FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data

Junhong Min[1], Cédric Vonesch[2*], Hagai Kirshner[2*], Lina Carlini[3*], Nicolas Olivier[3,4], Seamus Holden[3], Suliana Manley[3], Jong Chul Ye[1], Michael Unser[2]

[1]*Department of Bio and Brain Engineering, KAIST, Daejeon, Republic of Korea*

[2]*Institute of Microengineering, EPFL, Switzerland*

[3]*Institute of the Physics of Biological Systems, EPFL, Switzerland*

[4]*Department of Physics, King's College London, UK*

** Authors with equal contribution*

Super resolution microscopy such as STORM and (F)PALM is now well known method for biological study at the nanometer scale. However, conventional imaging scheme based on sparse activation of photo-switchable fluorescent probes has inherently slow temporal resolution which is a serious limitation when investigating live-cell dynamics. Here, we present an algorithm for high-density super-resolution microscopy which combines a sparsity-promoting formulation with Taylor series approximation of the PSF. Our algorithm is designed to provide unbiased localization on continuous space and high recall rates for high density imaging, and to have orders-of-magnitude faster run time compared to previous high-density algorithms. We validated our algorithm on both simulated and experimental data, and demonstrated live-cell imaging with temporal resolution of 2.5 seconds by recovering fast ER dynamics.

Introduction

Single-molecule localization microscopy methods, such as STORM [1] and (F)PALM [2,3], utilize sparse activation of photo-switchable fluorescent probes in both temporal and spatial domains. Each activated probe can be assimilated to an ideal point source so that the acquired images consist of isolated replicates of the point spread function of the microscope (PSF). This allows one to achieve sub-pixel accuracy on the order of tens of nanometers for the estimated location of each probe [1,3–6]. In general, reconstruction of sub-cellular structures relies on numerous localized probes, and the required acquisition time of these methods is therefore relatively long, i.e. on the order of minutes. This is a serious limitation when investigating live-cell dynamics.

One possible approach for overcoming this limitation is high-density imaging [7]. By increasing the density of activated probes, shorter acquisition times for a single super-resolution image can be achieved. However this complicates the localization task due to overlapping PSFs. DAOSTORM [7], for example, fits multiple overlapping PSFs in an iterative manner by analyzing pixel clusters in the residual image. The position of the probe is determined by minimizing a least-squares criterion. CSSTORM [8] (Compressed sensing STORM) and

deconvolution-STORM (deconSTORM)[9] impose a sparsity prior on the distribution of probes. In CSSTORM algorithm, the localization task is formulated as a convex optimization problem and solved by means of linear programming, while deconSTORM uses a modified Lucy-Richardson deconvolution algorithm by exploiting temporal correlation of activated probes. In general, these sparsity promoting methods provide increased recall rates compared to multi-emitter fitting at the expense of higher computation complexity. In a different approach, super resolution optical fluctuations imaging (SOFI)[10] and 3B analysis[11] utilize stochastic photon-emission processes such as photo-bleaching and blinking to reconstruct high-density data. For example, 3B analysis based on realistic models of photo-bleaching and blinking processes reconstructs the high-density data using Bayesian approach. Current implementation of 3B analysis has relatively slow reconstruction time but it can be made faster by using computationally efficient Bayesian algorithm or parallel computing[12].

In addition, all of these sparsity-promoting methods are based on similar discrete formulations. They reconstruct a high resolution image on a pre-defined sub-pixel grid, e.g., with a pixel size of 20 nm. Such formulations, however, have three inherent limitations. First, discrete-domain formulations can account

for only pre-defined locations, not all possible probe locations over a continuum. Therefore, these methods need to extract the localization information from the high resolution image. A partial solution to this problem is to compute local centers of mass on the reconstructed image[8], but this tends to result in a biased estimation of the probe locations. Second, using a finer sub-pixel grid increases the computational load, especially with the linear-programming approach used in CSSTORM. The final issue is the underestimation of photon counts. Indeed, many sparse solvers are based on iterative thresholding, which often truncates signals toward zero, resulting in spatial bias of recovered locations of the probes.

To address these limitations of current high-density localization algorithms, we introduce a **FA**st **L**ocalization algorithm based on a **CON**tinuous-space formulation (**FALCON**) for high density super-resolution microscopy data. In particular, to obtain a grid-free reconstruction, our approach combines a sparsity-promoting formulation with a Taylor approximation of the PSF. It consists of three main stages: deconvolution with a sparsity prior, deconvolution with a spatial support, and continuous-domain refinement (**Fig. 1 and 2, and Supplementary Fig. 1, Supplementary note**). The first step generates a high-resolution sparse image similarly to CSSTORM. The second deconvolution step corrects underestimated

pixel values. The last stage uses a Taylor series approximation of the PSF for refining the position of each probe over a continuum. It also refines the photon count of each probe. During Taylor expansions of PSF with respect to spatial offset from the grid location, we only used linear terms to make the optimization procedure faster. Importantly, this approximation is not limited to Gaussian functions. The final output of the algorithm are positions and photon counts of the localized probes. All these steps are implemented in a computationally efficient way by utilizing ADMM (Alternating Direction Method of Multipliers) [13] for fast reconstruction of the high resolution sparse images, and by using alternating refinements .

Results

We validated the performance of the proposed algorithm using both simulated and real STORM data. First, we confirmed that the linear approximation of PSF provides a good approximation for a PSF that is shifted by a few nanometers from the sub-pixel grid. Specifically, the relative approximation error is less than 2% for displacement of up to 30 nm for a Gaussian function with a FWHM (Full Width Half Maximum of 360 nm (**Fig. 2b**). Next, we investigated the localization per-

formance of our algorithm for a single probe. Starting from a uniform probability distribution over an area of a single sub-pixel of 33.3 nm (**Fig. 3a-c**), we localized $50,000$ individual probes using two methods: center-of-mass on a deconvolved image[8] and our continuous refinement by Taylor series approximation. The former method resulted in estimations that were strongly biased towards the sub-pixel grid points, while our approach produced a much more uniform distribution of the estimated particle locations. Similar results were observed for probes that were uniformly distributed along a diagonal line (**Fig. 3d-f**). These simulations were performed using high photon-emission rates ($5,000$ photons on average, standard deviation of $2,000$).

We also compared our algorithm with DAOSTORM and CSSTORM over a wide range of imaging densities. We simulated two different photon-emission rates: high ($5,000$ photons on average, standard deviation of $2,000$) and low (350 photons on average, standard deviation of 70). The FWHM of the PSF was fixed at 385 nm. For the high photon-emission rates simulations (**Fig. 4a and b**), FALCON and CSSTORM achieved two times higher recall rates than DAOSTORM did for densities larger than 6 $\mu\mathrm{m}^{-2}$. In terms of localization accuracy, the proposed algorithm outperforms both DAOSTORM and CSSTORM in high imaging

densities ($4 - 9.3\ \mu\mathrm{m}^{-2}$), and our results is compatible with least-square fitting showing only 5% difference in the lowest imaging density. For the low photon-emission rates simulation (**Fig. 4c and d**), the improvement is even more distinct in terms of recall rate, and FALCON demonstrates better localization accuracy over the whole range of imaging densities ($0.05 - 6.5\ \mu\mathrm{m}^{-2}$).

The performance dependency on the size of PSF was also investigated. Fixing the imaging density to $6\ \mu\mathrm{m}^{-2}$, we investigated the performance of FALCON, DAOSTORM and CSSTORM for various FWHMs in the range of $250 - 430$ nm (**Fig. 5**). As the FWHM becomes smaller, the performance of the both algorithms improve, demonstrating comparable results. Nevertheless, as the FWHM becomes larger, FALCON performs better than CSSTORM, demonstrating better accuracy and higher recall rates with a notable improvement in the low photon-emission cases (**Fig. 5c and d**). DAOSTORM always shows minimum recall rates and the differences from the others are much clear in the broader PSF.

In order to quantify the performance of our algorithm on specific geometric structures, we simulated circular structures whose radii varied between 150 nm and 400 nm (**Fig. 6**). In every frame, 4 activated probes are randomly placed on the circle, resulting in overlapping PSFs even at the largest radius. We estimated

the radius of the reconstructed circle by averaging the distances between all localized probes from 300 simulated images and the center of the circle. The results of the simulation demonstrate that our algorithm performs better for extracting this geometrically relevant measure, over the entire range of simulations. In contrast, the values by CSSTORM were underestimated by up to 18 nm as a result of underestimated photons and shows spurious peaks as artifacts in the histogram (**Fig. 6b**) because of the discrete formulation as shown in (**Fig. 3**). These artifacts were not shown in FALCON and DAOSTORM. In terms of the estimation bias, DAOSTORM was comparable to FALCON, but it had smaller recall rates and larger standard deviation of distances (**Fig. 6c**).

Our ADMM implementation of the two deconvolution stages exhibits fast convergence, and the refinement stage is computationally efficient thanks to the closed-form expression of each update step. For example, our stand-alone Matlab implementation run on an Intel i7, 3.4 GHz CPU, and it took FALCON 5 minutes, and on an GPU (Nvidia GTX Titan) took FALCON 40 seconds to reconstruct 100 frames high-density ($6\ \mu m^{-2}$) STORM images of 100 μm^2 (100x100 pixels). Since run time of our method mainly depends on the raw camera size and oversampling factor M for the reconstruction grid, computation cost will be further

reduced by using coarser sub-pixel grid for the deconvolution steps. In FALCON, role of the deconvolution steps is mainly to estimate number of sources and initial locations and these initial values can be refined. We demonstrated that FALCON with coarse grid of 50nm provides compatible results at the small expense of recall rates in comparison with finer grid case but only using reduced number of variables to estimate in **Supplementary Fig. 5**.

We demonstrate the performance of FALCON on fixed super-resolution data. FALCON reconstructed alpha tubulin subunits of microtubules (MTs) labeled with Alexa 647 in Cos-7 cells, by using 500 frames imaged at 25 Hz (**Fig. 7a**). For the quantitative analysis, we measured resolutions of the reconstructed images by using Fourier ring correlation (FRC) criteria [14]. For every algorithm, each half of localized locations were binned into image of 5 nm pitch and calculated cross correlations between two images, resulting FRC resolution. FALCON shows resolution improvement up to 20% over the other high-density algorithms (**Fig. 7b**). In other words, FALCON can achieve the same spatial resolution using 100 fewer frames. The superiority of FALCON is especially visible at the intersection of MTs, FALCON offers an unbiased estimation of distances between MTs (**Fig. 7c**), showing fewer spurious and noisy peaks than other high den-

sity algorithms (**Fig. 7c**). This is in accordance to our results concerning the radii of simulated circles. FALCON also provides reliable distribution of photon counts which are well matched to results of another low-density MT data, acquired under identical preparation protocols[15], whereas DAOSTORM (green line) loses many molecules with photon counts $<$ 5,000 in comparison and CSSTORM renders many false positives molecules with photon counts $<$ 500 (**Supplementary Fig. 12**). Furthermore, the proposed method showed improvements even in low-density data; especially around regions of complex MTs structures or inserted fiducial beads, in comparison with Gaussian fitting method (**Supplementary Fig. 13**).

We also applied FALCON to reconstruct live, PALM data. In particular, the dynamics of the endoplasmic reticulum (ER) were captured in U2OS cells labeled with tdEos fused to reticulon-4; reconstructed movies were made, with each frame being a single PALM image acquired over 2.5 s (160 frames of raw data). Over the course of 20 s, we observed several changing features of the ER network. In particular, expanding and shrinking tubules were visualized along with the dynamic motion of tubule junction points (**Fig. 8f-j**, white arrow). We also captured distinct features of this organelle reshaping itself: the disappearance of

tubule junctions (**Fig. 8f-j**, dotted circle) and emergence of tubules (**Fig. 8f-j**, blue arrow) to form new junction points are just a few examples of this. Importantly, we measured an average tubule thickness of approximately 60 nm (**Fig. 8g**), which is in line with previously reported diameter measurements[16,17]. Furthermore, we compared super-resolution images reconstructed by DAOSTORM, CSSTORM and FALCON by increasing accumulation time scales ($1 - 3$s) in **Supplementary Fig. 14**. FALCON shows better reconstruction of complex ER structures than the others, and detected more molecules up to 50% in comparison with DAOSTORM and CSSTORM, which can improve temporal resolution.

Discussion

We present a fast localization algorithm for high-density data which incorporates sparsity promoting formulation with Taylor series approximation of PSF for continuous localization. The robustness of the proposed method has been extensively studied by simulated and experimental data in comparison with previous high-density algorithms based on multi-emitter fitting (DAOSTORM) or deconvolution using sparsity promoting priors (CSSTORM and deconSTORM). Localization performances have been quantitatively analyzed not only in terms of recall

rates and localization accuracy, but also localization bias. Especially, even though localization bias can be easily observed in case of high-density imaging and is important for quantitative analysis in biological study, it has not been properly investigated before[7–9].

We have shown that localization bias can be induced by several factors. Low recall rates is one factor, which can occur, by counting multiple closely-spaced probes as one, and localizing centroid position. In general, deconvolution based approach utilizing sparsity promoting prior can achieve higher recall rates than multi-emitter fitting based methods by effectively resolving closely placed probes. However, this approach can introduce additional bias for location estimation mainly originating from its use of the discrete sparsity-promoting formulation. Our experiments have illustrated grid-pattern artifacts and spatial distortions appearing in the final image reconstructed when using the discrete sparsity-promoting formulation. In particular, spatial distortions can be problematic when imaging intracellular structures. For instance, quantifying the sizes of mitochondria, microtubules or the tubules of the ER would be done with less confidence, since structures would be falsely constricted. On the other hand, the proposed method based on the continuous formulation minimized these biased errors and

has shown almost bias-free localization results.

In live imaging, the variance of the localization accuracy is composed of both the algorithmic localization variance and the variance of the fluctuations. Indeed, the motion artifacts due to dynamic structures that move on time-scales shorter than our exposure time can be problematic, often resulting in artificially larger or smaller structures- as previously reported [18]. In the ER experiment, we sought to probe the dynamic motion of ER tubules. Previous studies [19,20], with similar exposure times demonstrated that motion of this organelle is dynamic on longer timescales than the one used here.

Our method is currently using a two dimensional PSF model which is appropriate to a relatively flat sample region having thickness below 1 μm. However, it must be noted that our continuous sparsity-promoting formulation is not limited to two-dimensional Gaussian PSF model. It can be easily extended to non-Gaussian models including three dimensional PSF models. For example, multi-plane imaging system with three-dimensional PSF engineering techniques can be accommodated with our formulation.

The proposed formulation with Gaussian noise assumption can be fur-

ther improved by considering more realistic noise statistics. Since many state-of-the-art cameras for localization microscopy, such as an EMCCD (Electron multiplying charge-coupled device) or sCMOS(scientific complementary metal–oxide–semiconductor) camera, have very small read-out noise[21], Poissonian shot noise becomes a dominant noise source to resolve. Basically, exploiting Poisson statistic [22] would be more advantageous especially for live-imaging data, which is likely to be achieved at low SNR due to low emission photons and high background. However, even with current implementation of the Gaussian noise model, at the SNR levels from our experimental conditions, the proposed algorithm is robust under Poisson noise statistics.

Concerning modeling of readout noise for our simulations, we assumed that CCD-type camera was used, where readout noise is known to be well approximated by normal distribution. The well-established high-density algorithms such as DAOSTORM and CSSTORM also used Gaussian readout noise of small variance for their simulations[7,8]. Although sCMOS camera has readout noise with pixel-to-pixel variation resulting in non-Gaussian distribution such as log-normal distribution[23], our simulation results indicate that the proposed algorithm is not sensitive to the choice of the noise model (**Supplementary Fig. 9**).

In general, it is difficult to make a fair comparison between algorithms in terms of calculation speed due to the difference in implementation environment. Since the speed of the algorithm is determined by various factors including optimization, programming language, operating system and usage of parallel implementation, the implementation environment must also be considered wisely to maximize calculation efficiency. For example, a computationally optimized CSSTORM implementation[24] is faster by two-order than the initial MATLAB-based implementation[8] which takes approximately one day to reconstruct 100 μm^2 region of 100 STORM images with high-density activation($6\ \mu m^{-2}$) on the Intel i7, 3.4 GHz CPU.

