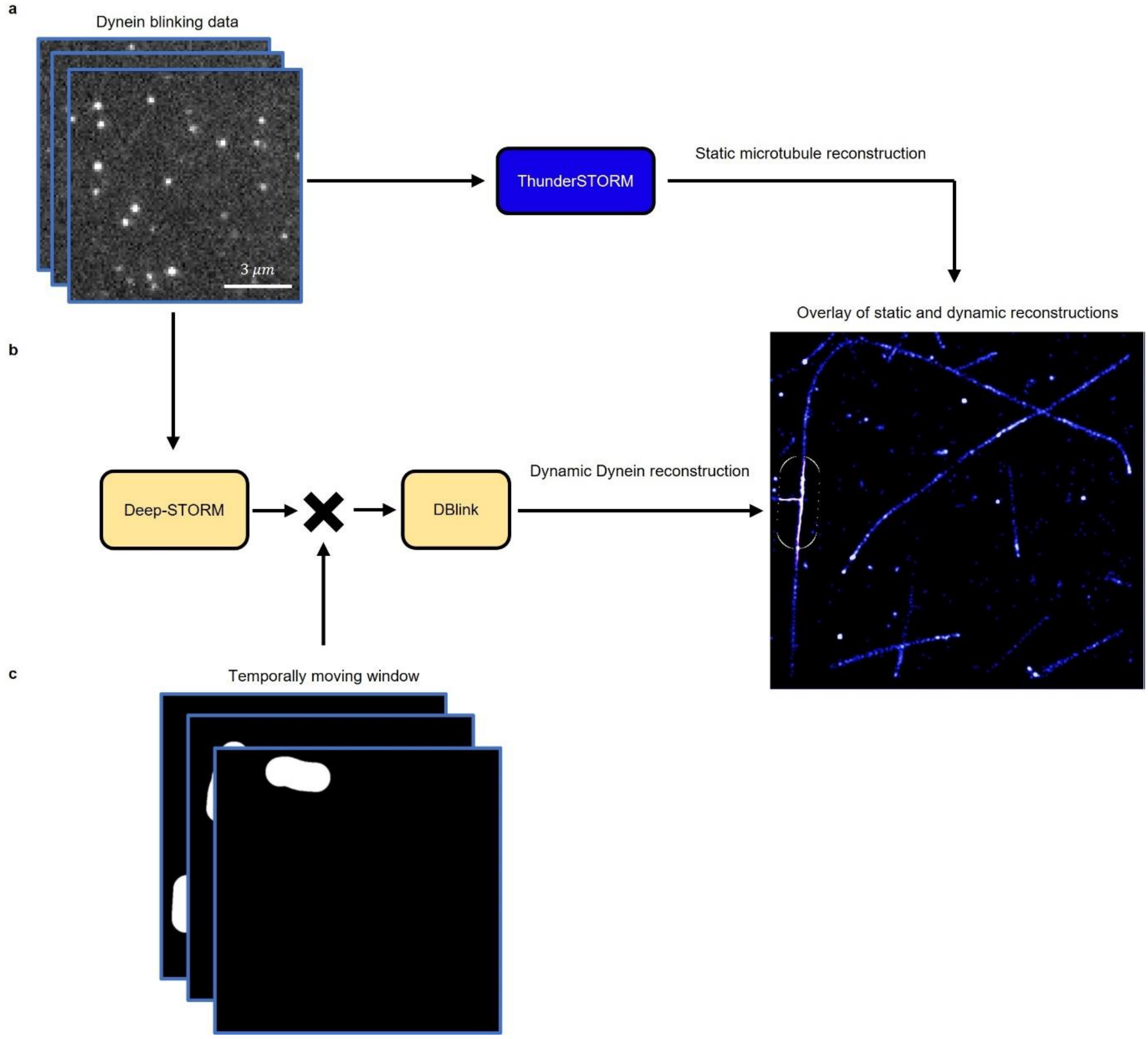
a
Dynein blinking data
3 μm
ThunderSTORM
Static microtubule reconstruction
Overlay of static and dynamic reconstructions
b
Deep-STORM
DBlink
Dynamic Dynein reconstruction
c
Temporally moving window

Figure 4: Tracking dynein motors dynamically moving on microtubules. **a** Fluorescently labeled microtubules are scattered in the field-of-view. Dynein motors labeled with HALO Alexa 488 are moving on the microtubules. The blinking data of the entire experiment is localized using ThunderSTORM to generate the ground truth structure in super-resolution. **b** Then, dynein motor localization is performed using Deep-STORM; the localization maps are multiplied by a temporally changing window to simulate the movement of a short filament-like structure. The localizations within the temporally moving window serve as the input to DBlink, that provides as output the reconstruction of the dynamic simulated filaments at super spatio-temporal resolution. **c** Overlay of the microtubule static reconstruction (blue) and the dynein motor dynamic reconstruction (white). The edges of the temporally moving window are marked in yellow. Temporal resolution = 50 s (25 frames with intervals of 2 seconds).

