Deep learning massively accelerates super-resolution localization microscopy

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

2   **ANNA-PALM: Deep learning accelerates super-resolution localization**

3   **microscopy by orders of magnitude**

4

5

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

7   Zimmer[1,2,3]

8

9   [1] Institut Pasteur, Unité Imagerie et Modélisation, Paris, France

10   [2] UMR 3691, CNRS

11   [3] C3BI, USR 3756, IP CNRS

12

13

14

15

16

17

18

19

20

21

22

23   Corresponding author: C. Zimmer

24 The speed of super-resolution microscopy methods based on single molecule

25 localization, e.g. PALM or STORM, is severely limited by the need to record

26 many thousands of frames with a low number of observed molecules in each.

27 Here, we present ANNA-PALM, a computational strategy that uses artificial

28 neural networks to reconstruct super-resolution views from sparse, rapidly

29 acquired localization images and/or widefield images. Simulations and

30 experimental imaging of microtubules, nuclear pores and mitochondria show

31 that high-quality super-resolution images can be reconstructed from up to two

32 orders of magnitude fewer frames than usually needed, without

33 compromising spatial resolution. Super-resolution reconstructions are even

34 possible from widefield images alone, though adding localization data

35 improves image quality. We demonstrate super-resolution imaging of >1,000

36 fields of view containing >1,000 cells in ~3 h, yielding an image spanning

37 spatial scales from ~20 nm to ~2 mm. The drastic reduction in acquisition

38 time and sample irradiation afforded by ANNA-PALM enables faster and

39 gentler high-throughput and live cell super-resolution imaging.

40

41 Fluorescence microscopy methods that overcome the diffraction limit of resolution

42 (∼200-300 nm) allow imaging of biological structures with molecular specificity

43 closer to the molecular scale. Among super-resolution microscopy approaches,

44 those based on single molecule localization, such as PALM[1] or STORM[2] (hereafter

45 referred to collectively as PALM) are particularly attractive owing to their exquisite

46 spatial resolution and ease of implementation. In these methods, random subsets of

47 fluorophores are imaged in many consecutive diffraction-limited frames,

48 computationally localized to high precision, and the combined localizations are used

49 to generate a super-resolution view. In practice, typically 10^3-10^5 diffraction-limited

50 frames are needed to assemble a single super-resolution image. This requirement

51 follows from two conditions that must be simultaneously satisfied to ensure high

52 spatial resolution: (i) a low number (∼10-10^2) of active fluorophores per frame, to

53 avoid overlaps between diffraction limited spots and enable precise localization of

54 individual molecules, and (ii) a large number of independent localizations to ensure

55 a sufficiently dense sampling of the underlying biological structures[3,4]. The large

56 number of required frames makes localization microscopy inherently slow, thereby

57 limiting its potential for high-throughput imaging, where many fields of view (FoVs)

58 are to be imaged, and for imaging live cell dynamics. As a result, most localization

59 microscopy studies are restricted to analyzing a small number of cells (typically less

60 than ten).

61 Multiple approaches have been explored to accelerate localization

62 microscopy. Using bright dyes with rapid switching kinetics, high power lasers and

63 fast cameras allows to minimize exposure time without losing signal to noise

64   ratio[5,6], but reaching sub-millisecond exposure remains challenging, and intense

65   irradiation exacerbates phototoxicity in live cell imaging[7,8]. Increasing the number

66   of active fluorophores per frame can reduce acquisition time, but despite algorithms

67   designed to handle overlapping fluorescent spots[9–13] this approach necessarily

68   compromises spatial resolution[14,15].

69   Here, we introduce a computational strategy that allows reducing the total

70   number of frames and independent localizations without losing spatial resolution.

71   Unlike previous approaches, our method leverages the structural redundancy of

72   most biological images to reconstruct high quality images from vastly under-

73   sampled localization microscopy data. Our method leverages deep learning, which

74   employs artificial neural networks (ANNs) to learn complex non-linear mappings

75   between numerical inputs and outputs[16]. Accordingly, we call it 'artificial neural

76   network accelerated PALM', or ANNA-PALM.

77

78   **RESULTS**

79   **A deep learning approach to super-resolution image reconstruction**

80   We aim to reconstruct a super-resolution image of approximately similar

81   information content as a standard PALM acquisition (with K frames and N

82   localizations) from a much smaller number of raw frames ($k<<K$) without changing

83   the average density of localizations, ρ, i.e. from a much smaller number of total

84   localizations ($n = \rho k \ll N = \rho K$). If PALM images are defined as 2D histograms of

85   independent localizations, this task can be formulated as restoring an image

86   corrupted by Poisson noise (and potentially additional forms of noise). Image

87   restoration is an ill-posed problem that has an infinity of solutions in the high-

88   dimensional space of all possible images, unless additional constraints (priors) are

89   imposed that restrict the solution to a lower dimensional subspace. Suitable

90   subspaces exist because most natural images are highly redundant, and can be

91   represented to very good approximation with a much smaller number of coefficients

92   than pixels, via appropriate functions that map feature space to pixel space[17,18]. In

93   recent years, ANNs with multiple layers (deep nets) have proven very successful at

94   learning meaningful features and non-linear mappings for image classification,

95   segmentation, restoration and many other tasks[16,18,19]. Inspired by these

96   developments, we designed ANNA-PALM, a deep learning approach for restoring

97   super-resolution views from under-sampled (sparse) localization microscopy data.

98   ANNA-PALM comprises a training stage and an inference stage (**Figure 1**).

99   For training (**Figure 1a**), a few super-resolution images representative of the

100   structure of interest (e.g. microtubules, nuclear pores, or mitochondria) are

101   obtained using standard PALM imaging, i.e. by acquiring long diffraction limited

102   image sequences (e.g. $K \sim 10^4 - 10^5$, $N \sim 10^5 - 10^7$) and processing them with

103   standard localization software[20], resulting in highly sampled (dense) PALM images.

104   In addition, a low resolution (widefield) image can also be acquired, as is commonly

105   done before single molecule imaging when bleaching out preactivated fluorophores.

106   Next, the dense PALM images are under-sampled by using a much smaller number

107   of input frames, $k \ll K$, thus yielding sparse PALM images from the same

108   localization data. Then, an ANN is trained to recover approximations of the dense

109   PALM images from these sparse PALM images (and the optional widefield image).

110   Once trained, the ANN is applied to new sparse PALM images (with or without a

111   widefield image), obtained from new image sequences with small numbers of

112   frames ($k \ll K$) -and hence in much shorter time- in order to reconstruct high

113   quality super-resolution images not previously seen (inference, **Figure 1b**).

114

115   **Neural net architecture and learning strategy**

116   Our ANN, hereafter called A-net, contains a total of 25 convolutional layers, and

117   roughly 42 million trainable parameters. A-net is adapted from the pix2pix

118   network[21], which itself builds on U-nets[22] and generative adversarial networks

119   (GANs)[23], two recent successful deep learning techniques. U-nets are special types of

120   convolutional neural networks (CNNs) that have proven effective at learning multi-

121   scale representations of images and accurate, pixel-wise mappings[22,24]. GANs can

122   generate new samples from real image distributions using a generator network that

123   outputs synthetic images, and a discriminator network that outputs the probability

124   that an input image is real or synthetic, both networks being trained simultaneously

125   to compete against each other[23]. Importantly, the generator can be conditioned on

126   input data (conditional GAN, or cGAN)[21,23], e.g. on images as in the pix2pix network.

127   We modified the pix2pix architecture to accept a computational switch as additional

128   input to handle multiple types of data, and introduced an additional network to

129   evaluate the consistency between the reconstructed image and the widefield input

130   image.

131   Training of our A-net proceeds as follows. Randomly under-sampled (i.e.

132   sparse) versions of PALM images are fed as input to the A-net, while the

133 corresponding dense PALM images are defined as the A-net's targets, i.e. desired

134 outputs (**Figure 1a**). Additional, optional inputs are widefield images, if available,

135 and the switch, which indicates the image type when multiple types of images (e.g.

136 microtubules and nuclear pores) are used during training. ANN training requires

137 defining an objective function (also called loss), which measures how well the

138 outputs match the targets. We implemented a loss function containing three terms.