Although the speed of our approach is also dependent on the implementation environment, a generally fast algorithm speed can be achieved since the algorithm starts from a coarser grid which inherently reduces the problem complexity. Specifically, the proposed method would be far beneficial for three dimensional localization problems with high-density data, where calculation time reduction is critical since discrete formulation requires numerous finer voxels to be estimated.

In conclusion, Taylor approximation of the PSF can be jointly utilized with sparsity-promoting formulation for high-density imaging to yield contin-

uous localization with reduced complexity up to several orders of computation.Furthermore,we have also investigated how spatial bias is introduced in the localization estimation of high-density imaging, and showed that the proposed continuous formulation with specially designed sparsity priors substantially reduces the bias and provides better localization accuracy and higher recall rates than other currently available methods over a wide variety of experimental conditions. Moreover, our method is efficient in terms of implementation, which reduces the amount of the computation time by orders of magnitude in comparison with the previous high density algorithms. These results were well confirmed through experimental data obtained using fixed microtubules sample. We also successfully reconstructed fast dynamic motions of ER and its reliable profiles.

Methods

Simulation & Evaluation

Photon-emission statistics We assume that the total number of photons emitted from a probe is a random variable that follows a log-normal distribution [8]. We considered two photons-emission rates: high and low. The former setting was primarily meant to simulate STORM data while the latter was closer to PALM

data. The log-normal parameters for the high photon-emission case are (mean, standard deviation) = $(5,000, 2,000)$, and for the low photon-emission case, they are $(350, 70)$. Constant background fluorescence photons of 100 or 10 are added to every camera pixel for high and low emission cases, respectively.

PSF model When reconstructing real STORM data, we used low-density STORM images for fitting the measured PSF. Our PSF model is the sum of two Gaussian functions[8]. Low-density data is used as a reference of the PSF model, in order to optimize their widths and weights. For our simulated data, we chose to have two Gaussian functions that have a standard deviation parametrization of σ and 2σ, and that are linearly combined with a ratio of $4:1$.

Noise statistics For simulating noise, we assumed that high performance CCD camera is used for data acquisition and considered two type of noise contribution, shot noise and small readout noise[21,23]. The shot noise follows Poisson statistics. Specifically, the mean of the detected photons is determined by the expected number of incoming photons and by the background fluorescence level at each pixel. We also added CCD-type readout noise which follows normal distribution with zero mean and variance of 2. In **Supplementary Fig. 9**, we also considered CMOS-type camera whose readout noise has intrinsic pixel to pixel variation,

resulting in non-Gaussian distribution. Specifically, CMOS-type readout noise follows log-normal distribution with unit mean and variance of 2.

Probe distribution and performance analysis For the experiment of **Fig. 4 and 5**, we randomly distributed N probes in the central region of the image; the size of the region is 100×100 pixels. The size of the image is slightly larger, 108×108 pixels. The camera pixel size is 100 nm. In **Fig. 4**, the imaging densities vary between $0.05\ \mu\mathrm{m}^{-2}$ and $9.3\ \mu\mathrm{m}^{-2}$ for the high photon-emission rates (**Fig. 4a and b**) and between $0.05\ \mu\mathrm{m}^{-2}$ and $6.5\ \mu\mathrm{m}^{-2}$ for the low photon-emission rates (**Fig. 4c and d**). The FWHM of the PSF was set to $385\ \mathrm{nm}$. We used 50 realizations for every density. **Fig. 5** depicts additional analysis with FWHM PSF between 250 nm and 430 nm. The imaging density is $6\ \mu\mathrm{m}^{-2}$ and we used 30 realizations for each FWHM.

Performance analysis was carried out by matching each localized probe with the closest ground-truth probe. Localization errors that are more than $300\ \mathrm{nm}$ were excluded from the analysis. The localization accuracy is expressed in terms of standard deviation or the FWHM of the histograms of localization errors. The recall rate is defined by the ratio between the number of matched probes and N.

Simulation of geometric structures We generated STORM images by distributing particles randomly on a circle of radius r. The radius varies between 150 nm and 400 nm. The probes are activated at different time instances, covering 300 frames. The number of probes at each frame is $\mathrm{N} = 4$ (**Fig. 6**). **Supplementary Fig. 4** depicts estimation results for various number of molecules N between $1-6$ with fixed $r = 200\,\mathrm{nm}$. For every localized probe, we calculated the distance from the probe to the center of the circle, and estimated the radius by averaging all distances. Then, we plot the differences between the true radii and the estimated values. Here, high photon-emission rates are used.

Image and Signal Processing We used Gaussian rendering to generate super-resolution images. For rendered image, every localized molecule is convolved with a Gaussian kernel, and then scaled by its estimated photon count. In **Fig. 8**, normalized Gaussian kernels were used without photon scaling. For the experimental STORM data of fixed microtubule (**Fig. 7**), stage drifts are corrected by using fiducial beads. To correct the drift of the sample in **Supplementary Fig. 13**, the mean vertical (resp. horizontal) position of a straight horizontal(resp. vertical) segment of microtubules was measured as a function of frame number, and the polynomial fit of this function was subtracted from the vertical (resp. horizontal) coordinates of all the peaks.

In order to determine the achievable temporal resolution of our live data, we used the Nyquist criteria ($\rho = (\frac{2}{R})^2$) as commonly applied to super-resolution imaging[25]. We counted the number of molecules on our structure of interest and obtained an estimate of its effective area by thresholding the super-resolution image; from this we had an estimate on our localization density. With a 2.5 s reconstructed super-resolution image, our localization density corresponds to R = 40 nm.

Cell culture COS-7 and U2OS cells were cultured in DMEM supplemented with 10% FBS (Sigma Aldrich) in a cell culture incubator (37°C and 5% CO2) and plated at low confluency on cleaned 25 mm size 1 cover-glass (Menzell).

Sample preparation for microtubule imaging in fixed cells Prior to fixation, all solutions were pre-warmed at 37°C: 24 hours after plating, Cos-7 cells were pre-extracted for 10 s in 0.5% Triton X-100 (Triton) in BRB80 (80 mM PIPES, 1 mM $MgCl_2$, 1 mM EGTA, adjusted to pH 6.8 with KOH) supplemented with 4 mM EGTA, washed in PBS, fixed for 10 min in - 20°C-Methanol (Sigma), and washed again in PBS. The samples were then blocked 30 minutes in 5% BSA, before being incubated for 1.5 hour at room temperature with 1:1000 mouse α-tubulin antibodies (Sigma, T5168) in PBS - 1% BSA - 0.2% Triton (PBST), followed by

3 washes with PBS-0.2% Triton, and then incubated for 45 min in PBST with 1:1000 goat anti-mouse Alexa-647 F(ab)2 secondary antibody fragments (Life Technologies, A-21237).

Sample preparation for live-cell imaging of the endoplasmic reticulum (ER) After letting cells grow to approximately 60% confluency, U2OS cells were transfected with 2 μg of the reticulon 4 (Rtn4)-tdEos plasmid using FuGENE 6 transfection reagent (Roche). Cells were washed 24 hours post transfection and imaged in Leibovitz.

Live and fixed super-resolution imaging Imaging of live and fixed was performed on a modified Olympus IX71 inverted microscope. In the case of Alexa-647 labeled α-tubulin, a laser at 641 nm (Coherent, CUBE 640-100C) was reflected by a multi-band dichroic (89100bs, Chroma) onto the back aperture of a 100x 1.3 NA oil objective (Olympus, UplanFL) mounted on a piezo objective scanner (P-725 PIFOC, Physicinstrument). The collected fluorescence from the sample was filtered by a band-pass emission filter (ET700/75, Chroma) and imaged onto an EMCCD camera (IxonEM+, Andor) with a 100 nm pixel size and using the conventional CCD amplifier at a frame rate of 25 fps. Laser intensity on the sample was $\approx$ 1-2 kW.cm^{-2}. A 405 nm laser was added to maintain a

high peak density. Blinking was induced according to[15]. Briefly, $30\ \mu$L of a mixture of 20% Vectashield, 80% Glycerol was pipeted on top of the sample, and a cover-glass was added on top of the buffer to spread it evenly on the sample. For live-cell imaging, transfected cells were identified using a 488 nm laser (Sapphire 488, Coherent) at an intensity < 1 kW cm^{-2} to prevent activation of tdEos. Subsequently, cells were imaged using a 561 nm laser (Sapphire 561, Coherent) with an intensity of approximately 3 kW cm^{-2}. A 405 nm laser was used to convert tdEos molecules from their green to red form and was left on throughout acquisition. Fluorescence was collected after passing through a ET605/70 (Chroma) emission filter. Imaging was performed at a frame rate of 64 fps.

References

1. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy STORM. *Nat. Methods* **3**, 793–796 (2006).

2. Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258 (2006).

3. Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J. & Hess, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).

4. Smith., C. S., Joseph., N., Rieger., B. & Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods* **7**, 373–375 (2010).

5. Henriques, R., Lelek, M., Fornasiero, E. F., Valtorta, F., Zimmer, C. & Mhlanga, M. M. Quickpalm: 3d real-time photoactivation nanoscopy image processing in imagej. *Nat. Methods* **7**, 339–340 (2010).

6. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).

7. Holden., S. J., Uphoff., S. & Kapanidis, A. N. DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nat. Methods* **8**, 279–280 (2011).

8. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster storm using compressed sensing. *Nat. Methods* **9**, 721–723 (2012).

9. Mukamel, E. A., Babcock, H. & Zhuang, X. Statistical deconvolution for superresolution fluorescence microscopy. *Biophys. J.* **102**, 2391–2400 (2012).

10. Dertinger, T., Colyer, R., Iyer, G., Weiss, S. & Enderlein, J. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22287 (2009).

11. Cox, S., Rosten, E., Monypenny, J., Jovanovic-Talisman, T., Burnette, D. T., Lippincott-Schwartz, J., Jones, G. E. & Heintzmann, R. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat. Methods* **9**, 195–200 (2011).

12. Hu, Y. S., Nan, X., Sengupta, P., Lippincott-Schwartz, J. & Cang, H. Accelerating 3b single-molecule super-resolution microscopy with cloud computing. *Nat. Methods* **10**, 96–97 (2013).

13. Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process* **19**, 2345–2356 (2010).

14. Nieuwenhuizen, R. P., Lidke, K. A., Bates, M., Puig, D. L., Grünwald, D., Stallinga, S. & Rieger, B. Measuring image resolution in optical nanoscopy. *Nat. Methods* **10**, 557–562 (2013).

15. Olivier, N., Keller, D., Rajan, V. S., Gönczy, P. & Manley, S. Simple buffers for 3d storm microscopy. *Biomed. Opt.* **4**, 885–899 (2013).

16. Hu, J., Prinz, W. A. & Rapoport, T. A. Weaving the web of er tubules. *Cell* **147**, 1226–1231 (2011).

17. Snapp, E. L. *ER biogenesis: proliferation and differentiation. The Biogenesis of Cellular Organelles* (New York: Landes Bioscience and KluwerAcademic/Plenum Publishers, 2004).

18. Frost, N. A., Lu, H. E. & Blanpied, T. A. Optimization of cell morphology measurement via single-molecule tracking palm. *PloS one* **7**, e36751 (2012).

19. Shim, S.-H. *et al.* Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13978–13983 (2012).

20. York, A. G., Chandris, P., Dalle Nogare, D., Head, J., Wawrzusin, P., Fischer, R. S., Chitnis, A. & Shroff, H. Instant super-resolution imaging in live cells and embryos via analog image processing. *Nat. Methods* **10**, 1122–1126 (2013).

21. Beier, H. T. & Ibey, B. L. Experimental comparison of the high-speed imaging performance of an em-ccd and scmos camera in a dynamic live-cell imaging test case. *PloS one* **9**, e84614 (2014).

22. Kim, K., Min, J., Carlini, L., Unser, M., Manley, S., Jeon, D. & Ye, J. C. Fast maximum likelihood high-density low-snr super-resolution localization microscopy. *10th International Conference on Sampling Theory and Applications* (2013).

23. Fowler, B., McGrath, D. & Bartkovjak, P. Read noise distribution modeling for cmos image sensors. *2013 INTERNATIONAL IMAGE SENSOR WORKSHOP* .

24. Babcock, H. P., Moffitt, J. R., Cao, Y. & Zhuang, X. Fast compressed sensing analysis for super-resolution imaging using l1-homotopy. *Opt. Express* **21**, 28583–28596 (2013).

25. Shroff, H., Galbraith, C. G., Galbraith, J. A. & Betzig, E. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nat. Methods* **5**, 417–423 (2008).

Acknowledgements This research was supported by the Korea Science and Engineer-

ing Foundation (KOSEF) grant funded by the Korea government (MEST) (No.2009-0081089)and the Swiss National Science foundation under grant (No.200020-144355)

Author Contributions M.U. supervised the project. M.U., J.M., C.V. and H.K. J.Y designed the algorithm. J.M. implemented the algorithm and ran the simulated experiments. J.M. C.V. and H.K designed and analysed the simulated data. N.O. prepared the sample of COS-7 cell and acquired STORM data of microtubule. L.C. prepared the sample of U2OS cell and J.M. and L.C acquired live PALM data of ER. J.M, L.C, N.O. and S.M. performed analysis on the experimental data of the microtubule and ER. S.H. performed the DAOSTORM analysis on simulated and experimental STORM data. J.M. generated the figures for all simulations, and J.M, L.C, N.O. generated the figures of the experimental data. All authors wrote the manuscript.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Michael Unser. (email: michael.unser@epfl.ch).

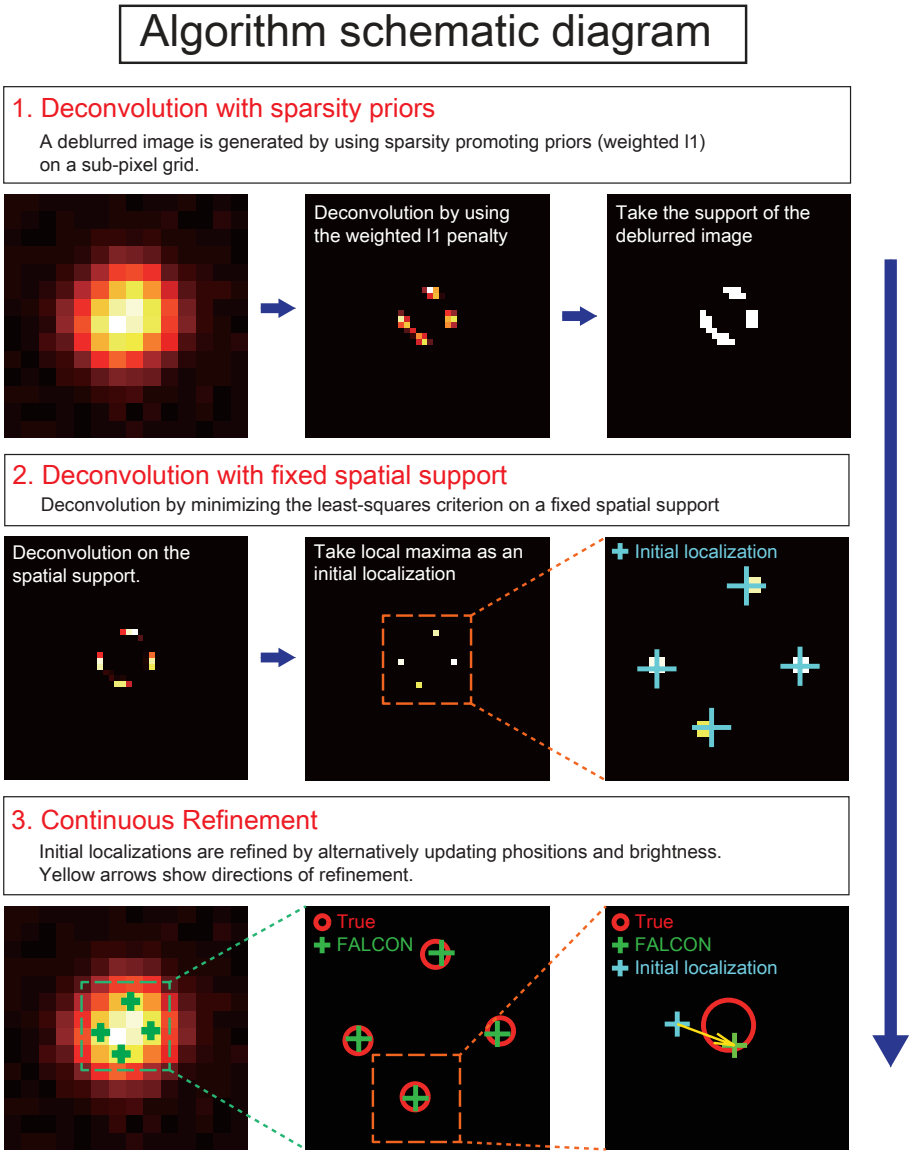

Figure 1 Schematic illustration for FALCON. For a detailed description, **see Supplementary note**.

