Deep learning massively accelerates super-resolution localization microscopy

Wei Ouyang, Andrey Aristov, Mickaël Lelek, Xian Hao, Christophe Zimmer

ANNA-PALM: Deep learning accelerates super-resolution localization microscopy by orders of magnitude

Wei Ouyang[1,2,3], Andrey Aristov[1,2,3], Mickaël Lelek[1,2,3], Xian Hao[1,2,3], Christophe Zimmer[1,2,3]

[1] Institut Pasteur, Unité Imagerie et Modélisation, Paris, France

[2] UMR 3691, CNRS

[3] C3BI, USR 3756, IP CNRS

Corresponding author: C. Zimmer

The speed of super-resolution microscopy methods based on single molecule localization, e.g. PALM or STORM, is severely limited by the need to record many thousands of frames with a low number of observed molecules in each. Here, we present ANNA-PALM, a computational strategy that uses artificial neural networks to reconstruct super-resolution views from sparse, rapidly acquired localization images and/or widefield images. Simulations and experimental imaging of microtubules, nuclear pores and mitochondria show that high-quality super-resolution images can be reconstructed from up to two orders of magnitude fewer frames than usually needed, without compromising spatial resolution. Super-resolution reconstructions are even possible from widefield images alone, though adding localization data improves image quality. We demonstrate super-resolution imaging of >1,000 fields of view containing >1,000 cells in ~3 h, yielding an image spanning spatial scales from ~20 nm to ~2 mm. The drastic reduction in acquisition time and sample irradiation afforded by ANNA-PALM enables faster and gentler high-throughput and live cell super-resolution imaging.

Fluorescence microscopy methods that overcome the diffraction limit of resolution (~200-300 nm) allow imaging of biological structures with molecular specificity closer to the molecular scale. Among super-resolution microscopy approaches, those based on single molecule localization, such as PALM[1] or STORM[2] (hereafter referred to collectively as PALM) are particularly attractive owing to their exquisite spatial resolution and ease of implementation. In these methods, random subsets of fluorophores are imaged in many consecutive diffraction-limited frames, computationally localized to high precision, and the combined localizations are used to generate a super-resolution view. In practice, typically 10^3-10^5 diffraction-limited frames are needed to assemble a single super-resolution image. This requirement follows from two conditions that must be simultaneously satisfied to ensure high spatial resolution: (i) a low number (~10-10^2) of active fluorophores per frame, to avoid overlaps between diffraction limited spots and enable precise localization of individual molecules, and (ii) a large number of independent localizations to ensure a sufficiently dense sampling of the underlying biological structures[3,4]. The large number of required frames makes localization microscopy inherently slow, thereby limiting its potential for high-throughput imaging, where many fields of view (FoVs) are to be imaged, and for imaging live cell dynamics. As a result, most localization microscopy studies are restricted to analyzing a small number of cells (typically less than ten).

Multiple approaches have been explored to accelerate localization microscopy. Using bright dyes with rapid switching kinetics, high power lasers and fast cameras allows to minimize exposure time without losing signal to noise

ratio[5,6], but reaching sub-millisecond exposure remains challenging, and intense irradiation exacerbates phototoxicity in live cell imaging[7,8]. Increasing the number of active fluorophores per frame can reduce acquisition time, but despite algorithms designed to handle overlapping fluorescent spots[9–13] this approach necessarily compromises spatial resolution[14,15].

Here, we introduce a computational strategy that allows reducing the total number of frames and independent localizations without losing spatial resolution. Unlike previous approaches, our method leverages the structural redundancy of most biological images to reconstruct high quality images from vastly under-sampled localization microscopy data. Our method leverages deep learning, which employs artificial neural networks (ANNs) to learn complex non-linear mappings between numerical inputs and outputs[16]. Accordingly, we call it 'artificial neural network accelerated PALM', or ANNA-PALM.

RESULTS

A deep learning approach to super-resolution image reconstruction

We aim to reconstruct a super-resolution image of approximately similar information content as a standard PALM acquisition (with K frames and N localizations) from a much smaller number of raw frames (k<<K) without changing the average density of localizations, ρ, i.e. from a much smaller number of total localizations ($n = \rho k \ll N = \rho K$). If PALM images are defined as 2D histograms of independent localizations, this task can be formulated as restoring an image corrupted by Poisson noise (and potentially additional forms of noise). Image

restoration is an ill-posed problem that has an infinity of solutions in the high-dimensional space of all possible images, unless additional constraints (priors) are imposed that restrict the solution to a lower dimensional subspace. Suitable subspaces exist because most natural images are highly redundant, and can be represented to very good approximation with a much smaller number of coefficients than pixels, via appropriate functions that map feature space to pixel space[17,18]. In recent years, ANNs with multiple layers (deep nets) have proven very successful at learning meaningful features and non-linear mappings for image classification, segmentation, restoration and many other tasks[16,18,19]. Inspired by these developments, we designed ANNA-PALM, a deep learning approach for restoring super-resolution views from under-sampled (sparse) localization microscopy data.

ANNA-PALM comprises a training stage and an inference stage (**Figure 1**). For training (**Figure 1a**), a few super-resolution images representative of the structure of interest (e.g. microtubules, nuclear pores, or mitochondria) are obtained using standard PALM imaging, i.e. by acquiring long diffraction limited image sequences (e.g. $K{\sim}10^4 - 10^5$, $N{\sim}10^5 - 10^7$) and processing them with standard localization software[20], resulting in highly sampled (dense) PALM images. In addition, a low resolution (widefield) image can also be acquired, as is commonly done before single molecule imaging when bleaching out preactivated fluorophores. Next, the dense PALM images are under-sampled by using a much smaller number of input frames, $k \ll K$, thus yielding sparse PALM images from the same localization data. Then, an ANN is trained to recover approximations of the dense PALM images from these sparse PALM images (and the optional widefield image).

Once trained, the ANN is applied to new sparse PALM images (with or without a widefield image), obtained from new image sequences with small numbers of frames ($k \ll K$) -and hence in much shorter time- in order to reconstruct high quality super-resolution images not previously seen (inference, **Figure 1b**).

Neural net architecture and learning strategy

Our ANN, hereafter called A-net, contains a total of 25 convolutional layers, and roughly 42 million trainable parameters. A-net is adapted from the pix2pix network[21], which itself builds on U-nets[22] and generative adversarial networks (GANs)[23], two recent successful deep learning techniques. U-nets are special types of convolutional neural networks (CNNs) that have proven effective at learning multi-scale representations of images and accurate, pixel-wise mappings[22,24]. GANs can generate new samples from real image distributions using a generator network that outputs synthetic images, and a discriminator network that outputs the probability that an input image is real or synthetic, both networks being trained simultaneously to compete against each other[23]. Importantly, the generator can be conditioned on input data (conditional GAN, or cGAN)[21,23], e.g. on images as in the pix2pix network. We modified the pix2pix architecture to accept a computational switch as additional input to handle multiple types of data, and introduced an additional network to evaluate the consistency between the reconstructed image and the widefield input image.

Training of our A-net proceeds as follows. Randomly under-sampled (i.e. sparse) versions of PALM images are fed as input to the A-net, while the

corresponding dense PALM images are defined as the A-net's targets, i.e. desired outputs (**Figure 1a**). Additional, optional inputs are widefield images, if available, and the switch, which indicates the image type when multiple types of images (e.g. microtubules and nuclear pores) are used during training. ANN training requires defining an objective function (also called loss), which measures how well the outputs match the targets. We implemented a loss function containing three terms. The first term measures the difference between the A-net output and the dense PALM image. Instead of the widely used mean squared error (MSE), which poorly reflects visual quality[25], we used a combination of the absolute difference (L1 norm) with a multi-scale version of the structural similarity index, a perceptually motivated quality metric shown to improve image restoration with deep learning (MS-SSIM)[26]. The second term measures the consistency between the A-net output and the widefield image. Although in theory the latter should simply be a blurred version of the PALM image, this is often not the case in practice[27]. Therefore, we introduced another CNN (with 4-layers), called 'low resolution estimator' to predict the widefield image from the super-resolution image. The corresponding loss was defined as the MS-SSIM between this CNN's output and the observed widefield image. (In absence of a widefield image, this loss is set to zero). The third term contains a cGAN discriminator loss[21], where the discriminator is a 5-layer CNN, whose inputs are the sparse PALM (and widefield) image(s) and either the dense PALM image or the output of the generator above; the discriminator's output is compared to 0s and 1s (for synthetic and real, respectively), respectively, via the MSE. We use dropout[28] and extensive data augmentation, including random

rotations, translations, elastic deformations and addition of noise in the input image to mimic false detections and unspecific labeling. As a result, only a few dense PALM images are required for successful training without overfitting. On graphical processing units (GPU), training ANNA-PALM from scratch takes on the order of hours to days, but when starting from a previously trained A-net, retraining can be done in an hour or less.

Once trained, the A-net can take sparse localization data with an optional widefield image as input(s), and output a reconstructed super-resolution image in less than a second (**Figure 1b**). In addition, the A-net produces an 'error map' that measures the consistency of this super-resolution image with the widefield image[27] (when available) and can be used to estimate the degree of reliability and highlight potential reconstruction artifacts. For more details, see **Online Methods** and **Supplementary Note 1**.

Validating ANNA-PALM on simulated images

We first sought to validate ANNA-PALM on synthetic data. For this, we used Brownian dynamics simulations[29] to generate 200 dense PALM images of semi-flexible filaments mimicking microtubules, with a resolution $R_{\mathrm{loc}} \approx 23$ nm. These represent "perfect" PALM images that would be obtained with an infinite number of localizations. We applied varying levels of Poisson noise to these perfect images to create sparse PALM images corresponding to finite numbers of localizations (**Supplementary Figure 1**). We then trained our A-net using the perfect images as

targets and the sparse images (and widefield image) as inputs, varying the number of localizations over a large range.

Next, we applied the trained A-net to a distinct set of PALM images generated by the same stochastic simulation (**Supplementary Figure 2**). **Figure 2a** shows a widefield image and **Figure 2b** a corresponding sparse PALM image obtained from n=6,834 localizations. Although curvilinear structures can be seen in this image despite its sparsity, small-scale features remain highly ambiguous (**Figure 2b,** inset), and the resolution according to a recently proposed five-fold Nyquist criterion[3] is limited by sampling to $R_{5\times\mathrm{Nyq}} \approx 85$ nm; according to this criterion, $N > N_{5\mathrm{xNyq}} = 60{,}000$ localizations are needed to achieve 23 nm resolution (**Supplementary Figure 3d**). **Figure 2c** shows the ANNA-PALM image reconstructed from the wide-field image alone, which exhibits clear and continuous filaments that were not previously recognizable. Most of the relatively isolated filaments roughly agree with the perfect PALM image (**Figure 2e**). In the denser regions, however, many small features are erroneous, e.g. filaments are incorrectly joined, displaced, split or merged (**Figure 2c,** blue arrows and **Supplementary Figure 4, top**). By contrast, the ANNA-PALM image reconstructed from the sparse PALM image alone or in combination with the widefield image exhibits continuous and sharp filaments in very good agreement with the perfect PALM image (**Supplementary Figure 3b,c** and **Figure 2d,e,f**). The spatial resolution of these reconstructed images is limited neither by diffraction nor sampling, but only by the localization precision, and is thus $R_{\mathrm{loc}} \approx 23$ nm, as in the perfect images (**Supplementary Figure 3e,f**). These results indicate that high quality super-

resolution images can be obtained from only a small fraction of the number of localizations traditionally required (here, approximately ~11% of N_{5xNyq} above; see **Supplementary Figure 3d**), hence enabling a strong reduction in acquisition time. Nevertheless, reconstruction errors can still occur in areas where the sparse localization data are most ambiguous, e.g. where filament density is highest (**Figure 2d,e,f,** white arrow). These errors can be reduced by increasing the localization number n, implying a trade-off between acquisition time and reconstruction quality (**Supplementary Figure 4**).

To quantify this trade-off, we computed the MS-SSIM between reconstructed ANNA-PALM and perfect PALM images ($n = \infty$) as function of localization number, from n~200 to n~ 2x10^6, in comparison with the standard PALM images (**Figure 2g**). The MS-SSIM ranges from 0 to 1 and reaches 1 for perfect reconstructions. For standard PALM images, the MS-SSIM increases monotonically, as expected, from <0.2 to >0.95 for n = 2x10^6 million localizations (**Figure 2g**, black curve). Using only the sparse image as input, ANNA-PALM reconstructions achieve MS-SSIM that are consistently higher and increase with localization number n much more rapidly than standard PALM, already exceeding 0.9 for $n \approx$ 10,000 localizations (**Figure 2**, dashed blue curve). ANNA-PALM achieves the same MS-SSIM as standard PALM at the five-fold Nyquist sampling level (≈0.65) with only n =2,248 localizations instead of n =58,588, suggesting a ~26-fold speed-up. If the widefield image is used as additional input, the MS-SSIM further increases, and dramatically so for low localization numbers (**Figure 2g**, solid blue curve). For example, with n =7,477 localizations, ANNA-PALM achieves a MS-SSIM (≈0.95)

similar to standard PALM with n =644,844, implying a speed-up of roughly two orders of magnitude. (Note that, if the perfect PALM image was not available for these quantifications, it could be replaced by the ANNA-PALM reconstruction of a dense PALM image with a large number of localizations, e.g. $n = 10^5$, with similar results- see **Supplementary Figure 5**).

As any image restoration method, ANNA-PALM can make errors. The low resolution error map described above (**Figure 1b**) provides a means to estimate where errors are most likely to occur. When applied to ANNA-PALM reconstructions of a sparse PALM image, this error map highlights regions containing the highest density of filament crossings, where reconstructions tend to be least accurate (**Supplementary Figure 6i,k**). If we artificially displace a small piece of filament in this image to simulate a false positive and a false negative in the reconstruction (**Supplementary Figure 6b,d,** white and blue arrows, respectively), the affected regions also light up in the error map (**Supplementary Figure 6j,l**). Thus, the error map offers a useful tool to highlight regions most likely to contain reconstruction errors, and conversely, to outline regions where reconstructions are most trustworthy. Thus, simulations suggest that ANNA-PALM can considerably reduce acquisition time in localization microscopy and also map reconstruction reliability.

ANNA-PALM reconstructions of immunostained microtubules

We next tested our method on real images of immunolabeled microtubules (**Figure 3**). We trained our A-net on seven dense PALM images (with corresponding widefield images) obtained during 10 minute long acquisitions (K=60,000;

247 Δt=10 ms exposure time) (not shown). We then considered a sparse PALM image of
248 microtubules in a distinct FoV obtained from only 9 s of acquisition (*k*=300; Δt=30
249 ms), together with a widefield image (Δt=2x50 ms) (**Figure 3a,b**). Whereas
250 microtubule filaments can already be seen in this sparse PALM image, structural
251 details below the diffraction limit are hard to discern, making it difficult to follow
252 the path of individual filaments in the denser regions and to identify features such
253 as filament crossings (**Figure 3b**). By contrast, the ANNA-PALM images, whether
254 reconstructed from the widefield image alone, the sparse PALM image alone, or
255 both, all display sharp and continuous filaments and clearly reveal many structural
256 details (**Figure 3d-f**). Their resolution is similar to or even better than the dense
257 PALM image (**Supplementary Figure 7a**). As for the simulations, in regions where
258 microtubule filaments are isolated, the ANNA-PALM image reconstructed from the
259 widefield image alone is in good agreement with the dense PALM image
260 (**Figure 3d,g**). However, it is often incorrect in areas of high microtubule density
261 (e.g. **Figure 3d**,**g** white and gray arrows). Most of these reconstruction errors are
262 corrected when applying ANNA-PALM to the sparse PALM image instead
263 (**Figure 3e,h**). For example, parallel sections of two microtubules unresolved in the
264 widefield image and incorrectly merged in **Figure 3d** are now clearly separated and
265 positioned correctly, and missed portions of other filaments are now recovered
266 (**Figure 3h,** white and gray arrows). Counter-intuitively, the sparse PALM image
267 exhibits high signal in some locations where the dense PALM image does not,
268 presumably because of spurious localizations due e.g. to unspecific binding
269 (**Figure 3b**, blue arrow). Such signal can lead to incorrect features in the ANNA-

PALM reconstruction from the sparse localization data alone (**Figure 3e,h**, blue arrows). However, when combining the widefield and sparse PALM data, these artifacts are largely removed and reconstructions agree very well with the dense PALM image (**Figure 3f,i**). Reconstruction quality increases with the number of frames *k* (**Figure 3j, Supplementary Figures 8-9** and **Supplementary Video 1**). More quantitatively, a MS-SSIM analysis similar to that for the simulated data above (with the ANNA-PALM output of the dense PALM image defined as ground truth; see **Supplementary Figures 5, 9f**) suggests that ANNA-PALM allows a hundred-fold reduction of acquisition time compared to standard PALM (**Supplementary Figure 9g**). **Supplementary Figure 10** shows other examples of sparse microtubule images reconstructed by ANNA-PALM.

As for simulations above, we used the widefield image to compute an error map (**Supplementary Figures 6,11**). Bright areas in this error map highlight regions where the reconstruction indeed disagrees with the dense PALM image; conversely, reconstructions are of high quality in the majority of regions where the error map is dim (**Supplementary Figure 11d-f**). These results demonstrate experimentally that ANNA-PALM can restore high quality approximations of super-resolution images from much shorter acquisition time than typical for PALM imaging, and also predict where reconstruction errors are most likely.

ANNA-PALM enables high-throughput super-resolution imaging

The drastic improvement in imaging efficiency afforded by ANNA-PALM permits super-resolution imaging of orders of magnitude more cells and FoVs per unit time.

To demonstrate this, we used an automated acquisition protocol to image >1,000 cells with immunolabeled microtubules in 1,089 (33x33), partly overlapping, FoVs of 55.3 µm x 55.3 µm each (**Figure 4**, **Supplementary Figure 12**). We first acquired widefield images at each of these positions, in a total of ~12 minutes, mostly consisting of stage stabilization delays (**Supplementary Figure 12a**). Next, we obtained 1,089 sparse PALM images using only 10 s of imaging time per FoV (k=1,000 frames, Δt = 10 ms), in a total of only ~3.1 hours (**Figure 4a**). Neither the widefield nor the sparse PALM images provided much small scale information (**Figure 4c, Supplementary Figure 12 c,e**). However, ANNA-PALM reconstructions led to high quality super-resolution images, allowing to visualize the microtubule network with clarity and to distinguish microtubule filaments in dense areas that appeared as unstructured regions in the sparse PALM image (**Figure 4b,d**). The FWHM across filaments in the reconstructed image was ~51 nm (**Figure 4d**), within the range measured for the training data (**Supplementary Figure 7a**). Similar images can be obtained by ANNA-PALM using the widefield images alone (**Supplementary Figure 12**), although of lower quality, as discussed above. Stitching the reconstructed images together yielded a single super-resolution image that contained approximately seven billion 20x20 nm pixels and covered an area of 1.8 mm x 1.8 mm, thereby spanning almost five orders of magnitude in spatial scales (**Figure 4b,d, Supplementary Figure 12b** and **Supplementary Video 2**).

ANNA-PALM is robust to experimental perturbations

315 ANNA-PALM can reconstruct accurate super-resolution images from sparse data
316 because it was trained on similar images before. This raises the question of whether
317 an ANN trained in one experimental condition can be successfully applied to
318 another condition. To test this, we used the A-net as trained above to analyze
319 microtubule images of cells subjected to drugs affecting the cytoskeletal network.
320 We first treated U373 cells with 1 μM of Taxol, an antimitotic agent that inhibits the
321 depolymerization of microtubules and increases their bending rigidity[30,31].
322 **Figure 5d** shows a sparse PALM image of these cells (k=800, kΔt=8 s), in which the
323 microtubule network is barely recognizable. By contrast, the ANNA-PALM
324 reconstructions clearly display a complex microtubule network and agree well with
325 the dense PALM image obtained from K=60,000 frames ($K\Delta$t=10 min)
326 (**Figure 5d,e,f**). These images display a larger density of straighter and more
327 parallel filaments with less frequent crossings than in the untreated cells
328 (**Figure 5a-c**), consistent with microtubule stabilization and increased rigidity[31].
329 Next, we treated cells with 1 μM of Nocodazole, a drug that, on the opposite,
330 promotes microtubule depolymerization and is expected to more dramatically alter
331 the cytoskeletal network[32]. Again, whereas the sparse image obtained from k=300
332 frames ($k\Delta$t=3 s) contained little exploitable information (**Figure 5g**), the ANNA-
333 PALM reconstruction offered clear and detailed views of the disrupted microtubule
334 network (**Figure 5h**), exhibiting a much smaller number of filaments, with higher
335 curvature, than in untreated cells. These reconstructions were in good (though not
336 perfect) agreement with dense PALM images obtained from K=60,000 frames
337 ($K\Delta$t= 10 min) (**Figure 5i**). Thus, ANNA-PALM, when trained on microtubules in

one experimental condition, may be successfully applied to new experimental conditions without retraining, thereby highlighting the method's robustness to biologically relevant structural alterations.

We further asked if ANNA-PALM is robust to changes in technical imaging conditions. To address this, we performed localization microscopy on microtubules by simultaneously changing multiple important imaging parameters relative to the training data. Instead of PALM/STORM, we used DNA-PAINT a technique where single molecule detection relies on transient binding of fluorophore-conjugated short DNA strands to complementary, antibody-conjugated, DNA strands[33], rather than on fluorophore blinking. The continuously emitting freely diffusing dyes lead to higher background noise in DNA-PAINT compared to PALM/STORM. Moreover, we used primary mouse antibodies instead of rat antibodies, Cy3 dyes instead of Alexa-647 dyes, and an EMCCD instead of a sCMOS camera. Despite all these differences, when ANNA-PALM was applied without retraining on a sparse microtubule image (k=400 frames, kΔt=12 s) , the reconstructed image still agreed very well with the dense DNA-PAINT image obtained from K=60,000 frames ($K\Delta$t=30 min) (**Figure 5j-l**). These data demonstrate the high robustness of ANNA-PALM to changes in experimental imaging conditions.