139 The first term measures the difference between the A-net output and the dense

140 PALM image. Instead of the widely used mean squared error (MSE), which poorly

141 reflects visual quality[25], we used a combination of the absolute difference (L1 norm)

142 with a multi-scale version of the structural similarity index, a perceptually

143 motivated quality metric shown to improve image restoration with deep learning

144 (MS-SSIM)[26]. The second term measures the consistency between the A-net output

145 and the widefield image. Although in theory the latter should simply be a blurred

146 version of the PALM image, this is often not the case in practice[27]. Therefore, we

147 introduced another CNN (with 4-layers), called 'low resolution estimator' to predict

148 the widefield image from the super-resolution image. The corresponding loss was

149 defined as the MS-SSIM between this CNN's output and the observed widefield

150 image. (In absence of a widefield image, this loss is set to zero). The third term

151 contains a cGAN discriminator loss[21], where the discriminator is a 5-layer CNN,

152 whose inputs are the sparse PALM (and widefield) image(s) and either the dense

153 PALM image or the output of the generator above; the discriminator's output is

154 compared to 0s and 1s (for synthetic and real, respectively), respectively, via the

155 MSE. We use dropout[28] and extensive data augmentation, including random

156 rotations, translations, elastic deformations and addition of noise in the input image

157 to mimic false detections and unspecific labeling. As a result, only a few dense PALM

158 images are required for successful training without overfitting. On graphical

159 processing units (GPU), training ANNA-PALM from scratch takes on the order of

160 hours to days, but when starting from a previously trained A-net, retraining can be

161 done in an hour or less.

162   Once trained, the A-net can take sparse localization data with an optional

163 widefield image as input(s), and output a reconstructed super-resolution image in

164 less than a second (**Figure 1b**). In addition, the A-net produces an 'error map' that

165 measures the consistency of this super-resolution image with the widefield image[27]

166 (when available) and can be used to estimate the degree of reliability and highlight

167 potential reconstruction artifacts. For more details, see **Online Methods** and

168 **Supplementary Note 1**.

169

170 **Validating ANNA-PALM on simulated images**

171 We first sought to validate ANNA-PALM on synthetic data. For this, we used

172 Brownian dynamics simulations[29] to generate 200 dense PALM images of semi-

173 flexible filaments mimicking microtubules, with a resolution $R_{loc} \approx 23$ nm. These

174 represent "perfect" PALM images that would be obtained with an infinite number of

175 localizations. We applied varying levels of Poisson noise to these perfect images to

176 create sparse PALM images corresponding to finite numbers of localizations

177 (**Supplementary Figure 1**). We then trained our A-net using the perfect images as

178   targets and the sparse images (and widefield image) as inputs, varying the number

179   of localizations over a large range.

180   Next, we applied the trained A-net to a distinct set of PALM images generated

181   by the same stochastic simulation (**Supplementary Figure 2**). **Figure 2a** shows a

182   widefield image and **Figure 2b** a corresponding sparse PALM image obtained from

183   n=6,834 localizations. Although curvilinear structures can be seen in this image

184   despite its sparsity, small-scale features remain highly ambiguous (**Figure 2b,**

185   inset), and the resolution according to a recently proposed five-fold Nyquist

186   criterion[3] is limited by sampling to $R_{5\times\mathrm{Nyq}}\approx 85\,\mathrm{nm}$; according to this criterion,

187    $N > N_{5\mathrm{xNyq}}$ =60,000 localizations are needed to achieve 23 nm resolution

188   (**Supplementary Figure 3d**). **Figure 2c** shows the ANNA-PALM image

189   reconstructed from the wide-field image alone, which exhibits clear and continuous

190   filaments that were not previously recognizable. Most of the relatively isolated

191   filaments roughly agree with the perfect PALM image (**Figure 2e**). In the denser

192   regions, however, many small features are erroneous, e.g. filaments are incorrectly

193   joined, displaced, split or merged (**Figure 2c,** blue arrows and **Supplementary**

194   **Figure 4, top**). By contrast, the ANNA-PALM image reconstructed from the sparse

195   PALM image alone or in combination with the widefield image exhibits continuous

196   and sharp filaments in very good agreement with the perfect PALM image

197   (**Supplementary Figure 3b,c** and **Figure 2d,e,f**). The spatial resolution of these

198   reconstructed images is limited neither by diffraction nor sampling, but only by the

199   localization precision, and is thus $R_{\mathrm{loc}}\approx 23\,\mathrm{nm}$, as in the perfect images

200   (**Supplementary Figure 3e,f**). These results indicate that high quality super-

201 resolution images can be obtained from only a small fraction of the number of

202 localizations traditionally required (here, approximately ~11% of N_{5xNyq} above; see

203 **Supplementary Figure 3d**), hence enabling a strong reduction in acquisition time.

204 Nevertheless, reconstruction errors can still occur in areas where the sparse

205 localization data are most ambiguous, e.g. where filament density is highest

206 (**Figure 2d,e,f,** white arrow). These errors can be reduced by increasing the

207 localization number n, implying a trade-off between acquisition time and

208 reconstruction quality (**Supplementary Figure 4**).

209 To quantify this trade-off, we computed the MS-SSIM between reconstructed

210 ANNA-PALM and perfect PALM images ($n = \infty$) as function of localization number,

211 from $n \sim 200$ to $n \sim 2x10^6$, in comparison with the standard PALM images

212 (**Figure 2g**). The MS-SSIM ranges from 0 to 1 and reaches 1 for perfect

213 reconstructions. For standard PALM images, the MS-SSIM increases monotonically,

214 as expected, from <0.2 to >0.95 for $n = 2x10^6$ million localizations (**Figure 2g**, black

215 curve). Using only the sparse image as input, ANNA-PALM reconstructions achieve

216 MS-SSIM that are consistently higher and increase with localization number n much

217 more rapidly than standard PALM, already exceeding 0.9 for $n \approx 10,000$

218 localizations (**Figure 2**, dashed blue curve). ANNA-PALM achieves the same MS-

219 SSIM as standard PALM at the five-fold Nyquist sampling level (≈ 0.65) with only

220 $n = 2,248$ localizations instead of $n = 58,588$, suggesting a ~26-fold speed-up. If the

221 widefield image is used as additional input, the MS-SSIM further increases, and

222 dramatically so for low localization numbers (**Figure 2g**, solid blue curve). For

223 example, with $n = 7,477$ localizations, ANNA-PALM achieves a MS-SSIM (≈ 0.95)

224   similar to standard PALM with n =644,844, implying a speed-up of roughly two

225   orders of magnitude. (Note that, if the perfect PALM image was not available for

226   these quantifications, it could be replaced by the ANNA-PALM reconstruction of a

227   dense PALM image with a large number of localizations, e.g. $n = 10^5$, with similar

228   results- see **Supplementary Figure 5**).

229   As any image restoration method, ANNA-PALM can make errors. The low

230   resolution error map described above (**Figure 1b**) provides a means to estimate

231   where errors are most likely to occur. When applied to ANNA-PALM reconstructions

232   of a sparse PALM image, this error map highlights regions containing the highest

233   density of filament crossings, where reconstructions tend to be least accurate

234   (**Supplementary Figure 6i,k**). If we artificially displace a small piece of filament in

235   this image to simulate a false positive and a false negative in the reconstruction

236   (**Supplementary Figure 6b,d,** white and blue arrows, respectively), the affected

237   regions also light up in the error map (**Supplementary Figure 6j,l**). Thus, the error

238   map offers a useful tool to highlight regions most likely to contain reconstruction

239   errors, and conversely, to outline regions where reconstructions are most

240   trustworthy. Thus, simulations suggest that ANNA-PALM can considerably reduce

241   acquisition time in localization microscopy and also map reconstruction reliability.

242

243   **ANNA-PALM reconstructions of immunostained microtubules**

244   We next tested our method on real images of immunolabeled microtubules

245   (**Figure 3**). We trained our A-net on seven dense PALM images (with corresponding

246   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-

270   PALM reconstruction from the sparse localization data alone (**Figure 3e,h**, blue

271   arrows). However, when combining the widefield and sparse PALM data, these

272   artifacts are largely removed and reconstructions agree very well with the dense

273   PALM image (**Figure 3f,i**). Reconstruction quality increases with the number of

274   frames k (**Figure 3j, Supplementary Figures 8-9** and **Supplementary Video 1**).