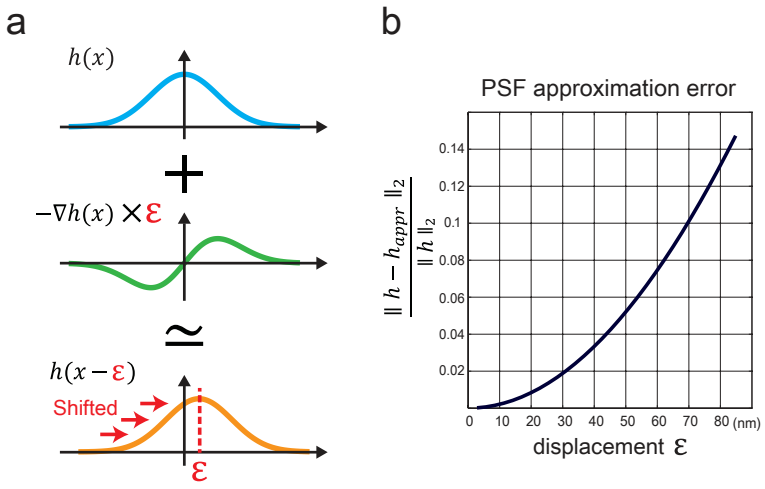

Figure 2 Taylor approximation of the PSF for continuous refinement of FALCON. (a) First-order Taylor approximation of the shifted PSF: the shifted PSF is approximated by the PSF centered on the sub-pixel and the derivative of the PSF. (b) Approximation error of Gaussian PSF of 360 nm FWHM. The error of the approximation does not exceed 2% up to 30 nm displacement.

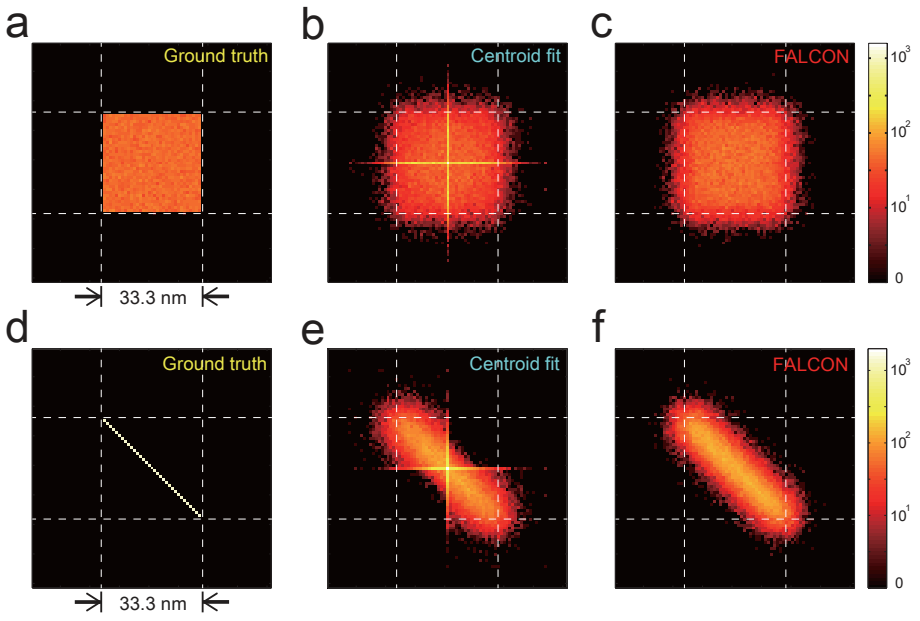

Figure 3 Numerical analysis of the continuous refinement on 50,000 simulated images. In each image, a single molecule is randomly placed with uniform probability distribution within a single sub-pixel area (a-c) or along a diagonal line (d-f). Histogram of ground-truth positions (a,d), histograms of the centroid fit that is applied to the sparsity-based deconvolved images (b,e) and histogram of FALCON (c,f).

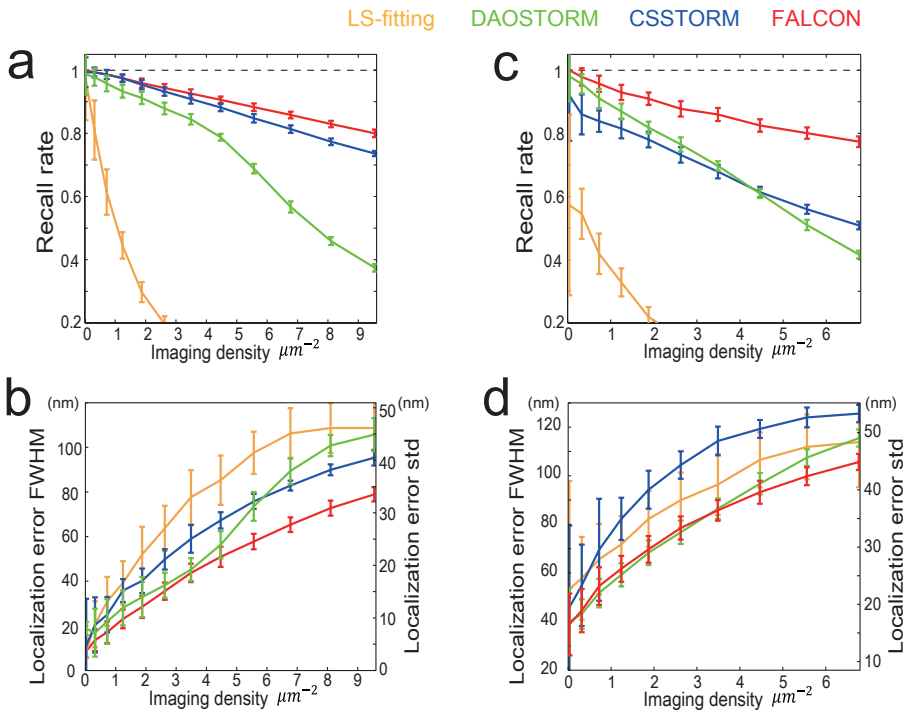

Figure 4 Performances of FALCON in comparison with Least-square fitting, DAOSTORM and CSSTORM wide range of imaging densities. Simulation on the random distribution of molecules over a wide range of imaging densities with high-photon emission rates: recall rates (a), localization accuracy (b) and low-photon emission rates: recall rates (c), localization accuracy (d). The error bars indicate standard deviations.

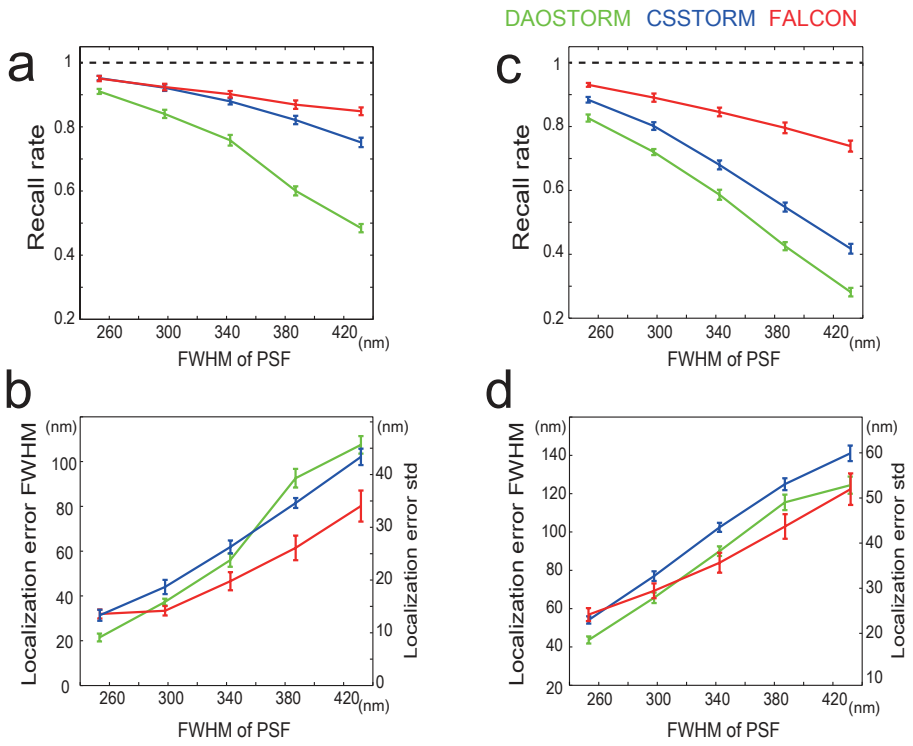

Figure 5 Performances of FALCON in comparison with DAOSTORM and CSSTORM over wide range of PSF widths from 250 nm to 430 nm. Simulated data with high-photon emission rates: (a) molecular recall rates, (b) localization accuracy and low-photon emission rates: (c) molecular recall rates, (d) localization accuracy. The error bars indicate standard deviations.

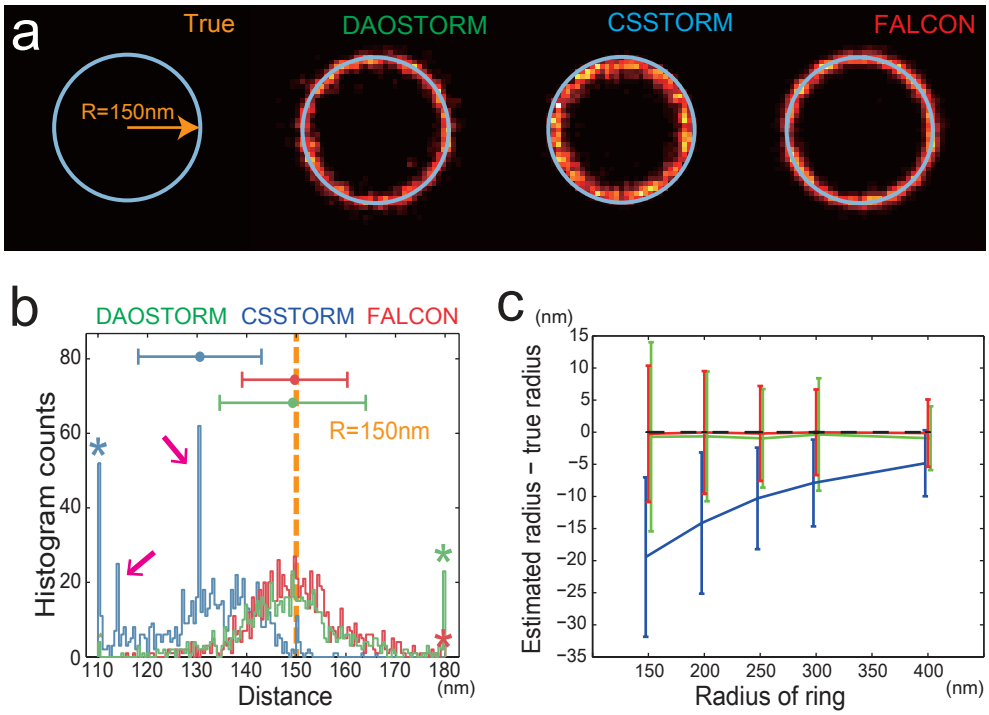

Figure 6 Performance analysis for synthetic "ring" phantoms with various radii. (a) Rendered images of the algorithms overlaid by a cyan circle with 150 nm radius as a ground truth. (b) Histogram of the distances from localized molecules by FALCON, DAOSTORM, and CSSTORM to center of the ring in (a). (c) The differences between the true radius and the distances on average along various radii from 150 nm to 400 nm with 4 fixed molecules on the circle in every simulated image. The error bars indicate standard deviations.

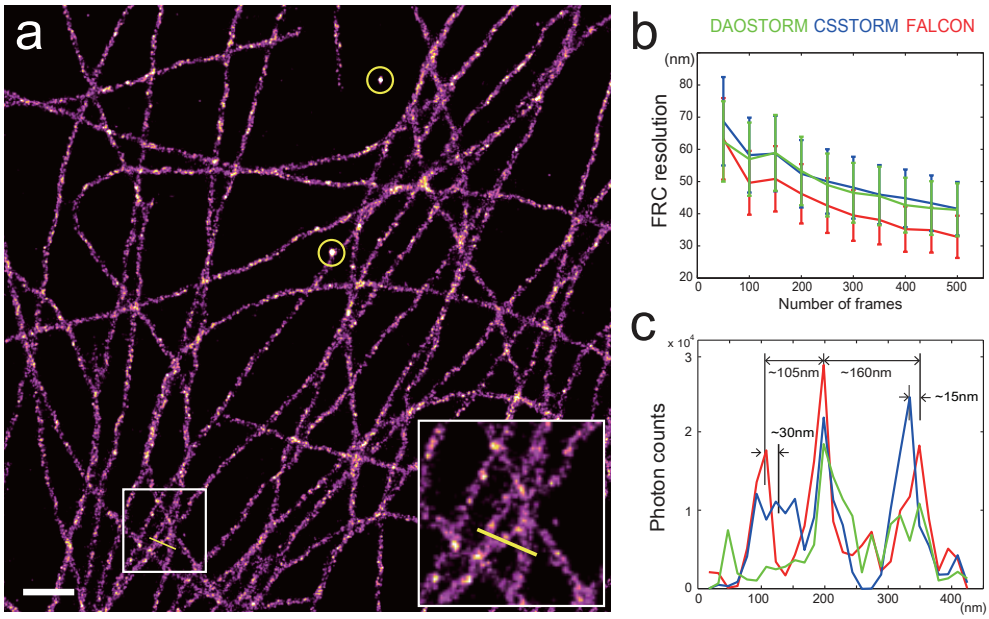

Figure 7 FALCON performance on fixed microtubules data. (a) FALCON reconstructed α-tubulin subunits of microtubules (MTs) labeled with Alexa 647 in Cos-7 cells; 500 raw frames were used to reconstruct an image. White box highlights the intersection of MTs reconstructed by FALCON. (b) Fourier ring correlation(FRC) analysis on the reconstructed images by DAOSTORM, CSSTORM and FALCON. For the FRC analysis, fiducial beads denoted by yellow circles are discarded. The error bars indicate standard deviations. Cross-sectional profiles across region in white box (yellow line) were measured by DAOSTORM, CSSTORM and FALCON, and plotted in (c). Scale bars are $1\ \mu m$ in (a)