ANNA-PALM adapts to different biological structures

To demonstrate that ANNA-PALM is not restricted to filamentary structure, we turned to nuclear pores, a very different biological structure, and another popular target of super-resolution imaging studies[34–36]. We retrained A-net simultaneously

on microtubule images and on a single PALM image of the nucleoporin gp210 in immunolabeled nuclear membranes of *Xenopus* frog eggs[34,36] (K=30,000). With the switch (**Figure 1**) set to microtubules ('MT'), this newly trained A-net can still reconstruct sparse images of microtubules as when trained exclusively on microtubule data (**Supplementary Figure 13a-c**). We then applied the same A-net with the switch set to nuclear pores ('NPC') to reconstruct a new sparse PALM image of gp210 obtained from the first k=3,000 frames (**Figure 6a**). The sparsity of this image makes it difficult to clearly distinguish individual nuclear pores. ANNA-PALM, however, reconstructs a much clearer image, containing many easily identifiable ring-like structures, as expected for nuclear pores[34] (**Figure 6b**), and in good agreement with the dense PALM image obtained from K=30,000 frames (even though the latter shows mostly incomplete, open rings, presumably due to suboptimal labeling) (**Figure 6c**). An automated procedure based on cross-correlation with a ring template indeed identified ~2.7 times more putative nuclear pores from the ANNA-PALM image than the sparse image (**Supplementary Figure 14a-c**). Moreover, computed pore locations were in good agreement with a PALM image of wheat germ agglutinin (WGA), a lectin that concentrates in the inner nuclear pore channel[36] (**Supplementary Figure 14 d-f**). These results show that ANNA-PALM can successfully analyze non-filamentary structures, when properly retrained, and that a single ANN, with a simple computational switch, can reconstruct very different types of structures.

Finally, we imaged TOM22, a protein of the mitochondrial outer membrane[37]. Whereas, at the resolution of our experiments, microtubules and nucleoporins are

essentially one-dimensional and zero-dimensional structures, mitochondrial membranes are two-dimensional surfaces. Furthermore, their complex 3D morphology might seem less predictable than filaments or nuclear pores, potentially hampering ANNA-PALM reconstruction. Despite these differences, after being trained on nine PALM images of TOM22 (with frame numbers ranging from K=24,000 to K=40,000), ANNA-PALM reconstructions of distinct sparse PALM images (k=400 frames) displayed mitochondrial morphologies in good agreement with the dense PALM images (**Figure 6d-f**) - although the protein's localization along the membrane was less well reproduced**.** Taken together, our results illustrate the versatility of ANNA-PALM and its applicability to images of very different structural content.

Discussion

We introduced ANNA-PALM, a computational method based on deep learning, that reconstructs high quality super-resolution images from sparse, rapidly acquired, single molecule localization data (and/or widefield images). Our method enables considerable gains in acquisition time compared to standard localization microscopy without increasing active fluorophore density, thereby preserving spatial resolution. In fact, ANNA-PALM even improves spatial resolution when applied to images of lower resolution than the training data (**Supplementary Figures 7a, 10**), and greatly diminishes the detrimental effect of drift. The improvement in imaging efficiency afforded by ANNA-PALM alleviates the incompatibility between high resolution and high-throughput microscopy by

enabling super-resolution imaging of thousands of cells within a few hours or even less (**Figure 4, Supplementary Figure 12**). This will facilitate super-resolution studies of rare events, cellular heterogeneity and of partly stochastic structures such as cytoskeletal polymers or chromosomes, whose characterization requires statistics on many configurations[38,39]. ANNA-PALM may also be beneficial for high-throughput imaging screens, e.g. of drug treatments or gene knock-outs[40–42]. In addition, we envision applications to super-resolution imaging of large samples by stitching together multiple images of spatially adjacent fields. The ability to generate images spanning many orders of magnitude in scale could be well adapted to expansion microscopy, a super-resolution technique that physically increases sample size, but often requires tiling many fields of view to image even a single cell[43,44]. With correlative microscopy[45], it might also be possible to train ANNA-PALM to reconstruct electron microscopy (EM) images from fluorescence images, potentially extending the method to molecular resolutions currently out of reach of localization microscopy. Adaptation of ANNA-PALM to 3D[15,46] and multi-color[47,48] localization microscopy should be relatively straightforward. Localization microscopy of cellular dynamics remains very challenging[3,49]. By using much fewer frames (or even only widefield images), ANNA-PALM could dramatically improve the temporal resolution of live cell localization microscopy without sacrificing spatial resolution or increasing phototoxicity and photobleaching. Thus, ANNA-PALM provides multiple novel avenues for multi-scale imaging beyond standard spatio-temporal resolution limits.

Neverthelesss, important caveats should be stressed. First, although ANNA-PALM can be applied successfully to very different types of images (**Figure 6**), the method fails in absence of statistical redundancies between molecular localizations, e.g. for entirely random distributions of molecules. Second, ANNA-PALM requires prior training on dense PALM images with structures similar to those in the images to be reconstructed. We showed that ANNA-PALM is robust, i.e. does not require retraining, for some experimentally induced changes in structures and variations in imaging parameters (**Figure 5**). Nevertheless, indiscriminate application of ANNA-PALM to very different structures without retraining, or incorrect setting of the switch, may result in artifacts (**Supplementary Figure 13h**). Third, even when applied to data similar to the training images, ANNA-PALM can produce errors -as any reconstruction method in a context of information scarcity. The frequency of errors can be reduced by increasing the number of recorded frames, at the cost of reduced acceleration (see **Figures 2,3** and **Supplementary Figures 4, 8, 9**). In addition, ANNA-PALM can use widefield images to estimate the reliability of reconstructions, thereby helping their interpretation, providing some protection against artifacts and indicating when retraining may be needed (**Supplementary Figures 11, 15**). Future work, e.g. using Bayesian deep learning[50], may provide additional assessments of reconstruction uncertainty and shed more light on the scope and limitations of our approach.

As a purely computational technique, ANNA-PALM does not necessitate any changes to existing microscopy systems, but only requires one or a few standard PALM images for training. To facilitate its adoption and future development, we

make our source code, an ImageJ plugin and a cloud-computing based web application available on https://annapalm.pasteur.fr/ along with sample data. Because the performance of deep learning methods improves with the amount and variety of training data, we designed our web application to enable sharing of data and trained ANNs. As ANNA-PALM will learn from an increasing large and diverse collection of images, we expect it to reach even higher accuracy or efficiency and to expand its scope of application in the future.

Acknowledgements

We thank the following colleagues for useful discussions and suggestions and/or critical reading of the manuscript: C. Leduc, S. Etienne-Manneville, S. Lévêque-Fort, N. Bourg, A. Echard, J-B. Masson, T. Rose, P. Hersen, F. Mueller, M. Cohen, Z. Zhang, and P. Kanchanawong. We also thank the four anonymous reviewers for their constructive criticism, which led to significant improvements of ANNA-PALM. We further thank O. Faklaris, J. Sellés and M. Penrad (Institut Jacques Monod), and F. Montel (Ecole Normale Supérieure de Lyon) for providing *Xenopus* nuclear pore data, B. Jian (Institut Pasteur) for TOM22 antibodies, and C. Leterrier for fixation protocols. We thank E. Rensen and C. Weber for help with experiments and suggestions, B. Lelandais for help with PALM image processing, J-B. Arbona for polymer simulations and J. Parmar for suggestions that led to the name ANNA-PALM. We thank the IT service of Institut Pasteur, including J-B. Denis, N. Joly, and S. Fournier, for access to the HPC cluster and relevant assistance, and T. Huynh for help with GPU computing. This work was funded by Institut Pasteur, Agence

Nationale de la Recherche grant (ANR 14 CE10 0018 02), Fondation pour la Recherche Médicale (Equipe FRM, DEQ 20150331762), and the Région Ile de France (DIM Malinf). We also acknowledge Investissement d'Avenir grant ANR-16-CONV-0005 for funding a GPU farm used in this work. A.A. and X.H. are recipients of Pasteur-Roux fellowships from Institut Pasteur. W.O. is a scholar in the Pasteur - Paris University (PPU) International PhD program.

Author contributions

W.O.: conceived method, developed ANNA-PALM software and web application, performed experiments and analyses. A.A., M.L., X.H.: performed experiments. C.Z.: conceived method, supervised project and wrote manuscript.

Competing financial interests statement

W.O. and C.Z. are listed as inventors on European patent application EP17306022 filed by Institut Pasteur.

REFERENCES

1. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science (80-.).* **313,** 1642–1645 (2006).
2. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3,** 793–795 (2006).
3. Legant, W. R. *et al.* High-density three-dimensional localization microscopy across large volumes. *Nat. Methods* **13,** 359–365 (2016).
4. Deschout, H. *et al.* Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11,** 253–266 (2014).
5. Jones, S. A., Shim, S. H., He, J. & Zhuang, X. Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods* **8,** 499–505 (2011).
6. Huang, F. *et al.* Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10,** 653–8 (2013).
7. Carlton, P. M. *et al.* Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16016–22 (2010).
8. Stelzer, E. H. K. Light-sheet fluorescence microscopy for quantitative biology. *Nat. Methods* **12,** 23–26 (2014).
9. Huang, F., Schwartz, S. L., Byars, J. M. & Lidke, K. A. Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomed. Opt. Express* **2,** 1377–93 (2011).
10. Burnette, D. T., Sengupta, P., Dai, Y., Lippincott-Schwartz, J. & Kachar, B. Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proc Natl Acad Sci U S A* **108,** 21081–21086 (2011).
11. Simonson, P. D., Rothenberg, E. & Selvin, P. R. Single-molecule-based super-resolution images in the presence of multiple fluorophores. *Nano Lett* **11,** 5090–5096 (2011).
12. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster STORM using compressed sensing. *Nat. Methods* **9,** 721–723 (2012).
13. Cox, S. *et al.* Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat Methods* **9,** 195–200 (2012).
14. Ram, S., Ward, E. S. & Ober, R. J. Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy. *Proc. Natl. Acad. Sci.* **103,** 4457–4462 (2006).
15. Deschout, H. *et al.* Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11,** 253–266 (2014).
16. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).
17. Michael, E. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. *Springer* (Springer Verlag, 2010).
18. Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science (80-.).* **313,** (2006).
19. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks* **61,** 85–117 (2014).
20. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12,** 717–724 (2015).

21. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with Conditional Adversarial Networks. *http://arxiv.org/abs/1611.07004* (2016).
22. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science,* **9351,** 234–241 (Springer International Publishing, 2015).
23. Goodfellow, I. *et al.* Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst. 27* 2672–2680 (2014). doi:10.1017/CBO9781139058452
24. Ciresan, D., Giusti Alessandro, Luca, G. & Schmidhuber, J. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. *Med. Image Comput. Comput. Interv. – MICCAI 2013* **8150,** (2013).
25. Wang, Z. W. Z. & Bovik, A. C. A. C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **26,** 98–117 (2009).
26. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss Functions for Image Restoration with Neural Networks. *IEEE Trans. Comput. IMAGING* **3,** 47–57 (2017).
27. Culley, S. *et al.* NanoJ-SQUIRREL: quantitative mapping and minimisation of super-resolution optical imaging artefacts. *bioRxiv* 158279 (2017). doi:10.1101/158279
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15,** (2014).
29. Arbona, J.-M., Herbert, S., Fabre, E. & Zimmer, C. Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations. *Genome Biol.* **18,** 81 (2017).
30. Arnal, I. & Wade, R. H. How does taxol stabilize microtubules? *Curr. Biol.* **5,** 900–908 (1995).
31. Wu, S. *et al.* Microtubule motors regulate ISOC activation necessary to increase endothelial cell permeability. *J. Biol. Chem.* **282,** 34801–8 (2007).
32. De Brabander, M., De May, J., Joniau, M. & Geuens, G. Ultrastructural immunocytochemical distribution of tubulin in cultured cells treated with microtubule inhibitors. *Cell Biol. Int. Rep.* **1,** 177–83 (1977).
33. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12,** 1198–1228 (2017).
34. Löschberger, A. *et al.* Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. *J. Cell Sci.* **125,** 570–5 (2012).
35. Szymborska, A. *et al.* Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341,** 655–8 (2013).
36. Sellés, J. *et al.* Nuclear pore complex plasticity during developmental process as revealed by super-resolution microscopy. *Sci. Rep.* **7,** 14732 (2017).
37. Bellot, G. *et al.* TOM22, a core component of the mitochondria outer membrane protein translocation pore, is a mitochondrial receptor for the proapoptotic protein Bax. *Cell Death Differ.* **14,** 785–794 (2007).
38. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin

folding for different epigenetic states. *Nature* **529,** 418–422 (2016).
39. Zhang, Z., Nishimura, Y. & Kanchanawong, P. Extracting microtubule networks from superresolution single-molecule localization microscopy data. *Mol. Biol. Cell* **28,** 333–345 (2017).
40. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464,** 721–727 (2010).
41. Beghin, A. *et al.* Localization-based super-resolution imaging meets high-content screening. *Nat. Methods* **14,** 1184–1190 (2017).
42. Ouyang, W. & Zimmer, C. The imaging tsunami: Computational opportunities and challenges. *Curr. Opin. Syst. Biol.* **4,** 105–113 (2017).
43. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science (80-.).* **347,** 543–548 (2015).
44. Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* (2017). doi:10.1038/nmeth.4261
45. de Boer, P., Hoogenboom, J. P. & Giepmans, B. N. G. Correlated light and electron microscopy: ultrastructure lights up! *Nat. Methods* **12,** 503–513 (2015).
46. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (80-.).* **319,** 810–813 (2008).
47. Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes. *Science (80-.).* 1146598v1 (2007).
48. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11,** 313–318 (2014).
49. 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).
50. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017) pre-proceedings* 5580–5590 (2017).

FIGURE LEGENDS

FIGURE 1: Overview of ANNA-PALM

ANNA-PALM consists of two main stages: (**a**) acquisition of training images using standard localization microscopy (PALM) followed by artificial neural network (ANN) training, and (**b**) reconstruction of super-resolution views and low resolution error maps from new sparse PALM and/or widefield images (inference). (**a**) Training images are obtained by acquiring one or a few long sequences, of $K \sim 10^3$-10^5 diffraction limited, single molecule image frames, as in standard PALM experiments; optionally, a widefield image W can also be acquired (top). The acquisition time for each image sequence is $K\Delta t$, where Δt is the single frame exposure time. Standard localization microscopy algorithms (grey 'localization' boxes) are used to generate super-resolution images. For each experiment, a highly sampled (dense) super-resolution image (PK) is generated using all (or in practice $\geq$95%) available K frames. Sparse PALM images (Pk) from the same experiment are obtained by using only $k << K$ frames. A switch (light blue) can be set to distinguish different types of structures, e.g. nuclear pore complexes ('NPC'), mitochondria ('Mito') or microtubules ('MT'). An ANN (labeled $\mathcal{G}$ for 'generator', orange) is trained by using the sparse PALM images Pk (plus an upsampled version L of the widefield image W and the switch setting) as inputs and the corresponding dense PALM image PK as target output. During training, the output of the generator $\mathcal{G}$ (Ak) is compared to the target image PK and the widefield image L (if available) via three loss, or error functions (gray bordered boxes): (i) the 'super-resolution reconstruction error'

measures the difference between the reconstructed image Ak and the target PK using a combination of the L1 norm and the MS-SSIM; (ii) the 'low resolution reconstruction error' measures the MS-SSIM between the low resolution image WA predicted from the reconstruction Ak and the low resolution image WP predicted from the target image PK. Images WA and WP are predicted using a second ANN, called low-resolution estimator (labeled Q, blue) that is trained to produce an approximation of the actual widefield image W based on the MS-SSIM metric; (iii) the 'conditional GAN error' uses a third ANN (labeled *D* for 'discriminator', red) that attempts to distinguish between real dense PALM images PK and the generator's output Ak. The combined loss functions are iteratively optimized using stochastic gradient descent. (**b**) A short sequence of diffraction limited single molecule images (with $k<<K$ frames, i.e. acquisition time $k\Delta t$), and an optional widefield image (W'), are acquired. Standard localization algorithms generate a sparse (under-sampled) PALM image (P'k). This sparse image P'k (and the upscaled widefield image L' and switch setting) are fed as inputs to the trained generator G, which outputs a reconstructed ANNA-PALM image (A'k). In addition, the low resolution estimator Q predicts a low resolution image WA', which can be compared to the input widefield image W' via the MS-SSIM to produce a low resolution error map (top).

FIGURE 2: Validation of ANNA-PALM on simulated images

(**a**) Simulated widefield image of microtubules. (**b**) Simulated sparse PALM image of microtubules with n= 6,834 localizations. (**c**) ANNA-PALM reconstruction using only the widefield image **a** as input. (**d**) ANNA-PALM reconstruction using both the

widefield image **a** and the sparse PALM **b** image as inputs. (**e**) Simulated "perfect" PALM image, equivalent to a PALM image with an infinite number of localizations (n=∞) and a resolution of 23 nm. This image was used to generate **a** (by convolution with a Gaussian kernel approximating the microscope point spread function) and **b** (by application of Poisson noise). (**f**) Merged image showing the perfect PALM image **e** in green and the ANNA-PALM reconstruction **d** in red. Note that the ANNA-PALM images **c,d** provide many high resolution details that are absent from the widefield image **a** and the sparse PALM image **b** and that are in good (**c**) or very good (**d**) agreement with the perfect PALM image **e**. Some reconstruction errors are highlighted by arrows. Blue arrows in panel **c** point to errors of ANNA-PALM reconstruction from the widefield image only, the white arrow in panel **d** points to an error of ANNA-PALM reconstruction from both widefield and sparse PALM images combined. Reconstruction errors diminish for larger numbers of localizations, n (**Supplementary Figure 4**). (**g**) Reconstruction quality of PALM and ANNA-PALM images, measured by the MS-SSIM with the perfect PALM image **e**, as function of localization number n. Black curve: reconstruction quality of the standard PALM images. Dashed blue curve: reconstruction quality of ANNA-PALM using the sparse PALM images as input. Solid blue curve: reconstruction quality of ANNA-PALM using both the sparse PALM and widefield images as inputs. Red dashed line: reconstruction quality of ANNA-PALM using the widefield images as input only. Dots are averages from 10 simulations; error bars show standard deviations. The vertical dashed orange line indicates the minimum number of localizations needed to achieve a resolution of $R_{5\times\mathrm{Nyq}}$ =23 nm according to the five-

fold Nyquist criterion[3]. The dashed grey line indicates the minimum number of localizations needed to achieve a double mean nearest neighbor distance less than 23 nm. ANNA-PALM reconstructions from sparse PALM images only (i.e. without widefield images) achieve the same average MS-SSIM as standard PALM at the five-fold Nyquist sampling limit with 26 times less localizations (blue double arrow). ANNA-PALM reconstruction quality is highest when using both widefield and sparse PALM images as inputs.

FIGURE 3: ANNA-PALM imaging of microtubules

ANNA-PALM reconstructions of a localization microscopy image of immunostained microtubules. (**a**) Widefield image. (**b**) Sparse PALM image obtained from the first 9 s of acquisition (k=300 frames, n=11,740 localizations). (**c**) Dense PALM image obtained from a 15 min long acquisition (K=30,000 frames, N=409,364 localizations). (**d**) ANNA-PALM reconstruction from the widefield image **a** only. (**e**) ANNA-PALM reconstruction from the sparse PALM image **b** only. (**f**) ANNA-PALM reconstruction from the widefield image **a** and sparse PLAM image **b** combined. In panels **b-f**, pixel values are linearly mapped to colors from the look-up table shown below. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, V_{max}=3, 24, 102, 102 and 102 for panels **b**, **c**, **d**, **e** and **f**, respectively. (**g-i**) Merged images comparing ANNA-PALM reconstructions from panels **d-f** to the dense PALM image **c**. ANNA-PALM reconstructions are shown in red, the dense PALM image in green. (**j**) Gradual improvement of image quality for increasing acquisition time $k\Delta t$, shown for the

area highlighted in the insets of panels **a-i**. Top row: sparse PALM images. Middle row: ANNA-PALM reconstructions from the sparse PALM images only (without widefield). Bottom row: ANNA-PALM reconstructions from the widefield and sparse PALM images combined. **Supplementary Video 1** shows the gradual increase in quality of PALM and ANNA-PALM images with increased acquisition time for the larger region of interest shown in panels **a-i.**

FIGURE 4: High-throughput imaging with ANNA-PALM

Application of ANNA-PALM to high-throughput imaging of a 1.8 mm x 1.8 mm area containing more than 1,000 cells. (**a**) Sparse PALM image of this area, constructed by assembling a mosaic of 33x33=1,089 sparse PALM images of individual fields of view, obtained from k=1,000 raw frames each (with Δt=10 ms exposure time per frame, i.e. in $k\Delta t$ =10 s). Total image acquisition time was 1,089x10s, i.e. ~3.1 hours. The sparsity of the image is not apparent at this large scale. (**b**) ANNA-PALM reconstruction of the image in **a**, obtained by assembling a mosaic of 1,089 individual reconstructions (one per field of view). (**c**) Magnified view of the green boxed region in **a**. The inset shows a further magnified view of the yellow boxed region, highlighting the sparsity of the image. (**d**) Same as **c**, but for the ANNA-PALM reconstruction. A line profile across a microtubule is shown, with a FWHM≈51 nm. Non-linear contrast adjustment was applied manually for panels **a** and **b,** with black corresponding to values of zero in both panels. In panels **c** and **d**, pixel values were linearly mapped to colors from the look-up table in **Figure 3;** Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, and

734 V_{max}=3 and 51 for panels **c** and **d**, respectively. See also **Supplementary Video 2** for
735 an animated 'zoom-in' highlighting the spatial scales covered by the assembled
736 image. See also **Supplementary Figure 12** for ANNA-PALM reconstructions of the
737 same area from the widefield images only.
738
739 **FIGURE 5: Robustness of ANNA-PALM to experimental perturbations**
740 This figure shows ANNA-PALM reconstructions using an ANN trained on PALM
741 images of microtubules in untreated cells and applied without retraining to sparse
742 localization images of microtubules in different experimental conditions: untreated
743 control cells (**a-c**); cells treated with 1 μM of Taxol (**d-f**); cells treated with 1 μM of
744 Nocodazole (**g-i**); untreated cells imaged with DNA-PAINT (**j-l**). (**a,d,g,j**) Sparse
745 localization images obtained from the first *k* frames of the acquired image sequence,
746 with *k*=500, 800, 300, and 400 for **a, d, g,** and **j**, respectively. (**b,e,h,k**) ANNA-PALM
747 reconstructions using the sparse localization images immediately to the left as
748 input. (**c,f,i,l**) Dense localization images obtained from *K*=60,000 frames. Pixel
749 values are linearly mapped to colors from the look-up table in **Figure 3**. Black and
750 white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels,
751 and V_{max}=10, 120, 90, 25, 150, 40, 18, 150, 50, 18, 120, and 200 for panels **a**, **b**, **c**, **d**,
752 **e**, **f**, **g, h**, **i**, **j**, **k**, and **l**, respectively.
753
754 **FIGURE 6: ANNA-PALM reconstructions of nuclear pores and mitochondria**
755 PALM and ANNA-PALM images of nuclear pores (**a**-**c**) and mitochondria (**d**-**f**).
756 (**a**) Sparse PALM image of the immunolabeled *Xenopus* nucleoporin gp210 obtained

from the first k=3,000 frames. Note that individual nuclear pores are hard to identify. (**b**) ANNA-PALM reconstruction of image **a**. (**c**) Dense PALM image obtained from all K=30,000 frames. (**d**) Sparse PALM image of the immunolabeled mitochondrial outer membrane protein TOM22, obtained from the first k=400 frames. (**e**) ANNA-PALM reconstruction of image **d**. (**f**) Dense PALM image obtained from all K=30,000 frames. Pixel values are linearly mapped to colors from the look-up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, and V_{max}=3, 51, 3, 3, 128, and 18 for panels **a**, **b**, **c**, **d**, **e**, and **f**, respectively

ONLINE METHODS

Artificial neural network

Our ANN, called 'A-net', is based on the pix2pix architecture[21], which is a special conditional generative adversarial network (cGAN)[23] for image to image "translation", i.e. mapping from one type of image to another. The A-net consists of three distinct neural networks: (i) a generator network G that produces the reconstructed super-resolution image, (ii) a network Q called 'low resolution estimator' that produces the low resolution error map, (iii) a cGAN discriminator network D that provides the adversarial loss (**Figure 1a**). The generator network G builds on the U-net architecture, and consists of an encoder-decoder network with skip connections[22] and 16 convolutional layers. Its inputs and outputs are image patches containing $(256m)$x$(256m)$ pixels, where m is an integer (we used m=1 or 2 but this can be adjusted for different sizes of CPU/GPU memory or input images). The input is a sparse PALM image, a widefield image upscaled to the same size (see below), and a computational switch number that allows the network to switch between different types of image structures (e.g. nuclear pores or microtubules). The switch setting is encoded numerically and coupled by convolutional operations into the A-net encoder. The output of the generator G is a reconstructed image (called ANNA-PALM reconstruction or ANNA-PALM image elsewhere) of the same size as the input images. The low resolution estimator Q has four convolutional layers. It takes the $(256m)$x$(256m)$ dense PALM image patch or the ANNA-PALM image patch as input and outputs a low resolution image with $(64m)$x$(64m)$ pixels.