Additional quantification of performance as a function of deviations between training data and testing data are reported in the additional performance quantification section in the SI.

Next, we tracked microtubule and endoplasmic reticulum dynamics in live cells. Because ground truth information is not available in this case, we compared our network reconstructions to four alternative solutions: a) Deep-STORM reconstructions based on short temporal windows (Supplementary Video S6); b) Deep-STORM reconstructions combined with a previously reported blind inpainting algorithm[9]; c) DECODE[14] reconstruction algorithm (Supplementary Video S7); d) eSRRF[21] reconstruction algorithm (Supplementary Video S8). The input to DBlink was the sum of localizations over windows of 40 frames for the ER experiment and 20 frames for the microtubules experiment.

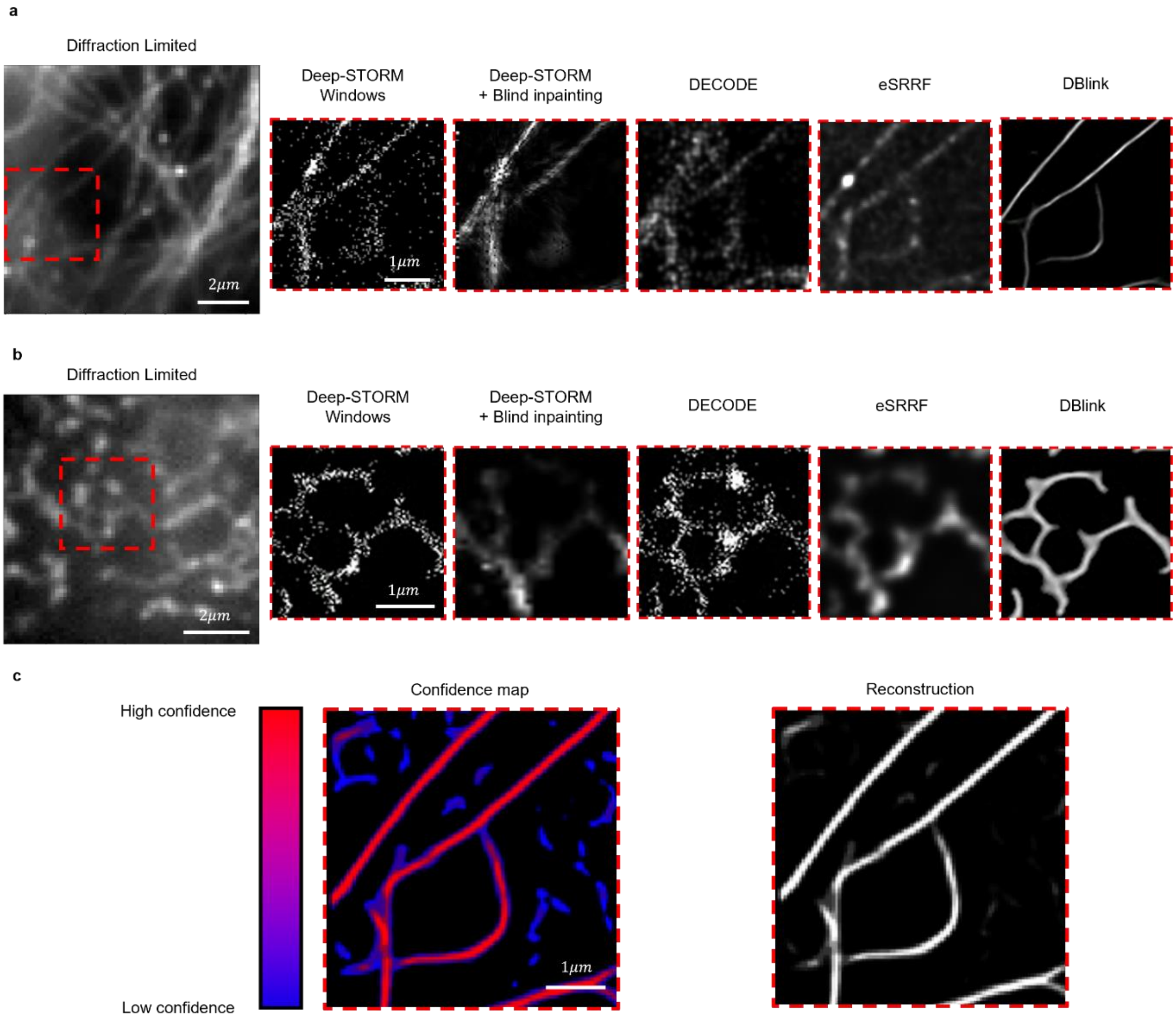

Figure 5: Qualitative comparison of DBlink to other state-of-the-art methods. **a** Left to right: A single diffraction limited frame of microtubules in live-cell experiment (ROI marked in red dashed rectangle); Deep-STORM reconstruction based on 300 frames; application of blind inpainting algorithm on Deep-STORM temporal-window; DECODE reconstruction based on 500 frames; eSRRF reconstruction based on 100 frames; DBlink reconstruction at temporal resolution of a single blinking frame. **b** Left to right: A single diffraction limited frame of ER in live-cell experiment (ROI marked in red dashed rectangle); Deep-STORM reconstruction based on 300 frames; application of blind inpainting on temporally-windowed Deep-STORM; DECODE reconstruction based on 300 frames; eSRRF reconstruction based on 250 frames; DBlink reconstruction at temporal resolution of a single blinking frame. **c** Left: the confidence map of a reconstructed ROI; blue colored pixel intensities are increased for better visualization. DBlink has higher confidence in red colored pixels than in blue colored pixels. Right: the actual reconstructed image.