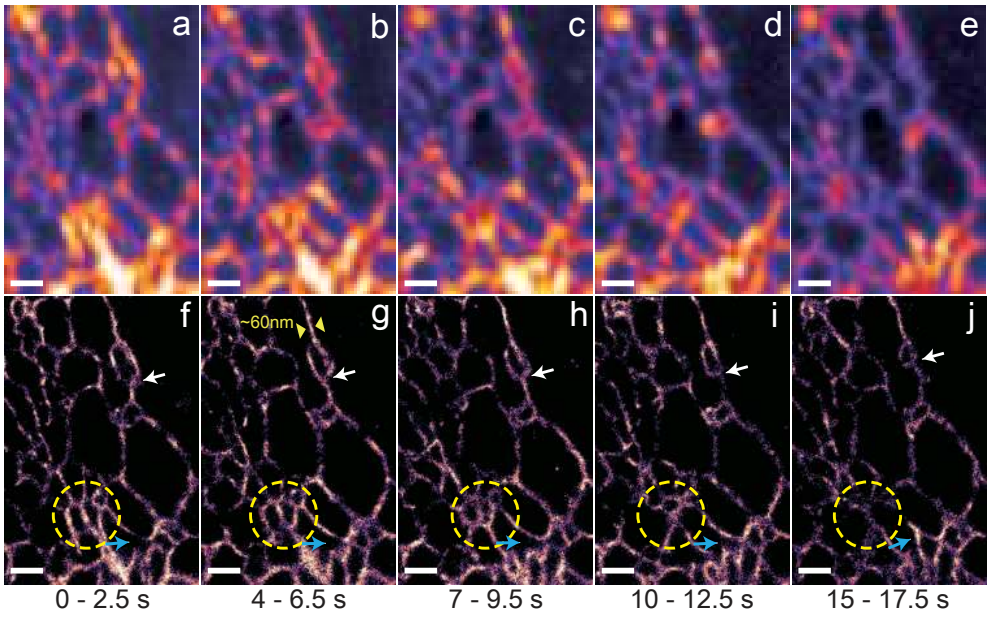

Figure 8 FALCON performance on live ER data. Live imaging of the endoplasmic reticulum protein, reticulon-4 fused to tdEos imaged over 20 s in a U2OS cell. Conventional (a-e) and super-resolution (f-j) snapshots are shown with a 2.5 s temporal resolution. The average size of tubules measured from the reconstruction is approximately 60 nm (at FWHM). Yellow markers in (g) highlight a representative tubule width. The dynamic motion of this structure is highlighted in (g-j) indicated by white and blue arrows and a dashed, yellow circle. All scale bars are $1\ \mu m$

Supplementary Note	Supplementary text for FALCON and other localization algorithms
Supplementary Figure 1	Background estimation of FALCON
Supplementary Figure 2	Performance analysis for continuous refinement on uniform distribution
Supplementary Figure 3	Performance analysis for continuous refinement on line distribution
Supplementary Figure 4	Performance analysis for synthetic ”ring” phantom with various imaging densities
Supplementary Figure 5	Performance analysis of FALCON in terms of size of sub-sampling grid
Supplementary Figure 6	Parameter choice analysis of FALCON with various sparsity levels
Supplementary Figure 7	Parameter-sensitivity analysis 1 of background estimation with various wavelet levels
Supplementary Figure 8	Parameter-sensitivity analysis 2 of background estimation with various wavelet levels
Supplementary Figure 9	Performance analysis of FALCON under Gaussian(CCD) and non-Gaussian(CMOS) readout noise statistics.
Supplementary Figure 10	Performance analysis 1 in comparison with deconSTORM
Supplementary Figure 11	Performance analysis 2 in comparison with deconSTORM
Supplementary Figure 12	Photon counts analysis on high-density STORM images of microtubules in COS-7 cell
Supplementary Figure 13	Super-resolution images reconstructed by least-square fitting and FALCON using low-density STORM data of fixed microtubules
Supplementary Figure 14	Super-resolution images reconstructed by DAOSTORM, CSSTORM and FALCON using high-density PALM data of live ER

1 Supplementary Note

1.1 General approach

1.1.1 Least-squares formulation

At a given time instant, the activated fluorescent probes can be described by a collection of point sources:

$$f(\mathbf{x}) = \sum_{k=1}^{K} c_k \, \delta(\mathbf{x} - \mathbf{x}_k).$$

Here, K is the number of the point sources, $\mathbf{x}_k$ is the position of the k-th source and c_k is its brightness in terms of the number of emitted photons. Formally, the camera image is obtained by convolving the sources with the PSF h of the microscope, adding a background signal b and introducing a random noise distortion represented by $\mathcal{N}$:

$$g[\mathbf{n}] \sim \mathcal{N}\left(b + \sum_{k=1}^{K} c_k \, h(\mathbf{n} - \mathbf{x}_k)\right)$$

where n denotes n-th pixel of the camera image. Given a set of measurements $g[\mathbf{n}]$, we would like to find the parameters K, c_k and $\mathbf{x}_k$ that minimize the least-squares functional

$$\sum_{\mathbf{n}} \left(g[\mathbf{n}] - b - \sum_{k=1}^{K} c_k \, h(\mathbf{n} - \mathbf{x}_k)\right)^2. \tag{1}$$

However the dependence of this functional on $\mathbf{x}_k$ is non-linear, which requires computationally expensive derivative of h at each iteration.

1.1.2 Taylor-series approximation

We thus propose to modify the problem by introducing a sub-pixel grid of the form $\mathbb{Z}^2/M$, where M is an integer oversampling factor. A given source position $\mathbf{x}_k$ can be expressed as the sum of the closest m-th point on the sub-pixel grid $\mathbf{m}_k$ and a small displacement $\boldsymbol{\epsilon}_k$:

$$\mathbf{x}_k = \frac{\mathbf{m}_k}{M} + \boldsymbol{\epsilon}_k. \tag{2}$$

Here, we consider linear approximation of h by Taylor series expansion as follows:

$$\begin{aligned} h(\mathbf{x}-\mathbf{c}) &= h(\mathbf{c}) + (\mathbf{x}-\mathbf{c})^T \nabla h(\mathbf{c}) + r(\mathbf{x}-\mathbf{c}) \\ h(\mathbf{x}-\mathbf{c}-\boldsymbol{\epsilon}) &= h(\mathbf{c}) + (\mathbf{x}-\mathbf{c}-\boldsymbol{\epsilon})^T \nabla h(\mathbf{c}) + r(\mathbf{x}-\mathbf{c}-\boldsymbol{\epsilon}) \\ &= (h(\mathbf{c}) + (\mathbf{x}-\mathbf{c})^T \nabla h(\mathbf{c}) + r(\mathbf{x}-\mathbf{c})) - \boldsymbol{\epsilon}^T \nabla h(\mathbf{c}) + (r(\mathbf{x}-\mathbf{c}-\epsilon) - r(\mathbf{x}-\mathbf{c})) \\ &\simeq h(\mathbf{x}-\mathbf{c}) - \boldsymbol{\epsilon}^T \nabla h(\mathbf{c}) \end{aligned}$$

where r denotes higher-order terms more than second-order and T is transpose. We can then apply a first-order Taylor approximation for every shifted PSF:

$$h(\mathbf{n}-\mathbf{x}_k) \simeq h(\mathbf{n} - \frac{\mathbf{m}_k}{M}) - \boldsymbol{\epsilon}_k^T \nabla h(\mathbf{n} - \frac{\mathbf{m}_k}{M}).$$

Inserting the latter expression into (1) and extending the inner sum to the entire sub-pixel grid leads to the functional

$$J(c, \boldsymbol{\epsilon}) = \sum_{\mathbf{n}} \left(g[\mathbf{n}] - b - \sum_{\mathbf{m}} c[\mathbf{m}] \left[h(\mathbf{n} - \frac{\mathbf{m}}{M}) - \boldsymbol{\epsilon}[\mathbf{m}]^T \nabla h(\mathbf{n} - \frac{\mathbf{m}}{M}) \right] \right)^2, \tag{3}$$

which becomes quadratic with respect to c or $\boldsymbol{\epsilon}$ respectively. Note that determining K and $\mathbf{m}_k$ now amounts to identifying the support of the non-zero coefficients $c[\mathbf{m}]$.

1.2 FALCON: FAst Localization algorithm based on CONtinuous-spatial formulation

Our algorithm can be divided into three main stages (**Fig. 1**). The first two are primarily aimed at determining the support of $c[\mathbf{m}]$. They consist in deconvolving the measured image on the sub-pixel grid using two subsequent types of sparsity constraints. Then, detection of local-maxima can be performed to isolate the sources and to compute initial estimates for their brightness and locations. These quantities are then refined iteratively in the later stage of the algorithm, where the position coordinates are allowed to take values over a continuum.

1.2.1 Deconvolution with a space-domain sparsity prior

If we neglect the displacements $\boldsymbol{\epsilon}[\mathbf{m}]$ in (3), we are left with a deconvolution problem where the object is known to comprise of a sparse set of sources. Thus, it is natural to first try to minimize

$$J_1(c) = J(c, \mathbf{0}) + \phi(c) + R_{\mathbb{R}^+}(c), \tag{4}$$

where

$$\phi(c) = \sum_{\mathbf{m}} w[\mathbf{m}] \, |c[\mathbf{m}]|$$

is a weighted ℓ_1 norm that favors sparse source distributions on the sub-pixel grid and $R_{\mathbb{R}^+}(c)$ constrains the coefficients to be positive, which is defined as

$$R_{\Omega}(c) = \begin{cases} 0 & \text{if } \forall \mathbf{m}, c[\mathbf{m}] \in \Omega, \\ +\infty & \text{otherwise.} \end{cases}$$

As we shall see in Section 1.3, each weight $w[\mathbf{m}]$ determines a threshold below which the coefficient $c[\mathbf{m}]$ is set to zero. In our experiments, we obtained good results by choosing it proportional to a local uncertainty of background noise:

$$w[\mathbf{m}] = \kappa \times \sqrt{b[\mathbf{m}] \times \gamma_{\text{EM}}}.$$

Here $b[\mathbf{m}]$ is the local background signal, which is interpolated from the background estimate described in Section 1.4. γ_{EM} denotes the EM-CCD gain factor; it is typically equal to 1.4 (EM gain on) or 1 (EM gain off). In general, the noise mainly follows Poisson statistics, hence the square root to estimate its the standard deviation of the background noise. We investigated various choices of κ over a wide range of simulation settings (**Supplementary Fig. 6**); based on these results we used $\kappa = 2$ in all of our experiments.

1.2.2 Least-squares deconvolution with a fixed spatial support

In our synthetic experiments (**Fig. 6 and Supplementary Fig. 4**), we observed that the first stage of our algorithm resulted in clusters of non-zero pixels around the true particle locations. However, the positions

of the local maxima of these clusters appeared to be biased, in the sense that the estimated distances among nearby particles were usually smaller than expected. This spatial bias seems to be a side-effect of the fact that the (weighted) ℓ_1 norm biases the pixel intensities towards zero.

We thus decided to incorporate a second stage in our algorithm where the sparsity constraint was replaced by a support constraint:

$$J_2(c) = J(c, \mathbf{0}) + S_{\Omega_1}(c) + R_{\mathbb{R}^+}(c). \tag{5}$$

Here

$$S_\Omega(c) = \begin{cases} 0 & \text{if } c[\mathbf{m}] \neq 0 \text{ only for } \mathbf{m} \in \Omega, \\ +\infty & \text{otherwise.} \end{cases}$$

and Ω_1 is the support of the coefficients $c[\mathbf{m}]$ returned by the first stage of the algorithm. The support is defined as the set of pixels whose value are above small threshold (e.g. 5% of the maximum image intensity). Minimizing (5) amounts to a simple least-squares fit with a positivity and support constraints, but without the typical bias induced by the ℓ_1 penalty.

1.2.3 Continuous-domain refinement of positions and coefficients

Let $\tilde{c}[\mathbf{m}]$ denote the deconvolved image obtained in the previous stage. Based on this outcome, we can compute a first estimate of the coefficients $c[\mathbf{m}]$ and of the displacements $\boldsymbol{\epsilon}[\mathbf{m}]$, which we will then refine.

We proceed as follows. The number of sources K is first set as the number of local maxima in $\tilde{c}[\mathbf{m}]$. Let $\widetilde{\mathbf{m}}_k$ be the positions of these local maxima. Then, the position of the k-th source is estimated as a center of mass in a small neighborhood around the corresponding local maximum:

$$\mathbf{x}_k = \frac{\sum_{\mathbf{m} \in N(\widetilde{\mathbf{m}}_k)} \tilde{c}[\mathbf{m}]\mathbf{m}}{\sum_{\mathbf{m} \in N(\widetilde{\mathbf{m}}_k)} \tilde{c}[\mathbf{m}]}.$$

In our implementation the neighbourhood $N(\widetilde{\mathbf{m}}_k)$ was chosen to be a 3×3 window centered on $\widetilde{\mathbf{m}}_k$. This gives

an initial estimate for $\mathbf{m}_k$ and $\boldsymbol{\epsilon}_k$ according to (2). From there on the support of the non-zero coefficients is fixed as

$$\Omega_2 = \{\mathbf{m}_k \,|\, k = 1, \ldots, K\}.$$

The initial value of these coefficients is simply given by the sum of the previous-stage coefficients in the local neighbourhood:

$$c[\mathbf{m}_k] = \sum_{\mathbf{m} \in N(\widetilde{\mathbf{m}}_k)} \tilde{c}[\mathbf{m}].$$

We can then return to the bivariate functional (3), augmented with three regularization terms:

$$J_3(c, \boldsymbol{\epsilon}) = J(c, \boldsymbol{\epsilon}) + S_{\Omega_2}(c) + R_{\mathbb{R}^+}(c) + R_{[-\Delta,\Delta]^2}(\boldsymbol{\epsilon}). \tag{6}$$

Here S_{Ω_2} and $R_{\mathbb{R}^+}$ express the support and positivity constraints on the coefficients. The additional term $R_{[-\Delta,\Delta]^2}$ simply prevents the displacement variables from becoming too large (in which case the Taylor approximation would become too inaccurate). In our implementation the distance parameter Δ is chosen to be $1/M$ in each lateral direction.

1.3 Implementation of FALCON

In general, (4), (5) and (6) involve millions of unknowns, which restricts the structure of the corresponding minimization algorithms. In the context of localization microscopy, sparsity-regularized functionals such as (4) and (5) have been tackled using linear programming on small image patches[1] or expectation-maximization (EM) methods[2]. Here we use the alternating direction method of multipliers (ADMM), which has recently been introduced in the field of image recovery and converges significantly faster[3]. To minimize (6) we have designed a custom alternating minimization scheme, taking advantage of the fact that $J(c, \boldsymbol{\epsilon})$ is essentially quadratic when one of its arguments is fixed.

Note that the step size of the sub-pixel grid has to be properly selected to balance computation complexity

and localization performance; empirically between 60 nm and 20 nm. In all our experiments, the size of the CCD pixels was 100 nm and we used an oversampling factor of $M = 3$, i.e. a sub-pixel size of 33 nm.

1.3.1 Sparse deconvolution using ADMM

Using matrix notations, the sparsity-based cost functions (4) and (5) have the following general structure:

$$\|g - b - DHc\|^2 + \psi(c),$$

where g is the measurement, b is the background signal, D is a down-sampling matrix, H is a convolution matrix corresponding to the PSF h and ψ is a sparsity and positivity-enforcing regularizers. In practice, we introduce an auxiliary variable v and apply ADMM to minimize $\|g - b - DHc\|^2 + \psi(v)$ under the constraint $v = c$. This leads to the following iterative update scheme:

$$\begin{aligned} c^{(i+1)} &= \arg\min_c \|g - b - DHc\|^2 + \mu\|c - v^{(i)} - d^{(i)}\|^2; \\ v^{(i+1)} &= \arg\min_v \psi(v) + \mu\|c^{(i+1)} - d^{(i)} - v\|^2; \\ d^{(i+1)} &= d^{(i)} + c^{(i+1)} - v^{(i+1)}. \end{aligned}$$

Note that the value of μ affects the convergence rate of the optimization. However, the results is not sensitive to the choice of μ in the range of $0.1 - 0.01$. We have chosen $\mu = 0.05$ as a universal setting for all our experiments.