275   More quantitatively, a MS-SSIM analysis similar to that for the simulated data above

276   (with the ANNA-PALM output of the dense PALM image defined as ground truth; see

277   **Supplementary Figures 5, 9f**) suggests that ANNA-PALM allows a hundred-fold

278   reduction of acquisition time compared to standard PALM (**Supplementary**

279   **Figure 9g**). **Supplementary Figure 10** shows other examples of sparse

280   microtubule images reconstructed by ANNA-PALM.

281   As for simulations above, we used the widefield image to compute an error

282   map (**Supplementary Figures 6,11**). Bright areas in this error map highlight

283   regions where the reconstruction indeed disagrees with the dense PALM image;

284   conversely, reconstructions are of high quality in the majority of regions where the

285   error map is dim (**Supplementary Figure 11d-f**). These results demonstrate

286   experimentally that ANNA-PALM can restore high quality approximations of super-

287   resolution images from much shorter acquisition time than typical for PALM

288   imaging, and also predict where reconstruction errors are most likely.

289

290   **ANNA-PALM enables high-throughput super-resolution imaging**

291   The drastic improvement in imaging efficiency afforded by ANNA-PALM permits

292   super-resolution imaging of orders of magnitude more cells and FoVs per unit time.

13

293   To demonstrate this, we used an automated acquisition protocol to image >1,000

294   cells with immunolabeled microtubules in 1,089 (33x33), partly overlapping, FoVs

295   of 55.3 μm x 55.3 μm each (**Figure 4**, **Supplementary Figure 12**). We first acquired

296   widefield images at each of these positions, in a total of ~12 minutes, mostly

297   consisting of stage stabilization delays (**Supplementary Figure 12a**). Next, we

298   obtained 1,089 sparse PALM images using only 10 s of imaging time per FoV

299   (k=1,000 frames, Δt = 10 ms), in a total of only ~3.1 hours (**Figure 4a**). Neither the

300   widefield nor the sparse PALM images provided much small scale information

301   (**Figure 4c, Supplementary Figure 12 c,e**). However, ANNA-PALM reconstructions

302   led to high quality super-resolution images, allowing to visualize the microtubule

303   network with clarity and to distinguish microtubule filaments in dense areas that

304   appeared as unstructured regions in the sparse PALM image (**Figure 4b,d**). The

305   FWHM across filaments in the reconstructed image was ~51 nm (**Figure 4d**), within

306   the range measured for the training data (**Supplementary Figure 7a**). Similar

307   images can be obtained by ANNA-PALM using the widefield images alone

308   (**Supplementary Figure 12**), although of lower quality, as discussed above.

309   Stitching the reconstructed images together yielded a single super-resolution image

310   that contained approximately seven billion 20x20 nm pixels and covered an area of

311   1.8 mm x 1.8 mm, thereby spanning almost five orders of magnitude in spatial scales

312   (**Figure 4b,d, Supplementary Figure 12b** and **Supplementary Video 2**).

313

314   **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

15

338   one experimental condition, may be successfully applied to new experimental

339   conditions without retraining, thereby highlighting the method's robustness to

340   biologically relevant structural alterations.

341   We further asked if ANNA-PALM is robust to changes in technical imaging

342   conditions. To address this, we performed localization microscopy on microtubules

343   by simultaneously changing multiple important imaging parameters relative to the

344   training data. Instead of PALM/STORM, we used DNA-PAINT a technique where

345   single molecule detection relies on transient binding of fluorophore-conjugated

346   short DNA strands to complementary, antibody-conjugated, DNA strands[33], rather

347   than on fluorophore blinking. The continuously emitting freely diffusing dyes lead to

348   higher background noise in DNA-PAINT compared to PALM/STORM. Moreover, we

349   used primary mouse antibodies instead of rat antibodies, Cy3 dyes instead of Alexa-

350   647 dyes, and an EMCCD instead of a sCMOS camera. Despite all these differences,

351   when ANNA-PALM was applied without retraining on a sparse microtubule image

352   (k=400 frames, kΔt=12 s) , the reconstructed image still agreed very well with the

353   dense DNA-PAINT image obtained from K=60,000 frames ($K\Delta$t=30 min) (**Figure 5j-**

354   **l**). These data demonstrate the high robustness of ANNA-PALM to changes in

355   experimental imaging conditions.

356

357   **ANNA-PALM adapts to different biological structures**

358   To demonstrate that ANNA-PALM is not restricted to filamentary structure, we

359   turned to nuclear pores, a very different biological structure, and another popular

360   target of super-resolution imaging studies[34–36]. We retrained A-net simultaneously

361   on microtubule images and on a single PALM image of the nucleoporin gp210 in

362   immunolabeled nuclear membranes of *Xenopus* frog eggs[34,36] (K=30,000). With the

363   switch (**Figure 1**) set to microtubules ('MT'), this newly trained A-net can still

364   reconstruct sparse images of microtubules as when trained exclusively on

365   microtubule data (**Supplementary Figure 13a-c**). We then applied the same A-net

366   with the switch set to nuclear pores ('NPC') to reconstruct a new sparse PALM

367   image of gp210 obtained from the first k=3,000 frames (**Figure 6a**). The sparsity of

368   this image makes it difficult to clearly distinguish individual nuclear pores. ANNA-

369   PALM, however, reconstructs a much clearer image, containing many easily

370   identifiable ring-like structures, as expected for nuclear pores[34] (**Figure 6b**), and in

371   good agreement with the dense PALM image obtained from K=30,000 frames (even

372   though the latter shows mostly incomplete, open rings, presumably due to

373   suboptimal labeling) (**Figure 6c**). An automated procedure based on cross-

374   correlation with a ring template indeed identified ~2.7 times more putative nuclear

375   pores from the ANNA-PALM image than the sparse image (**Supplementary Figure**

376   **14a-c**). Moreover, computed pore locations were in good agreement with a PALM

377   image of wheat germ agglutinin (WGA), a lectin that concentrates in the inner

378   nuclear pore channel[36] (**Supplementary Figure 14 d-f**).   These results show that

379   ANNA-PALM can successfully analyze non-filamentary structures, when properly

380   retrained, and that a single ANN, with a simple computational switch, can

381   reconstruct very different types of structures.

382   Finally, we imaged TOM22, a protein of the mitochondrial outer membrane[37].

383   Whereas, at the resolution of our experiments, microtubules and nucleoporins are

384　essentially one-dimensional and zero-dimensional structures, mitochondrial

385　membranes are two-dimensional surfaces. Furthermore, their complex 3D

386　morphology might seem less predictable than filaments or nuclear pores, potentially

387　hampering ANNA-PALM reconstruction. Despite these differences, after being

388　trained on nine PALM images of TOM22 (with frame numbers ranging from

389　K=24,000 to K=40,000), ANNA-PALM reconstructions of distinct sparse PALM

390　images (k=400 frames) displayed mitochondrial morphologies in good agreement

391　with the dense PALM images (**Figure 6d-f**) - although the protein's localization

392　along the membrane was less well reproduced. Taken together, our results illustrate

393　the versatility of ANNA-PALM and its applicability to images of very different

394　structural content.