Notably, the reconstructed video is at the same temporal resolution as the input video, which is achieved by using overlapping windows with one frame shifts. While this comes with a price of longer inference time ranging from few minutes up to few hours, the result is a super-resolved video at the temporal resolution of a single blinking frame. Importantly, the same exact network was used to recover both the ER data and the microtubule data, with no retraining; this serves as a demonstration of the generalizability of DBlink beyond its training data. The decision on the window sizes for each reconstruction algorithm is described in the window size optimization section in the SI.

Qualitatively, DBlink reconstructions consistently outperformed the other methods. Although blind inpainting has managed to filter most of the noise in Deep-STORM data, it performed poorly in densely labeled areas. Furthermore, rapid dynamics caused motion blur in temporally-windowed Deep-STORM and DECODE reconstructions (Fig. 5), while DBlink provided a more stable reconstruction in areas exhibiting rapid motion. Moreover, eSRRF has managed to overcome motion blur and provide a stable reconstruction; nonetheless, achieving the same temporal resolution as DBlink with eSRRF entails losing some structural information in the reconstruction (Supplementary Video S8). Notably, the temporal resolution achieved by DBlink in these experiments was 15 ms (66.6 frames per second), compared to eSRRF temporal resolution of 1.5 s and 3.75 s (0.66 and 0.26 frames per second) for the microtubule and the ER experiments respectively. The spatial resolution of DBlink reconstructions measured by decorrelation analysis was 30 nm.

In order to enable the evaluation of the quality of the reconstruction, we supply DBlink's confidence measure as an additional explainable output (Fig. 5c). The confidence map highlights areas in which the network reconstructions are more likely to miss the correct structure. To validate the usefulness of the confidence map, we measured the hallucination percentage in low confidence areas in comparison to high confidence areas. For more details see confidence hallucination calculation section in the SI. While the hallucination level is very low in both cases, the mean hallucination percentage was 1.5% in low confidence areas and 0.1% in high confidence areas, indicating that high confidence regions indeed correlate with fewer errors in the reconstructed frame.

Next, we reconstructed the dynamics of mitochondria in live cells from high-density single-molecule data recorded using cells labelled for the mitochondrial protein COX8. We extended the observation time in live cell imaging by using a HaloTag7 fusion in combination with a non-covalent, weak affinity fluorophore tag that binds to and unbinds from the target and acts as an exchangeable fluorophore label[30] (Supplementary Videos S9, S10). In this case, training required a model for mitochondrion size and shape, labeling density, motion type and speed, etc. For this purpose, we developed a dynamic simplified-mitochondria simulator (see mitochondria simulation section in the SI). After training the neural network, we analyzed a SMLM video of live cell imaging data of mitochondria labeled with HaloTag7 (see methods section) and followed the structural dynamics (Fig. 6). We could clearly visualize morphological changes of mitochondria, including fusion, fission, and drift, at different velocities. The temporal resolution we achieved in this experiment is 20 frames per second and the spatial resolution is 75 nm, determined from decorrelation analysis[28].

DBlink managed to detect mitochondrial dynamics with high fidelity to the observed data (Fig. 6), despite the fact that the training data contained simple pixelated-structures and motions (Supplementary Video S11). This validates the generalizability and applicability of our network for the analysis of various biological samples, contingent on appropriate training. The structural changes observed were similar to previously published work on mitochondrial dynamics[31], extended by a faster temporal resolution, longer observation time, and additional sub-diffraction structural information. Previous work on live-cell STED microscopy of mitochondrial dynamics has shown the potential of long-term observations of mitochondria, yet at the expense of extended time periods for fluorophore recovery that limits the temporal resolution[32]. DBlink in combination with exchangeable fluorophores increases the temporal resolution while not requiring time for signal recovery. This opens the door to study the relationship between ultra-structural organization and dynamics of mitochondria, which are key processes in the life cycle of cells and are tightly regulated in health and disease[33,34]. So far, many of the underlying mechanisms remain elusive, because of their inherent heterogeneity, and because of a limited spatiotemporal resolution[35].