More explicitly, the update of c is given by the following closed-form expression:

$$c^{(i+1)} = (H^T D^T DH + \mu I)^{-1}(H^T D^T (g - b) - \mu(v^{(i)} + d^{(i)})).$$

Note that the matrix inversion in this equation can be carried out directly using the Fast Fourier Transform algorithm. The update of v can be written as $v^{(i+1)} = \mathcal{M}_{\psi/\mu}(c^{(i+1)} - d^{(i)})$ where $\mathcal{M}_{\psi/\mu}$ is the so-called Moreau proximal mapping corresponding to ψ/μ. The two proximal mappings that are relevant to the functionals (4)

and (5) are defined by

$$\mathcal{M}_{(\phi+R_{\mathbb{R}^+})/\mu}(v)[\mathbf{m}] = \begin{cases} v[\mathbf{m}] - w[\mathbf{m}]/\mu & \text{if } v[\mathbf{m}] > w[\mathbf{m}]/\mu, \\ 0 & \text{otherwise;} \end{cases} \tag{7}$$

$$\mathcal{M}_{(S_\Omega+R_{\mathbb{R}^+})}(v)[\mathbf{m}] = \begin{cases} v[\mathbf{m}] & \text{if } v[\mathbf{m}] > 0 \text{ and } \mathbf{m} \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

1.3.2 Continuous-domain refinement using alternating minimization

When one of its arguments is fixed (say $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^{(i)}$ or $c = c^{(i+1)}$), the cost function for the refinement stage satisfies the following equalities (up to some additive constants):

$$J_3(c, \boldsymbol{\epsilon}^{(i)}) = \|g - b - (DH - D\nabla H\mathcal{E}^{(i)})c\|^2 + S_{\Omega_2}(c) + R_{\mathbb{R}^+}(c); \tag{8}$$

$$J_3(c^{(i+1)}, \boldsymbol{\epsilon}) = \|g - b - DHc^{(i+1)} + D\nabla HC^{(i+1)}\boldsymbol{\epsilon}\|^2 + R_{[-\Delta,\Delta]^2}(\boldsymbol{\epsilon}). \tag{9}$$

Here ∇H is a matrix that corresponds to the gradient of the PSF and $\mathcal{E}^{(i)}$ and $C^{(i+1)}$ are matrices that depend on $\boldsymbol{\epsilon}^{(i)}$ and $c^{(i+1)}$, respectively.

Equations (8) and (9) define quadratic minimization problems with convex constraints. Thus, to minimize J_3, we propose an algorithm that alternates between two projected gradient descents:

$$c^{(i+1)} = \mathcal{M}_{(S_\Omega+R_{\mathbb{R}^+})}\Big[c^{(i)} - \tau^{(i)}(DH - D\nabla H\mathcal{E}^{(i)})^T\Big((DH - D\nabla H\mathcal{E}^{(i)})c^{(i)} + b - g\Big)\Big];$$

$$\boldsymbol{\epsilon}^{(i+1)} = \mathcal{M}_{R_{[-\Delta,\Delta]^2}}\Big[\boldsymbol{\epsilon}^{(i)} - \theta^{(i)}(D\nabla H(C^{(i+1)}))^T\Big(D\nabla HC^{(i+1)}\boldsymbol{\epsilon}^{(i)} - DHc^{(i+1)} - b + g\Big)\Big].$$

Here the first projection operator has been defined in (7); the second is defined by

$$\mathcal{M}_{R_{[-\Delta,\Delta]^2}}(\boldsymbol{\epsilon})[\mathbf{m}] = \begin{cases} \boldsymbol{\epsilon}[\mathbf{m}] & \text{if } \boldsymbol{\epsilon}[\mathbf{m}] \in [-\Delta, \Delta]^2, \\ \Delta\epsilon[\mathbf{m}]/\|\boldsymbol{\epsilon}[\mathbf{m}]\| & \text{otherwise.} \end{cases}$$

At every iteration, the step sizes $\tau^{(i)}$ and $\theta^{(i)}$ are chosen so as to minimize the quadratic term in (8) and (9), respectively; this simply amounts to finding the roots of second-degree polynomials.

1.4 Background estimation of FALCON

In general, the background fluorescence signal varies smoothly over the field of view and the estimation of background is important to improve localization performance. We applied iterative wavelet transform method used in spectroscopy [4] to the step 1&2. In detail, we get a residual image by subtracting estimated photon contributions of molecules from a camera image, and then, the background is iteratively estimated from the lowest frequency wavelets bands of the residual image as shown in (**Supplementary Fig. 1**). In each iteration, all values of the residual image above current estimated background level are clipped. In the step 1&2, estimation of probes c and background b are conducted alternatively. The weighting factor w in the step 1 is also updated at every update of the background. More specifically, we used wavelet decomposition up to 8th level using Daubechies-6 wavelets. For an initial background level, we took a minimum value of median filtered camera image using 5×5 kernel.

We tested localization performances by chaining wavelet level from 6 to 9 with different number of iteration of the method (**Supplementary Fig. 7 and 8**). This simulation studies show that background is well estimated and localization performances are consistent. For all the other simulations in this paper, wavelet decomposition level is fixed at eight, and three iterations are used in the step 1 and one iteration is used in the step 2.

1.5 Other localization algorithms

1.5.1 Gaussian fitting by Non-linear least-squares minimization

In every local peak, we crop a small window centred on the peak. The cropped image is fitted with an two-dimensional symmetric Gaussian function:

$$I(x,y) = \frac{c}{2\pi\sigma^2} exp^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} + b$$

where c is an estimated number of photon counts from a single molecule. σ is width of Gaussian PSF and b is an uniform background. All variables are optimized via non-linear least-squares minimization[5,6]. Using shape-based filtering, the localized molecules are excluded for all our experiments if min(σ) $<$ 1 pixel, max(σ) $>$ 2 pixels. The window size is 9 $\times$ 9 pixels in **Fig. 4 and Supplementary Fig. 13**.

1.5.2 DAOSTORM

DAOSTORM[7] performs simultaneous multi-emitter fitting of an experimentally derived model PSF to the input data, iteratively identify new candidate localizations from a residuals image of the fits to the data. Fitting to the input data using new candidate localizations is repeated until new candidate localizations fall below a target signal to noise ratio, or until a maximum number of iterations is reached, whichever occurs first. Model PSFs were derived from low density frames of the dataset (simulations) or low density data imaged under the same experimental conditions (Tubulin Alexa-647 data).

The key parameters in the DAOSTORM algorithm are: sigmaBG, the estimated background noise; nSigma, the SNR detection threshold; and nIter, the maximum number of iterations. sigmaBG was estimated from empty subregions of images in each dataset. nSigma was chosen by running DAOSTORM several times on single frames of the datasets, and visually estimating the value of nSigma which gave the best results. For the Tubulin Alexa-647 data, sigmaBG = 17, nSigma = 5. For the high SNR simulated data, sigmaBG = 10, nSigma = 10. For the low SNR simulated data, sigmaBG = 4, nSigma = 5. For all datasets, nIter = 4.

1.5.3 Compressed sensing STORM (CSSTORM)

In CSSTORM[1], the localization is formulated as a convex optimization with weighted ℓ_1 norm penalty for sparsity priors, which is solved by linear programming:

$$\textit{Minimize } \ w^T c$$

$$\textit{Subject to: } \ c[m] \geq 0 \ \ and \ \ \|g - (Hc + b)\|^2 \leq \alpha \cdot (\sum g[n])^{1/2}$$

where g is a measured camera image, c is molecule distribution in terms of the fluorescence intensity, b is uniform background fluorescence. The matrix H is convolution matrix of PSF and w is weighting vector. The parameter α determines the noise level of the camera image g to be allowed in the optimization. Since linear program requires high computational complexity, the problem is divided into the several sub-problem of the overlapped patches. In the every patch of 12 $\times$ 12 μm^2, 14 $\times$ 14 μm^2 image is reconstructed on the fine sub-pixel grid of 20 nm, and then only central 0.8 $\times$ 0.8 μm^2 area is included in the final solution. The noise parameter α is 1.5 as default setting (without electron multiplying(EM) gain)[1] for all our experiments. For live ER data with EM gain, we used $\alpha = 1.7$ which is slightly lower value than $\alpha = 1.7$ used in [1] in order to detect more molecules.

1.5.4 deconvolution-STORM (deconSTORM)

In deconSTORM[2], deconSTORM uses the maximum likelihood estimator to solve localization problem. Specifically, it is solved by modified Richardson–Lucy deconvolution algorithm imposing sparse prior distribution for every pixel intensities of fine sub-pixel grid of 20 nm. More specifically, deconSTORM approximates switching events of molecules between activated and off state as a Markov process. All state transitions of molecules

are assumed to be statistically independent, which allows to update sparse prior distribution for each pixel independently. deconSTORM requires prior knowledge concerning constant background level and probabilities of the state transitions. We set all the required parameters as optimal values in our simulations.

References

1. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster storm using compressed sensing. *Nat. Methods* **9**, 721–723 (2012).

2. Mukamel, E. A., Babcock, H. & Zhuang, X. Statistical deconvolution for superresolution fluorescence microscopy. *Biophys. J.* **102**, 2391–2400 (2012).

3. Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process* **19**, 2345–2356 (2010).

4. Galloway, C., Ru, E. L. & Etchegoin, P. An iterative algorithm for background removal in spectroscopy by wavelet transforms. *Applied spectroscopy* **63**, 1370–1376 (2009).

5. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy STORM. *Nat. Methods* **3**, 793–796 (2006).

6. Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J. & Hess, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).

7. Holden., S. J., Uphoff., S. & Kapanidis, A. N. DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nat. Methods* **8**, 279–280 (2011).

8. Olivier, N., Keller, D., Rajan, V. S., Gönczy, P. & Manley, S. Simple buffers for 3d storm microscopy. *Biomed. Opt.* **4**, 885–899 (2013).

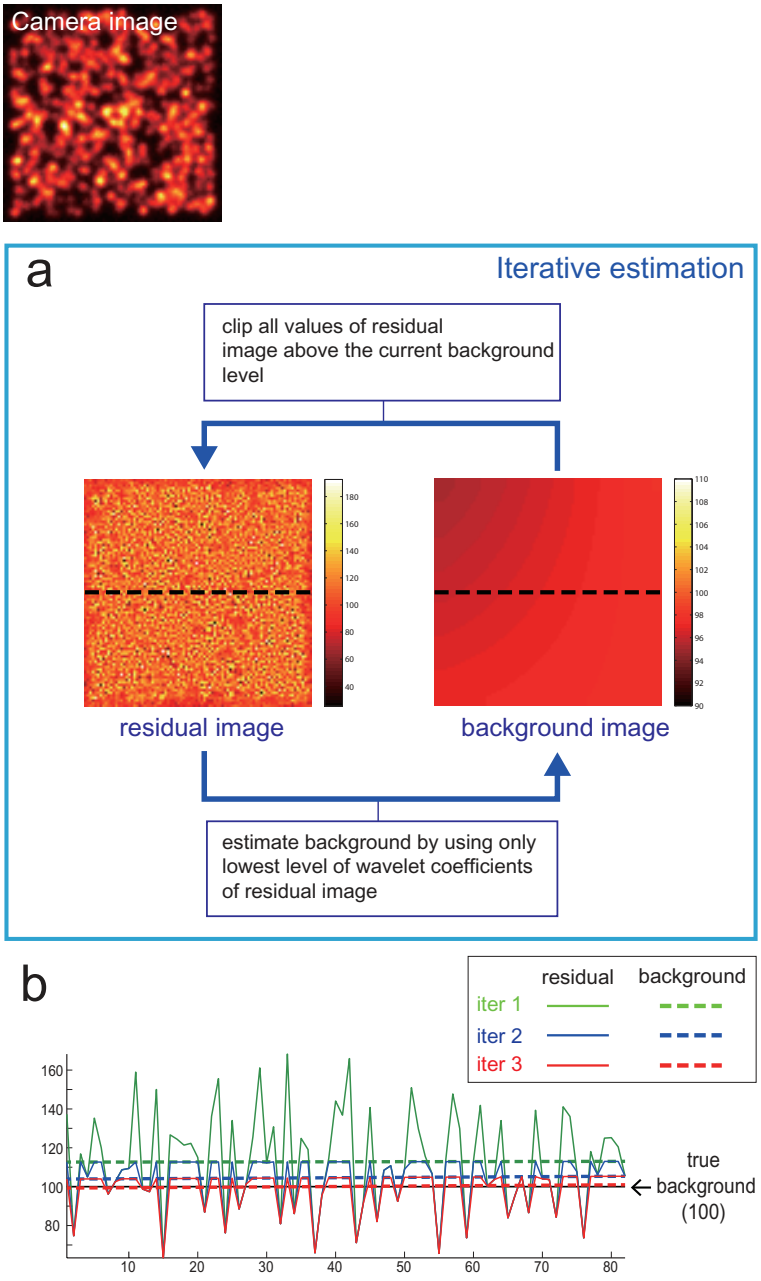

Figure 1 Background estimation of FALCON. (a) schematic diagram of the iterative background estimation. Background is iteratively estimated by using only lowest level of wavelet coefficients of residual image. (b) line profiles on dotted lines in (a) from estimated background and clipped residual image at $1-3$ iteration. The estimated background level at 3 iteration is close to true background level 100. For a detailed description, **see Supplementary text, section 1.4**.

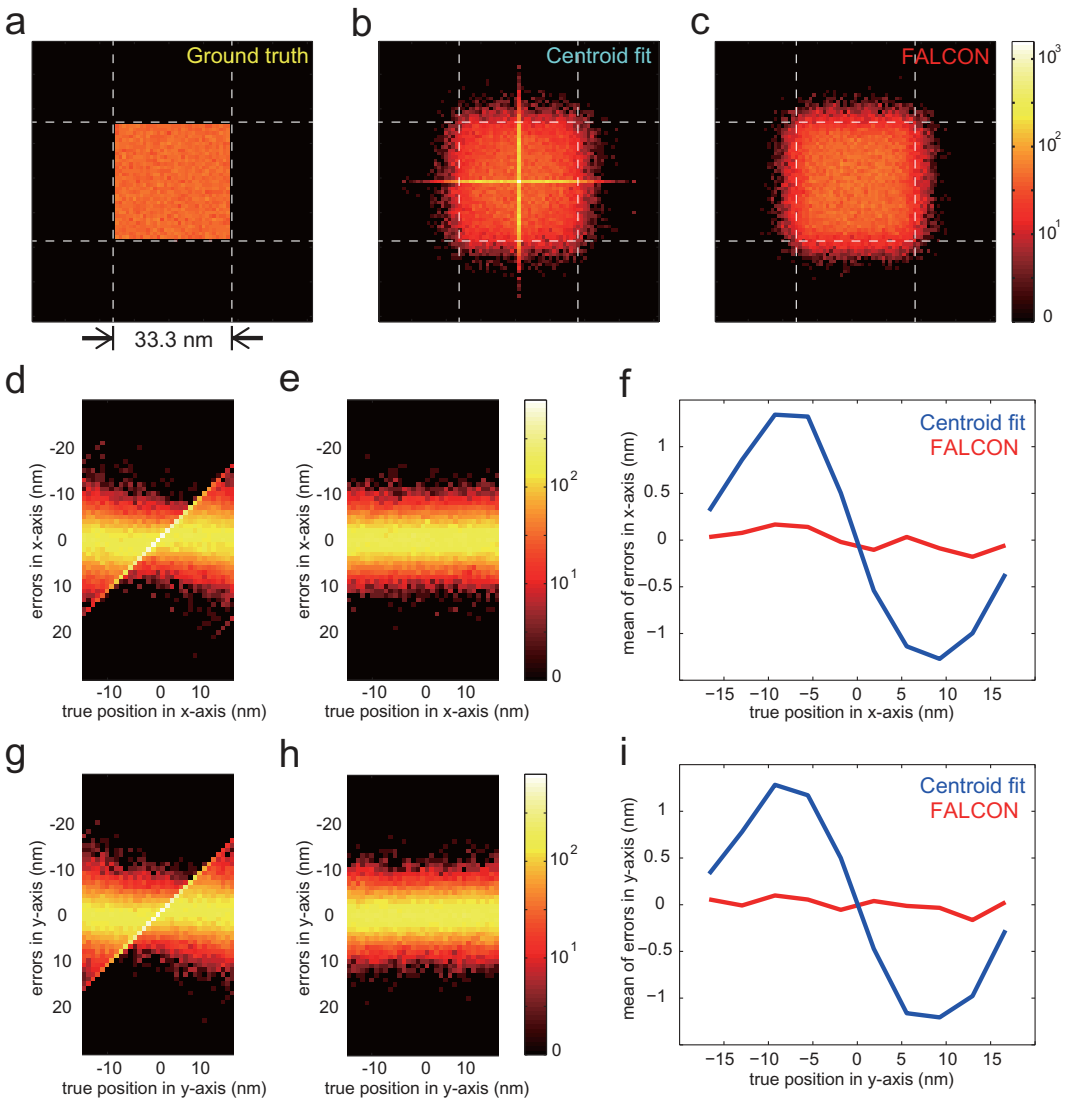

Figure 2 Performance analysis for continuous refinement on 50,000 simulated images with uniform distribution. A single molecule is randomly placed within a single sub-pixel area in each frame. (a) Histogram of the true distribution, (b) Histograms of the centroid fit applied to the deconvolved images by ℓ_1 minimization, (c) Histogram of FALCON. (d,e) Histograms of x-axis errors of centroid fit and FALCON respectively corresponding to positions of true molecules in x-axis, and (f) average errors of the histograms in x-axis. (g,h) Histograms of y-axis errors of centroid fit and FALCON respectively corresponding to positions of true molecules in y-axis, and (i) average errors of the histograms in y-axis.