395

396　**Discussion**

397　We introduced ANNA-PALM, a computational method based on deep learning, that

398　reconstructs high quality super-resolution images from sparse, rapidly acquired,

399　single molecule localization data (and/or widefield images). Our method enables

400　considerable gains in acquisition time compared to standard localization

401　microscopy without increasing active fluorophore density, thereby preserving

402　spatial resolution. In fact, ANNA-PALM even improves spatial resolution when

403　applied to images of lower resolution than the training data

404　(**Supplementary Figures 7a, 10**), and greatly diminishes the detrimental effect of

405　drift. The improvement in imaging efficiency afforded by ANNA-PALM alleviates the

406　incompatibility between high resolution and high-throughput microscopy by

407 enabling super-resolution imaging of thousands of cells within a few hours or even

408 less (**Figure 4, Supplementary Figure 12**). This will facilitate super-resolution

409 studies of rare events, cellular heterogeneity and of partly stochastic structures such

410 as cytoskeletal polymers or chromosomes, whose characterization requires

411 statistics on many configurations[38,39]. ANNA-PALM may also be beneficial for high-

412 throughput imaging screens, e.g. of drug treatments or gene knock-outs[40–42]. In

413 addition, we envision applications to super-resolution imaging of large samples by

414 stitching together multiple images of spatially adjacent fields. The ability to generate

415 images spanning many orders of magnitude in scale could be well adapted to

416 expansion microscopy, a super-resolution technique that physically increases

417 sample size, but often requires tiling many fields of view to image even a single

418 cell[43,44]. With correlative microscopy[45], it might also be possible to train ANNA-

419 PALM to reconstruct electron microscopy (EM) images from fluorescence images,

420 potentially extending the method to molecular resolutions currently out of reach of

421 localization microscopy. Adaptation of ANNA-PALM to 3D[15,46] and multi-color[47,48]

422 localization microscopy should be relatively straightforward. Localization

423 microscopy of cellular dynamics remains very challenging[3,49]. By using much fewer

424 frames (or even only widefield images), ANNA-PALM could dramatically improve

425 the temporal resolution of live cell localization microscopy without sacrificing

426 spatial resolution or increasing phototoxicity and photobleaching. Thus, ANNA-

427 PALM provides multiple novel avenues for multi-scale imaging beyond standard

428 spatio-temporal resolution limits.

429   Neverthelesss, important caveats should be stressed. First, although ANNA-

430   PALM can be applied successfully to very different types of images (**Figure 6**), the

431   method fails in absence of statistical redundancies between molecular localizations,

432   e.g. for entirely random distributions of molecules. Second, ANNA-PALM requires

433   prior training on dense PALM images with structures similar to those in the images

434   to be reconstructed. We showed that ANNA-PALM is robust, i.e. does not require

435   retraining, for some experimentally induced changes in structures and variations in

436   imaging parameters (**Figure 5**). Nevertheless, indiscriminate application of ANNA-

437   PALM to very different structures without retraining, or incorrect setting of the

438   switch, may result in artifacts (**Supplementary Figure 13h**). Third, even when

439   applied to data similar to the training images, ANNA-PALM can produce errors -as

440   any reconstruction method in a context of information scarcity. The frequency of

441   errors can be reduced by increasing the number of recorded frames, at the cost of

442   reduced acceleration (see **Figures 2,3** and **Supplementary Figures 4, 8, 9**). In

443   addition, ANNA-PALM can use widefield images to estimate the reliability of

444   reconstructions, thereby helping their interpretation, providing some protection

445   against artifacts and indicating when retraining may be needed (**Supplementary**

446   **Figures 11, 15**). Future work, e.g. using Bayesian deep learning[50], may provide

447   additional assessments of reconstruction uncertainty and shed more light on the

448   scope and limitations of our approach.

449   As a purely computational technique, ANNA-PALM does not necessitate any

450   changes to existing microscopy systems, but only requires one or a few standard

451   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.

484   **Author contributions**

485   W.O.: conceived method, developed ANNA-PALM software and web application,

486   performed experiments and analyses. A.A., M.L., X.H.: performed experiments. C.Z.:

487   conceived method, supervised project and wrote manuscript.

488
489
490   **Competing financial interests statement**

491   W.O. and C.Z. are listed as inventors on European patent application EP17306022

492   filed by Institut Pasteur.

493
494

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).

541 21. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with
542   Conditional Adversarial Networks. *http://arxiv.org/abs/1611.07004* (2016).
543 22. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for
544   biomedical image segmentation. in *Medical Image Computing and Computer-*
545   *Assisted Intervention − MICCAI 2015. Lecture Notes in Computer Science,* **9351,**
546   234–241 (Springer International Publishing, 2015).
547 23. Goodfellow, I. *et al.* Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.*
548   *27* 2672–2680 (2014). doi:10.1017/CBO9781139058452
549 24. Ciresan, D., Giusti Alessandro, Luca, G. & Schmidhuber, J. Mitosis Detection in
550   Breast Cancer Histology Images with Deep Neural Networks. *Med. Image*
551   *Comput. Comput. Interv. − MICCAI 2013* **8150,** (2013).
552 25. Wang, Z. W. Z. & Bovik, A. C. A. C. Mean squared error: Love it or leave it? A
553   new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **26,** 98–117
554   (2009).
555 26. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss Functions for Image Restoration
556   with Neural Networks. *IEEE Trans. Comput. IMAGING* **3,** 47–57 (2017).
557 27. Culley, S. *et al.* NanoJ-SQUIRREL: quantitative mapping and minimisation of
558   super-resolution optical imaging artefacts. *bioRxiv* 158279 (2017).
559   doi:10.1101/158279
560 28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R.
561   Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach.*
562   *Learn. Res.* **15,** (2014).
563 29. Arbona, J.-M., Herbert, S., Fabre, E. & Zimmer, C. Inferring the physical
564   properties of yeast chromatin through Bayesian analysis of whole nucleus
565   simulations. *Genome Biol.* **18,** 81 (2017).
566 30. Arnal, I. & Wade, R. H. How does taxol stabilize microtubules? *Curr. Biol.* **5,**
567   900–908 (1995).
568 31. Wu, S. *et al.* Microtubule motors regulate ISOC activation necessary to
569   increase endothelial cell permeability. *J. Biol. Chem.* **282,** 34801–8 (2007).
570 32. De Brabander, M., De May, J., Joniau, M. & Geuens, G. Ultrastructural
571   immunocytochemical distribution of tubulin in cultured cells treated with
572   microtubule inhibitors. *Cell Biol. Int. Rep.* **1,** 177–83 (1977).
573 33. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R.
574   Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12,** 1198–1228
575   (2017).
576 34. Löschberger, A. *et al.* Super-resolution imaging visualizes the eightfold
577   symmetry of gp210 proteins around the nuclear pore complex and resolves
578   the central channel with nanometer resolution. *J. Cell Sci.* **125,** 570–5 (2012).
579 35. Szymborska, A. *et al.* Nuclear pore scaffold structure analyzed by super-
580   resolution microscopy and particle averaging. *Science* **341,** 655–8 (2013).
581 36. Sellés, J. *et al.* Nuclear pore complex plasticity during developmental process
582   as revealed by super-resolution microscopy. *Sci. Rep.* **7,** 14732 (2017).
583 37. Bellot, G. *et al.* TOM22, a core component of the mitochondria outer
584   membrane protein translocation pore, is a mitochondrial receptor for the
585   proapoptotic protein Bax. *Cell Death Differ.* **14,** 785–794 (2007).
586 38. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin

587   folding for different epigenetic states. *Nature* **529,** 418–422 (2016).

588 39. Zhang, Z., Nishimura, Y. & Kanchanawong, P. Extracting microtubule networks
589   from superresolution single-molecule localization microscopy data. *Mol. Biol.*
590   *Cell* **28,** 333–345 (2017).

591 40. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse
592   microscopy reveals cell division genes. *Nature* **464,** 721–727 (2010).

593 41. Beghin, A. *et al.* Localization-based super-resolution imaging meets high-
594   content screening. *Nat. Methods* **14,** 1184–1190 (2017).

595 42. Ouyang, W. & Zimmer, C. The imaging tsunami: Computational opportunities
596   and challenges. *Curr. Opin. Syst. Biol.* **4,** 105–113 (2017).

597 43. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science (80-.).*
598   **347,** 543–548 (2015).

599 44. Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* (2017).
600   doi:10.1038/nmeth.4261

601 45. de Boer, P., Hoogenboom, J. P. & Giepmans, B. N. G. Correlated light and
602   electron microscopy: ultrastructure lights up! *Nat. Methods* **12,** 503–513
603   (2015).

604 46. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-
605   resolution imaging by stochastic optical reconstruction microscopy. *Science*
606   *(80-.).* **319,** 810–813 (2008).

607 47. Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. Multicolor Super-Resolution
608   Imaging with Photo-Switchable Fluorescent Probes. *Science (80-.).*
609   1146598v1 (2007).

610 48. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with
611   DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11,** 313–318 (2014).

612 49. Shroff, H., Galbraith, C. G., Galbraith, J. A. & Betzig, E. Live-cell photoactivated
613   localization microscopy of nanoscale adhesion dynamics. *Nat Methods* **5,** 417–
614   423 (2008).

615 50. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep
616   Learning for Computer Vision? *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017)*
617   *pre-proceedings* 5580–5590 (2017).