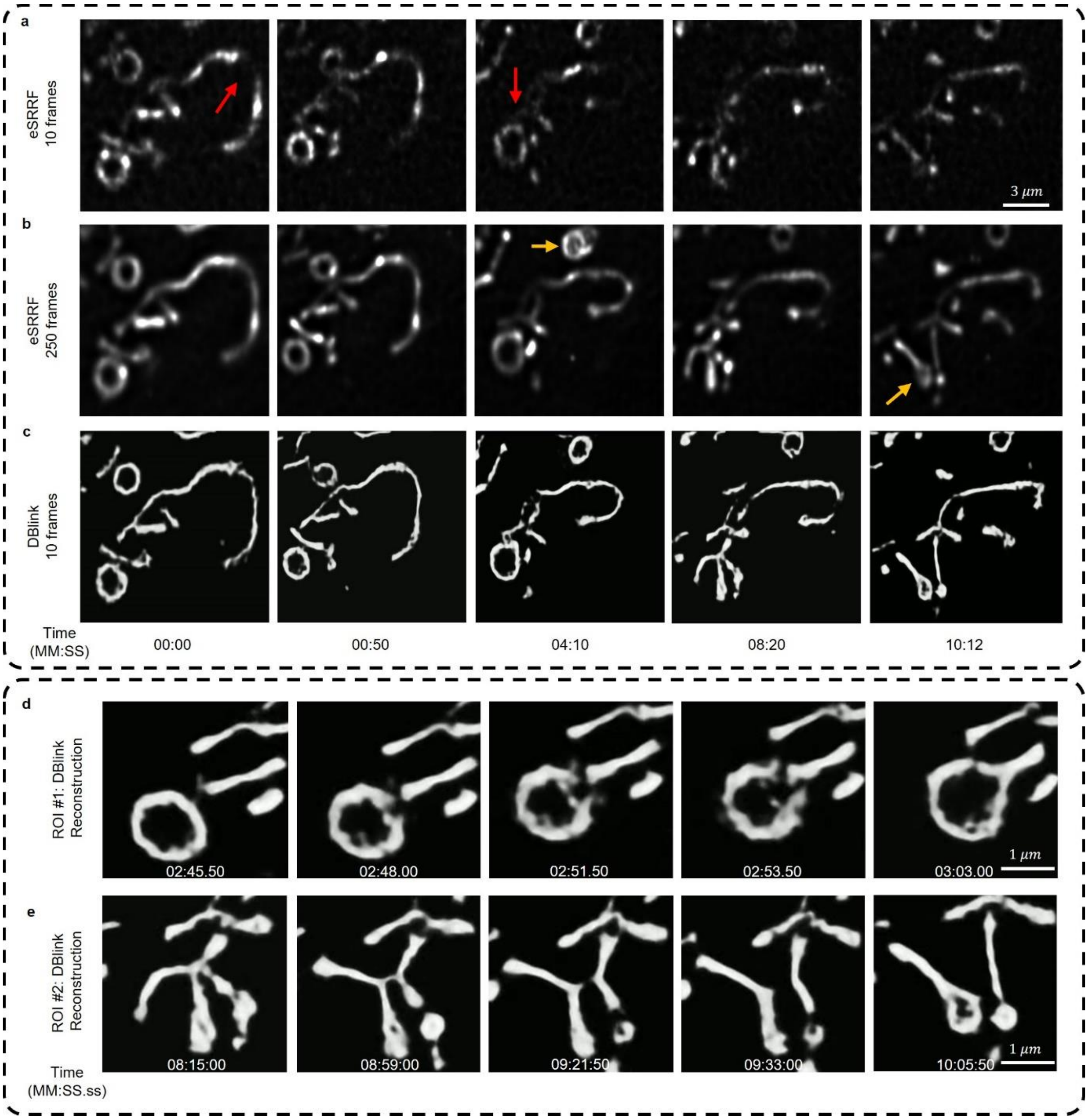

Figure 6: Reconstruction of 12.5 minutes long video of mitochondria dynamics in a live cell. **a** eSRRF[21] reconstruction of mitochondrial dynamics using temporal windows of 10 frames. **b** eSRRF reconstruction using temporal windows of 250 frames. **c** DBlink reconstruction at the temporal resolution of the blinking data, namely, 20 frames per second. Each column represents a different timepoint. Yellow arrow marks the rapid formation of a circular mitochondria structure that could not be clearly observed by analyzing long temporal windows, due to motion blur. Red arrows mark missing structure not resolved by eSRRF due to the short analyzed temporal window. Scale bar = 3 μm. **d** ROI #1: DBlink reconstruction enables tracking of mitochondrial fusion and fission dynamics in high spatiotemporal resolution. Scale bar = 1 μm. **e** ROI #2: DBlink reconstruction containing previously mitochondria thinning prior to a fission event[36,37]. Scale bar = 1 μm.

Finally, we compared our reconstructed video to eSRRF. The reconstructions of DBlink and eSRRF agreed spatially in areas containing slow dynamics; however, eSRRF reconstruction presented motion blur due to the fast dynamics even using the optimal temporal window (250 frames, based on ImageJ plugin the parameter sweep in ImageJ plugin).