The cGAN discriminator network D has five convolutional layers. Its inputs are three $(256m)$x$(256m)$ pixel image patches (the sparse PALM image, the widefield image, and either the ANNA-PALM reconstruction or the corresponding dense PALM image), plus the upscaled widefield image, and its output is a $(30m)$x$(30m)$ image whose pixel values indicate whether the corresponding input patch is real (i.e. an experimentally obtained conventional PALM image) or produced by the generator $\mathcal{G}$. All convolutional layers are followed by batch normalization[51]. Dropout layers[28] (with dropout probability p=0.5) are introduced in the central layers of the A-net generator and turned on during training, but switched off during inference. Activation functions are rectified linear units (ReLU) $x \to \sup(x, 0)$ or "leaky" ReLUs $x \to \sup(x, 0) + \inf(\varepsilon x, 0)$ with ε =0.2[52], except for the last layer of $\mathcal{G}$, which uses the hyperbolic tangent $x \to \tanh(x)$ and the last layer of Q, which uses a sigmoid function $x \to (1 + \exp(-x))^{-1}$. The A-net architecture is fully described in **Supplementary Note 1** and **Supplementary Tables 1-4**.

Training objectives and error map

Each of the three networks mentioned above ($\mathcal{G}$, Q, and D) is associated to a distinct objective function - also called loss- and hereafter noted $\mathcal{L}_{\mathcal{G}}$, $\mathcal{L}_{Q}$, and $\mathcal{L}_{D}$ respectively. These loss functions are specified in detail below. In the following equations, for notational simplicity, we designate the sparse input image as $\boldsymbol{S}$, the low resolution (widefield) input image as $\boldsymbol{W}$, the corresponding dense PALM image (i.e. the target) as $\boldsymbol{T}$, and the A-net's output as $\boldsymbol{A} = \mathcal{G}(\boldsymbol{S}, \boldsymbol{W})$ (in **Figure 1a**, images $\boldsymbol{S}$, $\boldsymbol{T}$ and $\boldsymbol{A}$ are labeled Pk, PK and Ak, respectively); the computational switch indicating the image

type is noted $\boldsymbol{M}$. Low resolution images produced by the low resolution estimator network Q from $\boldsymbol{A}$ and $\boldsymbol{T}$ are designated as $\boldsymbol{W}_A = Q(\boldsymbol{A})$ and $\boldsymbol{W}_T = Q(\boldsymbol{T})$, respectively.

The generator loss function $\mathscr{L}_{\mathcal{G}}$ is the sum of three terms. The first term of $\mathscr{L}_{\mathcal{G}}$ is the super-resolution reconstruction error, hereafter called $\mathscr{L}_{\text{SuperRes}}(\mathcal{G})$. This term penalizes the difference between the generator output $\boldsymbol{A}$ and the target image $\boldsymbol{T}$. Based on a recent analysis of loss functions for image restoration with neural networks[26], we defined this difference as a weighted average of two quantities: (i) the multi-scale structural similarity index (MS-SSIM) between $\boldsymbol{A}$ and $\boldsymbol{T}$ and (ii) a modification of the L1 norm, where the absolute difference between $\boldsymbol{A}$ and $\boldsymbol{T}$ is smoothed by a Gaussian kernel:

$$\begin{aligned}\mathscr{L}_{\text{SuperRes}}(\mathcal{G}) &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[\rho\left(1 - MS\_SSIM(\boldsymbol{A},\boldsymbol{T})\right) + (1\right. \\ &\left. - \rho)\langle G_\sigma * |\boldsymbol{A} - \boldsymbol{T}|\rangle\right] \\ &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[\rho\left(1 - MS\_SSIM(\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}),\boldsymbol{T})\right) + (1\right. \\ &\left. - \rho)\langle G_\sigma * |\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}) - \boldsymbol{T}|\rangle\right]\end{aligned}$$

where $\mathbb{E}$ denotes expectation; $p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})$ is the joint probability density of the sparse PALM images $\boldsymbol{S}$, dense PALM images $\boldsymbol{T}$, widefield images $\boldsymbol{W}$ and switch settings $\boldsymbol{M}$ from the training data set; $MS\_SSIM(\boldsymbol{A},\boldsymbol{T})$ is the multi-scale structural similarity index between $\boldsymbol{A}$ and $\boldsymbol{T}$; G_σ is a Gaussian smoothing kernel; $*$ denotes convolution; $|\boldsymbol{A} - \boldsymbol{T}|$ is the absolute difference image (i.e. pixel (i,j) has value $|\boldsymbol{A}(\text{i},\text{j}) - \boldsymbol{T}(\text{i},\text{j})|$ and $\rho \in [0,1]$ is a scalar weight that balances the relative contributions of MS-SSIM and the modified L1 norm and is set to $\rho = 0.84$ as in ref. [26].

The second term of $\mathscr{L}_{\mathcal{G}}$ is called $\mathscr{L}_{\text{LowRes}}(\mathcal{G}, Q)$ and measures the consistency between the low resolution images $\boldsymbol{W}_A$ and $\boldsymbol{W}_T$ predicted by the low resolution estimator network Q:

$$\mathscr{L}_{\text{LowRes}}(\mathcal{G}, Q) = \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[1 - MS\_SSIM(\boldsymbol{W}_A, \boldsymbol{W}_T)\right]$$

$$= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[1 - MS\_SSIM\big(Q\big(\mathcal{G}(\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{M})\big), Q(\boldsymbol{T})\big)\right]$$

Alternatively, in the above objective function, $\boldsymbol{W}_T$ can be replaced by the actually observed widefield image $\boldsymbol{W}$, although with our data this led to slightly lower reconstruction quality. The low resolution estimator network Q is trained simultaneously with the generator $\mathcal{G}$ to produce a low resolution image from the dense PALM image $\boldsymbol{T}$ that is consistent with the observed low resolution image $\boldsymbol{W}$. This training is done based on the following objective function:

$$\mathscr{L}_Q(Q) = \mathbb{E}_{(\boldsymbol{T},\boldsymbol{W})\sim p_{\text{data}}(\boldsymbol{T},\boldsymbol{W})}\left[1 - MS\_SSIM(Q(\boldsymbol{T}), \boldsymbol{W})\right]$$

Note that the reconstructed low resolution image $Q(\boldsymbol{T})$ is four times smaller than the dense PALM image $\boldsymbol{T}$, as described in **Supplementary Note 1**. Because the input widefield image $\boldsymbol{W}$ can have a different size, we use bilinear interpolation to resize $\boldsymbol{W}$ to the same size as $Q(\boldsymbol{T})$. (If needed, a scaling factor different from four can be obtained by adding or removing downsample layers in network Q). At inference, the low resolution estimator Q is also used to produce the error map, as shown in **Figure 1b**, and **Supplementary Figures 6, 11, 15**. This error map is defined as:

$$\boldsymbol{E}_Q(\boldsymbol{A}, \boldsymbol{W}) = \big(1 - MS\_SSIM(Q(\boldsymbol{A}), \boldsymbol{W})\big)(Q(\boldsymbol{A}) + \boldsymbol{W})$$

High (respectively low) values of the error map indicate large (respectively small) inconsistencies, between the reconstructed super-resolution image $\boldsymbol{A}$ and the observed widefield image $\boldsymbol{W}$.

The third term of $\mathscr{L}_{\mathcal{G}}$ draws from recent work on generative adversarial networks (GAN)[21,23,53] and is noted $\mathscr{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})$. In a GAN, a generator network $\mathcal{G}$ learns to transform random input vectors $\boldsymbol{z}$ (drawn from a probability density $p_{\mathrm{z}}(\boldsymbol{z})$) into new samples of a data probability density $p_{\mathrm{data}}(\boldsymbol{x})$. In our case, the data samples $\boldsymbol{x}$ are the dense PALM images $\boldsymbol{T}$. The generator $\mathcal{G}$ learns by working against a discriminator network $\mathcal{D}$ that simultaneously learns to discriminate between original data samples and samples generated by $\mathcal{G}$. Adversarial training thus consists in playing a minmax game such that $(\mathcal{G}^*,\mathcal{D}^*)=\arg\min_{\mathcal{G}}\max_{\mathcal{D}}\mathcal{L}_{\mathrm{GAN}}(\mathcal{G},\mathcal{D})$, with an objective function of the form[23]: $\mathcal{L}_{\mathrm{GAN}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\big[1-\mathcal{D}\big(\mathcal{G}(\boldsymbol{z})\big)\big]$, or equivalently by simultaneous optimization of two coupled loss functions:

$$\begin{cases}\mathcal{D}^*=\arg\max\limits_{\mathcal{D}}\big(\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\big[1-\mathcal{D}\big(\mathcal{G}(\boldsymbol{z})\big)\big]\big)\\ \mathcal{G}^*=\arg\min\limits_{G}\big(\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\big[1-\mathcal{D}\big(\mathcal{G}(\boldsymbol{z})\big)\big]\big)\end{cases}$$

In a conditional GAN (cGAN), the generator and the discriminator have an extra input vector $\boldsymbol{c}$ and the first objective function above becomes: $\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\mathrm{data}}(\boldsymbol{c},\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{c},\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\big[1-\mathcal{D}\big(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z})\big)\big]$, such that the generator learns a conditional probability density $p_{\mathrm{data}}(\boldsymbol{x}|\boldsymbol{c})$; and the second objective function likewise becomes $\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\big[1-\mathcal{D}\big(\mathcal{G}(\boldsymbol{z})\big)\big]$. In our A-net, we replaced the logarithmic losses above by least square losses[53], as they empirically yielded better results. Thus, we used the objective functions:

$\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\mathrm{data}}(\boldsymbol{c},\boldsymbol{x})}(\mathcal{D}(\boldsymbol{c},\boldsymbol{x})-1)^2+\mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\big[\mathcal{D}\big(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z})\big)\big]^2$ and

$\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\left[1-\mathcal{D}\left(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z})\right)\right]^2$. In our case, the input $\boldsymbol{c}$ is the sparse PALM image $\boldsymbol{S}$ combined with the upsampled version $\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ of the widefield image $\boldsymbol{W}$, where $\mathcal{B}$ denotes bilinear interpolation. Note that in practice the noise $\boldsymbol{z}$ in our ANN was introduced only through the use of dropout layers, as in the pix2pix implementation[21]. Thus, the objective functions are:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{D}) = \mathbb{E}_{\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M}\sim p_{\mathrm{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}(\mathcal{D}(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\boldsymbol{T})-1)^2 + \mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z}),\boldsymbol{S}\sim p_{\mathrm{data}}(\boldsymbol{S})}\left(\mathcal{D}\left(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})\right)\right)^2$$

and:

$$\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\mathrm{data}}(S,T,W,M)}\left(\mathcal{D}\left(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})\right)-1\right)^2$$

In the end, combining the three loss terms described above, we implemented the following optimization problem :

$$\begin{cases} Q^* = \arg\min_{Q} \mathcal{L}_Q(Q) \\ \mathcal{D}^* = \arg\max_{\mathcal{D}} \mathcal{L}_{\mathcal{D}}(\mathcal{D}) \\ \mathcal{G}^* = \arg\min_{\mathcal{G}}\left[\alpha\mathcal{L}_{\mathrm{SuperRes}}(\mathcal{G}) + \beta\mathcal{L}_{\mathrm{LowRes}}(\mathcal{G},Q) + \gamma\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})\right] \end{cases}$$

The weights α, β and γ are hyperparameters, which we set manually to $\alpha = 50$, $\beta = 25$ and $\gamma = 1$ for most experiments. In absence of widefield images $\boldsymbol{W}$, β was simply set to zero. The reported results are not very sensitive to these parameters.

We trained the A-net end-to-end using stochastic gradient descent (SGD) with Adam[54] and a batch size of 1 with 200,000 or more iterations (backpropagation steps). Our implementation was adapted from affinelayer's TensorFlow[55] implementation, which is ported from the Torch implementation of pix2pix[21]. Both network training and inference were performed on Tesla P100, Tesla M40, Tesla

K80 or GTX TitanXP graphical processing units (GPUs) from Nvidia. A-net training from scratch typically takes from hours to days on a single GPU. Once trained, the A-net takes only ~1 second or less to reconstruct a super-resolution image of 2560x2560 pixels (corresponding to an entire FoV). Training time could be further reduced by pretraining (or transfer learning), use of GPU clusters, or optimized data augmentation.

Experimental training images and data augmentation

Experimental training data are obtained from standard localization microscopy data (dense PALM images). To achieve good performance, ANNs generally necessitate large amounts of training data. However, ANNA-PALM typically requires PALM images from no more than 10 FoVs (of 55 µm x 55 µm each) and can even be trained with a single FoV. This is possible thanks to an extensive on-the-fly data augmentation strategy, as described below. Each of the dense PALM training images corresponds to a list of localizations $(f_i, x_i, y_i)_{i=1..n}$, where $f_i \in [1, K]$ is the index of the diffraction limited frame from which localization x_i, y_i originates, and K is the total frame number. PALM images are obtained as plain 2D histograms of these localizations with typical pixel sizes of 10-20 nm. From each list of localizations (corresponding to a dense PALM image of a single FoV), we generate 10-30 pairs of input and target images ($\boldsymbol{S}$,$\boldsymbol{T}$) for training. To define the target image $\boldsymbol{T}$, we take a random consecutive subset $[k_T, k_T + 0.95K]$ of 95% of all available K frames (k_T is chosen randomly between 0 and $0.05K$) and create the 2D histogram image based on localizations from those frames only, i.e. from all (x_j, y_j) such that $f_j \in$

$[k_T, k_T + 0.95K]$. To define the sparse input image $\boldsymbol{S}$, we take random subsets of 300-500 consecutive frames from the first half of the image sequence and similarly create a 2D histogram of the localizations from those frames only. When a widefield image $\boldsymbol{W}$ is available, this image must first be aligned with the corresponding dense PALM image $\boldsymbol{T}$. This is done using an FFT-based phase correlation algorithm[56] after histogram equalization of image $\boldsymbol{T}$, smoothing by convolution with a Gaussian Kernel of standard deviation 6 pixels, and resizing with bilinear interpolation to the same size as image $\boldsymbol{W}$. The registered widefield image $\boldsymbol{W}$ is then scaled up using bilinear interpolation to an image $\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ with the same size as image $\boldsymbol{T}$.

During training, for each iteration of SGD, we crop the images $\boldsymbol{S}$, $\boldsymbol{L}$ and $\boldsymbol{T}$ with a randomly placed 712x712 pixel sized region $\mathcal{R} = [x_{\min}, x_{\min} + 712] \times [y_{\min}, y_{\min} + 712]$. We then use random geometric transformations and apply them identically to the three images. Specifically, we rotate the images by a random angle between 0 and 360 degrees, apply elastic transformations [57], and then crop the center region of size 512x512 pixels. In addition to geometric transformations, we also introduce realistic noise from experimental background images. This is done by manually outlining regions of background in selected PALM training images, splitting these regions into small patches of 40x40 pixels, grouping them according to their summed pixel values, then assembling them into a larger image with the same size as $\boldsymbol{T}$. During training, these semi-synthetic noise images are randomly selected and added to the input image $\boldsymbol{S}$ without altering $\boldsymbol{T}$ or $\boldsymbol{W}$. Finally, we normalize the input image $\boldsymbol{S}$ by subtracting its mean and dividing by standard deviation. If a widefield image $\boldsymbol{W}$ is provided, its pixel values are scaled to a

minimum of 0 and a maximum of 1. Otherwise, ***W*** is replaced by an image containing zeros only. The target image ***T*** is truncated at a maximum value of 255 and then scaled to have a minimum of 0 and a maximum of 1. For the switch ***M***, we used an integer number to define the type of training images, e.g. 0 for microtubules, 1 for nucleoporins, 2 for mitochondria. When training on different types of images, e.g. microtubules and nucleoporins (**Supplementary Figure 13**), we assign the corresponding switch value to ***M*** and use it as additional input to the A-net together with images ***S*** and ***W***, as described in **Supplementary Note 1.**

Image simulations

Our procedure to simulate localization microscopy (PALM) images of microtubules is illustrated in **Supplementary Figure 1**. To simulate microtubule filaments, we used a Langevin dynamics simulation[29] that generates random configurations of semiflexible curves with a specified rigidity (persistence length), starting from a random initial configuration (**Supplementary Figure 1a,b**). The initial configurations were generated with a Python library named cpolymer and the Langevin dynamics was implemented using the molecular dynamics code LAMMPS[58]. Although the simulation generates 3D polymer chains, we only considered their 2D projections, consisting of N_p connected positions $(x_k, y_k)_{k=1..N_p}$. To obtain smooth filaments we further interpolated these connected segments using spline functions with the Scipy function scipy.interpolate.splev. Next, we turned these 2D curves into a grey scale image of 800x800 pixels, with an assumed pixel size of 7.8 nm, using the Python library Matplotlib. This image was further

convolved with a Gaussian kernel of standard deviation 1.5 pixels, resulting in a smooth image $I(i,j)$ as shown in **Supplementary Figure 1c,** and normalized to a probability density ($\sum\sum_{i,j} I(i,j) = 1$, with all $I(i,j) \geq 0$). This image was used to mimic a "perfect" PALM image of filaments corresponding to an infinite number of localizations ($n = \infty$). Such perfect images were used as targets during ANN training for simulated data and defined as ground truth for the quantification of reconstruction quality by MS-SSIM (**Figure 2g**). During training, we applied the same rotations and elastic transformations described for experimental data in the previous section.

Localization microscopy images obtained from a finite number of localizations $n < \infty$ (sparse PALM images $\boldsymbol{S}$), can be considered as a sampling of the probability density $I(i,j)$ with n samples. These images can therefore be simulated by applying Poisson noise to a rescaled version of the perfect PALM image, i.e.: $\boldsymbol{S}(\lambda, I) = \mathcal{P}(\lambda\, I/I_{\max})$, where $I_{\max}$ is the maximum value of I, $\mathcal{P}(\mu)$ denotes the Poisson probability distribution of mean μ and where the peak parameter λ controls the level of sampling. In order to simulate sparse PALM images for various levels of sampling, we varied the peak value λ following a log-normal distribution where $\ln(\lambda)$ has mean -0.5 and standard deviation 0.001 and applied Poisson noise using the numpy library function random.poisson. An example of a simulated sparse PALM image is shown in **Supplementary Figure 1d**. Besides finite sampling, localization microscopy images are corrupted by additional noise sources such as false detections from background noise due to out-of-focus light or unspecific binding of antibodies. To mimic this, we first created a probability density

$I_b = I * G_{\sigma_b}$ for the background noise by convolving I with a Gaussian kernel G_{σ_b} of large standard deviation σ_b=25 pixels, and applied Poisson noise with $\lambda = 0.06$. To create training images, we added this background noise image to the sparse PALM image $\boldsymbol{S}$ above (**Supplementary Figure 2**). We did not add background noise to the test images used during inference (**Figure 2**). To simulate the widefield images (**Figure 2a**), we first blurred the perfect PALM image by convolution with a Gaussian kernel of standard deviation 8 pixels, then added Gaussian noise with zero mean and standard deviation chosen randomly between 0.5 and 1.5.

For simulations of nuclear pore images (**Supplementary Figure 15**), we applied a similar procedure, except that the perfect PALM images were obtained by randomly distributing circles of diameter 150 nm in the plane (avoiding overlaps) and placing eight Gaussian spots (of standard deviation 1.7 pixels) at equal distance from each other on each circle to mimic the octogonal shape of nuclear pores.

Sample preparation

For microtubule imaging experiments (**Figures 3-5** and **Supplementary Figures 8-10**), except those using DNA-PAINT (**Figure 5j,l**), U-373 MG (Uppsala) cells were cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12; Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v) penicillin-streptomycin (Gibco), in a 5% CO_2 environment at 37°C on 18-mm cleaned coverslips in 12-well plates. 24 hours after plating, cells were pre-extracted for 10 s in 0.25% (v/v) Triton X-100 (Triton) in BRB80 (80 mM PIPES, 1 mM MgCl2, 1 mM EGTA, adjusted to pH 6.8 with KOH) supplemented with 4 mM EGTA, and

immediately fixed for 10 min with 0.25% (v/v) Triton + 0.5% Glutaraldehyde in BRB80, followed by reduction for 7 min with 0.1% NaBH4 solution in PBS and another washing step in PBS. Cells were directly incubated for 1h at room temperature in PBS with 1:500 rat alpha-tubulin antibodies (Bio-Rad MCA77G), followed by 3 washing steps with PBS, and then incubated for 45min in PBS with 1:500 anti-rat Alexa-647 conjugated secondary antibodies from donkey (Jackson ImmunoResearch Laboratories, ref. 712-605-153).