618

FIGURE LEGENDS

620

621 **FIGURE 1: Overview of ANNA-PALM**

622 ANNA-PALM consists of two main stages: (**a**) acquisition of training images using

623 standard localization microscopy (PALM) followed by artificial neural network

624 (ANN) training, and (**b**) reconstruction of super-resolution views and low resolution

625 error maps from new sparse PALM and/or widefield images (inference).

626 (**a**) Training images are obtained by acquiring one or a few long sequences, of

627 $K\sim10^3$-10^5 diffraction limited, single molecule image frames, as in standard PALM

628 experiments; optionally, a widefield image W can also be acquired (top). The

629 acquisition time for each image sequence is $K\Delta t$, where Δt is the single frame

630 exposure time. Standard localization microscopy algorithms (grey 'localization'

631 boxes) are used to generate super-resolution images. For each experiment, a highly

632 sampled (dense) super-resolution image (PK) is generated using all (or in practice

633 ≥95%) available K frames. Sparse PALM images (Pk) from the same experiment are

634 obtained by using only $k<<K$ frames. A switch (light blue) can be set to distinguish

635 different types of structures, e.g. nuclear pore complexes ('NPC'), mitochondria

636 ('Mito') or microtubules ('MT'). An ANN (labeled G for 'generator', orange) is trained

637 by using the sparse PALM images Pk (plus an upsampled version L of the widefield

638 image W and the switch setting) as inputs and the corresponding dense PALM image

639 PK as target output. During training, the output of the generator G (Ak) is compared

640 to the target image PK and the widefield image L (if available) via three loss, or error

641 functions (gray bordered boxes): (i) the 'super-resolution reconstruction error'

642   measures the difference between the reconstructed image Ak and the target PK

643   using a combination of the L1 norm and the MS-SSIM; (ii) the 'low resolution

644   reconstruction error' measures the MS-SSIM between the low resolution image WA

645   predicted from the reconstruction Ak and the low resolution image WP predicted

646   from the target image PK. Images WA and WP are predicted using a second ANN,

647   called low-resolution estimator (labeled Q, blue) that is trained to produce an

648   approximation of the actual widefield image W based on the MS-SSIM metric; (iii)

649   the 'conditional GAN error' uses a third ANN (labeled D for 'discriminator', red) that

650   attempts to distinguish between real dense PALM images PK and the generator's

651   output Ak. The combined loss functions are iteratively optimized using stochastic

652   gradient descent. (**b**) A short sequence of diffraction limited single molecule images

653   (with $k<<K$ frames, i.e. acquisition time $k\Delta t$), and an optional widefield image (W'),

654   are acquired. Standard localization algorithms generate a sparse (under-sampled)

655   PALM image (P'k). This sparse image P'k (and the upscaled widefield image L' and

656   switch setting) are fed as inputs to the trained generator G, which outputs a

657   reconstructed ANNA-PALM image (A'k). In addition, the low resolution estimator Q

658   predicts a low resolution image WA', which can be compared to the input widefield

659   image W' via the MS-SSIM to produce a low resolution error map (top).

660

661   **FIGURE 2: Validation of ANNA-PALM on simulated images**

662   (**a**) Simulated widefield image of microtubules. (**b**) Simulated sparse PALM image of

663   microtubules with $n=$ 6,834 localizations. (**c**) ANNA-PALM reconstruction using only

664   the widefield image **a** as input. (**d**) ANNA-PALM reconstruction using both the

665 widefield image **a** and the sparse PALM **b** image as inputs. (**e**) Simulated "perfect"

666 PALM image, equivalent to a PALM image with an infinite number of localizations

667 ($n=\infty$) and a resolution of 23 nm. This image was used to generate **a** (by convolution

668 with a Gaussian kernel approximating the microscope point spread function) and **b**

669 (by application of Poisson noise). (**f**) Merged image showing the perfect PALM image

670 **e** in green and the ANNA-PALM reconstruction **d** in red. Note that the ANNA-PALM

671 images **c,d** provide many high resolution details that are absent from the widefield

672 image **a** and the sparse PALM image **b** and that are in good (**c**) or very good (**d**)

673 agreement with the perfect PALM image **e**. Some reconstruction errors are

674 highlighted by arrows. Blue arrows in panel **c** point to errors of ANNA-PALM

675 reconstruction from the widefield image only, the white arrow in panel **d** points to

676 an error of ANNA-PALM reconstruction from both widefield and sparse PALM

677 images combined. Reconstruction errors diminish for larger numbers of

678 localizations, n (**Supplementary Figure 4**). (**g**) Reconstruction quality of PALM and

679 ANNA-PALM images, measured by the MS-SSIM with the perfect PALM image **e**, as

680 function of localization number n. Black curve: reconstruction quality of the

681 standard PALM images. Dashed blue curve: reconstruction quality of ANNA-PALM

682 using the sparse PALM images as input. Solid blue curve: reconstruction quality of

683 ANNA-PALM using both the sparse PALM and widefield images as inputs. Red

684 dashed line: reconstruction quality of ANNA-PALM using the widefield images as

685 input only. Dots are averages from 10 simulations; error bars show standard

686 deviations. The vertical dashed orange line indicates the minimum number of

687 localizations needed to achieve a resolution of $R_{5\times\mathrm{Nyq}} = 23$ nm according to the five-

688 fold Nyquist criterion[3]. The dashed grey line indicates the minimum number of

689 localizations needed to achieve a double mean nearest neighbor distance less than

690 23 nm. ANNA-PALM reconstructions from sparse PALM images only (i.e. without

691 widefield images) achieve the same average MS-SSIM as standard PALM at the five-

692 fold Nyquist sampling limit with 26 times less localizations (blue double arrow).

693 ANNA-PALM reconstruction quality is highest when using both widefield and sparse

694 PALM images as inputs.

695

696 **FIGURE 3: ANNA-PALM imaging of microtubules**

697 ANNA-PALM reconstructions of a localization microscopy image of immunostained

698 microtubules. (**a**) Widefield image. (**b**) Sparse PALM image obtained from the first

699 9 s of acquisition (k=300 frames, n=11,740 localizations). (**c**) Dense PALM image

700 obtained from a 15 min long acquisition (K=30,000 frames, N=409,364

701 localizations). (**d**) ANNA-PALM reconstruction from the widefield image **a** only.

702 (**e**) ANNA-PALM reconstruction from the sparse PALM image **b** only. (**f**) ANNA-

703 PALM reconstruction from the widefield image **a** and sparse PLAM image **b**

704 combined. In panels **b-f**, pixel values are linearly mapped to colors from the look-up

705 table shown below. Black and white correspond to values V_{min} and V_{max},

706 respectively, with V_{min}=0 for all panels, V_{max}=3, 24, 102, 102 and 102 for panels **b**, **c**,

707 **d**, **e** and **f**, respectively. (**g-i**) Merged images comparing ANNA-PALM

708 reconstructions from panels **d-f** to the dense PALM image **c**. ANNA-PALM

709 reconstructions are shown in red, the dense PALM image in green. (**j**) Gradual

710 improvement of image quality for increasing acquisition time $k\Delta t$, shown for the

711   area highlighted in the insets of panels **a-i**. Top row: sparse PALM images. Middle

712   row: ANNA-PALM reconstructions from the sparse PALM images only (without

713   widefield). Bottom row: ANNA-PALM reconstructions from the widefield and sparse

714   PALM images combined. **Supplementary Video 1** shows the gradual increase in

715   quality of PALM and ANNA-PALM images with increased acquisition time for the

716   larger region of interest shown in panels **a-i.**

717

718   **FIGURE 4: High-throughput imaging with ANNA-PALM**

719   Application of ANNA-PALM to high-throughput imaging of a 1.8 mm x 1.8 mm area

720   containing more than 1,000 cells. (**a**) Sparse PALM image of this area, constructed

721   by assembling a mosaic of 33x33=1,089 sparse PALM images of individual fields of

722   view, obtained from k=1,000 raw frames each (with Δt=10 ms exposure time per

723   frame, i.e. in $k\Delta t$ =10 s). Total image acquisition time was 1,089x10s, i.e. ~3.1 hours.