Discussion

In this paper, we present a method for super spatiotemporal resolution reconstruction of dynamic SMLM data. Our solution utilizes two main assumptions: 1) the imaged-sample class is known (e.g. filaments, mitochondria), and 2) dynamically varying objects maintain some degree of structural similarity over time, allowing the network to exploit this information. In other words – the information used by our network to recover a SMLM video with a temporal window of 20 frames is not contained in these 20 frames alone – but rather also in a window of hundreds of frames around it. Notably, both assumptions are necessary to achieve the spatiotemporal resolutions demonstrated in this work and must hold for accurate reconstructions.

To overcome the challenge of verifying the network results, we tested several cases in which estimation of ground truth position was possible, including numerical simulation, whole-sample motion, i.e. sample-drift and camera rotation, and motion of dynein motors on static microtubules. In simulations, the network could reconstruct nanoscale rapid movement of simulated filaments with 90% of the structure correctly classified per video, assuming certain SMLM conditions (e.g. ~20 nm localization precision, ~1 emitter per μm^2); naturally, prediction accuracy varies as a function of fluorophore density, motion speed and other experimental parameters.

Ultimately, a main goal of our method is to enable live SMLM. For this purpose, we have analyzed SMLM videos of live-cell microtubule dynamics, provided by R. Tachibana, et al.[38] as well as videos of ER dynamics previously analyzed by DECODE[14]. Since no ground truth structure is available in such an experiment, we have qualitatively assessed the reconstruction accuracy by comparison to other state-of-the-art solutions for dynamic SMLM. While state-of-the-art methods suffered from loss of structural information due to the analysis of short temporal windows, DBlink reconstructions presented a more complete description of the entire structure.

As in all model-based neural-net reconstruction algorithms, the network's ability to generalize will always be limited by the training data, and caution should be exercised when applying the method; specifically, training data must resemble the experimental structure to avoid hallucinations[39–41]. To guide DBlink users in choosing the appropriate SMLM training data and in the interpretation of the reconstructions we have reported multiple ablation studies, the confidence map, and further discussion in the additional performance quantification section and the confidence hallucination calculation in the SI. Moreover, to validate the applicability of DBlink to structures with higher structural complexity than filaments, we tracked mitochondrial dynamics. First, we trained the neural network on simulated mitochondria-like structures, drifting and wobbling in time. Then, we used weak-affinity, non-covalent fluorophore labels that allow extended observation time[42]. The combination of these dyes with our high spatiotemporal reconstruction enables tracking dynamics in live cells at high spatiotemporal resolution in SMLM imaging over long observation times.

Future work can include extensions to other structures in live cells, expanding the types of motion in the training data to simulate elongation, contraction, wobbling and more complex dynamics, and systematic parameter optimization, e.g. optimal sample densities, sample-motion-rate to acquisition-rate ratio, and more. Additionally, although in this work we used the same network architecture to reconstruct different types of samples, this is not always the optimal solution. For example, in case that the imaged sample is relatively large, increased receptive fields in the convolutional layers might be beneficial. Moreover, reducing the inference time of our network would also be very useful; novel self-attention-based neural network architectures such as Transformers[43,44] might be exploited for this task. Also, expanding the ability of the network to analyze 3D information is also a desirable extension for this work. This would require good 4D (3D+time) models of different biological structures to generate the training data.

Finally, our LSTM-based framework is versatile in terms of its input data; here, we have used single-molecule localization maps as inputs; however, other inputs may provide even better performance. For example, DBlink can be used in combination with eSRRF, receiving as input radial gradient maps, which is suitable for analysis of densely labeled samples and may further improve performance. Our neural network based spatiotemporal interpolation could enable higher quality observation and ultimately facilitate discovery in various applications including cellular dynamics[45], colocalization of nanoparticles with organelles[46,47], synthetic materials[48], and more.

Acknowledgements

Live cell data of dynamic microtubules was generously shared with us by Yasuteru Urano and Mako Kamiya (The University of Tokyo). The data containing dynein moving on microtubule filaments was generously shared with us by Stefan Niekamp and Ronald Vale (University of California, San Francisco). We thank Ricardo Henriques (Instituto Gulbenkian de Ciência, Portugal and UCL Honorary Chair of Computational and Optical Biophysics) for his valuable input. We thank Julian Kompa and Kai Johnsson (MPI Medical Research, Heidelberg, and EPFL) for kindly providing the plasmid COX8A-HaloTag7. M.H. and S.J. gratefully acknowledge funding by LOEWE (FCI) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB1177; INST 161/926-1 FUGG. A.S. has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 802567 -ERC- Five-Dimensional Localization Microscopy for Sub-Cellular Dynamics. Y.S. is supported by the Zuckerman Foundation.

Author Contributions Statement

A. S. and Y. S. devised the main manuscript concept, developed the neural network architecture, planned the manuscript experiments, and wrote the paper with the help of all contributing authors. O. A. has prepared microtubule samples for static STORM imaging. N. O. contributed to the design and execution of the static STORM experiments. S. J. and M. H. performed dynamic super-resolution imaging in living cells.