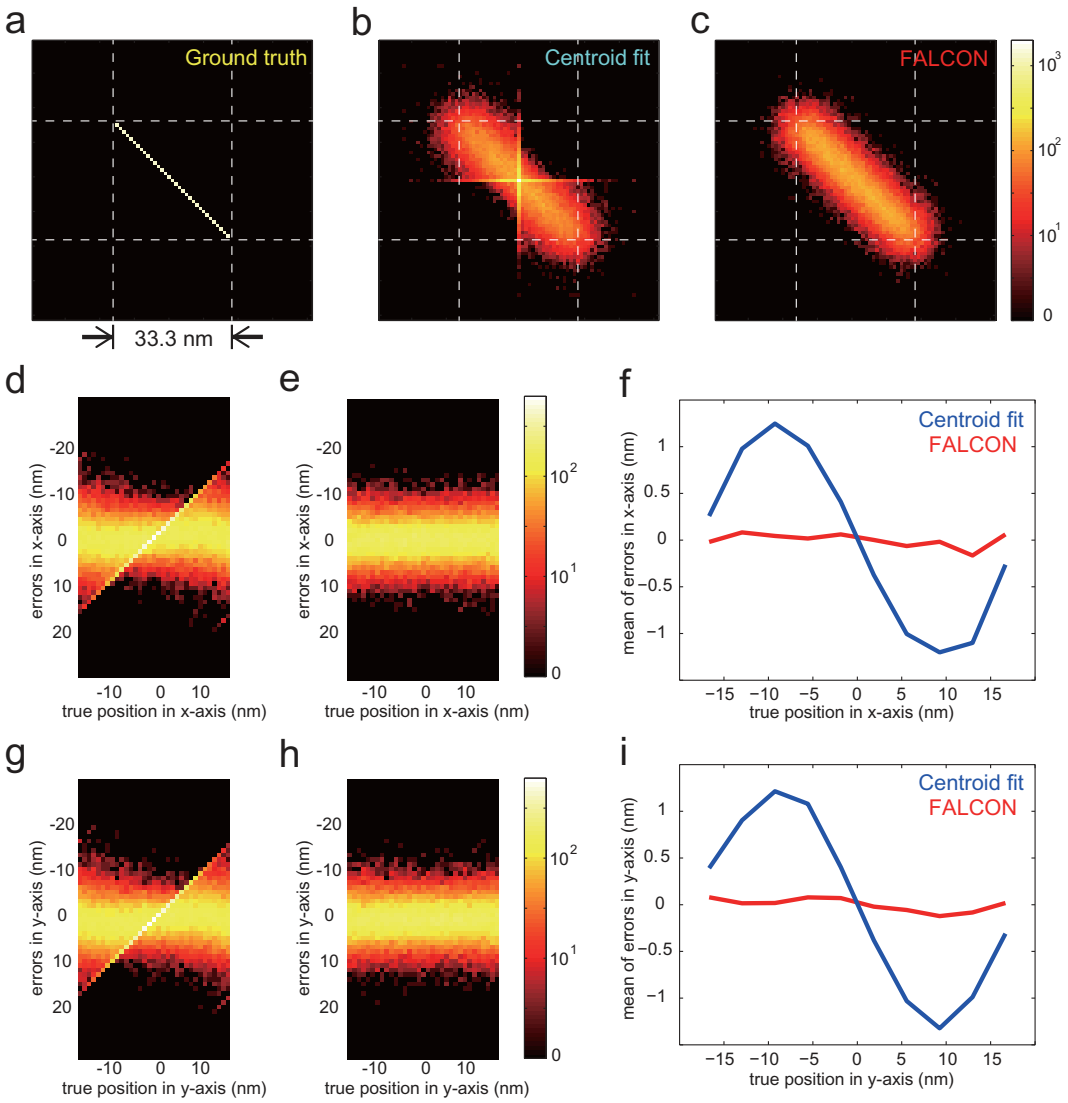

Figure 3 Performance analysis for continuous refinement on 50,000 simulated images of line distributions. A single molecule is randomly placed on a diagonal of a single sub-pixel in each frame. (a) Histogram of the true distribution, (b) Histograms of the centroid fit applied to the deconvolved images by ℓ_1 minimization, (c) Histogram of FALCON. (d,e) Histograms of x-axis errors of centroid fit and FALCON respectively corresponding to positions of true molecules in x-axis, and (f) average errors of the histograms in x-axis. (g,h) Histograms of y-axis errors of centroid fit and FALCON respectively corresponding to positions of true molecules in y-axis, and (i) average errors of the histograms in y-axis.

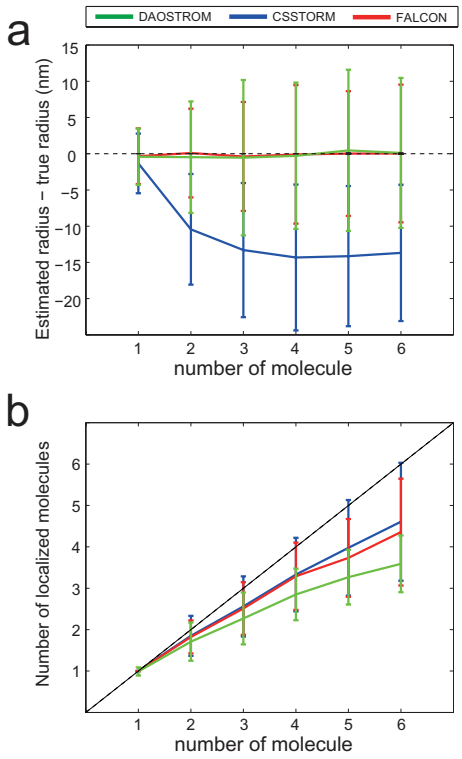

Figure 4 Performance analysis for synthetic ”ring” phantoms with various imaging densities. (a) Differences between true radius and estimated radius for various number of molecules on the circle of 200 nm radius in every simulated image. (b) Numbers of localized molecules by FALCON, DAOSTORM, and CSSTORM. The estimated radius is calculated by averaging the distances of localized molecules to the center of the circle. The error bars indicate standard deviations.

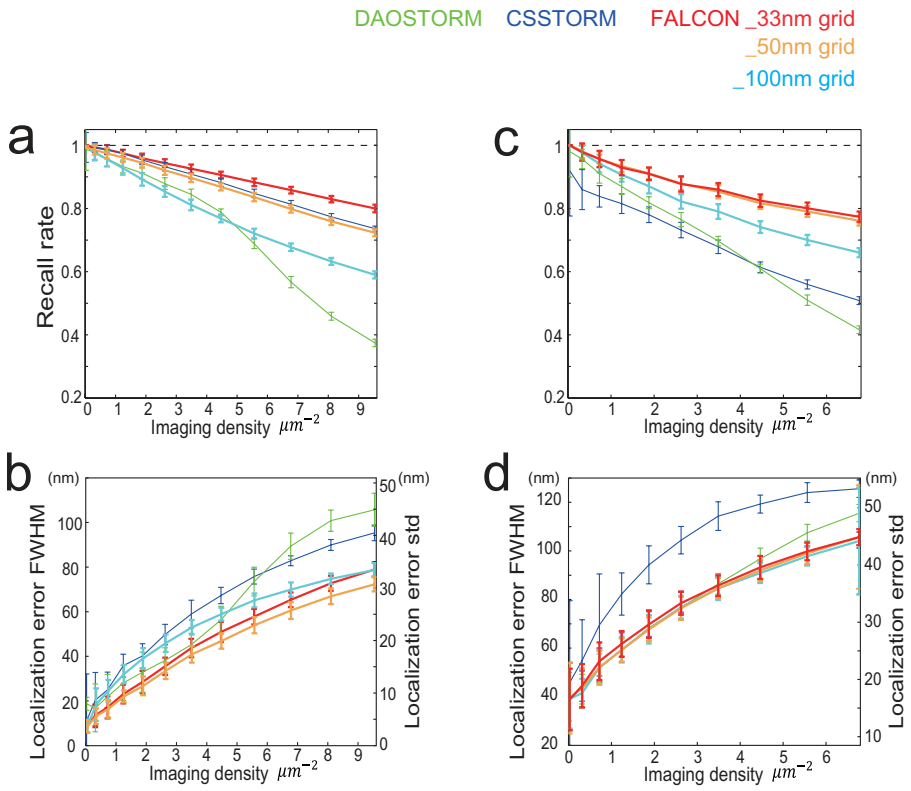

Figure 5 Performances of FALCON with different sub-pixel size for deconvolution in comparison of DAOSTORM and CSSTORM. The sub-pixel size for step1 and 2 of FALCON varies from 33.3 nm (up-sampling factor = 3) to 100 nm (without up-sampling). Simulated analysis on the random distribution of molecules over a wide range of imaging densities with high-photon emission rates: recall rates (a), localization accuracy (b) and low-photon emission rates: recall rates (c), localization accuracy (d). The error bars indicate standard deviations.

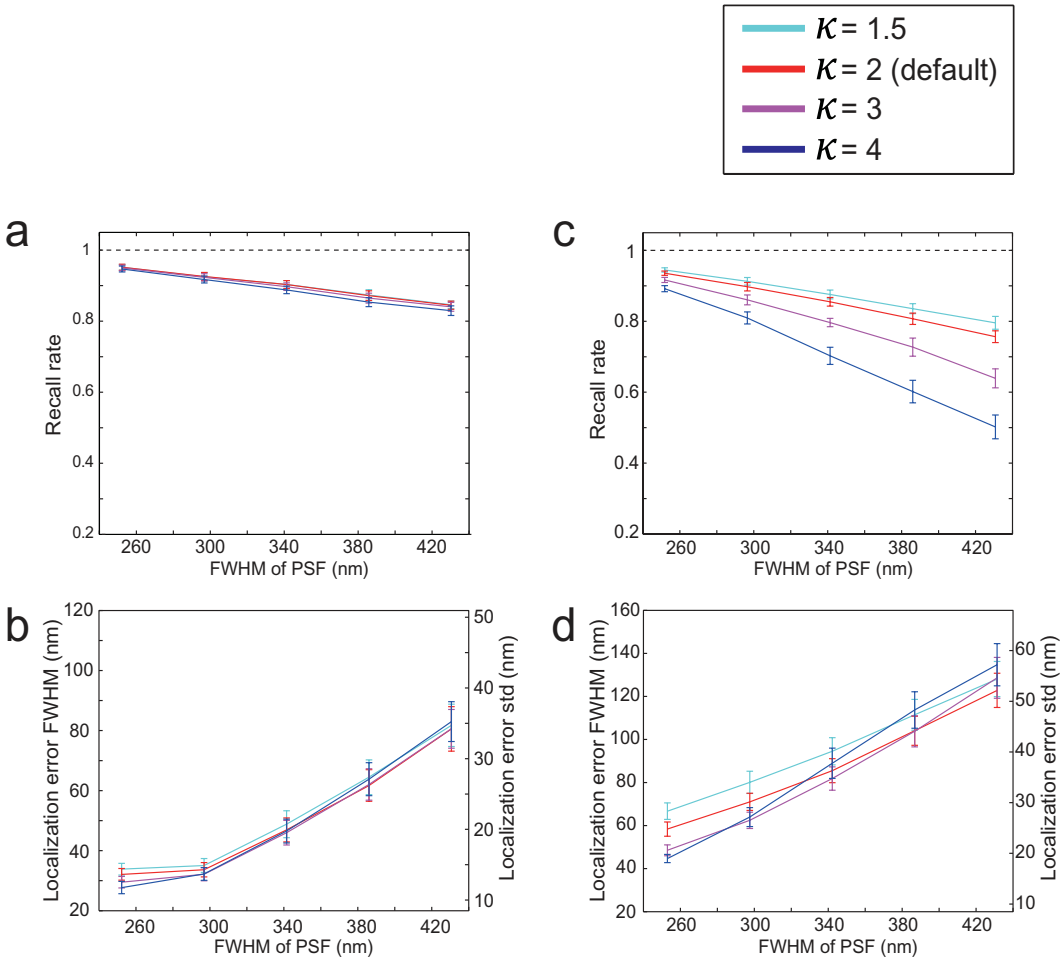

Figure 6 Parameter choice analysis with high & low SNR simulated data. κ determines minimum level of brightness coefficient to be counted in the first deconvolution step of FALCON. (a) Molecular recall rates, (b) Localization accuracy with κ ranging from 1.5 to 4 for high SNR simulated data and (c) Molecular recall rates, (d) Localization accuracy with κ ranging from 1.5 to 4 for low SNR simulated data. Imaging density is fixed to 6 μm^{-2} in all simulated image, and FWHM of PSF varies from 250 nm to 430 nm. The simulated image was generated by the following settings: on average, 5000 photons per molecule (with 2000 s.d.) and 100 uniform background photons for high SNR, and on average, 350 photons per molecule (with 70 s.d.) and 10 uniform background photons for low SNR. The error bars indicate standard deviations.