724   The sparsity of the image is not apparent at this large scale. (**b**) ANNA-PALM

725   reconstruction of the image in **a**, obtained by assembling a mosaic of 1,089

726   individual reconstructions (one per field of view). (**c**) Magnified view of the green

727   boxed region in **a**. The inset shows a further magnified view of the yellow boxed

728   region, highlighting the sparsity of the image. (**d**) Same as **c**, but for the ANNA-PALM

729   reconstruction. A line profile across a microtubule is shown, with a FWHM≈51 nm.

730   Non-linear contrast adjustment was applied manually for panels **a** and **b,** with black

731   corresponding to values of zero in both panels. In panels **c** and **d**, pixel values were

732   linearly mapped to colors from the look-up table in **Figure 3;** Black and white

733   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

757   from the first k=3,000 frames. Note that individual nuclear pores are hard to

758   identify. (**b**) ANNA-PALM reconstruction of image **a**. (**c**) Dense PALM image obtained

759   from all K=30,000 frames. (**d**) Sparse PALM image of the immunolabeled

760   mitochondrial outer membrane protein TOM22, obtained from the first k=400

761   frames. (**e**) ANNA-PALM reconstruction of image **d**. (**f**) Dense PALM image obtained

762   from all K=30,000 frames. Pixel values are linearly mapped to colors from the look-

763   up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max},

764   respectively, with V_{min}=0 for all panels, and V_{max}=3, 51, 3, 3, 128, and 18 for panels **a**,

765   **b**, **c**, **d**, **e**, and **f**, respectively

766   **ONLINE METHODS**

767

768   **Artificial neural network**

769   Our ANN, called 'A-net', is based on the pix2pix architecture[21], which is a special

770   conditional generative adversarial network (cGAN)[23] for image to image

771   "translation", i.e. mapping from one type of image to another. The A-net consists of

772   three distinct neural networks: (i) a generator network G that produces the

773   reconstructed super-resolution image, (ii) a network Q called 'low resolution

774   estimator' that produces the low resolution error map, (iii) a cGAN discriminator

775   network D that provides the adversarial loss (**Figure 1a**). The generator network G

776   builds on the U-net architecture, and consists of an encoder-decoder network with

777   skip connections[22] and 16 convolutional layers. Its inputs and outputs are image

778   patches containing $(256m)$x$(256m)$ pixels, where m is an integer (we used m=1 or 2

779   but this can be adjusted for different sizes of CPU/GPU memory or input images).

780   The input is a sparse PALM image, a widefield image upscaled to the same size (see

781   below), and a computational switch number that allows the network to switch

782   between different types of image structures (e.g. nuclear pores or microtubules).

783   The switch setting is encoded numerically and coupled by convolutional operations

784   into the A-net encoder. The output of the generator G is a reconstructed image

785   (called ANNA-PALM reconstruction or ANNA-PALM image elsewhere) of the same

786   size as the input images. The low resolution estimator Q has four convolutional

787   layers. It takes the $(256m)$x$(256m)$ dense PALM image patch or the ANNA-PALM

788   image patch as input and outputs a low resolution image with $(64m)$x$(64m)$ pixels.

789   The cGAN discriminator network D has five convolutional layers. Its inputs are three

790    $(256m)$x$(256m)$ pixel image patches (the sparse PALM image, the widefield image,

791   and either the ANNA-PALM reconstruction or the corresponding dense PALM

792   image), plus the upscaled widefield image, and its output is a $(30m)$x$(30m)$ image

793   whose pixel values indicate whether the corresponding input patch is real (i.e. an

794   experimentally obtained conventional PALM image) or produced by the generator $\mathcal{G}$.

795   All convolutional layers are followed by batch normalization[51]. Dropout layers[28]

796   (with dropout probability p=0.5) are introduced in the central layers of the A-net

797   generator and turned on during training, but switched off during inference.

798   Activation functions are rectified linear units (ReLU) $x \rightarrow \sup(x, 0)$ or "leaky" ReLUs

799    $x \rightarrow \sup(x, 0) + \inf(\varepsilon x, 0)$ with ε =0.2[52], except for the last layer of $\mathcal{G}$, which uses the

800   hyperbolic tangent $x \rightarrow \tanh(x)$ and the last layer of Q, which uses a sigmoid

801   function $x \rightarrow (1 + \exp(-x))^{-1}$. The A-net architecture is fully described in

802   **Supplementary Note 1** and **Supplementary Tables 1-4**.

803

804   **Training objectives and error map**

805   Each of the three networks mentioned above ($\mathcal{G}, Q, \text{and } D$) is associated to a distinct

806   objective function - also called loss- and hereafter noted $\mathscr{L}_{\mathcal{G}}$, $\mathscr{L}_{Q}$, and $\mathscr{L}_{D}$ respectively.

807   These loss functions are specified in detail below. In the following equations, for

808   notational simplicity, we designate the sparse input image as $\boldsymbol{S}$, the low resolution

809   (widefield) input image as $\boldsymbol{W}$, the corresponding dense PALM image (i.e. the target)

810   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

811   labeled Pk, PK and Ak, respectively); the computational switch indicating the image

812   type is noted $\boldsymbol{M}$. Low resolution images produced by the low resolution estimator

813   network Q from $\boldsymbol{A}$ and $\boldsymbol{T}$ are designated as $\boldsymbol{W}_A = Q(\boldsymbol{A})$ and $\boldsymbol{W}_T = Q(\boldsymbol{T})$,

814   respectively.

815   The generator loss function $\mathscr{L}_G$ is the sum of three terms. The first term of $\mathscr{L}_G$

816   is the super-resolution reconstruction error, hereafter called $\mathscr{L}_{\text{SuperRes}}(\mathcal{G})$. This term

817   penalizes the difference between the generator output $\boldsymbol{A}$ and the target image $\boldsymbol{T}$.

818   Based on a recent analysis of loss functions for image restoration with neural

819   networks[26], we defined this difference as a weighted average of two quantities:

820   (i) the multi-scale structural similarity index (MS-SSIM) between $\boldsymbol{A}$ and $\boldsymbol{T}$ and (ii) a

821   modification of the L1 norm, where the absolute difference between $\boldsymbol{A}$ and $\boldsymbol{T}$ is

822   smoothed by a Gaussian kernel:

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

823   where $\mathbb{E}$ denotes expectation; $p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})$ is the joint probability density of the

824   sparse PALM images $\boldsymbol{S}$, dense PALM images $\boldsymbol{T}$, widefield images $\boldsymbol{W}$ and switch

825   settings $\boldsymbol{M}$ from the training data set; $MS\_SSIM(\boldsymbol{A},\boldsymbol{T})$ is the multi-scale structural

826   similarity index between $\boldsymbol{A}$ and $\boldsymbol{T}$; G_σ is a Gaussian smoothing kernel; $*$ denotes

827   convolution; $|\boldsymbol{A} - \boldsymbol{T}|$ is the absolute difference image (i.e. pixel (i,j) has value

828    $|\boldsymbol{A}(\text{i},\text{j}) - \boldsymbol{T}(\text{i},\text{j})|$ and $\rho \in [0,1]$ is a scalar weight that balances the relative

829   contributions of MS-SSIM and the modified L1 norm and is set to $\rho = 0.84$ as in ref.

830   [26].

831   The second term of $\mathscr{L}_{\mathcal{G}}$ is called $\mathscr{L}_{\text{LowRes}}(\mathcal{G}, Q)$ and measures the consistency

832   between the low resolution images $\boldsymbol{W}_A$ and $\boldsymbol{W}_T$ predicted by the low resolution

833   estimator network Q:

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

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

834   Alternatively, in the above objective function, $\boldsymbol{W}_T$ can be replaced by the actually

835   observed widefield image $\boldsymbol{W}$, although with our data this led to slightly lower

836   reconstruction quality. The low resolution estimator network Q is trained

837   simultaneously with the generator $\mathcal{G}$ to produce a low resolution image from the

838   dense PALM image $\boldsymbol{T}$ that is consistent with the observed low resolution image $\boldsymbol{W}$.