Competing Interests Statement

The authors declare no competing interests.

References

1. Hell, S. W. & Wichmann, J. Breaking the Diffraction Resolution Limit By Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt Lett* **19**, 780–782 (1994).

2. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J Microsc* **198**, 82–87 (2000).

3. Sauer, M. & Heilemann, M. Single-Molecule Localization Microscopy in Eukaryotes. *Chem Rev* **117**, 7478–7509 (2017).

4. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science (1979)* **313**, 1642–1645 (2006).

5. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793–795 (2006).

6. Sharonov, A. & Hochstrasser, R. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Sciences* **103**, 18911–18916 (2006).

7. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat Protoc* **12**, 1198–1228 (2017).

8. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat Biotechnol* **36**, 460–468 (2018).

9. Wang, Y. *et al.* Blind sparse inpainting reveals cytoskeletal filaments with sub-Nyquist localization. *Optica* **4**, 1277–1284 (2017).

10. Barentine, A. E. S. *et al.* An Integrated Platform for High-Throughput Nanoscopy. *bioRxiv* 606954 (2022).

11. Liu, Z., Lavis, L. D. & Betzig, E. Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. *Mol Cell* **58**, 644–659 (2015).

12. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* (2018).

13. Nehme, E. *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nat Methods* (2020) doi:10.1038/s41592-020-0853-5.

14. Speiser, A. *et al.* Deep learning enables fast and dense single-molecule localization with high accuracy. *Nat Methods* **18**, 1082–1090 (2021).

15. Nixon-Abell, J. *et al.* Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER. *Science (1979)* **354**, aaf3928 (2016).

16. Priessner, M. *et al.* Content-aware frame interpolation (CAFI): Deep Learning-based temporal super-resolution for fast bioimaging. *bioRxiv* 2021.11.02.466664 (2021).

17. Chen, R. *et al.* Deep-Learning Super-Resolution Microscopy Reveals Nanometer-Scale Intracellular Dynamics at the Millisecond Temporal Resolution. *bioRxiv* 2021.10.08.463746 (2021).

18. Dertinger, T., Colyer, R., Iyer, G., Weiss, S. & Enderlein, J. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences* **106**, 22287–22292 (2009).

19. Gustafsson, N. *et al.* Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nat Commun* **7**, 12471 (2016).

20. Agarwal, K. & Macháň, R. Multiple signal classification algorithm for super-resolution fluorescence microscopy. *Nat Commun* **7**, 13752 (2016).

21. Laine, R. F. *et al.* High-fidelity 3D live-cell nanoscopy through data-driven enhanced super-resolution radial fluctuation. *bioRxiv* 2022.04.07.487490 (2022).

22. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: A comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).

23. Valli, J. *et al.* Seeing beyond the limit: A guide to choosing the right super-resolution microscopy technique. *Journal of Biological Chemistry* **297**, (2021).

24. Su, Y.-T., Lu, Y., Chen, M. & Liu, A.-A. Spatiotemporal Joint Mitosis Detection Using CNN-LSTM Network in Time-Lapse Phase Contrast Microscopy Images. *IEEE Access* **5**, 18033–18041 (2017).

25. Yu, Y., Si, X., Hu, C. & Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput* **31**, 1235–1270 (2019).

26. Shariff, A., Murphy, R. F. & Rohde, G. K. A generative model of microtubule distributions, and indirect estimation of its parameters from fluorescence microscopy images. *Cytometry Part A* **77**, 457–466 (2010).

27. Banterle, N., Bui, K. H., Lemke, E. A. & Beck, M. Fourier ring correlation as a resolution criterion for super-resolution microscopy. *J Struct Biol* **183**, 363–367 (2013).

28. Descloux, A., Grußmayer, K. S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nat Methods* **16**, 918–924 (2019).

29. Niekamp, S., Coudray, N., Zhang, N., Vale, R. D. & Bhabha, G. Coupling of ATPase activity, microtubule binding, and mechanics in the dynein motor domain. *EMBO J* **38**, e101414 (2019).

30. Kompa, J. *et al.* Exchangeable HaloTag Ligands (xHTLs) for multi-modal super-resolution fluorescence microscopy. *bioRxiv* 2022.06.20.496706 (2022).

31. Lefebvre, A. E. Y. T., Ma, D., Kessenbrock, K., Lawson, D. A. & Digman, M. A. Automated segmentation and tracking of mitochondria in live-cell time-lapse images. *Nat Methods* **18**, 1091–1102 (2021).

32. Yang, X. *et al.* Mitochondrial dynamics quantitatively revealed by STED nanoscopy with an enhanced squaraine variant probe. *Nat Commun* **11**, 3699 (2020).

33. Friedman, J. R. & Nunnari, J. Mitochondrial form and function. *Nature* **505**, 335–343 (2014).

34. Quiles, J. M. & Gustafsson, Å. B. The role of mitochondrial fission in cardiovascular health and disease. *Nat Rev Cardiol* **19**, 723–736 (2022).