For the DNA-PAINT experiment on microtubules (**Figure 5k-m**), U-373 cells stuck on 18 mm diameter coverslips were fixed at 37°C with 4% PFA in PHEM buffer and permeabilized in 0.2% glutaldehyde. Next, cells were incubated for 1 h with 1:500 primary mouse antibodies against alpha-tubulin. The sample was washed 3 times in PBS, then incubated with 1:100 anti-mouse oligo-conjugated antibodies from Ultivue Kit 2 for DNA-PAINT imaging[33]. After washing the sample 3 times in PBS, and just before imaging, 2nM of complementary oligos coupled to Cy3 fluorophores were added to the sample.

Nuclear pore imaging data of gp210 and WGA (**Figure 6a-c** and **Supplementary Figure 14**) were kindly provided by J. Sellés and O. Falklaris and obtained from nuclear membranes of *Xenopus* frog eggs prepared as described previously[36].

For mitochondria imaging experiments (**Figure 6e-f**), COS7 cells were cultured under the same conditions as U-373 cells above using phenol-red free DMEM medium and fixed with 4% PFA in PBS for 10 min. The sample was blocked with 3% BSA in PBS for 20 min and immunostained with 1:500 mouse antibodies

against TOM22 (Sigma, ref. T6319) in wash buffer (PBS with 0.5% BSA) for 1 h. After extensive washing with wash buffer, the sample was incubated with 1:500 anti-mouse secondary antibodies from donkey conjugated to Alexa-647 dyes (Jackson ImmunoResearch Laboratories, ref. 715-605-151) in wash buffer for 30 min. After washing 5 times with wash buffer and 2 times with PBS, samples were post-fixed with 2% PFA in PBS for 10 min and washed 5 times with PBS.

For all localization microscopy experiments except DNA-PAINT, we used a photoswitching buffer[59] composed of 50 mM Tris-HCl + 10 mM NaCl + 10% (w/v) glucose + 168 AU/mL Glucose-Oxidase + 1404 AU/mL Catalase + 1% 2-Mercaptoethanol. For microtubule imaging experiments, we used this buffer to fill a square hole that was manually cut in a parafilm sheet, which was deposited on a rectangular coverglass. The round coverslips were sealed with nail polish.

Image acquisition in localization and high-throughput microscopy

We performed single molecule localization microscopy experiments (PALM/STORM and DNA-PAINT) on custom built microscopy systems, as previously described[59–61]. The system used for PALM/STORM imaging of microtubules is based on an inverted microscope body (Nikon Ti Eclipse) equipped with a either a 60x 1.49 NA oil immersion objective (Nikon) or a 60x 1.2 NA water immersion objective (Nikon) and with the Perfect Focus System active. A 642 nm wavelength laser with 500 mW power was used to excite Alexa-647 fluorophores and an AOTF (AA optics) was used to modulate laser excitation. Sequences of diffraction limited single molecule image frames were acquired either on a sCMOS camera (Hamamatsu ORCA-Flash4.0),

which can capture images of 2,042x2,042 pixels (for **Figures 4, 5a-i**), or on an EMCCD (Andor IXON ULTRA 897) with 512x512 pixels (for **Figures 3, 6d-f**). Both cameras were controlled by MicroManager software[62]. For experiments using the sCMOS camera, the effective pixel size was 108 nm and we used a 512x512 region of interest, which resulted in an imaged FoV of 55.3 µm x 55.3 µm. For experiments using the EMCCD camera, we used a 2x telescope and the effective pixel size was 107 nm, resulting in a FoV of 54.8 µm x 54.8 µm. The exposure time was set to Δt =10 ms or 30 ms per frame. The number of frames acquired ranged from k=1,000 (**Figure 4**) to K=60,000 (e.g. **Figure 5c**) per FoV.

For the DNA-PAINT experiment (**Figure 5j,l**), we used an inverted Nikon Ti-E Eclipse microscope equipped with a 100x 1.49 NA TIRF objective and with the Perfect Focus System active. A 561 nm wavelength laser with 500 mW power was used to excite Cy3 dyes. Highly inclined laser illumination was used to reduce out-of-focus background signal. Images were acquired on an EMCCD camera as above, with a 1.5 x telescope, resulting in an effective pixel size of 106 nm and a FoV of 54 µm x 54 µm. The sample was mounted in a magnetic sample holder filled with the imaging buffer provided with the Ultivue kit. Exposure time was set to Δt =30 ms and the EM gain of the EMCCD was set to 300. The laser power was increased until isolated fluorescent spots were observed. For the experiment shown in **Figure 5g,i**, K=60,000 frames were acquired.

The *Xenopus* nuclear pore data (**Figure 6a-c** and **Supplementary Figure 14**), were acquired on a Zeiss Elyra P.S.1 microscope as described previously[36].

For high-throughput imaging of microtubules (**Figure 4**), we used the Multi-Dimensional Acquisition tool in Micro-manager to define the positions of 1,089 FoVs of 55.3 µm x 55.3 µm on a 33x33 grid, with overlaps of 1 µm; the stage was automatically shifted to each of these 1,089 positions. We first acquired only widefield images, taking five frames at each of these positions (the first two were ignored because of motion blur), in a total acquisition time of 12 minutes. Then, the laser power was raised to bleach out preactivated molecules and k=1,000 frames of single molecule images were acquired at each of the 1,089 positions, in a total acquisition time of 3 hours and 8 minutes. Raw image frames were written directly to a remote storage server via Samba networking protocol.

Localization microscopy image analysis

The input to ANNA-PALM reconstruction is a localization image, defined as a 2D histogram of n single molecule positions $(x_i, y_i)_{i=1..n}$. The histogram bin, i.e. the pixel size of the localization image, was set to 7.8 nm for the simulated data (**Figure 2** and **Supplementary Figures 1-6,15**) and 20 nm for the experimental data (**Figures 3-6** and **Supplementary Figures 7-14**). The positions $(x_i, y_i)_{i=1..n}$ were obtained by analyzing sequences of diffraction limited frames using standard single molecule localization algorithms. For experimental microtubule images, we used the ThunderSTORM[63] plugin of ImageJ, applying wavelet filters for detection and weighted least squares Gaussian fitting for precise estimation of subpixelic positions. We used the cross-correlation feature in ThunderSTORM for drift correction, and filtered out the least certain localizations based on the fitted

Gaussian's standard deviation and the χ^2 of the residual. Localizations in consecutive frames separated by less than 20 nm were assumed to originate from the same molecule and merged into a single localization. The final number n of localizations was ~7 million for the full 55 µm x 55 µm FoV of the images shown in **Figure 3** and **Supplementary Figures 8,9** (obtained from K=30,000 frames). For the high-throughput experiment (**Figure 4** and **Supplementary Figure 12**), the number of localizations per 55 µm x 55 µm FoV ranged from n=2,949 to n=1,442,048 with an average $\langle n \rangle = 610{,}983$ and standard deviation $\sigma(n) = 273{,}606$. The total number of localizations across all 1,089 FoVs was ≈665 million. ThunderSTORM analyses were performed either on high end workstations or on Institut Pasteur's high performance computer (HPC) cluster. For the high-throughput experiments, we used Python scripts to run ThunderSTORM in batch mode (without user intervention) on the HPC cluster and assembled mosaic images (**Figure 4a,b** and **Supplementary Figure 12a,b**) using a stitching plugin of ImageJ[64].

The nuclear pore images were analyzed using the ZEN software from Zeiss as previously described[36]. For the DNA-PAINT experiments, we used PALMTT, a modified version of the single molecule tracking algorithm MTT[65], based on Matlab (Mathworks). This algorithm uses Gaussian smoothing and thresholding for detection, and Gaussian fitting for precise estimation of subpixelic positions. Drift correction was performed computationally by tracking fluorescent beads used as fiducial markers.

Quality metrics and sampling resolution

In order to quantitatively assess the quality of PALM images and ANNA-PALM reconstructions, we calculated the multi-scale structural similarity index (MS-SSIM) between either image and the ground truth (**Figure 2g** and **Supplementary Figures 5,9**). For the simulated data, the ground truth was simply defined as the "perfect" PALM image, corresponding to an infinite number of localizations (see 'Image simulations' above, **Figure 2e** and **Supplementary Figure 1c**). For the experimental data, the ground truth was defined as the ANNA-PALM reconstruction of a dense PALM image obtained from all available frames (e.g. **Supplementary Figure 9c**). Before calculation of the MS-SSIM, all simulated images were linearly normalized without clipping to a maximum value of 255.

In order to evaluate the effect of sampling on the resolution of PALM images, we computed the double mean distance, $R_{\mathrm{Nyq}} = 2\langle d \rangle$ between nearest neighbors in the underlying sets of localizations[49] as function of localization number (**Supplementary Figure 3d**). For the simulated sparse PALM data, sets of localizations were obtained by interpreting each image $\boldsymbol{S}$ as a 2D histogram of localizations, and creating a random subpixelic position $(x_k, y_k)_{k=1..m}$ within each pixel (i, j), as many times as given by the pixel value $m = S(i, j)$ (therefore resulting in a set of $n = \sum_{i,j} S(i, j)$ localizations). The quantity $R_{\mathrm{Nyq}}(n) = 2\langle d(n) \rangle$ decreases towards zero with increasing number of localizations, n. A Nyquist criterion introduced in ref.[49] suggests that the resolution is limited by sampling to no less than R_{Nyq}, implying that at least $n \geq N_{\mathrm{Nyq}} = R_{\mathrm{Nyq}}^{-1}(R)$ localizations are needed to achieve a given resolution R. However, a more stringent and realistic criterion[3]

prescribes a five-fold larger number of localizations to reach resolution R: $n \geq N_{5\times\mathrm{Nyq}} = 5 \times N_{\mathrm{Nyq}}$, implying that the sampling limit to resolution is $R_{5\times\mathrm{Nyq}} = 2\langle d(n/5)\rangle$. Accordingly, if this condition is met, the resolution is no longer limited by sampling, but by the localization precision, $R_{\mathrm{loc}} \approx 2.3\sigma_{\mathrm{loc}}$, where σ_{loc} is the standard deviation of localization errors along each coordinate. In general, the resolution of a PALM image, as limited by both sampling and localization precision, can be written: $R = \max\left(R_{5\times\mathrm{Nyq}}, R_{\mathrm{loc}}\right) = \max(2\langle d(n/5)\rangle, 2.3\sigma_{\mathrm{loc}})$.

Life Sciences Reporting Summary.

Further information is available in the Life Sciences Reporting Summary.

Code availability statement

The source code of ANNA-PALM is available from https://annapalm.pasteur.fr/ .

Data availability statement

The localization data used in this paper can be downloaded directly from https://annapalm.pasteur.fr/.

Methods-only references

51. Xu, L., Choy, C. S. & Li, Y. W. Deep sparse rectifier neural networks for speech denoising. in *2016 International Workshop on Acoustic Signal Enhancement, IWAENC 2016* (2016). doi:10.1109/IWAENC.2016.7602891
52. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proc. 30 th Int. Conf. Mach. Learn.* (2013).
53. Mao, X. *et al.* Least Squares Generative Adversarial Networks. in *ICCV* 2794–2802 (2017). doi:10.1109/ICCV.2017.304

54. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ICLR* 1–15 (2015). doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503
55. Rampasek, L. & Goldenberg, A. TensorFlow: Biology's Gateway to Deep Learning? *Cell Syst.* **2,** 12–14 (2016).
56. Srinivasa Reddy, B. & Chatterji, B. N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5,** 1266–1271 (1996).
57. Simard, P. Y., Steinkraus, D. & Platt, J. C. Best practices for convolutional neural networks applied to visual document analysis. *Seventh Int. Conf. Doc. Anal. Recognition, 2003. Proceedings.* **1,** 958–963 (2003).
58. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117,** 1–19 (1995).
59. van de Linde, S. *et al.* Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat. Protoc.* **6,** 991–1009 (2011).
60. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat. Methods* **7,** 339–340 (2010).
61. Lelek, M. *et al.* Superresolution imaging of HIV in infected cells with FlAsH-PALM. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8564–9 (2012).
62. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using manager. *Current Protocols in Molecular Biology* (2010). doi:10.1002/0471142727.mb1420s92
63. Ovesny, M., Kiek, P., Borkovec, J., Vindrych, 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).
64. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25,** 1463–1465 (2009).
65. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods* **5,** 687–694 (2008).

ANNA-PALM: Deep learning accelerates super-resolution localization microscopy by orders of magnitude

Wei Ouyang[1,2,3], Andrey Aristov[1,2,3], Mickaël Lelek[1,2,3], Xian Hao[1,2,3], Christophe Zimmer[1,2,3]

[1] Institut Pasteur, Unité Imagerie et Modélisation, Paris, France

[2] UMR 3691, CNRS

[3] C3BI, USR 3756, IP CNRS

Corresponding author: C. Zimmer

The speed of super-resolution microscopy methods based on single molecule localization, e.g. PALM or STORM, is severely limited by the need to record many thousands of frames with a low number of observed molecules in each. Here, we present ANNA-PALM, a computational strategy that uses artificial neural networks to reconstruct super-resolution views from sparse, rapidly acquired localization images and/or widefield images. Simulations and experimental imaging of microtubules, nuclear pores and mitochondria show that high-quality super-resolution images can be reconstructed from up to two orders of magnitude fewer frames than usually needed, without compromising spatial resolution. Super-resolution reconstructions are even possible from widefield images alone, though adding localization data improves image quality. We demonstrate super-resolution imaging of >1,000 fields of view containing >1,000 cells in ~3 h, yielding an image spanning spatial scales from ~20 nm to ~2 mm. The drastic reduction in acquisition time and sample irradiation afforded by ANNA-PALM enables faster and gentler high-throughput and live cell super-resolution imaging.

Fluorescence microscopy methods that overcome the diffraction limit of resolution (~200-300 nm) allow imaging of biological structures with molecular specificity closer to the molecular scale. Among super-resolution microscopy approaches, those based on single molecule localization, such as PALM[1] or STORM[2] (hereafter referred to collectively as PALM) are particularly attractive owing to their exquisite spatial resolution and ease of implementation. In these methods, random subsets of fluorophores are imaged in many consecutive diffraction-limited frames, computationally localized to high precision, and the combined localizations are used to generate a super-resolution view. In practice, typically 10^3-10^5 diffraction-limited frames are needed to assemble a single super-resolution image. This requirement follows from two conditions that must be simultaneously satisfied to ensure high spatial resolution: (i) a low number (~10-10^2) of active fluorophores per frame, to avoid overlaps between diffraction limited spots and enable precise localization of individual molecules, and (ii) a large number of independent localizations to ensure a sufficiently dense sampling of the underlying biological structures[3,4]. The large number of required frames makes localization microscopy inherently slow, thereby limiting its potential for high-throughput imaging, where many fields of view (FoVs) are to be imaged, and for imaging live cell dynamics. As a result, most localization microscopy studies are restricted to analyzing a small number of cells (typically less than ten).

Multiple approaches have been explored to accelerate localization microscopy. Using bright dyes with rapid switching kinetics, high power lasers and fast cameras allows to minimize exposure time without losing signal to noise

ratio[5,6], but reaching sub-millisecond exposure remains challenging, and intense irradiation exacerbates phototoxicity in live cell imaging[7,8]. Increasing the number of active fluorophores per frame can reduce acquisition time, but despite algorithms designed to handle overlapping fluorescent spots[9–13] this approach necessarily compromises spatial resolution[14,15].

Here, we introduce a computational strategy that allows reducing the total number of frames and independent localizations without losing spatial resolution. Unlike previous approaches, our method leverages the structural redundancy of most biological images to reconstruct high quality images from vastly under-sampled localization microscopy data. Our method leverages deep learning, which employs artificial neural networks (ANNs) to learn complex non-linear mappings between numerical inputs and outputs[16]. Accordingly, we call it 'artificial neural network accelerated PALM', or ANNA-PALM.

RESULTS

A deep learning approach to super-resolution image reconstruction

We aim to reconstruct a super-resolution image of approximately similar information content as a standard PALM acquisition (with K frames and N localizations) from a much smaller number of raw frames (k<<K) without changing the average density of localizations, ρ, i.e. from a much smaller number of total localizations ($n = \rho k \ll N = \rho K$). If PALM images are defined as 2D histograms of independent localizations, this task can be formulated as restoring an image corrupted by Poisson noise (and potentially additional forms of noise). Image

restoration is an ill-posed problem that has an infinity of solutions in the high-dimensional space of all possible images, unless additional constraints (priors) are imposed that restrict the solution to a lower dimensional subspace. Suitable subspaces exist because most natural images are highly redundant, and can be represented to very good approximation with a much smaller number of coefficients than pixels, via appropriate functions that map feature space to pixel space[17,18]. In recent years, ANNs with multiple layers (deep nets) have proven very successful at learning meaningful features and non-linear mappings for image classification, segmentation, restoration and many other tasks[16,18,19]. Inspired by these developments, we designed ANNA-PALM, a deep learning approach for restoring super-resolution views from under-sampled (sparse) localization microscopy data.

ANNA-PALM comprises a training stage and an inference stage (**Figure 1**). For training (**Figure 1a**), a few super-resolution images representative of the structure of interest (e.g. microtubules, nuclear pores, or mitochondria) are obtained using standard PALM imaging, i.e. by acquiring long diffraction limited image sequences (e.g. $K \sim 10^4 - 10^5$, $N \sim 10^5 - 10^7$) and processing them with standard localization software[20], resulting in highly sampled (dense) PALM images. In addition, a low resolution (widefield) image can also be acquired, as is commonly done before single molecule imaging when bleaching out preactivated fluorophores. Next, the dense PALM images are under-sampled by using a much smaller number of input frames, $k \ll K$, thus yielding sparse PALM images from the same localization data. Then, an ANN is trained to recover approximations of the dense PALM images from these sparse PALM images (and the optional widefield image).

Once trained, the ANN is applied to new sparse PALM images (with or without a widefield image), obtained from new image sequences with small numbers of frames ($k \ll K$) -and hence in much shorter time- in order to reconstruct high quality super-resolution images not previously seen (inference, **Figure 1b**).

Neural net architecture and learning strategy

Our ANN, hereafter called A-net, contains a total of 25 convolutional layers, and roughly 42 million trainable parameters. A-net is adapted from the pix2pix network[21], which itself builds on U-nets[22] and generative adversarial networks (GANs)[23], two recent successful deep learning techniques. U-nets are special types of convolutional neural networks (CNNs) that have proven effective at learning multi-scale representations of images and accurate, pixel-wise mappings[22,24]. GANs can generate new samples from real image distributions using a generator network that outputs synthetic images, and a discriminator network that outputs the probability that an input image is real or synthetic, both networks being trained simultaneously to compete against each other[23]. Importantly, the generator can be conditioned on input data (conditional GAN, or cGAN)[21,23], e.g. on images as in the pix2pix network. We modified the pix2pix architecture to accept a computational switch as additional input to handle multiple types of data, and introduced an additional network to evaluate the consistency between the reconstructed image and the widefield input image.

Training of our A-net proceeds as follows. Randomly under-sampled (i.e. sparse) versions of PALM images are fed as input to the A-net, while the

corresponding dense PALM images are defined as the A-net's targets, i.e. desired outputs (**Figure 1a**). Additional, optional inputs are widefield images, if available, and the switch, which indicates the image type when multiple types of images (e.g. microtubules and nuclear pores) are used during training. ANN training requires defining an objective function (also called loss), which measures how well the outputs match the targets. We implemented a loss function containing three terms. The first term measures the difference between the A-net output and the dense PALM image. Instead of the widely used mean squared error (MSE), which poorly reflects visual quality[25], we used a combination of the absolute difference (L1 norm) with a multi-scale version of the structural similarity index, a perceptually motivated quality metric shown to improve image restoration with deep learning (MS-SSIM)[26]. The second term measures the consistency between the A-net output and the widefield image. Although in theory the latter should simply be a blurred version of the PALM image, this is often not the case in practice[27]. Therefore, we introduced another CNN (with 4-layers), called 'low resolution estimator' to predict the widefield image from the super-resolution image. The corresponding loss was defined as the MS-SSIM between this CNN's output and the observed widefield image. (In absence of a widefield image, this loss is set to zero). The third term contains a cGAN discriminator loss[21], where the discriminator is a 5-layer CNN, whose inputs are the sparse PALM (and widefield) image(s) and either the dense PALM image or the output of the generator above; the discriminator's output is compared to 0s and 1s (for synthetic and real, respectively), respectively, via the MSE. We use dropout[28] and extensive data augmentation, including random

rotations, translations, elastic deformations and addition of noise in the input image to mimic false detections and unspecific labeling. As a result, only a few dense PALM images are required for successful training without overfitting. On graphical processing units (GPU), training ANNA-PALM from scratch takes on the order of hours to days, but when starting from a previously trained A-net, retraining can be done in an hour or less.

Once trained, the A-net can take sparse localization data with an optional widefield image as input(s), and output a reconstructed super-resolution image in less than a second (**Figure 1b**). In addition, the A-net produces an 'error map' that measures the consistency of this super-resolution image with the widefield image[27] (when available) and can be used to estimate the degree of reliability and highlight potential reconstruction artifacts. For more details, see **Online Methods** and **Supplementary Note 1**.

Validating ANNA-PALM on simulated images

We first sought to validate ANNA-PALM on synthetic data. For this, we used Brownian dynamics simulations[29] to generate 200 dense PALM images of semi-flexible filaments mimicking microtubules, with a resolution $R_{\mathrm{loc}} \approx 23$ nm. These represent "perfect" PALM images that would be obtained with an infinite number of localizations. We applied varying levels of Poisson noise to these perfect images to create sparse PALM images corresponding to finite numbers of localizations (**Supplementary Figure 1**). We then trained our A-net using the perfect images as

177 targets and the sparse images (and widefield image) as inputs, varying the number
178 of localizations over a large range.

179 Next, we applied the trained A-net to a distinct set of PALM images generated
180 by the same stochastic simulation (**Supplementary Figure 2**). **Figure 2a** shows a
181 widefield image and **Figure 2b** a corresponding sparse PALM image obtained from
182 n=6,834 localizations. Although curvilinear structures can be seen in this image
183 despite its sparsity, small-scale features remain highly ambiguous (**Figure 2b,**
184 inset), and the resolution according to a recently proposed five-fold Nyquist
185 criterion[3] is limited by sampling to $R_{5\times\mathrm{Nyq}}$ ≈85 nm; according to this criterion,
186 N>$N_{5\mathrm{xNyq}}$ =60,000 localizations are needed to achieve 23 nm resolution
187 (**Supplementary Figure 3d**). **Figure 2c** shows the ANNA-PALM image
188 reconstructed from the wide-field image alone, which exhibits clear and continuous
189 filaments that were not previously recognizable. Most of the relatively isolated
190 filaments roughly agree with the perfect PALM image (**Figure 2e**). In the denser
191 regions, however, many small features are erroneous, e.g. filaments are incorrectly
192 joined, displaced, split or merged (**Figure 2c,** blue arrows and **Supplementary**
193 **Figure 4, top**). By contrast, the ANNA-PALM image reconstructed from the sparse
194 PALM image alone or in combination with the widefield image exhibits continuous
195 and sharp filaments in very good agreement with the perfect PALM image
196 (**Supplementary Figure 3b,c** and **Figure 2d,e,f**). The spatial resolution of these
197 reconstructed images is limited neither by diffraction nor sampling, but only by the
198 localization precision, and is thus R_{loc} ≈23 nm, as in the perfect images
199 (**Supplementary Figure 3e,f**). These results indicate that high quality super-

resolution images can be obtained from only a small fraction of the number of localizations traditionally required (here, approximately ~11% of N_{5xNyq} above; see **Supplementary Figure 3d**), hence enabling a strong reduction in acquisition time. Nevertheless, reconstruction errors can still occur in areas where the sparse localization data are most ambiguous, e.g. where filament density is highest (**Figure 2d,e,f,** white arrow). These errors can be reduced by increasing the localization number n, implying a trade-off between acquisition time and reconstruction quality (**Supplementary Figure 4**).