839   This training is done based on the following objective function:

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

840   Note that the reconstructed low resolution image $Q(\boldsymbol{T})$ is four times smaller than

841   the dense PALM image $\boldsymbol{T}$, as described in **Supplementary Note 1**. Because the input

842   widefield image $\boldsymbol{W}$ can have a different size, we use bilinear interpolation to resize

843    $\boldsymbol{W}$ to the same size as $Q(\boldsymbol{T})$. (If needed, a scaling factor different from four can be

844   obtained by adding or removing downsample layers in network Q). At inference, the

845   low resolution estimator Q is also used to produce the error map, as shown in

846   **Figure 1b**, and **Supplementary Figures 6, 11, 15**. This error map is defined as:

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

847   High (respectively low) values of the error map indicate large (respectively small)

848   inconsistencies, between the reconstructed super-resolution image $\boldsymbol{A}$ and the

849   observed widefield image $\boldsymbol{W}$.

850   The third term of $\mathcal{L}_G$ draws from recent work on generative adversarial

851   networks (GAN)[21,23,53] and is noted $\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D})$. In a GAN, a generator network $\mathcal{G}$

852   learns to transform random input vectors $\boldsymbol{z}$ (drawn from a probability density

853    $p_{\mathrm{z}}(\boldsymbol{z})$) into new samples of a data probability density $p_{\mathrm{data}}(\boldsymbol{x})$. In our case, the data

854   samples $\boldsymbol{x}$ are the dense PALM images $\boldsymbol{T}$. The generator $\mathcal{G}$ learns by working against

855   a discriminator network $\mathcal{D}$ that simultaneously learns to discriminate between

856   original data samples and samples generated by $\mathcal{G}$. Adversarial training thus

857   consists in playing a minmax game such that $(\mathcal{G}^*, \mathcal{D}^*) = \arg\min_{\mathcal{G}} \max_{\mathcal{D}} \mathcal{L}_{\mathrm{GAN}}(\mathcal{G},\mathcal{D})$,

858   with an objective function of the form[23]: $\mathcal{L}_{\mathrm{GAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{x})] +$

859    $\mathbb{E}_{\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log[1 - \mathcal{D}(\mathcal{G}(\boldsymbol{z}))]$, or equivalently by simultaneous optimization of two

860   coupled loss functions:

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

861   In a conditional GAN (cGAN), the generator and the discriminator have an extra

862   input vector $\boldsymbol{c}$ and the first objective function above becomes: $\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D}) =$

863    $\mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\mathrm{data}}(\boldsymbol{c},\boldsymbol{x})}[\log\mathcal{D}(\boldsymbol{c},\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log[1 - \mathcal{D}(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z}))]$, such that the

864   generator learns a conditional probability density $p_{\mathrm{data}}(\boldsymbol{x}|\boldsymbol{c})$; and the second

865   objective function likewise becomes $\mathcal{L}_{\mathrm{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{c}\sim p_{\mathrm{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_{\mathrm{z}}(\boldsymbol{z})}\log[1 -$

866    $\mathcal{D}(\mathcal{G}(\boldsymbol{z}))]$. In our A-net, we replaced the logarithmic losses above by least square

867   losses[53], as they empirically yielded better results. Thus, we used the objective

868   functions:

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

870    $\mathcal{L}_{cGAN}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{c\sim p_{data}(c),z\sim p_z(z)}\big[1 - \mathcal{D}\big(c,\mathcal{G}(c,z)\big)\big]^2$. In our case, the input c is the

871   sparse PALM image S combined with the upsampled version $L = \mathcal{B}(W)$ of the

872   widefield image W, where $\mathcal{B}$ denotes bilinear interpolation. Note that in practice the

873   noise z in our ANN was introduced only through the use of dropout layers, as in the

874   pix2pix implementation[21]. Thus, the objective functions are:

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

$$+ \mathbb{E}_{z\sim p_z(z),S\sim p_{data}(S)}\Big(\mathcal{D}\big(\mathcal{B}(W),S,\mathcal{G}(S,W,M)\big)\Big)^2$$

875   and:

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

876   In the end, combining the three loss terms described above, we implemented

877   the following optimization problem :

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

878   The weights α, β and γ are hyperparameters, which we set manually to $\alpha = 50$,

879    $\beta = 25$ and $\gamma = 1$ for most experiments. In absence of widefield images W, β was

880   simply set to zero. The reported results are not very sensitive to these parameters.

881   We trained the A-net end-to-end using stochastic gradient descent (SGD)

882   with Adam[54] and a batch size of 1 with 200,000 or more iterations (backpropagation

883   steps). Our implementation was adapted from affinelayer's TensorFlow[55]

884   implementation, which is ported from the Torch implementation of pix2pix[21]. Both

885   network training and inference were performed on Tesla P100, Tesla M40, Tesla

886   K80 or GTX TitanXP graphical processing units (GPUs) from Nvidia. A-net training

887   from scratch typically takes from hours to days on a single GPU. Once trained, the A-

888   net takes only ~1 second or less to reconstruct a super-resolution image of

889   2560x2560 pixels (corresponding to an entire FoV). Training time could be further

890   reduced by pretraining (or transfer learning), use of GPU clusters, or optimized data

891   augmentation.

892

893   **Experimental training images and data augmentation**

894   Experimental training data are obtained from standard localization microscopy data

895   (dense PALM images). To achieve good performance, ANNs generally necessitate

896   large amounts of training data. However, ANNA-PALM typically requires PALM

897   images from no more than 10 FoVs (of 55 μm x 55 μm each) and can even be trained

898   with a single FoV. This is possible thanks to an extensive on-the-fly data

899   augmentation strategy, as described below. Each of the dense PALM training images

900   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

901   the diffraction limited frame from which localization x_i, y_i originates, and K is the

902   total frame number. PALM images are obtained as plain 2D histograms of these

903   localizations with typical pixel sizes of 10-20 nm. From each list of localizations

904   (corresponding to a dense PALM image of a single FoV), we generate 10-30 pairs of

905   input and target images (***S***,***T***) for training. To define the target image ***T***, we take a

906   random consecutive subset $[k_T, k_T + 0.95K]$ of 95% of all available K frames (k_T is

907   chosen randomly between 0 and $0.05K$) and create the 2D histogram image based

908   on localizations from those frames only, i.e. from all (x_j, y_j) such that $f_j \in$

909 $[k_T, k_T + 0.95K]$. To define the sparse input image $\boldsymbol{S}$, we take random subsets of

910 300-500 consecutive frames from the first half of the image sequence and similarly

911 create a 2D histogram of the localizations from those frames only. When a widefield

912 image $\boldsymbol{W}$ is available, this image must first be aligned with the corresponding dense

913 PALM image $\boldsymbol{T}$. This is done using an FFT-based phase correlation algorithm[56] after

914 histogram equalization of image $\boldsymbol{T}$, smoothing by convolution with a Gaussian

915 Kernel of standard deviation 6 pixels, and resizing with bilinear interpolation to the

916 same size as image $\boldsymbol{W}$. The registered widefield image $\boldsymbol{W}$ is then scaled up using

917 bilinear interpolation to an image $\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ with the same size as image $\boldsymbol{T}$.

918   During training, for each iteration of SGD, we crop the images $\boldsymbol{S}$, $\boldsymbol{L}$ and $\boldsymbol{T}$ with

919 a randomly placed 712x712 pixel sized region $\mathcal{R} = [x_{\min}, x_{\min} + 712] \times$

920 $[y_{\min}, y_{\min} + 712]$. We then use random geometric transformations and apply them

921 identically to the three images. Specifically, we rotate the images by a random angle

922 between 0 and 360 degrees, apply elastic transformations [57], and then crop the

923 center region of size 512x512 pixels. In addition to geometric transformations, we

924 also introduce realistic noise from experimental background images. This is done by

925 manually outlining regions of background in selected PALM training images,

926 splitting these regions into small patches of 40x40 pixels, grouping them according

927 to their summed pixel values, then assembling them into a larger image with the

928 same size as $\boldsymbol{T}$. During training, these semi-synthetic noise images are randomly

929 selected and added to the input image $\boldsymbol{S}$ without altering $\boldsymbol{T}$ or $\boldsymbol{W}$. Finally, we

930 normalize the input image $\boldsymbol{S}$ by subtracting its mean and dividing by standard

931 deviation. If a widefield image $\boldsymbol{W}$ is provided, its pixel values are scaled to a

932 minimum of 0 and a maximum of 1. Otherwise, W is replaced by an image containing

933 zeros only. The target image T is truncated at a maximum value of 255 and then

934 scaled to have a minimum of 0 and a maximum of 1. For the switch M, we used an

935 integer number to define the type of training images, e.g. 0 for microtubules, 1 for

936 nucleoporins, 2 for mitochondria. When training on different types of images, e.g.