35. Kleele, T. *et al.* Distinct fission signatures predict mitochondrial degradation or biogenesis. *Nature* **593**, 435–439 (2021).

36. Mahecic, D. *et al.* Mitochondrial membrane tension governs fission. *Cell Rep* **35**, 108947 (2021).

37. Mahecic, D. *et al.* Event-driven acquisition for content-enriched microscopy. *Nat Methods* **19**, 1262–1267 (2022).

38. Tachibana, R. *et al.* Design of spontaneously blinking fluorophores for live-cell super-resolution imaging based on quantum-chemical calculations. *Chemical Communications* **56**, 13173–13176 (2020).

39. 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).

40. Matlock, A. & Tian, L. Physical model simulator-trained neural network for computational 3d phase imaging of multiple-scattering samples. *arXiv preprint arXiv:2103.15795* (2021).

41. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat Methods* (2019).

42. Spahn, C., Grimm, J. B., Lavis, L. D., Lampe, M. & Heilemann, M. Whole-Cell, 3D, and Multicolor STED Imaging with Exchangeable Fluorophores. *Nano Lett* **19**, 500–505 (2019).

43. Vaswani, A. *et al.* Attention is all you need. in *Advances in neural information processing systems* 5998–6008 (2017).

44. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

45. Wensel, T. G., Potter, V. L., Moye, A., Zhang, Z. & Robichaux, M. A. Structure and dynamics of photoreceptor sensory cilia. *Pflugers Arch* **473**, 1517–1537 (2021).

46. Guggenheim, E. J. *et al.* Comparison of confocal and super-resolution reflectance imaging of metal oxide nanoparticles. *PLoS One* **11**, e0159980 (2016).

47. van der Zwaag, D. *et al.* Super Resolution Imaging of Nanoparticles Cellular Uptake and Trafficking. *ACS Appl Mater Interfaces* **8**, 6391–6399 (2016).

48. Pujals, S., Feiner-Gracia, N., Delcanale, P., Voets, I. & Albertazzi, L. Super-resolution microscopy as a powerful tool to study complex synthetic materials. *Nat Rev Chem* **3**, 68–84 (2019).

Methods

Training scheme

The network architecture is described in detail in the neural network architecture section in the SI. To train the neural network we used 1000 pairs of localization videos and ground truth structure videos (see filament/ mitochondria simulation sections in the SI) as a training set, and 250 pairs as validation set. Our loss function is comprised of three main terms: (i) Mean Squared Error (MSE); (ii) consistency loss; (iii) total variation loss. The consistency loss is calculated by the sum of pixelwise distance between every two adjacent frames:

$$Consistency\ Loss = \sum_{i=1}^{Video\ Length} \sum_{x,y} frame_i[x,y] - frame_{i-1}[x,y] \qquad (1)$$

Additionally, we used the Adam optimizer (betas = [0.99, 0.999]) with reduce on plateau mechanism (patience = 3). The network was trained on a single Titan RTX GPU for approximately 3 days.

Dynamic data acquisition

Sample preparation and imaging of global motion

To prepare fixed cells for imaging, we cleaned cover glasses (#1.5H, 22X22 mm, Marienfeld) in an ultrasonic bath (DCG-120H, mrc) with 5% Decon90 at 60 °C for 30 min. Then we washed the cover glasses with water, incubated them in ethanol absolute for 30 min, and sterilized them with 70% filtered ethanol for 30 min. The cover glasses were then seeded with COS7 cells and grown for 24 h in a six-well plate using phenol-free Dulbecco's Modified Eagle's medium (Gibco) with 1 g/l D-Glucose (i.e., low glucose), supplemented with fetal bovine serum, penicillin–streptomycin and glutamine (Biological industries) at 37 °C and 5% CO_2. The Cells were fixed with 4%

paraformaldehyde and 0.2% glutaraldehyde in PBS (37 °C, pH 7.3) for 45 min, washed, and incubated in 0.3 M glycine/PBS solution for 15 min. The cover glasses were transferred into a clean six-well plate and incubated in a blocking solution for 2 h (10% goat serum, 3% BSA, 2.2% glycine, and 0.1% Triton-X in 1X PBS, filtered with 0.45-µm PVDF filter unit, Millex). The cells were then immunostained overnight with anti-α-tubulin-AF647 (ab190573, Abcam) and anti-β-tubulin-AF647 (ab235759, Abcam) diluted 1:250 in the blocking buffer. After staining, the samples were washed five times with PBS. To prevent detachment of the anti-tubulin antibodies, the sample was again treated with 4% paraformaldehyde and 0.2% glutaraldehyde in PBS (pH 7.3) for 10 min at room temperature, washed, and incubated in 0.3 M glycine/PBS solution for 10 min.