To quantify this trade-off, we computed the MS-SSIM between reconstructed ANNA-PALM and perfect PALM images ($n = \infty$) as function of localization number, from n~200 to n~ 2x10^6, in comparison with the standard PALM images (**Figure 2g**). The MS-SSIM ranges from 0 to 1 and reaches 1 for perfect reconstructions. For standard PALM images, the MS-SSIM increases monotonically, as expected, from <0.2 to >0.95 for n = 2x10^6 million localizations (**Figure 2g**, black curve). Using only the sparse image as input, ANNA-PALM reconstructions achieve MS-SSIM that are consistently higher and increase with localization number n much more rapidly than standard PALM, already exceeding 0.9 for $n \approx$ 10,000 localizations (**Figure 2**, dashed blue curve). ANNA-PALM achieves the same MS-SSIM as standard PALM at the five-fold Nyquist sampling level (≈0.65) with only n =2,248 localizations instead of n =58,588, suggesting a ~26-fold speed-up. If the widefield image is used as additional input, the MS-SSIM further increases, and dramatically so for low localization numbers (**Figure 2g**, solid blue curve). For example, with n =7,477 localizations, ANNA-PALM achieves a MS-SSIM (≈0.95)

similar to standard PALM with n =644,844, implying a speed-up of roughly two orders of magnitude. (Note that, if the perfect PALM image was not available for these quantifications, it could be replaced by the ANNA-PALM reconstruction of a dense PALM image with a large number of localizations, e.g. $n = 10^5$, with similar results- see **Supplementary Figure 5**).

As any image restoration method, ANNA-PALM can make errors. The low resolution error map described above (**Figure 1b**) provides a means to estimate where errors are most likely to occur. When applied to ANNA-PALM reconstructions of a sparse PALM image, this error map highlights regions containing the highest density of filament crossings, where reconstructions tend to be least accurate (**Supplementary Figure 6i,k**). If we artificially displace a small piece of filament in this image to simulate a false positive and a false negative in the reconstruction (**Supplementary Figure 6b,d,** white and blue arrows, respectively), the affected regions also light up in the error map (**Supplementary Figure 6j,l**). Thus, the error map offers a useful tool to highlight regions most likely to contain reconstruction errors, and conversely, to outline regions where reconstructions are most trustworthy. Thus, simulations suggest that ANNA-PALM can considerably reduce acquisition time in localization microscopy and also map reconstruction reliability.

ANNA-PALM reconstructions of immunostained microtubules

We next tested our method on real images of immunolabeled microtubules (**Figure 3**). We trained our A-net on seven dense PALM images (with corresponding widefield images) obtained during 10 minute long acquisitions (K=60,000;

246 Δt=10 ms exposure time) (not shown). We then considered a sparse PALM image of
247 microtubules in a distinct FoV obtained from only 9 s of acquisition (*k*=300; Δt=30
248 ms), together with a widefield image (Δt=2x50 ms) (**Figure 3a,b**). Whereas
249 microtubule filaments can already be seen in this sparse PALM image, structural
250 details below the diffraction limit are hard to discern, making it difficult to follow
251 the path of individual filaments in the denser regions and to identify features such
252 as filament crossings (**Figure 3b**). By contrast, the ANNA-PALM images, whether
253 reconstructed from the widefield image alone, the sparse PALM image alone, or
254 both, all display sharp and continuous filaments and clearly reveal many structural
255 details (**Figure 3d-f**). Their resolution is similar to or even better than the dense
256 PALM image (**Supplementary Figure 7a**). As for the simulations, in regions where
257 microtubule filaments are isolated, the ANNA-PALM image reconstructed from the
258 widefield image alone is in good agreement with the dense PALM image
259 (**Figure 3d,g**). However, it is often incorrect in areas of high microtubule density
260 (e.g. **Figure 3d**,**g** white and gray arrows). Most of these reconstruction errors are
261 corrected when applying ANNA-PALM to the sparse PALM image instead
262 (**Figure 3e,h**). For example, parallel sections of two microtubules unresolved in the
263 widefield image and incorrectly merged in **Figure 3d** are now clearly separated and
264 positioned correctly, and missed portions of other filaments are now recovered
265 (**Figure 3h,** white and gray arrows). Counter-intuitively, the sparse PALM image
266 exhibits high signal in some locations where the dense PALM image does not,
267 presumably because of spurious localizations due e.g. to unspecific binding
268 (**Figure 3b**, blue arrow). Such signal can lead to incorrect features in the ANNA-

PALM reconstruction from the sparse localization data alone (**Figure 3e,h**, blue arrows). However, when combining the widefield and sparse PALM data, these artifacts are largely removed and reconstructions agree very well with the dense PALM image (**Figure 3f,i**). Reconstruction quality increases with the number of frames *k* (**Figure 3j, Supplementary Figures 8-9** and **Supplementary Video 1**). More quantitatively, a MS-SSIM analysis similar to that for the simulated data above (with the ANNA-PALM output of the dense PALM image defined as ground truth; see **Supplementary Figures 5, 9f**) suggests that ANNA-PALM allows a hundred-fold reduction of acquisition time compared to standard PALM (**Supplementary Figure 9g**). **Supplementary Figure 10** shows other examples of sparse microtubule images reconstructed by ANNA-PALM.

As for simulations above, we used the widefield image to compute an error map (**Supplementary Figures 6,11**). Bright areas in this error map highlight regions where the reconstruction indeed disagrees with the dense PALM image; conversely, reconstructions are of high quality in the majority of regions where the error map is dim (**Supplementary Figure 11d-f**). These results demonstrate experimentally that ANNA-PALM can restore high quality approximations of super-resolution images from much shorter acquisition time than typical for PALM imaging, and also predict where reconstruction errors are most likely.

ANNA-PALM enables high-throughput super-resolution imaging

The drastic improvement in imaging efficiency afforded by ANNA-PALM permits super-resolution imaging of orders of magnitude more cells and FoVs per unit time.

To demonstrate this, we used an automated acquisition protocol to image >1,000 cells with immunolabeled microtubules in 1,089 (33x33), partly overlapping, FoVs of 55.3 µm x 55.3 µm each (**Figure 4**, **Supplementary Figure 12**). We first acquired widefield images at each of these positions, in a total of ~12 minutes, mostly consisting of stage stabilization delays (**Supplementary Figure 12a**). Next, we obtained 1,089 sparse PALM images using only 10 s of imaging time per FoV (k=1,000 frames, Δt = 10 ms), in a total of only ~3.1 hours (**Figure 4a**). Neither the widefield nor the sparse PALM images provided much small scale information (**Figure 4c, Supplementary Figure 12 c,e**). However, ANNA-PALM reconstructions led to high quality super-resolution images, allowing to visualize the microtubule network with clarity and to distinguish microtubule filaments in dense areas that appeared as unstructured regions in the sparse PALM image (**Figure 4b,d**). The FWHM across filaments in the reconstructed image was ~51 nm (**Figure 4d**), within the range measured for the training data (**Supplementary Figure 7a**). Similar images can be obtained by ANNA-PALM using the widefield images alone (**Supplementary Figure 12**), although of lower quality, as discussed above. Stitching the reconstructed images together yielded a single super-resolution image that contained approximately seven billion 20x20 nm pixels and covered an area of 1.8 mm x 1.8 mm, thereby spanning almost five orders of magnitude in spatial scales (**Figure 4b,d, Supplementary Figure 12b** and **Supplementary Video 2**).

ANNA-PALM is robust to experimental perturbations

ANNA-PALM can reconstruct accurate super-resolution images from sparse data because it was trained on similar images before. This raises the question of whether an ANN trained in one experimental condition can be successfully applied to another condition. To test this, we used the A-net as trained above to analyze microtubule images of cells subjected to drugs affecting the cytoskeletal network. We first treated U373 cells with 1 μM of Taxol, an antimitotic agent that inhibits the depolymerization of microtubules and increases their bending rigidity[30,31]. **Figure 5d** shows a sparse PALM image of these cells (k=800, kΔt=8 s), in which the microtubule network is barely recognizable. By contrast, the ANNA-PALM reconstructions clearly display a complex microtubule network and agree well with the dense PALM image obtained from K=60,000 frames ($K\Delta$t=10 min) (**Figure 5d,e,f**). These images display a larger density of straighter and more parallel filaments with less frequent crossings than in the untreated cells (**Figure 5a-c**), consistent with microtubule stabilization and increased rigidity[31].

Next, we treated cells with 1 μM of Nocodazole, a drug that, on the opposite, promotes microtubule depolymerization and is expected to more dramatically alter the cytoskeletal network[32]. Again, whereas the sparse image obtained from k=300 frames ($k\Delta$t=3 s) contained little exploitable information (**Figure 5g**), the ANNA-PALM reconstruction offered clear and detailed views of the disrupted microtubule network (**Figure 5h**), exhibiting a much smaller number of filaments, with higher curvature, than in untreated cells. These reconstructions were in good (though not perfect) agreement with dense PALM images obtained from K=60,000 frames ($K\Delta$t= 10 min) (**Figure 5i**). Thus, ANNA-PALM, when trained on microtubules in

one experimental condition, may be successfully applied to new experimental conditions without retraining, thereby highlighting the method's robustness to biologically relevant structural alterations.

We further asked if ANNA-PALM is robust to changes in technical imaging conditions. To address this, we performed localization microscopy on microtubules by simultaneously changing multiple important imaging parameters relative to the training data. Instead of PALM/STORM, we used DNA-PAINT a technique where single molecule detection relies on transient binding of fluorophore-conjugated short DNA strands to complementary, antibody-conjugated, DNA strands[33], rather than on fluorophore blinking. The continuously emitting freely diffusing dyes lead to higher background noise in DNA-PAINT compared to PALM/STORM. Moreover, we used primary mouse antibodies instead of rat antibodies, Cy3 dyes instead of Alexa-647 dyes, and an EMCCD instead of a sCMOS camera. Despite all these differences, when ANNA-PALM was applied without retraining on a sparse microtubule image (k=400 frames, kΔt=12 s) , the reconstructed image still agreed very well with the dense DNA-PAINT image obtained from K=60,000 frames ($K\Delta$t=30 min) (**Figure 5j-l**). These data demonstrate the high robustness of ANNA-PALM to changes in experimental imaging conditions.

ANNA-PALM adapts to different biological structures

To demonstrate that ANNA-PALM is not restricted to filamentary structure, we turned to nuclear pores, a very different biological structure, and another popular target of super-resolution imaging studies[34–36]. We retrained A-net simultaneously

on microtubule images and on a single PALM image of the nucleoporin gp210 in immunolabeled nuclear membranes of *Xenopus* frog eggs[34,36] (K=30,000). With the switch (**Figure 1**) set to microtubules ('MT'), this newly trained A-net can still reconstruct sparse images of microtubules as when trained exclusively on microtubule data (**Supplementary Figure 13a-c**). We then applied the same A-net with the switch set to nuclear pores ('NPC') to reconstruct a new sparse PALM image of gp210 obtained from the first k=3,000 frames (**Figure 6a**). The sparsity of this image makes it difficult to clearly distinguish individual nuclear pores. ANNA-PALM, however, reconstructs a much clearer image, containing many easily identifiable ring-like structures, as expected for nuclear pores[34] (**Figure 6b**), and in good agreement with the dense PALM image obtained from K=30,000 frames (even though the latter shows mostly incomplete, open rings, presumably due to suboptimal labeling) (**Figure 6c**). An automated procedure based on cross-correlation with a ring template indeed identified ~2.7 times more putative nuclear pores from the ANNA-PALM image than the sparse image (**Supplementary Figure 14a-c**). Moreover, computed pore locations were in good agreement with a PALM image of wheat germ agglutinin (WGA), a lectin that concentrates in the inner nuclear pore channel[36] (**Supplementary Figure 14 d-f**). These results show that ANNA-PALM can successfully analyze non-filamentary structures, when properly retrained, and that a single ANN, with a simple computational switch, can reconstruct very different types of structures.

Finally, we imaged TOM22, a protein of the mitochondrial outer membrane[37]. Whereas, at the resolution of our experiments, microtubules and nucleoporins are

essentially one-dimensional and zero-dimensional structures, mitochondrial membranes are two-dimensional surfaces. Furthermore, their complex 3D morphology might seem less predictable than filaments or nuclear pores, potentially hampering ANNA-PALM reconstruction. Despite these differences, after being trained on nine PALM images of TOM22 (with frame numbers ranging from K=24,000 to K=40,000), ANNA-PALM reconstructions of distinct sparse PALM images (k=400 frames) displayed mitochondrial morphologies in good agreement with the dense PALM images (**Figure 6d-f**) - although the protein's localization along the membrane was less well reproduced**.** Taken together, our results illustrate the versatility of ANNA-PALM and its applicability to images of very different structural content.

Discussion

We introduced ANNA-PALM, a computational method based on deep learning, that reconstructs high quality super-resolution images from sparse, rapidly acquired, single molecule localization data (and/or widefield images). Our method enables considerable gains in acquisition time compared to standard localization microscopy without increasing active fluorophore density, thereby preserving spatial resolution. In fact, ANNA-PALM even improves spatial resolution when applied to images of lower resolution than the training data (**Supplementary Figures 7a, 10**), and greatly diminishes the detrimental effect of drift. The improvement in imaging efficiency afforded by ANNA-PALM alleviates the incompatibility between high resolution and high-throughput microscopy by

enabling super-resolution imaging of thousands of cells within a few hours or even less (**Figure 4, Supplementary Figure 12**). This will facilitate super-resolution studies of rare events, cellular heterogeneity and of partly stochastic structures such as cytoskeletal polymers or chromosomes, whose characterization requires statistics on many configurations[38,39]. ANNA-PALM may also be beneficial for high-throughput imaging screens, e.g. of drug treatments or gene knock-outs[40–42]. In addition, we envision applications to super-resolution imaging of large samples by stitching together multiple images of spatially adjacent fields. The ability to generate images spanning many orders of magnitude in scale could be well adapted to expansion microscopy, a super-resolution technique that physically increases sample size, but often requires tiling many fields of view to image even a single cell[43,44]. With correlative microscopy[45], it might also be possible to train ANNA-PALM to reconstruct electron microscopy (EM) images from fluorescence images, potentially extending the method to molecular resolutions currently out of reach of localization microscopy. Adaptation of ANNA-PALM to 3D[15,46] and multi-color[47,48] localization microscopy should be relatively straightforward. Localization microscopy of cellular dynamics remains very challenging[3,49]. By using much fewer frames (or even only widefield images), ANNA-PALM could dramatically improve the temporal resolution of live cell localization microscopy without sacrificing spatial resolution or increasing phototoxicity and photobleaching. Thus, ANNA-PALM provides multiple novel avenues for multi-scale imaging beyond standard spatio-temporal resolution limits.

Neverthelesss, important caveats should be stressed. First, although ANNA-PALM can be applied successfully to very different types of images (**Figure 6**), the method fails in absence of statistical redundancies between molecular localizations, e.g. for entirely random distributions of molecules. Second, ANNA-PALM requires prior training on dense PALM images with structures similar to those in the images to be reconstructed. We showed that ANNA-PALM is robust, i.e. does not require retraining, for some experimentally induced changes in structures and variations in imaging parameters (**Figure 5**). Nevertheless, indiscriminate application of ANNA-PALM to very different structures without retraining, or incorrect setting of the switch, may result in artifacts (**Supplementary Figure 13h**). Third, even when applied to data similar to the training images, ANNA-PALM can produce errors -as any reconstruction method in a context of information scarcity. The frequency of errors can be reduced by increasing the number of recorded frames, at the cost of reduced acceleration (see **Figures 2,3** and **Supplementary Figures 4, 8, 9**). In addition, ANNA-PALM can use widefield images to estimate the reliability of reconstructions, thereby helping their interpretation, providing some protection against artifacts and indicating when retraining may be needed (**Supplementary Figures 11, 15**). Future work, e.g. using Bayesian deep learning[50], may provide additional assessments of reconstruction uncertainty and shed more light on the scope and limitations of our approach.

As a purely computational technique, ANNA-PALM does not necessitate any changes to existing microscopy systems, but only requires one or a few standard PALM images for training. To facilitate its adoption and future development, we

make our source code, an ImageJ plugin and a cloud-computing based web application available on https://annapalm.pasteur.fr/ along with sample data. Because the performance of deep learning methods improves with the amount and variety of training data, we designed our web application to enable sharing of data and trained ANNs. As ANNA-PALM will learn from an increasing large and diverse collection of images, we expect it to reach even higher accuracy or efficiency and to expand its scope of application in the future.

Acknowledgements

We thank the following colleagues for useful discussions and suggestions and/or critical reading of the manuscript: C. Leduc, S. Etienne-Manneville, S. Lévêque-Fort, N. Bourg, A. Echard, J-B. Masson, T. Rose, P. Hersen, F. Mueller, M. Cohen, Z. Zhang, and P. Kanchanawong. We also thank the four anonymous reviewers for their constructive criticism, which led to significant improvements of ANNA-PALM. We further thank O. Faklaris, J. Sellés and M. Penrad (Institut Jacques Monod), and F. Montel (Ecole Normale Supérieure de Lyon) for providing *Xenopus* nuclear pore data, B. Jian (Institut Pasteur) for TOM22 antibodies, and C. Leterrier for fixation protocols. We thank E. Rensen and C. Weber for help with experiments and suggestions, B. Lelandais for help with PALM image processing, J-B. Arbona for polymer simulations and J. Parmar for suggestions that led to the name ANNA-PALM. We thank the IT service of Institut Pasteur, including J-B. Denis, N. Joly, and S. Fournier, for access to the HPC cluster and relevant assistance, and T. Huynh for help with GPU computing. This work was funded by Institut Pasteur, Agence

Nationale de la Recherche grant (ANR 14 CE10 0018 02), Fondation pour la Recherche Médicale (Equipe FRM, DEQ 20150331762), and the Région Ile de France (DIM Malinf). We also acknowledge Investissement d'Avenir grant ANR-16-CONV-0005 for funding a GPU farm used in this work. A.A. and X.H. are recipients of Pasteur-Roux fellowships from Institut Pasteur. W.O. is a scholar in the Pasteur - Paris University (PPU) International PhD program.

Author contributions

W.O.: conceived method, developed ANNA-PALM software and web application, performed experiments and analyses. A.A., M.L., X.H.: performed experiments. C.Z.: conceived method, supervised project and wrote manuscript.

Competing financial interests statement

W.O. and C.Z. are listed as inventors on European patent application EP17306022 filed by Institut Pasteur.

REFERENCES

1. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science (80-.).* **313,** 1642–1645 (2006).
2. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3,** 793–795 (2006).
3. Legant, W. R. *et al.* High-density three-dimensional localization microscopy across large volumes. *Nat. Methods* **13,** 359–365 (2016).
4. Deschout, H. *et al.* Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11,** 253–266 (2014).
5. Jones, S. A., Shim, S. H., He, J. & Zhuang, X. Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods* **8,** 499–505 (2011).
6. Huang, F. *et al.* Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10,** 653–8 (2013).
7. Carlton, P. M. *et al.* Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16016–22 (2010).
8. Stelzer, E. H. K. Light-sheet fluorescence microscopy for quantitative biology. *Nat. Methods* **12,** 23–26 (2014).
9. Huang, F., Schwartz, S. L., Byars, J. M. & Lidke, K. A. Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomed. Opt. Express* **2,** 1377–93 (2011).
10. Burnette, D. T., Sengupta, P., Dai, Y., Lippincott-Schwartz, J. & Kachar, B. Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proc Natl Acad Sci U S A* **108,** 21081–21086 (2011).
11. Simonson, P. D., Rothenberg, E. & Selvin, P. R. Single-molecule-based super-resolution images in the presence of multiple fluorophores. *Nano Lett* **11,** 5090–5096 (2011).
12. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster STORM using compressed sensing. *Nat. Methods* **9,** 721–723 (2012).
13. Cox, S. *et al.* Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat Methods* **9,** 195–200 (2012).
14. Ram, S., Ward, E. S. & Ober, R. J. Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy. *Proc. Natl. Acad. Sci.* **103,** 4457–4462 (2006).
15. Deschout, H. *et al.* Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11,** 253–266 (2014).
16. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).
17. Michael, E. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. *Springer* (Springer Verlag, 2010).
18. Hinton, G. E. & Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science (80-.).* **313,** (2006).
19. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks* **61,** 85–117 (2014).
20. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12,** 717–724 (2015).

21. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with Conditional Adversarial Networks. *http://arxiv.org/abs/1611.07004* (2016).
22. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science,* **9351,** 234–241 (Springer International Publishing, 2015).
23. Goodfellow, I. *et al.* Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst. 27* 2672–2680 (2014). doi:10.1017/CBO9781139058452
24. Ciresan, D., Giusti Alessandro, Luca, G. & Schmidhuber, J. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. *Med. Image Comput. Comput. Interv. – MICCAI 2013* **8150,** (2013).
25. Wang, Z. W. Z. & Bovik, A. C. A. C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **26,** 98–117 (2009).
26. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss Functions for Image Restoration with Neural Networks. *IEEE Trans. Comput. IMAGING* **3,** 47–57 (2017).
27. Culley, S. *et al.* NanoJ-SQUIRREL: quantitative mapping and minimisation of super-resolution optical imaging artefacts. *bioRxiv* 158279 (2017). doi:10.1101/158279
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15,** (2014).
29. Arbona, J.-M., Herbert, S., Fabre, E. & Zimmer, C. Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations. *Genome Biol.* **18,** 81 (2017).
30. Arnal, I. & Wade, R. H. How does taxol stabilize microtubules? *Curr. Biol.* **5,** 900–908 (1995).
31. Wu, S. *et al.* Microtubule motors regulate ISOC activation necessary to increase endothelial cell permeability. *J. Biol. Chem.* **282,** 34801–8 (2007).
32. De Brabander, M., De May, J., Joniau, M. & Geuens, G. Ultrastructural immunocytochemical distribution of tubulin in cultured cells treated with microtubule inhibitors. *Cell Biol. Int. Rep.* **1,** 177–83 (1977).
33. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12,** 1198–1228 (2017).
34. Löschberger, A. *et al.* Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. *J. Cell Sci.* **125,** 570–5 (2012).
35. Szymborska, A. *et al.* Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341,** 655–8 (2013).
36. Sellés, J. *et al.* Nuclear pore complex plasticity during developmental process as revealed by super-resolution microscopy. *Sci. Rep.* **7,** 14732 (2017).
37. Bellot, G. *et al.* TOM22, a core component of the mitochondria outer membrane protein translocation pore, is a mitochondrial receptor for the proapoptotic protein Bax. *Cell Death Differ.* **14,** 785–794 (2007).
38. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin

folding for different epigenetic states. *Nature* **529,** 418–422 (2016).
39. Zhang, Z., Nishimura, Y. & Kanchanawong, P. Extracting microtubule networks from superresolution single-molecule localization microscopy data. *Mol. Biol. Cell* **28,** 333–345 (2017).
40. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464,** 721–727 (2010).
41. Beghin, A. *et al.* Localization-based super-resolution imaging meets high-content screening. *Nat. Methods* **14,** 1184–1190 (2017).
42. Ouyang, W. & Zimmer, C. The imaging tsunami: Computational opportunities and challenges. *Curr. Opin. Syst. Biol.* **4,** 105–113 (2017).
43. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science (80-.).* **347,** 543–548 (2015).
44. Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* (2017). doi:10.1038/nmeth.4261
45. de Boer, P., Hoogenboom, J. P. & Giepmans, B. N. G. Correlated light and electron microscopy: ultrastructure lights up! *Nat. Methods* **12,** 503–513 (2015).
46. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (80-.).* **319,** 810–813 (2008).
47. Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes. *Science (80-.).* 1146598v1 (2007).
48. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11,** 313–318 (2014).
49. 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).
50. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017) pre-proceedings* 5580–5590 (2017).

FIGURE LEGENDS

FIGURE 1: Overview of ANNA-PALM

ANNA-PALM consists of two main stages: (**a**) acquisition of training images using standard localization microscopy (PALM) followed by artificial neural network (ANN) training, and (**b**) reconstruction of super-resolution views and low resolution error maps from new sparse PALM and/or widefield images (inference). (**a**) Training images are obtained by acquiring one or a few long sequences, of $K \sim 10^3$-10^5 diffraction limited, single molecule image frames, as in standard PALM experiments; optionally, a widefield image W can also be acquired (top). The acquisition time for each image sequence is $K\Delta t$, where Δt is the single frame exposure time. Standard localization microscopy algorithms (grey 'localization' boxes) are used to generate super-resolution images. For each experiment, a highly sampled (dense) super-resolution image (PK) is generated using all (or in practice $\geq$95%) available K frames. Sparse PALM images (Pk) from the same experiment are obtained by using only $k << K$ frames. A switch (light blue) can be set to distinguish different types of structures, e.g. nuclear pore complexes ('NPC'), mitochondria ('Mito') or microtubules ('MT'). An ANN (labeled $\mathcal{G}$ for 'generator', orange) is trained by using the sparse PALM images Pk (plus an upsampled version L of the widefield image W and the switch setting) as inputs and the corresponding dense PALM image PK as target output. During training, the output of the generator $\mathcal{G}$ (Ak) is compared to the target image PK and the widefield image L (if available) via three loss, or error functions (gray bordered boxes): (i) the 'super-resolution reconstruction error'

measures the difference between the reconstructed image Ak and the target PK using a combination of the L1 norm and the MS-SSIM; (ii) the 'low resolution reconstruction error' measures the MS-SSIM between the low resolution image WA predicted from the reconstruction Ak and the low resolution image WP predicted from the target image PK. Images WA and WP are predicted using a second ANN, called low-resolution estimator (labeled Q, blue) that is trained to produce an approximation of the actual widefield image W based on the MS-SSIM metric; (iii) the 'conditional GAN error' uses a third ANN (labeled *D* for 'discriminator', red) that attempts to distinguish between real dense PALM images PK and the generator's output Ak. The combined loss functions are iteratively optimized using stochastic gradient descent. (**b**) A short sequence of diffraction limited single molecule images (with $k<<K$ frames, i.e. acquisition time $k\Delta t$), and an optional widefield image (W'), are acquired. Standard localization algorithms generate a sparse (under-sampled) PALM image (P'k). This sparse image P'k (and the upscaled widefield image L' and switch setting) are fed as inputs to the trained generator G, which outputs a reconstructed ANNA-PALM image (A'k). In addition, the low resolution estimator Q predicts a low resolution image WA', which can be compared to the input widefield image W' via the MS-SSIM to produce a low resolution error map (top).

FIGURE 2: Validation of ANNA-PALM on simulated images

(**a**) Simulated widefield image of microtubules. (**b**) Simulated sparse PALM image of microtubules with n= 6,834 localizations. (**c**) ANNA-PALM reconstruction using only the widefield image **a** as input. (**d**) ANNA-PALM reconstruction using both the

widefield image **a** and the sparse PALM **b** image as inputs. (**e**) Simulated "perfect" PALM image, equivalent to a PALM image with an infinite number of localizations (n=∞) and a resolution of 23 nm. This image was used to generate **a** (by convolution with a Gaussian kernel approximating the microscope point spread function) and **b** (by application of Poisson noise). (**f**) Merged image showing the perfect PALM image **e** in green and the ANNA-PALM reconstruction **d** in red. Note that the ANNA-PALM images **c,d** provide many high resolution details that are absent from the widefield image **a** and the sparse PALM image **b** and that are in good (**c**) or very good (**d**) agreement with the perfect PALM image **e**. Some reconstruction errors are highlighted by arrows. Blue arrows in panel **c** point to errors of ANNA-PALM reconstruction from the widefield image only, the white arrow in panel **d** points to an error of ANNA-PALM reconstruction from both widefield and sparse PALM images combined. Reconstruction errors diminish for larger numbers of localizations, n (**Supplementary Figure 4**). (**g**) Reconstruction quality of PALM and ANNA-PALM images, measured by the MS-SSIM with the perfect PALM image **e**, as function of localization number n. Black curve: reconstruction quality of the standard PALM images. Dashed blue curve: reconstruction quality of ANNA-PALM using the sparse PALM images as input. Solid blue curve: reconstruction quality of ANNA-PALM using both the sparse PALM and widefield images as inputs. Red dashed line: reconstruction quality of ANNA-PALM using the widefield images as input only. Dots are averages from 10 simulations; error bars show standard deviations. The vertical dashed orange line indicates the minimum number of localizations needed to achieve a resolution of $R_{5\times\mathrm{Nyq}}$ =23 nm according to the five-

fold Nyquist criterion[3]. The dashed grey line indicates the minimum number of localizations needed to achieve a double mean nearest neighbor distance less than 23 nm. ANNA-PALM reconstructions from sparse PALM images only (i.e. without widefield images) achieve the same average MS-SSIM as standard PALM at the five-fold Nyquist sampling limit with 26 times less localizations (blue double arrow). ANNA-PALM reconstruction quality is highest when using both widefield and sparse PALM images as inputs.

FIGURE 3: ANNA-PALM imaging of microtubules

ANNA-PALM reconstructions of a localization microscopy image of immunostained microtubules. (**a**) Widefield image. (**b**) Sparse PALM image obtained from the first 9 s of acquisition (k=300 frames, n=11,740 localizations). (**c**) Dense PALM image obtained from a 15 min long acquisition (K=30,000 frames, N=409,364 localizations). (**d**) ANNA-PALM reconstruction from the widefield image **a** only. (**e**) ANNA-PALM reconstruction from the sparse PALM image **b** only. (**f**) ANNA-PALM reconstruction from the widefield image **a** and sparse PLAM image **b** combined. In panels **b-f**, pixel values are linearly mapped to colors from the look-up table shown below. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, V_{max}=3, 24, 102, 102 and 102 for panels **b**, **c**, **d**, **e** and **f**, respectively. (**g-i**) Merged images comparing ANNA-PALM reconstructions from panels **d-f** to the dense PALM image **c**. ANNA-PALM reconstructions are shown in red, the dense PALM image in green. (**j**) Gradual improvement of image quality for increasing acquisition time $k\Delta t$, shown for the

area highlighted in the insets of panels **a-i**. Top row: sparse PALM images. Middle row: ANNA-PALM reconstructions from the sparse PALM images only (without widefield). Bottom row: ANNA-PALM reconstructions from the widefield and sparse PALM images combined. **Supplementary Video 1** shows the gradual increase in quality of PALM and ANNA-PALM images with increased acquisition time for the larger region of interest shown in panels **a-i.**

FIGURE 4: High-throughput imaging with ANNA-PALM

Application of ANNA-PALM to high-throughput imaging of a 1.8 mm x 1.8 mm area containing more than 1,000 cells. (**a**) Sparse PALM image of this area, constructed by assembling a mosaic of 33x33=1,089 sparse PALM images of individual fields of view, obtained from k=1,000 raw frames each (with Δt=10 ms exposure time per frame, i.e. in $k\Delta t$ =10 s). Total image acquisition time was 1,089x10s, i.e. ~3.1 hours. The sparsity of the image is not apparent at this large scale. (**b**) ANNA-PALM reconstruction of the image in **a**, obtained by assembling a mosaic of 1,089 individual reconstructions (one per field of view). (**c**) Magnified view of the green boxed region in **a**. The inset shows a further magnified view of the yellow boxed region, highlighting the sparsity of the image. (**d**) Same as **c**, but for the ANNA-PALM reconstruction. A line profile across a microtubule is shown, with a FWHM≈51 nm. Non-linear contrast adjustment was applied manually for panels **a** and **b,** with black corresponding to values of zero in both panels. In panels **c** and **d**, pixel values were linearly mapped to colors from the look-up table in **Figure 3;** Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, and

V_{max}=3 and 51 for panels **c** and **d**, respectively. See also **Supplementary Video 2** for an animated 'zoom-in' highlighting the spatial scales covered by the assembled image. See also **Supplementary Figure 12** for ANNA-PALM reconstructions of the same area from the widefield images only.

FIGURE 5: Robustness of ANNA-PALM to experimental perturbations

This figure shows ANNA-PALM reconstructions using an ANN trained on PALM images of microtubules in untreated cells and applied without retraining to sparse localization images of microtubules in different experimental conditions: untreated control cells (**a-c**); cells treated with 1 μM of Taxol (**d-f**); cells treated with 1 μM of Nocodazole (**g-i**); untreated cells imaged with DNA-PAINT (**j-l**). (**a,d,g,j**) Sparse localization images obtained from the first *k* frames of the acquired image sequence, with *k*=500, 800, 300, and 400 for **a, d, g,** and **j**, respectively. (**b,e,h,k**) ANNA-PALM reconstructions using the sparse localization images immediately to the left as input. (**c,f,i,l**) Dense localization images obtained from *K*=60,000 frames. Pixel values are linearly mapped to colors from the look-up table in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, and V_{max}=10, 120, 90, 25, 150, 40, 18, 150, 50, 18, 120, and 200 for panels **a**, **b**, **c**, **d**, **e**, **f**, **g, h**, **i**, **j**, **k**, and **l**, respectively.

FIGURE 6: ANNA-PALM reconstructions of nuclear pores and mitochondria

PALM and ANNA-PALM images of nuclear pores (**a**-**c**) and mitochondria (**d**-**f**). (**a**) Sparse PALM image of the immunolabeled *Xenopus* nucleoporin gp210 obtained

from the first k=3,000 frames. Note that individual nuclear pores are hard to identify. (**b**) ANNA-PALM reconstruction of image **a**. (**c**) Dense PALM image obtained from all K=30,000 frames. (**d**) Sparse PALM image of the immunolabeled mitochondrial outer membrane protein TOM22, obtained from the first k=400 frames. (**e**) ANNA-PALM reconstruction of image **d**. (**f**) Dense PALM image obtained from all K=30,000 frames. Pixel values are linearly mapped to colors from the look-up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, and V_{max}=3, 51, 3, 3, 128, and 18 for panels **a**, **b**, **c**, **d**, **e**, and **f**, respectively

ONLINE METHODS

Artificial neural network

Our ANN, called 'A-net', is based on the pix2pix architecture[21], which is a special conditional generative adversarial network (cGAN)[23] for image to image "translation", i.e. mapping from one type of image to another. The A-net consists of three distinct neural networks: (i) a generator network $\mathcal{G}$ that produces the reconstructed super-resolution image, (ii) a network Q called 'low resolution estimator' that produces the low resolution error map, (iii) a cGAN discriminator network D that provides the adversarial loss (**Figure 1a**). The generator network $\mathcal{G}$ builds on the U-net architecture, and consists of an encoder-decoder network with skip connections[22] and 16 convolutional layers. Its inputs and outputs are image patches containing $(256m)$x$(256m)$ pixels, where m is an integer (we used m=1 or 2 but this can be adjusted for different sizes of CPU/GPU memory or input images). The input is a sparse PALM image, a widefield image upscaled to the same size (see below), and a computational switch number that allows the network to switch between different types of image structures (e.g. nuclear pores or microtubules). The switch setting is encoded numerically and coupled by convolutional operations into the A-net encoder. The output of the generator $\mathcal{G}$ is a reconstructed image (called ANNA-PALM reconstruction or ANNA-PALM image elsewhere) of the same size as the input images. The low resolution estimator Q has four convolutional layers. It takes the $(256m)$x$(256m)$ dense PALM image patch or the ANNA-PALM image patch as input and outputs a low resolution image with $(64m)$x$(64m)$ pixels.

The cGAN discriminator network D has five convolutional layers. Its inputs are three $(256m)$x$(256m)$ pixel image patches (the sparse PALM image, the widefield image, and either the ANNA-PALM reconstruction or the corresponding dense PALM image), plus the upscaled widefield image, and its output is a $(30m)$x$(30m)$ image whose pixel values indicate whether the corresponding input patch is real (i.e. an experimentally obtained conventional PALM image) or produced by the generator $\mathcal{G}$. All convolutional layers are followed by batch normalization[51]. Dropout layers[28] (with dropout probability p=0.5) are introduced in the central layers of the A-net generator and turned on during training, but switched off during inference. Activation functions are rectified linear units (ReLU) $x \to \sup(x, 0)$ or "leaky" ReLUs $x \to \sup(x, 0) + \inf(\varepsilon x, 0)$ with ε =0.2[52], except for the last layer of $\mathcal{G}$, which uses the hyperbolic tangent $x \to \tanh(x)$ and the last layer of Q, which uses a sigmoid function $x \to (1 + \exp(-x))^{-1}$. The A-net architecture is fully described in **Supplementary Note 1** and **Supplementary Tables 1-4**.

Training objectives and error map

Each of the three networks mentioned above ($\mathcal{G}$, Q, and D) is associated to a distinct objective function - also called loss- and hereafter noted $\mathcal{L}_{\mathcal{G}}$, $\mathcal{L}_{Q}$, and $\mathcal{L}_{D}$ respectively. These loss functions are specified in detail below. In the following equations, for notational simplicity, we designate the sparse input image as $\boldsymbol{S}$, the low resolution (widefield) input image as $\boldsymbol{W}$, the corresponding dense PALM image (i.e. the target) as $\boldsymbol{T}$, and the A-net's output as $\boldsymbol{A} = \mathcal{G}(\boldsymbol{S}, \boldsymbol{W})$ (in **Figure 1a**, images $\boldsymbol{S}$, $\boldsymbol{T}$ and $\boldsymbol{A}$ are labeled Pk, PK and Ak, respectively); the computational switch indicating the image

type is noted $\boldsymbol{M}$. Low resolution images produced by the low resolution estimator network Q from $\boldsymbol{A}$ and $\boldsymbol{T}$ are designated as $\boldsymbol{W}_A = Q(\boldsymbol{A})$ and $\boldsymbol{W}_T = Q(\boldsymbol{T})$, respectively.

The generator loss function $\mathscr{L}_{\mathcal{G}}$ is the sum of three terms. The first term of $\mathscr{L}_{\mathcal{G}}$ is the super-resolution reconstruction error, hereafter called $\mathscr{L}_{\text{SuperRes}}(\mathcal{G})$. This term penalizes the difference between the generator output $\boldsymbol{A}$ and the target image $\boldsymbol{T}$. Based on a recent analysis of loss functions for image restoration with neural networks[26], we defined this difference as a weighted average of two quantities: (i) the multi-scale structural similarity index (MS-SSIM) between $\boldsymbol{A}$ and $\boldsymbol{T}$ and (ii) a modification of the L1 norm, where the absolute difference between $\boldsymbol{A}$ and $\boldsymbol{T}$ is smoothed by a Gaussian kernel:

$$\begin{aligned}\mathscr{L}_{\text{SuperRes}}(\mathcal{G}) &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[\rho\left(1 - MS\_SSIM(\boldsymbol{A},\boldsymbol{T})\right) + (1\right. \\ &\left. - \rho)\langle G_\sigma * |\boldsymbol{A} - \boldsymbol{T}|\rangle\right] \\ &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[\rho\left(1 - MS\_SSIM(\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}),\boldsymbol{T})\right) + (1\right. \\ &\left. - \rho)\langle G_\sigma * |\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}) - \boldsymbol{T}|\rangle\right]\end{aligned}$$

where $\mathbb{E}$ denotes expectation; $p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})$ is the joint probability density of the sparse PALM images $\boldsymbol{S}$, dense PALM images $\boldsymbol{T}$, widefield images $\boldsymbol{W}$ and switch settings $\boldsymbol{M}$ from the training data set; $MS\_SSIM(\boldsymbol{A},\boldsymbol{T})$ is the multi-scale structural similarity index between $\boldsymbol{A}$ and $\boldsymbol{T}$; G_σ is a Gaussian smoothing kernel; $*$ denotes convolution; $|\boldsymbol{A} - \boldsymbol{T}|$ is the absolute difference image (i.e. pixel (i,j) has value $|\boldsymbol{A}(\text{i},\text{j}) - \boldsymbol{T}(\text{i},\text{j})|$ and $\rho \in [0,1]$ is a scalar weight that balances the relative contributions of MS-SSIM and the modified L1 norm and is set to $\rho = 0.84$ as in ref. [26].

The second term of $\mathcal{L}_{\mathcal{G}}$ is called $\mathcal{L}_{\text{LowRes}}(\mathcal{G}, Q)$ and measures the consistency between the low resolution images $\boldsymbol{W}_A$ and $\boldsymbol{W}_T$ predicted by the low resolution estimator network Q:

$$\mathcal{L}_{\text{LowRes}}(\mathcal{G}, Q) = \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[1 - MS\_SSIM(\boldsymbol{W}_A, \boldsymbol{W}_T)\right]$$

$$= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\left[1 - MS\_SSIM\big(Q\big(\mathcal{G}(\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{M})\big), Q(\boldsymbol{T})\big)\right]$$

Alternatively, in the above objective function, $\boldsymbol{W}_T$ can be replaced by the actually observed widefield image $\boldsymbol{W}$, although with our data this led to slightly lower reconstruction quality. The low resolution estimator network Q is trained simultaneously with the generator $\mathcal{G}$ to produce a low resolution image from the dense PALM image $\boldsymbol{T}$ that is consistent with the observed low resolution image $\boldsymbol{W}$. This training is done based on the following objective function:

$$\mathcal{L}_Q(Q) = \mathbb{E}_{(\boldsymbol{T},\boldsymbol{W})\sim p_{\text{data}}(\boldsymbol{T},\boldsymbol{W})}\left[1 - MS\_SSIM(Q(\boldsymbol{T}), \boldsymbol{W})\right]$$

Note that the reconstructed low resolution image $Q(\boldsymbol{T})$ is four times smaller than the dense PALM image $\boldsymbol{T}$, as described in **Supplementary Note 1**. Because the input widefield image $\boldsymbol{W}$ can have a different size, we use bilinear interpolation to resize $\boldsymbol{W}$ to the same size as $Q(\boldsymbol{T})$. (If needed, a scaling factor different from four can be obtained by adding or removing downsample layers in network Q). At inference, the low resolution estimator Q is also used to produce the error map, as shown in **Figure 1b**, and **Supplementary Figures 6, 11, 15**. This error map is defined as:

$$\boldsymbol{E}_Q(\boldsymbol{A}, \boldsymbol{W}) = \big(1 - MS\_SSIM(Q(\boldsymbol{A}), \boldsymbol{W})\big)(Q(\boldsymbol{A}) + \boldsymbol{W})$$

High (respectively low) values of the error map indicate large (respectively small) inconsistencies, between the reconstructed super-resolution image $\boldsymbol{A}$ and the observed widefield image $\boldsymbol{W}$.