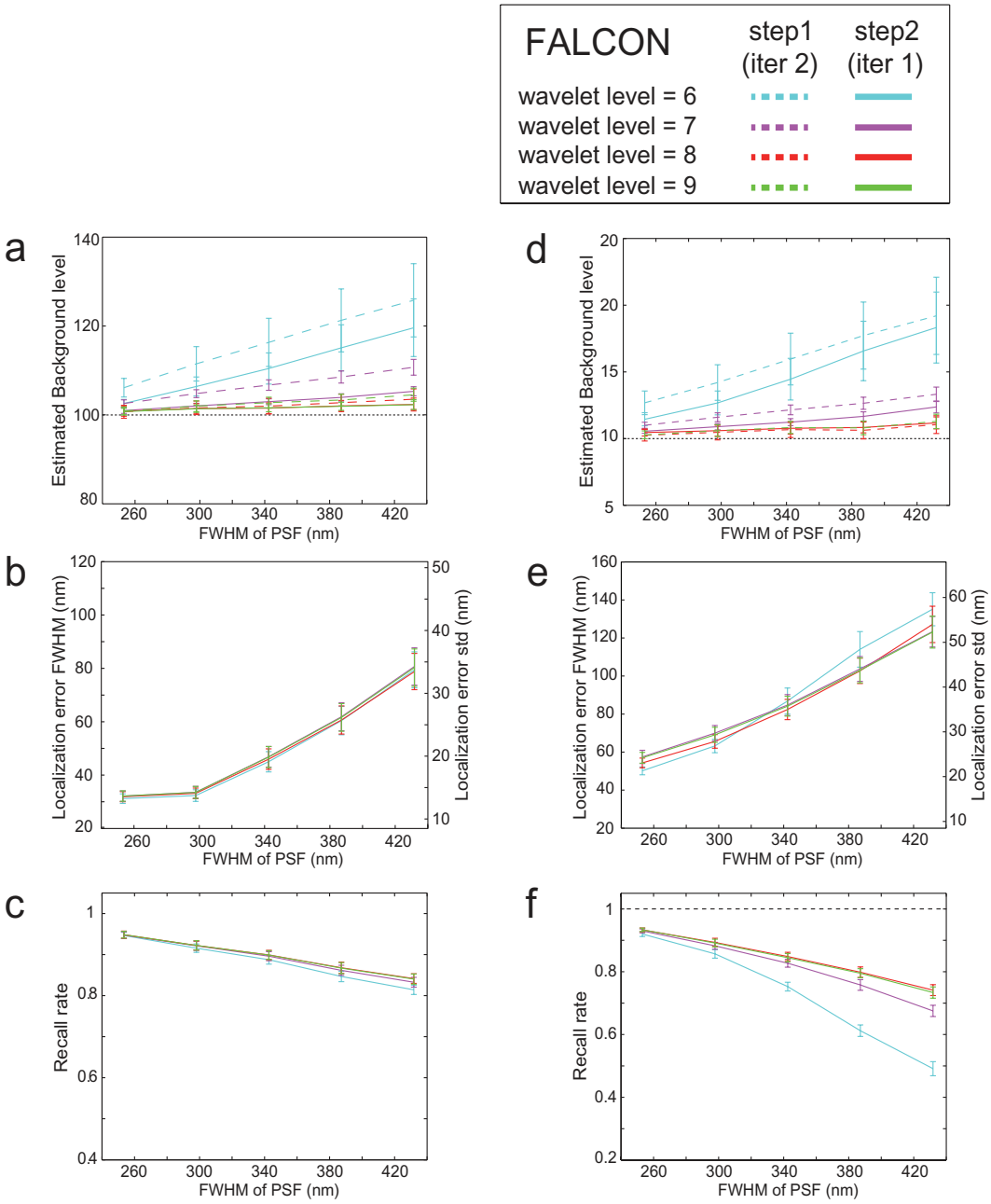

Figure 7 Parameter choice analysis for background estimation. Two iterations are used in the step 1 and one iteration is used in the step 2 at every update of the background. Background is estimated by only using coefficients of the highest wavelet level which varies from 6 to 9. (a) Estimated background level (true: 100) (b) Localization accuracy, (c) Molecular recall rates with various wavelet levels for high SNR simulated data, (d) Estimated background level (true: 10) (e) Localization accuracy, (f) Molecular recall rates with various wavelet levels for low SNR simulated

data. The simulated image was generated by the following settings: 100 uniform background photons and 5000 photons per molecule (with 2000 s.d.) on average for high SNR, and 10 uniform background photons and 350 photons per molecule (with 70 s.d.) on average for low SNR. The error bars indicate standard deviations.

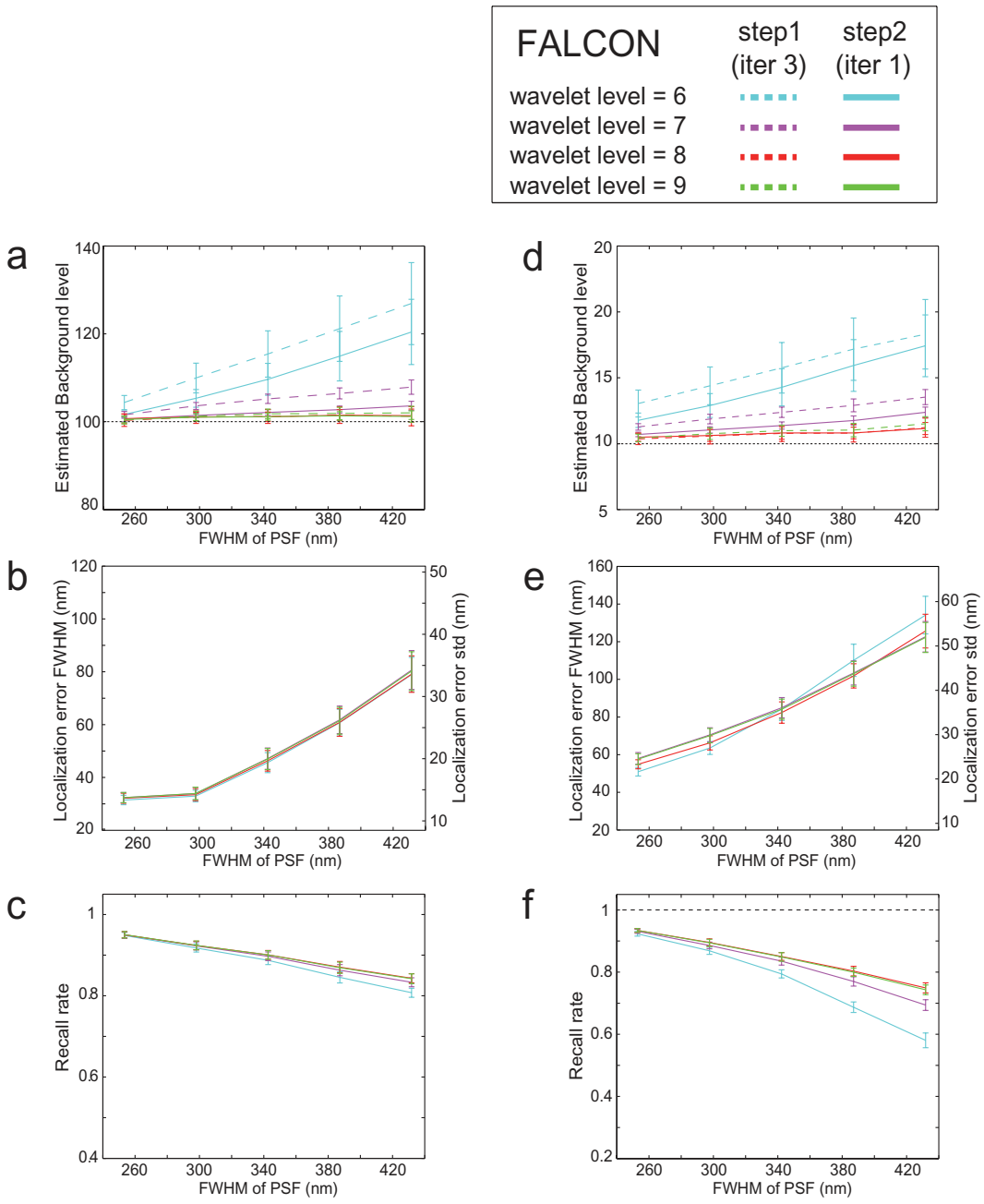

Figure 8 Parameter choice analysis for background estimation. Three iterations are used in the step 1 and one iteration is used in the step 2 at every update of the background. Background is estimated by only using coefficients of the highest wavelet level which varies from 6 to 9. (a) Estimated background level (true: 100) (b) Localization accuracy, (c) Molecular recall rates with various wavelet levels for high SNR simulated data, (d) Estimated background level (true: 10) (e) Localization accuracy, (f) Molecular recall rates with various wavelet levels for low SNR simulated

data. The simulated image was generated by the following settings: 100 uniform background photons and 5000 photons per molecule (with 2000 s.d.) on average for high SNR, and 10 uniform background photons and 350 photons per molecule (with 70 s.d.) on average for low SNR. The error bars indicate standard deviations.

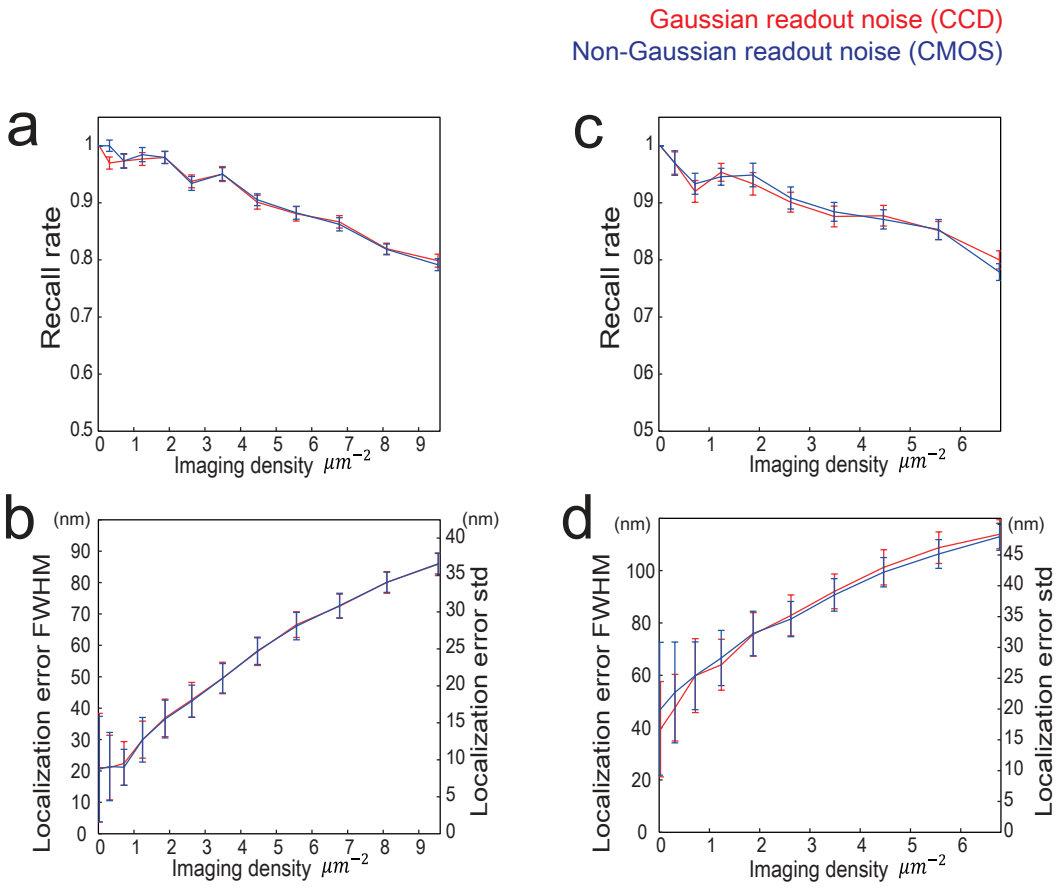

Figure 9 Performance analysis of FALCON under Gaussian(CCD) and non-Gaussian(CMOS) readout noise statistics. In order to approximate readout noise of CCD or CMOS type camera, simulated images are generated with Gaussian (zero mean with 2 s.d.) or log-normal distribution (unit mean with 2 s.d.). The simulated images are reconstructed by FALCON. Simulation results on the random distribution of molecules with high-photon emission rates: recall rates (a), localization accuracy (b) and low-photon emission rates: recall rates (c), localization accuracy (d). Simulation results under Gaussian and non-Gaussian readout noise are denoted by red and blue color respectively. The error bars indicate standard deviations.

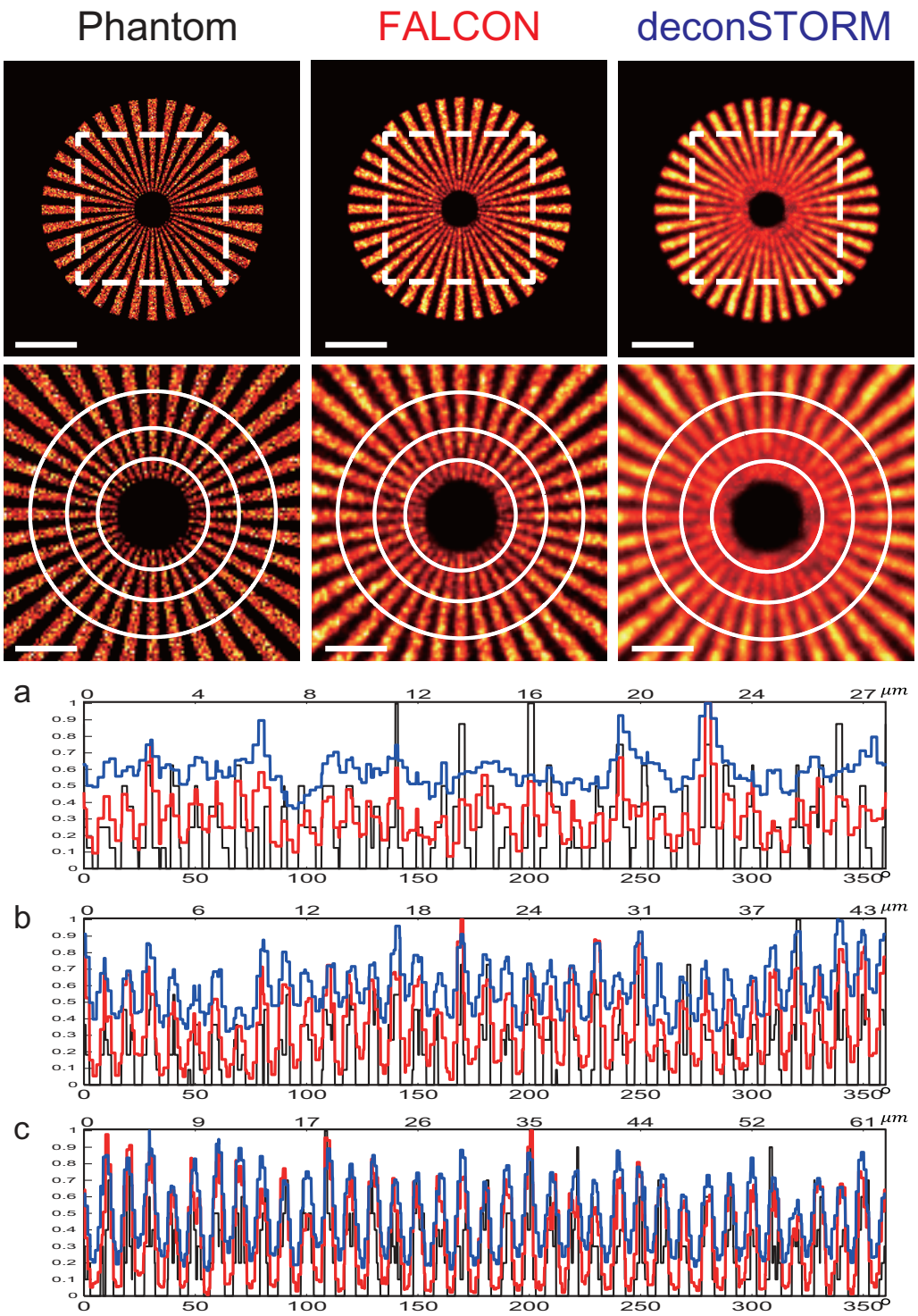

Figure 10 Performance analysis in comparison with deconSTORM in effective molecular density (on average, number of activated molecules/area of phantom) 8.6 μm^{-2}. Each activated molecule is shown in two consecutive frames on average. The 2000 simulated images were generated by the

following photon-emission settings: on average, 5000 photons per molecule (with 2000 s.d.) and 100 uniform background photons per camera pixel. The simulated phantom image and reconstructed images by FALCON and deconSTORM are shown in top row, and their close-up images in central region denoted by white box are presented in second row. For deconSTORM, required parameters of the algorithm such as average molecular activation time and background level set equal to simulation settings. Line profiles on circles from the inner to outer (a-c). profiles of the simulated phantom, FALCON and deconSTORM are denoted by black, red, blue colors respectively. Scale bars are $1\ \mu m$ in top row and 500 nm in close-up images.