For super-resolution imaging, a PDMS chamber was attached to a glass coverslip holding fluorescently labeled COS7 cells. Blinking buffer, containing 100 mM β-mercaptoethylamine hydrochloride, 20% sodium lactate, and 3% OxyFluor (Sigma Aldrich) in 1X PBS and pH 8–8.5, was added, and a clean coverslip was placed on top while minimizing any residual air bubbles in the chamber. In the microtubule experiments the sample was illuminated with 640 nm laser (~ 300 $\frac{W}{cm^2}$), and 10,000 images were captured at 50 ms exposure. For the drift correction experiment we used Photometrics Prime 95B camera and for the camera rotation experiment we used an EMCCD (iXon, Andor) camera.

Cell Culture – mitochondria experiment

U2OS cells were cultured in T-75 flasks (Greiner) at 37°C and 5% CO_2 in Dulbecco`s Modified Eagle Medium (DMEM) / F-12 (Gibco, Thermo Fisher, USA) containing 10% (v/v) fetal bovine serum (FBS) (Corning, USA), 1% penicillin-streptomycin (w/v) (Gibco, ThermoFisher, USA) and 1% GlutaMAX (v/v) (Gibco, USA).

Cells were transiently transfected with the plasmid pCDNA5/FRT/TO- COX8A-HaloTag7. For this purpose, 2 x 10^4 U2OS cells were seeded on fibronectin-coated 8-well chamber (Sarstedt, Germany). After 24 h incubation (37°C, 5% CO_2), cells were transfected using Lipofectamine 3000 transfection reagent (Gibco, ThermoFisher, USA). Briefly, 0.31 µL Lipofectamine 3000 was diluted in 10.42 µL OptiMEM medium (Gibco, Thermo Fisher, USA), and 210 ng vector DNA was diluted in 10.42 µL OptiMEM medium with 0.42 µL P3000 reagent (Gibco, ThermoFisher, USA). Diluted DNA solution was added to Lipofectamine diluent in a 1:1 ratio and incubated for 20 min at RT. After adding the DNA-lipid complex, cells were further incubated for 16-24 h at 37°C and 5% CO_2.

Prior to imaging, the cells were washed with pre-warmed live cell imaging solution (LCIS, ThermoFisher) and temperature adjusted to avoid lateral and axial drift.

Dynein motors on microtubules imaging

The dynein data was provided by Stefan Niekamp and Ronald Vale, acquired as described previously[29].

Live cell imaging

Live cell data of microtubules was generously provided as reported previously in R. Tachibana et al[38]. Briefly, SMLM imaging was carried out using an inverted fluorescence microscope (Eclipse Ti-E; Nikon) with an oil-immersion objective (CFI Apo TIRF 100X Oil, NA 1.49; Nikon), and irradiation laser at wavelength of 561 nm (Sapphire 561 LP; Coherent). The microtubules were labeled by 4(5)-Halo-HMCR550 conjugated to HaloTag proteins. For more details on sample preparation see R. Tachibana et al[35]. For live-cell confocal imaging of U2OS expressing COX8-HaloTag7, the exchangeable HT ligand HSAm carrying the fluorophore JF_{635} was added to the live cell imaging solution (LCIS, ThermoFisher, USA) at a final concentration of 500 nM. After an incubation time of 10 minutes, confocal microscopy was carried out on a Leica SP8 (Leica, Germany) equipped with an oil immersion objective (HC PL APO CS2 63x, NA 1.4) and an 633nm HeNe laser. Fluorophores were excited with an intensity setting of 1% 633 nm and a pinhole diameter of 1 Airy Unit. 300 frames were acquired using the HyD detector with a gain of 100 and at a scan speed of 400 Hz in xyt acquisition mode. Leica LASX software was used for the microscope control and data acquisition.

The imaging setup and the sample preparations of the live cell endoplasmic reticulum data are described in detail online[14].

For live-cell SMLM imaging of U2OS expressing COX8-HaloTag7, the exchangeable HT ligand HSAm carrying the fluorophore JF_{635} was added to the pre-warmed live cell imaging solution (LCIS, ThermoFisher, USA) at a final concentration of 1 nM. After an incubation time of 10 minutes, imaging was carried out on a N-STORM microscope (Nikon, Japan) equipped with an oil immersion objective (Apo, 100x, NA 1.49) and an EMCCD camera (DU-897U-CS0-#BV, Andor Technology, Ireland). Fluorophores were excited with a collimated 647 nm laser beam at an intensity of 400 $\frac{W}{cm^2}$ (measured at the objective) at highly inclined and laminated optical sheet (HILO) mode. 20,000-60,000 consecutive frames were acquired at 50 Hz in active frame transfer mode with an EMCCD gain of 200, a pre amp gain of 3, readout mode of 17MHz and at an effective pixel size of 158 nm. NIS Elements (Nikon, Japan), LCControl (Agilent, USA), and Micro-Manager were used for the optical setup and the data acquisition.

Data availability

The datasets analyzed during the current study are available from the corresponding author upon request. The datasets generated during this study are available online:

https://doi.org/10.5281/zenodo.7023414.

Code availability

The code is available online at: https://github.com/alonsaguy/DBlink.