The third term of $\mathcal{L}_{\mathcal{G}}$ draws from recent work on generative adversarial networks (GAN)[21,23,53] and is noted $\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})$. In a GAN, a generator network $\mathcal{G}$ learns to transform random input vectors $\boldsymbol{z}$ (drawn from a probability density $p_{\mathrm{z}}(\boldsymbol{z})$) into new samples of a data probability density $p_{\mathrm{data}}(\boldsymbol{x})$. In our case, the data samples $\boldsymbol{x}$ are the dense PALM images $\boldsymbol{T}$. The generator $\mathcal{G}$ learns by working against a discriminator network $\mathcal{D}$ that simultaneously learns to discriminate between original data samples and samples generated by $\mathcal{G}$. Adversarial training thus consists in playing a minmax game such that $(\mathcal{G}^*,\mathcal{D}^*)=\arg\min_{\mathcal{G}}\max_{\mathcal{D}}\mathcal{L}_{\mathrm{GAN}}(\mathcal{G},\mathcal{D})$, with an objective function of the form[23]: $\mathcal{L}_{\mathrm{GAN}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\left[1-\mathcal{D}\left(\mathcal{G}(\boldsymbol{z})\right)\right]$, or equivalently by simultaneous optimization of two coupled loss functions:

$$\begin{cases}\mathcal{D}^*=\arg\max\limits_{\mathcal{D}}\left(\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\left[1-\mathcal{D}\left(\mathcal{G}(\boldsymbol{z})\right)\right]\right)\\ \mathcal{G}^*=\arg\min\limits_{G}\left(\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\left[1-\mathcal{D}\left(\mathcal{G}(\boldsymbol{z})\right)\right]\right)\end{cases}$$

In a conditional GAN (cGAN), the generator and the discriminator have an extra input vector $\boldsymbol{c}$ and the first objective function above becomes: $\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\mathrm{data}}(\boldsymbol{c},\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{c},\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\left[1-\mathcal{D}\left(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z})\right)\right]$, such that the generator learns a conditional probability density $p_{\mathrm{data}}(\boldsymbol{x}|\boldsymbol{c})$; and the second objective function likewise becomes $\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log\left[1-\mathcal{D}\left(\mathcal{G}(\boldsymbol{z})\right)\right]$. In our A-net, we replaced the logarithmic losses above by least square losses[53], as they empirically yielded better results. Thus, we used the objective functions:

$\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D})=\mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\mathrm{data}}(\boldsymbol{c},\boldsymbol{x})}(\mathcal{D}(\boldsymbol{c},\boldsymbol{x})-1)^2+\mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\left[\mathcal{D}\left(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z})\right)\right]^2$ and

$\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\left[1-\mathcal{D}\left(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z})\right)\right]^2$. In our case, the input $\boldsymbol{c}$ is the sparse PALM image $\boldsymbol{S}$ combined with the upsampled version $\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ of the widefield image $\boldsymbol{W}$, where $\mathcal{B}$ denotes bilinear interpolation. Note that in practice the noise $\boldsymbol{z}$ in our ANN was introduced only through the use of dropout layers, as in the pix2pix implementation[21]. Thus, the objective functions are:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{D}) = \mathbb{E}_{\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M}\sim p_{\mathrm{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}(\mathcal{D}(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\boldsymbol{T})-1)^2 + \mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z}),\boldsymbol{S}\sim p_{\mathrm{data}}(\boldsymbol{S})}\left(\mathcal{D}\left(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})\right)\right)^2$$

and:

$$\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\mathrm{data}}(S,T,W,M)}\left(\mathcal{D}\left(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})\right)-1\right)^2$$

In the end, combining the three loss terms described above, we implemented the following optimization problem :

$$\begin{cases} Q^* = \arg\min_Q \mathcal{L}_Q(Q) \\ \mathcal{D}^* = \arg\max_{\mathcal{D}} \mathcal{L}_{\mathcal{D}}(\mathcal{D}) \\ \mathcal{G}^* = \arg\min_{\mathcal{G}}\left[\alpha\mathcal{L}_{\mathrm{SuperRes}}(\mathcal{G}) + \beta\mathcal{L}_{\mathrm{LowRes}}(\mathcal{G},Q) + \gamma\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})\right] \end{cases}$$

The weights α, β and γ are hyperparameters, which we set manually to $\alpha = 50$, $\beta = 25$ and $\gamma = 1$ for most experiments. In absence of widefield images $\boldsymbol{W}$, β was simply set to zero. The reported results are not very sensitive to these parameters.

We trained the A-net end-to-end using stochastic gradient descent (SGD) with Adam[54] and a batch size of 1 with 200,000 or more iterations (backpropagation steps). Our implementation was adapted from affinelayer's TensorFlow[55] implementation, which is ported from the Torch implementation of pix2pix[21]. Both network training and inference were performed on Tesla P100, Tesla M40, Tesla

K80 or GTX TitanXP graphical processing units (GPUs) from Nvidia. A-net training from scratch typically takes from hours to days on a single GPU. Once trained, the A-net takes only ~1 second or less to reconstruct a super-resolution image of 2560x2560 pixels (corresponding to an entire FoV). Training time could be further reduced by pretraining (or transfer learning), use of GPU clusters, or optimized data augmentation.

Experimental training images and data augmentation

Experimental training data are obtained from standard localization microscopy data (dense PALM images). To achieve good performance, ANNs generally necessitate large amounts of training data. However, ANNA-PALM typically requires PALM images from no more than 10 FoVs (of 55 µm x 55 µm each) and can even be trained with a single FoV. This is possible thanks to an extensive on-the-fly data augmentation strategy, as described below. Each of the dense PALM training images corresponds to a list of localizations $(f_i, x_i, y_i)_{i=1..n}$, where $f_i \in [1, K]$ is the index of the diffraction limited frame from which localization x_i, y_i originates, and K is the total frame number. PALM images are obtained as plain 2D histograms of these localizations with typical pixel sizes of 10-20 nm. From each list of localizations (corresponding to a dense PALM image of a single FoV), we generate 10-30 pairs of input and target images ($\boldsymbol{S}$,$\boldsymbol{T}$) for training. To define the target image $\boldsymbol{T}$, we take a random consecutive subset $[k_T, k_T + 0.95K]$ of 95% of all available K frames (k_T is chosen randomly between 0 and $0.05K$) and create the 2D histogram image based on localizations from those frames only, i.e. from all (x_j, y_j) such that $f_j \in$

$[k_T, k_T + 0.95K]$. To define the sparse input image $\boldsymbol{S}$, we take random subsets of 300-500 consecutive frames from the first half of the image sequence and similarly create a 2D histogram of the localizations from those frames only. When a widefield image $\boldsymbol{W}$ is available, this image must first be aligned with the corresponding dense PALM image $\boldsymbol{T}$. This is done using an FFT-based phase correlation algorithm[56] after histogram equalization of image $\boldsymbol{T}$, smoothing by convolution with a Gaussian Kernel of standard deviation 6 pixels, and resizing with bilinear interpolation to the same size as image $\boldsymbol{W}$. The registered widefield image $\boldsymbol{W}$ is then scaled up using bilinear interpolation to an image $\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ with the same size as image $\boldsymbol{T}$.

During training, for each iteration of SGD, we crop the images $\boldsymbol{S}$, $\boldsymbol{L}$ and $\boldsymbol{T}$ with a randomly placed 712x712 pixel sized region $\mathcal{R} = [x_{\min}, x_{\min} + 712] \times [y_{\min}, y_{\min} + 712]$. We then use random geometric transformations and apply them identically to the three images. Specifically, we rotate the images by a random angle between 0 and 360 degrees, apply elastic transformations [57], and then crop the center region of size 512x512 pixels. In addition to geometric transformations, we also introduce realistic noise from experimental background images. This is done by manually outlining regions of background in selected PALM training images, splitting these regions into small patches of 40x40 pixels, grouping them according to their summed pixel values, then assembling them into a larger image with the same size as $\boldsymbol{T}$. During training, these semi-synthetic noise images are randomly selected and added to the input image $\boldsymbol{S}$ without altering $\boldsymbol{T}$ or $\boldsymbol{W}$. Finally, we normalize the input image $\boldsymbol{S}$ by subtracting its mean and dividing by standard deviation. If a widefield image $\boldsymbol{W}$ is provided, its pixel values are scaled to a

minimum of 0 and a maximum of 1. Otherwise, ***W*** is replaced by an image containing zeros only. The target image ***T*** is truncated at a maximum value of 255 and then scaled to have a minimum of 0 and a maximum of 1. For the switch ***M***, we used an integer number to define the type of training images, e.g. 0 for microtubules, 1 for nucleoporins, 2 for mitochondria. When training on different types of images, e.g. microtubules and nucleoporins (**Supplementary Figure 13**), we assign the corresponding switch value to ***M*** and use it as additional input to the A-net together with images ***S*** and ***W***, as described in **Supplementary Note 1.**

Image simulations

Our procedure to simulate localization microscopy (PALM) images of microtubules is illustrated in **Supplementary Figure 1**. To simulate microtubule filaments, we used a Langevin dynamics simulation[29] that generates random configurations of semiflexible curves with a specified rigidity (persistence length), starting from a random initial configuration (**Supplementary Figure 1a,b**). The initial configurations were generated with a Python library named cpolymer and the Langevin dynamics was implemented using the molecular dynamics code LAMMPS[58]. Although the simulation generates 3D polymer chains, we only considered their 2D projections, consisting of N_p connected positions $(x_k, y_k)_{k=1..N_p}$. To obtain smooth filaments we further interpolated these connected segments using spline functions with the Scipy function scipy.interpolate.splev. Next, we turned these 2D curves into a grey scale image of 800x800 pixels, with an assumed pixel size of 7.8 nm, using the Python library Matplotlib. This image was further

convolved with a Gaussian kernel of standard deviation 1.5 pixels, resulting in a smooth image $I(i,j)$ as shown in **Supplementary Figure 1c,** and normalized to a probability density ($\sum\sum_{i,j} I(i,j) = 1$, with all $I(i,j) \geq 0$). This image was used to mimic a "perfect" PALM image of filaments corresponding to an infinite number of localizations ($n = \infty$). Such perfect images were used as targets during ANN training for simulated data and defined as ground truth for the quantification of reconstruction quality by MS-SSIM (**Figure 2g**). During training, we applied the same rotations and elastic transformations described for experimental data in the previous section.

Localization microscopy images obtained from a finite number of localizations $n < \infty$ (sparse PALM images $\boldsymbol{S}$), can be considered as a sampling of the probability density $I(i,j)$ with n samples. These images can therefore be simulated by applying Poisson noise to a rescaled version of the perfect PALM image, i.e.: $\boldsymbol{S}(\lambda, I) = \mathcal{P}(\lambda\, I/I_{\max})$, where $I_{\max}$ is the maximum value of I, $\mathcal{P}(\mu)$ denotes the Poisson probability distribution of mean μ and where the peak parameter λ controls the level of sampling. In order to simulate sparse PALM images for various levels of sampling, we varied the peak value λ following a log-normal distribution where $\ln(\lambda)$ has mean -0.5 and standard deviation 0.001 and applied Poisson noise using the numpy library function random.poisson. An example of a simulated sparse PALM image is shown in **Supplementary Figure 1d**. Besides finite sampling, localization microscopy images are corrupted by additional noise sources such as false detections from background noise due to out-of-focus light or unspecific binding of antibodies. To mimic this, we first created a probability density

$I_b = I * G_{\sigma_b}$ for the background noise by convolving I with a Gaussian kernel G_{σ_b} of large standard deviation σ_b=25 pixels, and applied Poisson noise with $\lambda = 0.06$. To create training images, we added this background noise image to the sparse PALM image $\boldsymbol{S}$ above (**Supplementary Figure 2**). We did not add background noise to the test images used during inference (**Figure 2**). To simulate the widefield images (**Figure 2a**), we first blurred the perfect PALM image by convolution with a Gaussian kernel of standard deviation 8 pixels, then added Gaussian noise with zero mean and standard deviation chosen randomly between 0.5 and 1.5.

For simulations of nuclear pore images (**Supplementary Figure 15**), we applied a similar procedure, except that the perfect PALM images were obtained by randomly distributing circles of diameter 150 nm in the plane (avoiding overlaps) and placing eight Gaussian spots (of standard deviation 1.7 pixels) at equal distance from each other on each circle to mimic the octogonal shape of nuclear pores.

Sample preparation

For microtubule imaging experiments (**Figures 3-5** and **Supplementary Figures 8-10**), except those using DNA-PAINT (**Figure 5j,l**), U-373 MG (Uppsala) cells were cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12; Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v) penicillin-streptomycin (Gibco), in a 5% CO_2 environment at 37°C on 18-mm cleaned coverslips in 12-well plates. 24 hours after plating, cells were pre-extracted for 10 s in 0.25% (v/v) Triton X-100 (Triton) in BRB80 (80 mM PIPES, 1 mM MgCl2, 1 mM EGTA, adjusted to pH 6.8 with KOH) supplemented with 4 mM EGTA, and

immediately fixed for 10 min with 0.25% (v/v) Triton + 0.5% Glutaraldehyde in BRB80, followed by reduction for 7 min with 0.1% NaBH4 solution in PBS and another washing step in PBS. Cells were directly incubated for 1h at room temperature in PBS with 1:500 rat alpha-tubulin antibodies (Bio-Rad MCA77G), followed by 3 washing steps with PBS, and then incubated for 45min in PBS with 1:500 anti-rat Alexa-647 conjugated secondary antibodies from donkey (Jackson ImmunoResearch Laboratories, ref. 712-605-153).

For the DNA-PAINT experiment on microtubules (**Figure 5k-m**), U-373 cells stuck on 18 mm diameter coverslips were fixed at 37°C with 4% PFA in PHEM buffer and permeabilized in 0.2% glutaldehyde. Next, cells were incubated for 1 h with 1:500 primary mouse antibodies against alpha-tubulin. The sample was washed 3 times in PBS, then incubated with 1:100 anti-mouse oligo-conjugated antibodies from Ultivue Kit 2 for DNA-PAINT imaging[33]. After washing the sample 3 times in PBS, and just before imaging, 2nM of complementary oligos coupled to Cy3 fluorophores were added to the sample.

Nuclear pore imaging data of gp210 and WGA (**Figure 6a-c** and **Supplementary Figure 14**) were kindly provided by J. Sellés and O. Falklaris and obtained from nuclear membranes of *Xenopus* frog eggs prepared as described previously[36].

For mitochondria imaging experiments (**Figure 6e-f**), COS7 cells were cultured under the same conditions as U-373 cells above using phenol-red free DMEM medium and fixed with 4% PFA in PBS for 10 min. The sample was blocked with 3% BSA in PBS for 20 min and immunostained with 1:500 mouse antibodies

against TOM22 (Sigma, ref. T6319) in wash buffer (PBS with 0.5% BSA) for 1 h. After extensive washing with wash buffer, the sample was incubated with 1:500 anti-mouse secondary antibodies from donkey conjugated to Alexa-647 dyes (Jackson ImmunoResearch Laboratories, ref. 715-605-151) in wash buffer for 30 min. After washing 5 times with wash buffer and 2 times with PBS, samples were post-fixed with 2% PFA in PBS for 10 min and washed 5 times with PBS.

For all localization microscopy experiments except DNA-PAINT, we used a photoswitching buffer[59] composed of 50 mM Tris-HCl + 10 mM NaCl + 10% (w/v) glucose + 168 AU/mL Glucose-Oxidase + 1404 AU/mL Catalase + 1% 2-Mercaptoethanol. For microtubule imaging experiments, we used this buffer to fill a square hole that was manually cut in a parafilm sheet, which was deposited on a rectangular coverglass. The round coverslips were sealed with nail polish.

Image acquisition in localization and high-throughput microscopy

We performed single molecule localization microscopy experiments (PALM/STORM and DNA-PAINT) on custom built microscopy systems, as previously described[59–61]. The system used for PALM/STORM imaging of microtubules is based on an inverted microscope body (Nikon Ti Eclipse) equipped with a either a 60x 1.49 NA oil immersion objective (Nikon) or a 60x 1.2 NA water immersion objective (Nikon) and with the Perfect Focus System active. A 642 nm wavelength laser with 500 mW power was used to excite Alexa-647 fluorophores and an AOTF (AA optics) was used to modulate laser excitation. Sequences of diffraction limited single molecule image frames were acquired either on a sCMOS camera (Hamamatsu ORCA-Flash4.0),

which can capture images of 2,042x2,042 pixels (for **Figures 4, 5a-i**), or on an EMCCD (Andor IXON ULTRA 897) with 512x512 pixels (for **Figures 3, 6d-f**). Both cameras were controlled by MicroManager software[62]. For experiments using the sCMOS camera, the effective pixel size was 108 nm and we used a 512x512 region of interest, which resulted in an imaged FoV of 55.3 µm x 55.3 µm. For experiments using the EMCCD camera, we used a 2x telescope and the effective pixel size was 107 nm, resulting in a FoV of 54.8 µm x 54.8 µm. The exposure time was set to Δt =10 ms or 30 ms per frame. The number of frames acquired ranged from k=1,000 (**Figure 4**) to K=60,000 (e.g. **Figure 5c**) per FoV.

For the DNA-PAINT experiment (**Figure 5j,l**), we used an inverted Nikon Ti-E Eclipse microscope equipped with a 100x 1.49 NA TIRF objective and with the Perfect Focus System active. A 561 nm wavelength laser with 500 mW power was used to excite Cy3 dyes. Highly inclined laser illumination was used to reduce out-of-focus background signal. Images were acquired on an EMCCD camera as above, with a 1.5 x telescope, resulting in an effective pixel size of 106 nm and a FoV of 54 µm x 54 µm. The sample was mounted in a magnetic sample holder filled with the imaging buffer provided with the Ultivue kit. Exposure time was set to Δt =30 ms and the EM gain of the EMCCD was set to 300. The laser power was increased until isolated fluorescent spots were observed. For the experiment shown in **Figure 5g,i**, K=60,000 frames were acquired.

The *Xenopus* nuclear pore data (**Figure 6a-c** and **Supplementary Figure 14**), were acquired on a Zeiss Elyra P.S.1 microscope as described previously[36].

For high-throughput imaging of microtubules (**Figure 4**), we used the Multi-Dimensional Acquisition tool in Micro-manager to define the positions of 1,089 FoVs of 55.3 µm x 55.3 µm on a 33x33 grid, with overlaps of 1 µm; the stage was automatically shifted to each of these 1,089 positions. We first acquired only widefield images, taking five frames at each of these positions (the first two were ignored because of motion blur), in a total acquisition time of 12 minutes. Then, the laser power was raised to bleach out preactivated molecules and k=1,000 frames of single molecule images were acquired at each of the 1,089 positions, in a total acquisition time of 3 hours and 8 minutes. Raw image frames were written directly to a remote storage server via Samba networking protocol.

Localization microscopy image analysis

The input to ANNA-PALM reconstruction is a localization image, defined as a 2D histogram of n single molecule positions $(x_i, y_i)_{i=1..n}$. The histogram bin, i.e. the pixel size of the localization image, was set to 7.8 nm for the simulated data (**Figure 2** and **Supplementary Figures 1-6,15**) and 20 nm for the experimental data (**Figures 3-6** and **Supplementary Figures 7-14**). The positions $(x_i, y_i)_{i=1..n}$ were obtained by analyzing sequences of diffraction limited frames using standard single molecule localization algorithms. For experimental microtubule images, we used the ThunderSTORM[63] plugin of ImageJ, applying wavelet filters for detection and weighted least squares Gaussian fitting for precise estimation of subpixelic positions. We used the cross-correlation feature in ThunderSTORM for drift correction, and filtered out the least certain localizations based on the fitted

Gaussian's standard deviation and the χ^2 of the residual. Localizations in consecutive frames separated by less than 20 nm were assumed to originate from the same molecule and merged into a single localization. The final number n of localizations was ~7 million for the full 55 µm x 55 µm FoV of the images shown in **Figure 3** and **Supplementary Figures 8,9** (obtained from K=30,000 frames). For the high-throughput experiment (**Figure 4** and **Supplementary Figure 12**), the number of localizations per 55 µm x 55 µm FoV ranged from n=2,949 to n=1,442,048 with an average $\langle n \rangle = 610{,}983$ and standard deviation $\sigma(n) = 273{,}606$. The total number of localizations across all 1,089 FoVs was ≈665 million. ThunderSTORM analyses were performed either on high end workstations or on Institut Pasteur's high performance computer (HPC) cluster. For the high-throughput experiments, we used Python scripts to run ThunderSTORM in batch mode (without user intervention) on the HPC cluster and assembled mosaic images (**Figure 4a,b** and **Supplementary Figure 12a,b**) using a stitching plugin of ImageJ[64].

The nuclear pore images were analyzed using the ZEN software from Zeiss as previously described[36]. For the DNA-PAINT experiments, we used PALMTT, a modified version of the single molecule tracking algorithm MTT[65], based on Matlab (Mathworks). This algorithm uses Gaussian smoothing and thresholding for detection, and Gaussian fitting for precise estimation of subpixelic positions. Drift correction was performed computationally by tracking fluorescent beads used as fiducial markers.

Quality metrics and sampling resolution

In order to quantitatively assess the quality of PALM images and ANNA-PALM reconstructions, we calculated the multi-scale structural similarity index (MS-SSIM) between either image and the ground truth (**Figure 2g** and **Supplementary Figures 5,9**). For the simulated data, the ground truth was simply defined as the "perfect" PALM image, corresponding to an infinite number of localizations (see 'Image simulations' above, **Figure 2e** and **Supplementary Figure 1c**). For the experimental data, the ground truth was defined as the ANNA-PALM reconstruction of a dense PALM image obtained from all available frames (e.g. **Supplementary Figure 9c**). Before calculation of the MS-SSIM, all simulated images were linearly normalized without clipping to a maximum value of 255.

In order to evaluate the effect of sampling on the resolution of PALM images, we computed the double mean distance, $R_{\mathrm{Nyq}} = 2\langle d \rangle$ between nearest neighbors in the underlying sets of localizations[49] as function of localization number (**Supplementary Figure 3d**). For the simulated sparse PALM data, sets of localizations were obtained by interpreting each image $\boldsymbol{S}$ as a 2D histogram of localizations, and creating a random subpixelic position $(x_k, y_k)_{k=1..m}$ within each pixel (i, j), as many times as given by the pixel value $m = S(i, j)$ (therefore resulting in a set of $n = \sum_{i,j} S(i, j)$ localizations). The quantity $R_{\mathrm{Nyq}}(n) = 2\langle d(n) \rangle$ decreases towards zero with increasing number of localizations, n. A Nyquist criterion introduced in ref.[49] suggests that the resolution is limited by sampling to no less than R_{Nyq}, implying that at least $n \geq N_{\mathrm{Nyq}} = R_{\mathrm{Nyq}}^{-1}(R)$ localizations are needed to achieve a given resolution R. However, a more stringent and realistic criterion[3]

prescribes a five-fold larger number of localizations to reach resolution R: $n \geq N_{5\times\mathrm{Nyq}} = 5 \times N_{\mathrm{Nyq}}$, implying that the sampling limit to resolution is $R_{5\times\mathrm{Nyq}} = 2\langle d(n/5)\rangle$. Accordingly, if this condition is met, the resolution is no longer limited by sampling, but by the localization precision, $R_{\mathrm{loc}} \approx 2.3\sigma_{\mathrm{loc}}$, where σ_{loc} is the standard deviation of localization errors along each coordinate. In general, the resolution of a PALM image, as limited by both sampling and localization precision, can be written: $R = \max\left(R_{5\times\mathrm{Nyq}}, R_{\mathrm{loc}}\right) = \max(2\langle d(n/5)\rangle, 2.3\sigma_{\mathrm{loc}})$.

Life Sciences Reporting Summary.

Further information is available in the Life Sciences Reporting Summary.

Code availability statement

The source code of ANNA-PALM is available from https://annapalm.pasteur.fr/ .