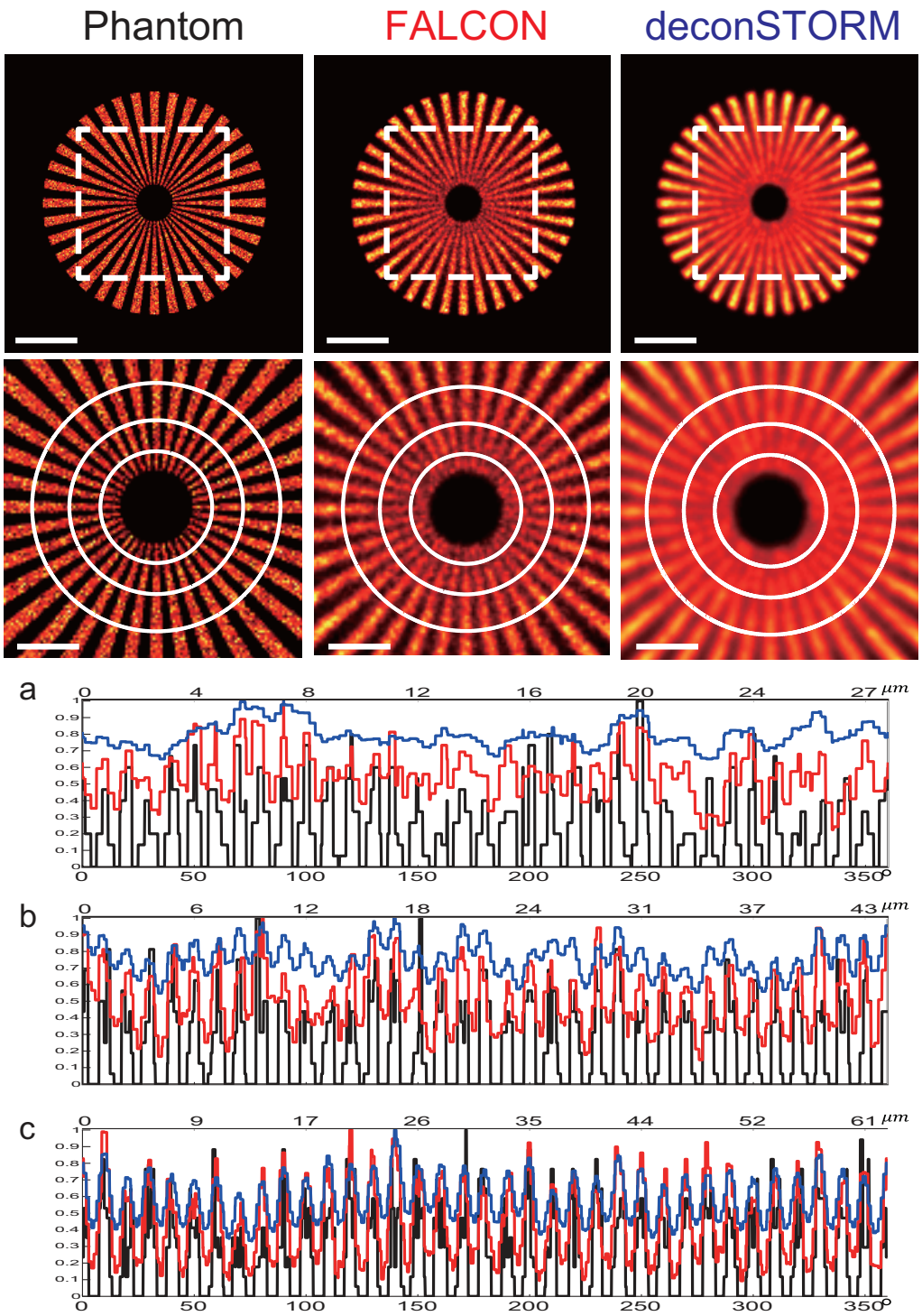

Figure 11 Performance analysis in comparison with deconSTORM in effective molecular density (on average, number of activated molecules/area of phantom) 17.2 μm^{-2}. Each activated molecule is shown in two consecutive frames on average. The 2000 simulated images were generated by the

following photon-emission settings: on average, 5000 photons per molecule (with 2000 s.d.) and 100 uniform background photons per camera pixel. The simulated phantom image and reconstructed images by FALCON and deconSTORM are shown in top row, and their close-up images in central region denoted by white box are presented in second row. For deconSTORM, required parameters of the algorithm such as average molecular activation time and background level set equal to simulation settings. Line profiles on circles from the inner to outer (a-c). profiles of the simulated phantom, FALCON and deconSTORM are denoted by black, red, blue colors respectively. Scale bars are $1\ \mu m$ in top row and 500 nm in close-up images.

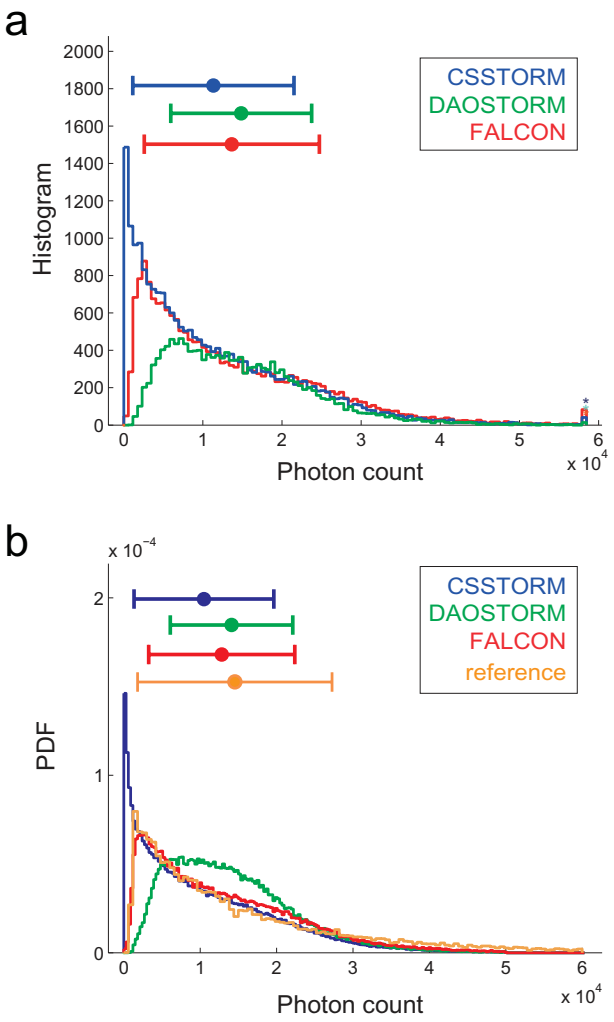

Figure 12 Photon counts analysis on high-density STORM images of microtubules (MTs) in a COS-7 cell. (a) Histograms of the photon-counts measured on the dataset used for figure 7, with the mean value and standard deviation displayed on top, for CSSTORM (blue), DAOSTORM (green) and FALCON (red). The three algorithms show significant differences for the range of low photon counts ($<$ 5,000). (b) Probability distributions of (a) in comparison with a reference distribution obtained from the low-density data under the identical preparation protocols [8]. FALCON shows similar distribution with that of the reference measurement, while DAOSTORM misses a lot of probes having low photon counts and CSSTORM overcounts probes with low photon counts ($<$ 500). These are likely false positive localizations. The error bars indicate standard deviations.

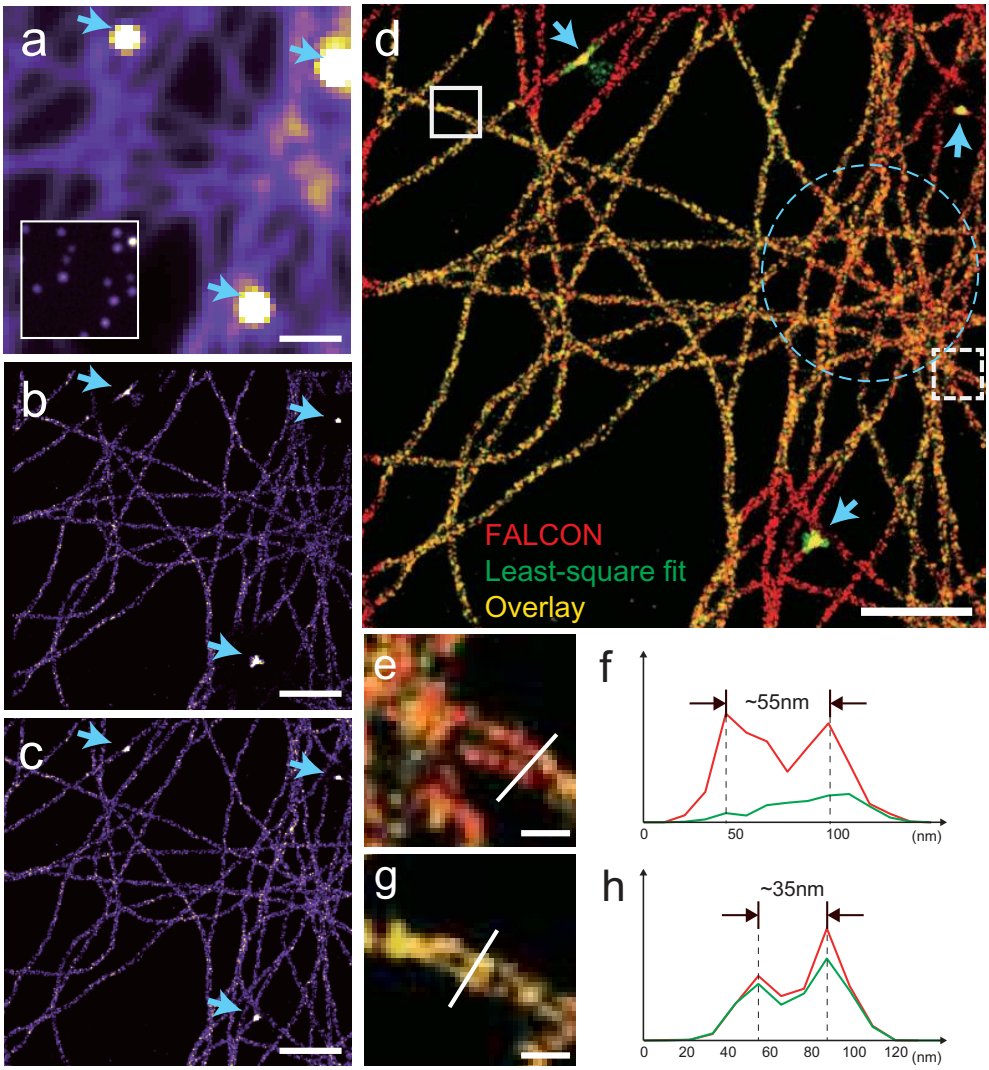

Figure 13 Low-density STORM images of microtubules (MTs) in a COS-7 cell. STORM dataset of 64 x 64 pixels x 10,000 frames recorded from microtubules (α-tubulin) stained with Alexa-647, and gold fiducial beads highlighted with blue arrows for de-drifting. (a) Wide-field image generated by summing all the frames. (b) STORM image reconstructed by Least-square (LS) fitting. (c) STORM image reconstructed by FALCON . (d) Overlay between the two images: LS fitting is missing molecules in the neighbouring areas of fiducial beads because the molecules too close to the fiducial beads are rejected, while the information is reserved by FALCON (red area around the blue arrows). Moreover, in the densest area of MTs without a fiducial bead denoted dotted blue circle, localization density by LS fitting is noticeably reduced due to the higher probability of PSF. However,

FALCON well resolved complex structures of MTs and has a more clear line profile (red curve) of two neighbouring MTs as can be seen in (e-f). We also took a line profile through MTs in lower density area where the LS fitting algorithm should perform optimally, and could resolve in both cases the hollowness of the structure, with two peaks distant from 35 nm consistent with a raw diameter of 25 nm augmented by the presence of the antibodies as can be seen in (h) where the green curve corresponds to LS fitting and the red curve to FALCON. Scale-bars are 1 μm in (a-d) and 100 nm in (e,g), respectively.

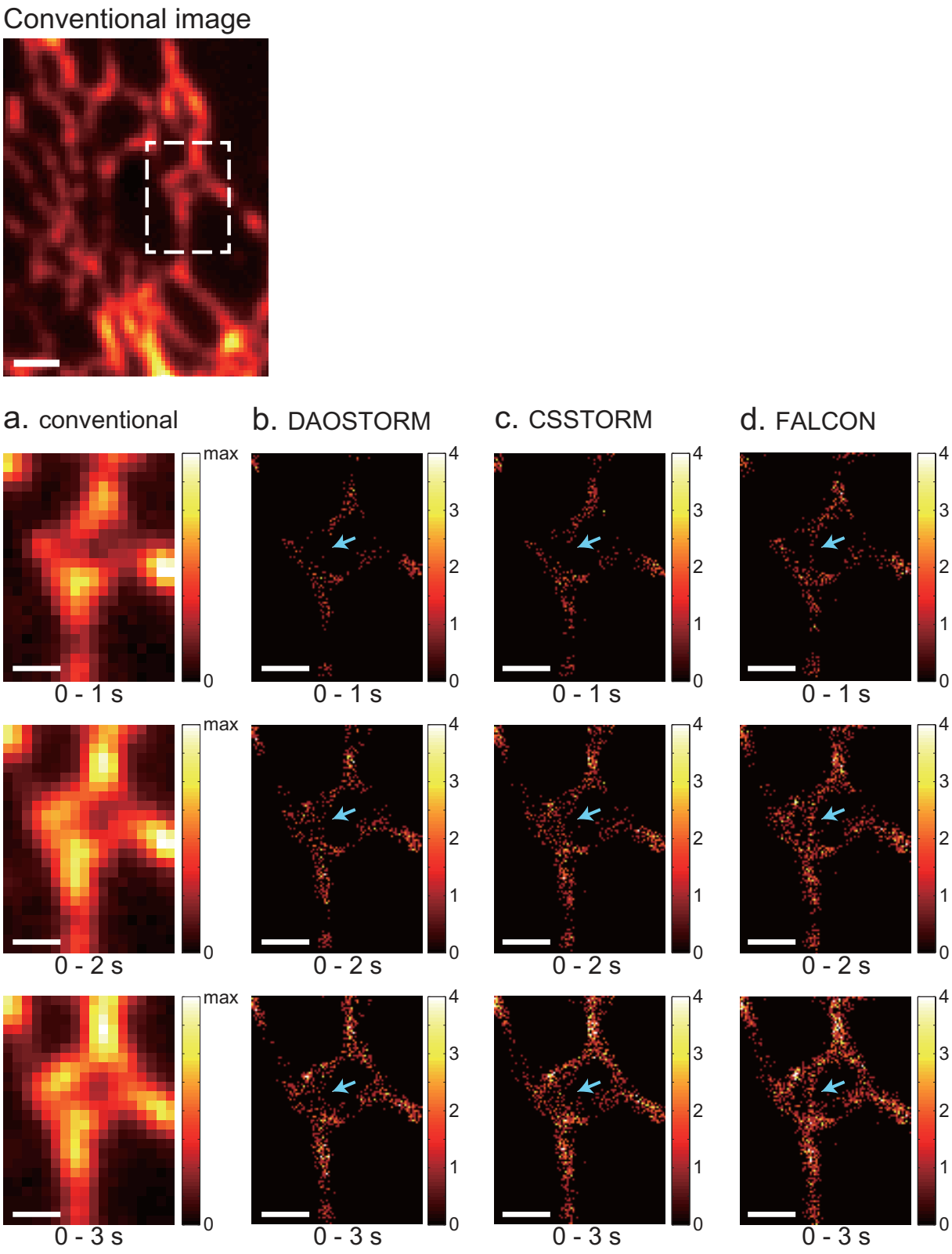

Figure 14 Super-resolution images reconstructed by DAOSTORM, CSSTORM and FALCON with live-imaging data of ER in U2OS cell. (a) conventional images and (b-d) reconstructed images by

DAOSTORM, CSSTORM and FALCON respectively with different time scales ($1s-3s$) from the region of dotted white box in the large conventional image on top. The reconstructed images generated as 2D histograms of 20 nm bins in (b-d). In comparison with DAOSTORM and CSSTORM, FALCON detected more molecules up to 50$\%$, which can improve temporal resolution. The tubular structures of ER only well preserved by FALCON are highlighted by cyanine arrows. Imaging of ER was performed at a frame rate of 64 fps. Scale-bars are 1 μm in the large conventional image and 500 nm in (a-d).