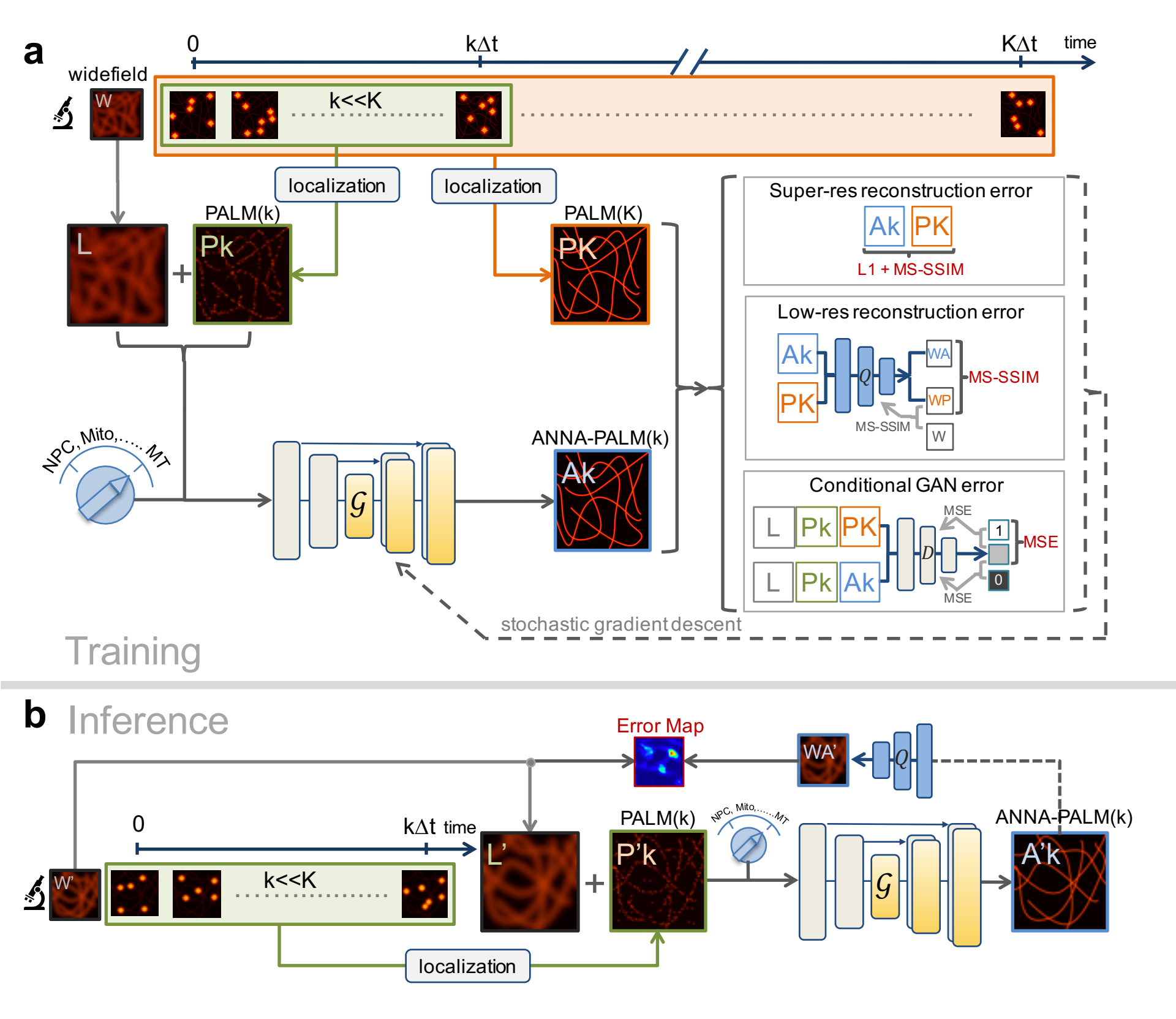
Data availability statement

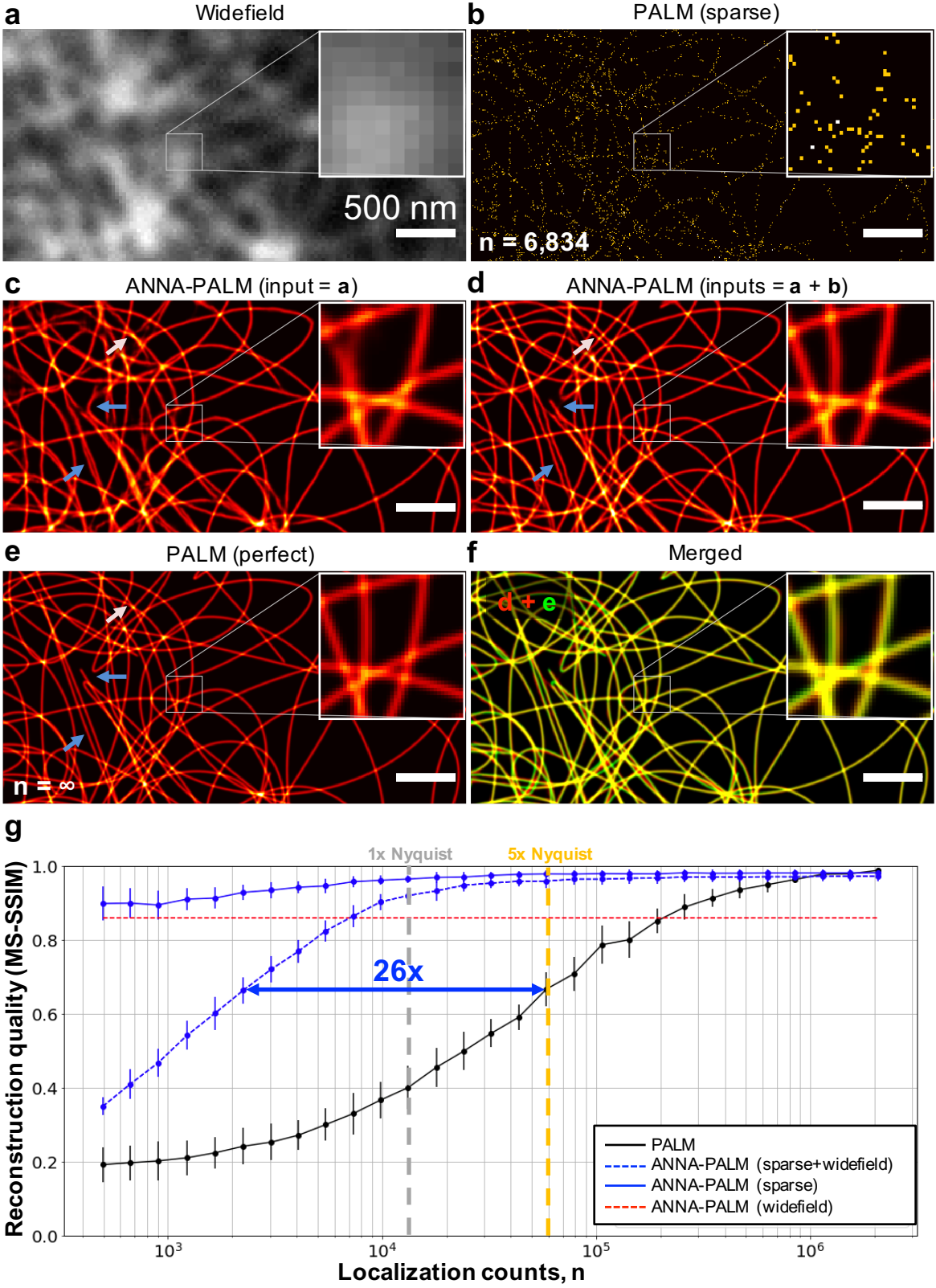
The localization data used in this paper can be downloaded directly from https://annapalm.pasteur.fr/.

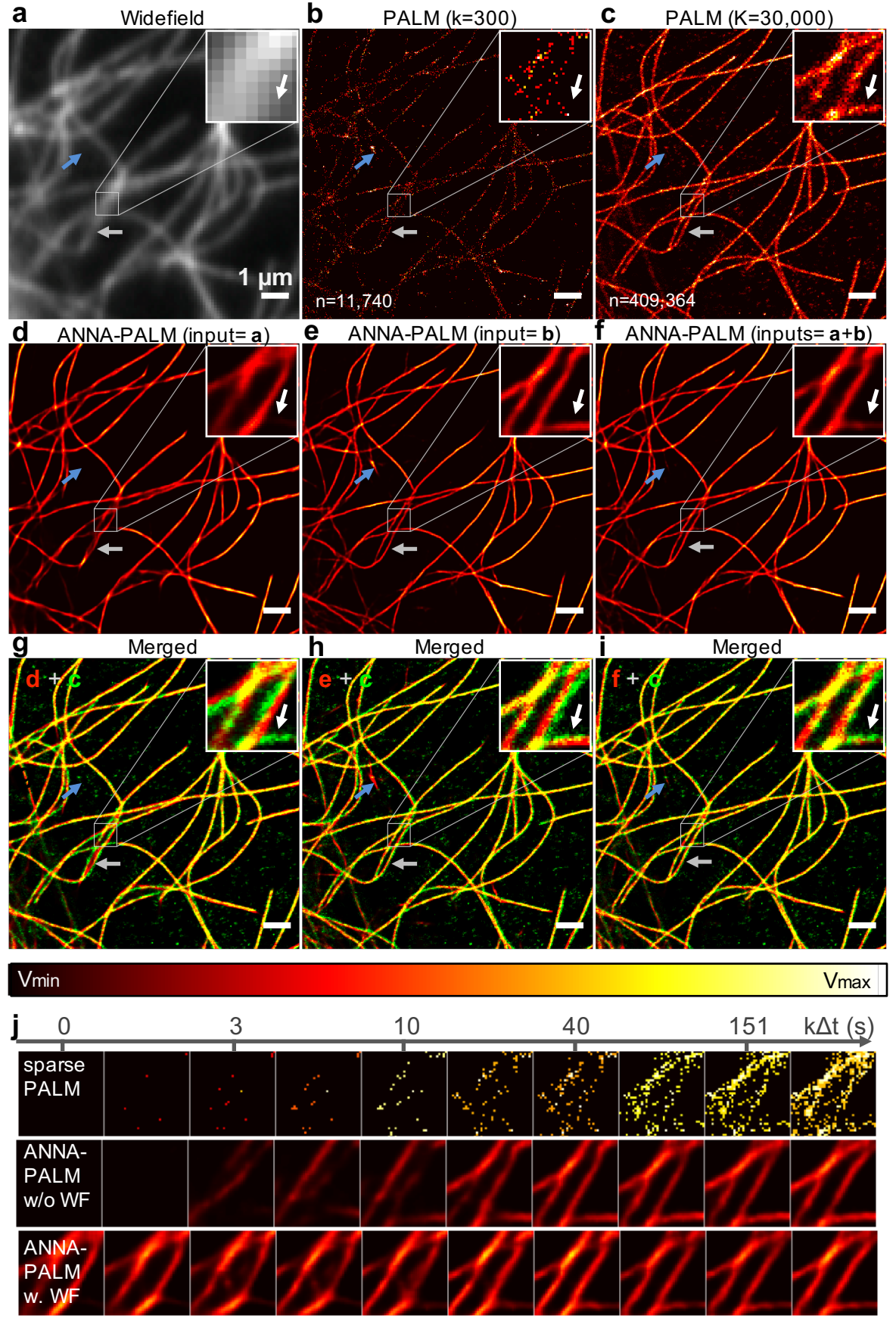
Methods-only references

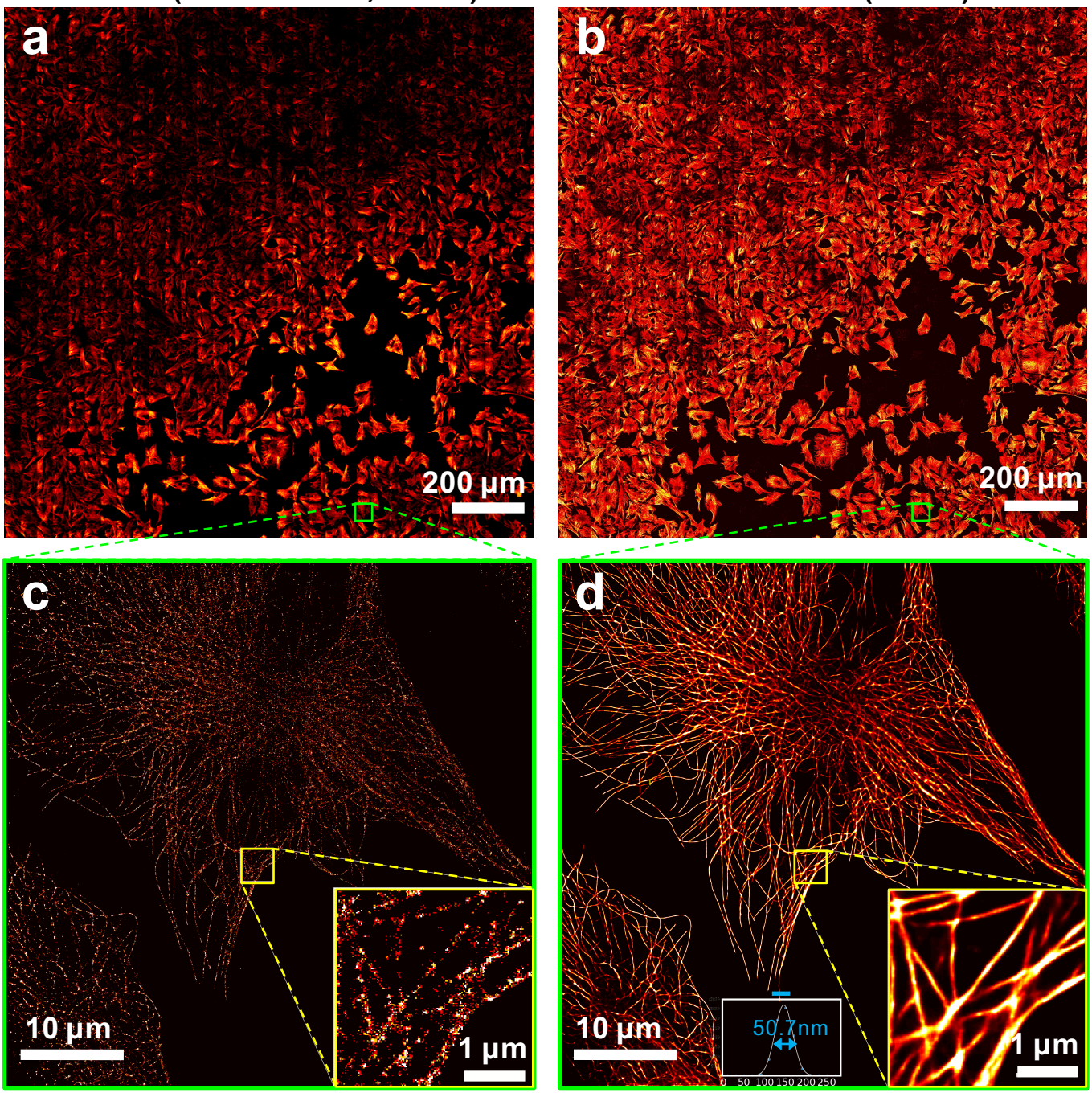
51. Xu, L., Choy, C. S. & Li, Y. W. Deep sparse rectifier neural networks for speech denoising. in *2016 International Workshop on Acoustic Signal Enhancement, IWAENC 2016* (2016). doi:10.1109/IWAENC.2016.7602891
52. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proc. 30 th Int. Conf. Mach. Learn.* (2013).
53. Mao, X. *et al.* Least Squares Generative Adversarial Networks. in *ICCV* 2794–2802 (2017). doi:10.1109/ICCV.2017.304

54. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ICLR* 1–15 (2015). doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503
55. Rampasek, L. & Goldenberg, A. TensorFlow: Biology's Gateway to Deep Learning? *Cell Syst.* **2,** 12–14 (2016).
56. Srinivasa Reddy, B. & Chatterji, B. N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5,** 1266–1271 (1996).
57. Simard, P. Y., Steinkraus, D. & Platt, J. C. Best practices for convolutional neural networks applied to visual document analysis. *Seventh Int. Conf. Doc. Anal. Recognition, 2003. Proceedings.* **1,** 958–963 (2003).
58. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117,** 1–19 (1995).
59. van de Linde, S. *et al.* Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat. Protoc.* **6,** 991–1009 (2011).
60. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat. Methods* **7,** 339–340 (2010).
61. Lelek, M. *et al.* Superresolution imaging of HIV in infected cells with FlAsH-PALM. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8564–9 (2012).
62. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using manager. *Current Protocols in Molecular Biology* (2010). doi:10.1002/0471142727.mb1420s92
63. Ovesny, M., Kiek, P., Borkovec, J., Vindrych, 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).
64. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25,** 1463–1465 (2009).
65. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods* **5,** 687–694 (2008).

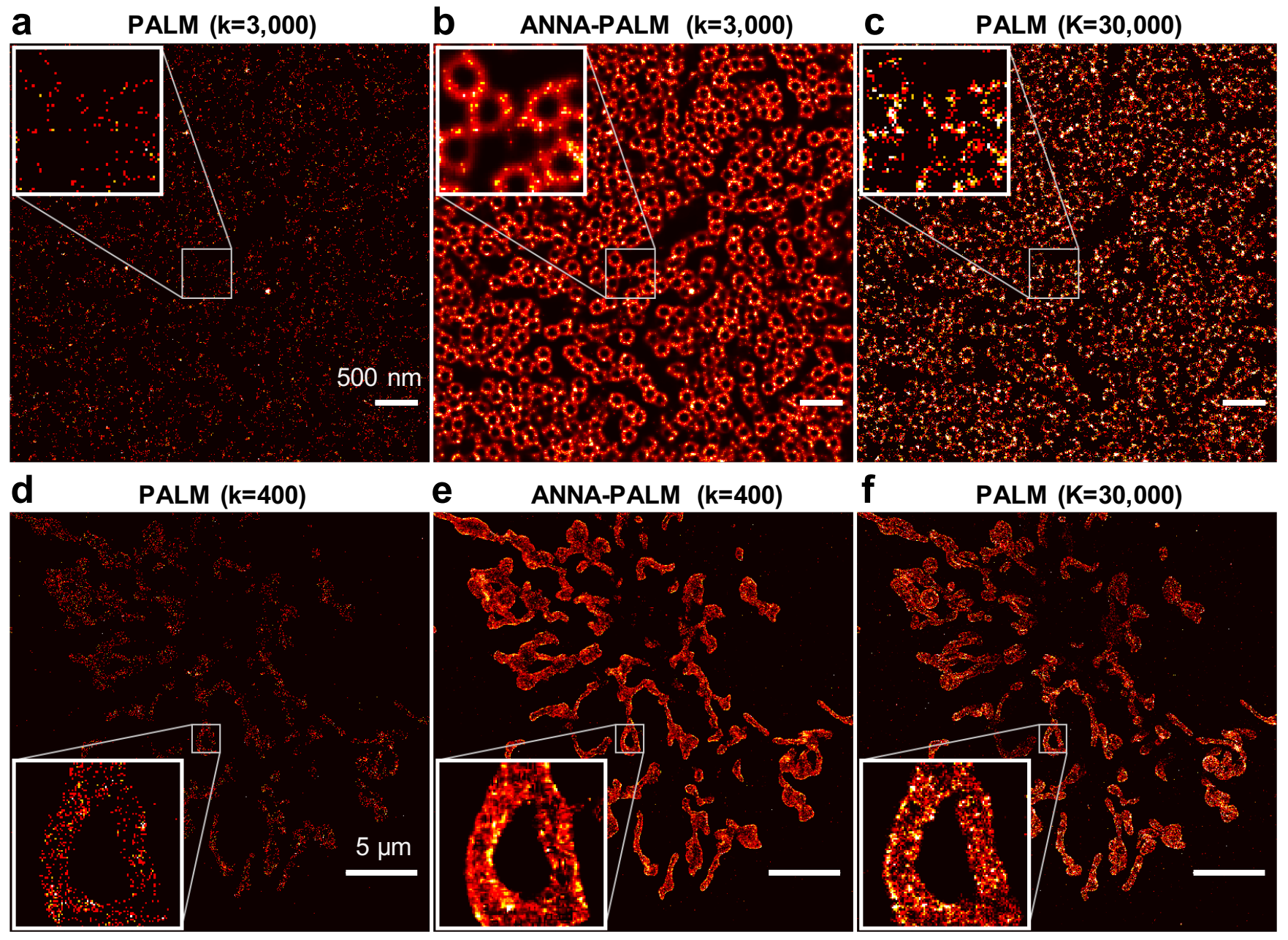

a
0
kΔt
KΔt
time
widefield
W
k<<K
localization
localization
PALM(k)
PALM(K)
L
Pk
PK
Super-res reconstruction error
Ak
PK
L1 + MS-SSIM
Low-res reconstruction error
Ak
PK
Q
WA
WP
W
MS-SSIM
MS-SSIM
NPC, Mito,..... MT
G
ANNA-PALM(k)
Ak
Conditional GAN error
L
Pk
PK
L
Pk
Ak
D
MSE
1
0
MSE
MSE
stochastic gradient descent
Training
b
Inference
Error Map
WA'
Q
0
kΔt
time
W'
k<<K
L'
PALM(k)
P'k
NPC, Mito,.....MT
G
ANNA-PALM(k)
A'k
localization


a
Widefield
500 nm
b
PALM (sparse)
n = 6,834
c
ANNA-PALM (input = a)
d
ANNA-PALM (inputs = a + b)
e
PALM (perfect)
n = ∞
f
Merged
d + e
g
1x Nyquist
5x Nyquist
26x
Reconstruction quality (MS-SSIM)
1.0
0.8
0.6
0.4
0.2
0.0
10^3
10^4
10^5
10^6
Localization counts, n
PALM
ANNA-PALM (sparse+widefield)
ANNA-PALM (sparse)
ANNA-PALM (widefield)


a
Widefield
1 µm
b
PALM (k=300)
n=11,740
c
PALM (K=30,000)
n=409,364
d
ANNA-PALM (input= a)
e
ANNA-PALM (input= b)
f
ANNA-PALM (inputs= a+b)
g
Merged
d + c
h
Merged
e + c
i
Merged
f + c
Vmin
Vmax
j
0
3
10
40
151
kΔt (s)
sparse PALM
ANNA-PALM w/o WF
ANNA-PALM w. WF


PALM (k=1000x1089, t=3.1h)
ANNA-PALM (t=3.1h)
a
b
200 μm
200 μm
c
d
10 μm
1 μm
10 μm
50.7nm
1 μm
0 50 100 150 200 250

a PALM (k=500)
b ANNA-PALM (k=500)
c PALM (K=60,000)
Control
5 μm
d PALM (k=800)
e ANNA-PALM (k=800)
f PALM (K=60,000)
Taxol (1uM, 20min)
5 μm
g PALM (k=300)
h ANNA-PALM (k=300)
i PALM (K=60,000)
Nocodazole (1uM, 20min)
5 μm
j DNA-PAINT(k=400)
k ANNA-PALM (k=400)
l DNA-PAINT (K=60,000)
DNA-PAINT
5 μm


a
PALM (k=3,000)
500 nm
b
ANNA-PALM (k=3,000)
c
PALM (K=30,000)
d
PALM (k=400)
5 μm
e
ANNA-PALM (k=400)
f
PALM (K=30,000)