937 microtubules and nucleoporins (**Supplementary Figure 13**), we assign the

938 corresponding switch value to M and use it as additional input to the A-net together

939 with images S and W, as described in **Supplementary Note 1.**

940

941 **Image simulations**

942 Our procedure to simulate localization microscopy (PALM) images of microtubules

943 is illustrated in **Supplementary Figure 1**. To simulate microtubule filaments, we

944 used a Langevin dynamics simulation[29] that generates random configurations of

945 semiflexible curves with a specified rigidity (persistence length), starting from a

946 random initial configuration (**Supplementary Figure 1a,b**). The initial

947 configurations were generated with a Python library named cpolymer and the

948 Langevin dynamics was implemented using the molecular dynamics code

949 LAMMPS[58]. Although the simulation generates 3D polymer chains, we only

950 considered their 2D projections, consisting of N_p connected positions $(x_k, y_k)_{k=1..N_p}$.

951 To obtain smooth filaments we further interpolated these connected segments using

952 spline functions with the Scipy function scipy.interpolate.splev. Next, we turned

953 these 2D curves into a grey scale image of 800x800 pixels, with an assumed pixel

954 size of 7.8 nm, using the Python library Matplotlib. This image was further

955   convolved with a Gaussian kernel of standard deviation 1.5 pixels, resulting in a

956   smooth image $I(i,j)$ as shown in **Supplementary Figure 1c,** and normalized to a

957   probability density ($\sum \sum_{i,j} I(i,j) = 1$, with all $I(i,j) \geq 0$). This image was used to

958   mimic a "perfect" PALM image of filaments corresponding to an infinite number of

959   localizations ($n = \infty$). Such perfect images were used as targets during ANN

960   training for simulated data and defined as ground truth for the quantification of

961   reconstruction quality by MS-SSIM (**Figure 2g**). During training, we applied the

962   same rotations and elastic transformations described for experimental data in the

963   previous section.

964   Localization microscopy images obtained from a finite number of

965   localizations $n < \infty$ (sparse PALM images **S**), can be considered as a sampling of the

966   probability density $I(i,j)$ with n samples. These images can therefore be simulated

967   by applying Poisson noise to a rescaled version of the perfect PALM image, i.e.:

968    $\boldsymbol{S}(\lambda, I) = \mathcal{P}(\lambda I/I_{\max})$, where $I_{\max}$ is the maximum value of I, $\mathcal{P}(\mu)$ denotes the

969   Poisson probability distribution of mean μ and where the peak parameter λ controls

970   the level of sampling. In order to simulate sparse PALM images for various levels of

971   sampling, we varied the peak value λ following a log-normal distribution where

972    $\ln(\lambda)$ has mean -0.5 and standard deviation 0.001 and applied Poisson noise using

973   the numpy library function random.poisson. An example of a simulated sparse

974   PALM image is shown in **Supplementary Figure 1d**. Besides finite sampling,

975   localization microscopy images are corrupted by additional noise sources such as

976   false detections from background noise due to out-of-focus light or unspecific

977   binding of antibodies. To mimic this, we first created a probability density

42

978    $I_b = I * G_{\sigma_b}$ for the background noise by convolving I with a Gaussian kernel G_{σ_b} of

979   large standard deviation σ_b=25 pixels, and applied Poisson noise with $\lambda = 0.06$. To

980   create training images, we added this background noise image to the sparse PALM

981   image S above (**Supplementary Figure 2**). We did not add background noise to the

982   test images used during inference (**Figure 2**). To simulate the widefield images

983   (**Figure 2a**), we first blurred the perfect PALM image by convolution with a

984   Gaussian kernel of standard deviation 8 pixels, then added Gaussian noise with zero

985   mean and standard deviation chosen randomly between 0.5 and 1.5.

986   For simulations of nuclear pore images (**Supplementary Figure 15**), we

987   applied a similar procedure, except that the perfect PALM images were obtained by

988   randomly distributing circles of diameter 150 nm in the plane (avoiding overlaps)

989   and placing eight Gaussian spots (of standard deviation 1.7 pixels) at equal distance

990   from each other on each circle to mimic the octogonal shape of nuclear pores.

991

992   **Sample preparation**

993   For microtubule imaging experiments (**Figures 3-5** and **Supplementary Figures 8-**

994   **10**), except those using DNA-PAINT (**Figure 5j,l**), U-373 MG (Uppsala) cells were

995   cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12;

996   Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v)

997   penicillin-streptomycin (Gibco), in a 5% CO_2 environment at 37°C on 18-mm

998   cleaned coverslips in 12-well plates. 24 hours after plating, cells were pre-extracted

999   for 10 s in 0.25% (v/v) Triton X-100 (Triton) in BRB80 (80 mM PIPES, 1 mM MgCl2,

1000   1 mM EGTA, adjusted to pH 6.8 with KOH) supplemented with 4 mM EGTA, and

1001 immediately fixed for 10 min with 0.25% (v/v) Triton + 0.5% Glutaraldehyde in

1002 BRB80, followed by reduction for 7 min with 0.1% NaBH4 solution in PBS and

1003 another washing step in PBS. Cells were directly incubated for 1h at room

1004 temperature in PBS with 1:500 rat alpha-tubulin antibodies (Bio-Rad MCA77G),

1005 followed by 3 washing steps with PBS, and then incubated for 45min in PBS with

1006 1:500 anti-rat Alexa-647 conjugated secondary antibodies from donkey (Jackson

1007 ImmunoResearch Laboratories, ref. 712-605-153).

1008   For the DNA-PAINT experiment on microtubules (**Figure 5k-m**), U-373 cells

1009 stuck on 18 mm diameter coverslips were fixed at 37°C with 4% PFA in PHEM

1010 buffer and permeabilized in 0.2% glutaldehyde. Next, cells were incubated for 1 h

1011 with 1:500 primary mouse antibodies against alpha-tubulin. The sample was

1012 washed 3 times in PBS, then incubated with 1:100 anti-mouse oligo-conjugated

1013 antibodies from Ultivue Kit 2 for DNA-PAINT imaging[33]. After washing the sample 3

1014 times in PBS, and just before imaging, 2nM of complementary oligos coupled to Cy3

1015 fluorophores were added to the sample.

1016   Nuclear pore imaging data of gp210 and WGA (**Figure 6a-c** and

1017 **Supplementary Figure 14**) were kindly provided by J. Sellés and O. Falklaris and

1018 obtained from nuclear membranes of *Xenopus* frog eggs prepared as described

1019 previously[36]*.*

1020   For mitochondria imaging experiments (**Figure 6e-f**), COS7 cells were

1021 cultured under the same conditions as U-373 cells above using phenol-red free

1022 DMEM medium and fixed with 4% PFA in PBS for 10 min. The sample was blocked

1023 with 3% BSA in PBS for 20 min and immunostained with 1:500 mouse antibodies

1024 against TOM22 (Sigma, ref. T6319) in wash buffer (PBS with 0.5% BSA) for 1 h.

1025 After extensive washing with wash buffer, the sample was incubated with 1:500

1026 anti-mouse secondary antibodies from donkey conjugated to Alexa-647 dyes

1027 (Jackson ImmunoResearch Laboratories, ref. 715-605-151) in wash buffer for 30

1028 min. After washing 5 times with wash buffer and 2 times with PBS, samples were

1029 post-fixed with 2% PFA in PBS for 10 min and washed 5 times with PBS.

1030   For all localization microscopy experiments except DNA-PAINT, we used a

1031 photoswitching buffer[59] composed of 50 mM Tris-HCl + 10 mM NaCl + 10% (w/v)

1032 glucose + 168 AU/mL Glucose-Oxidase + 1404 AU/mL Catalase + 1% 2-

1033 Mercaptoethanol. For microtubule imaging experiments, we used this buffer to fill a

1034 square hole that was manually cut in a parafilm sheet, which was deposited on a

1035 rectangular coverglass. The round coverslips were sealed with nail polish.

1036

1037 **Image acquisition in localization and high-throughput microscopy**

1038 We performed single molecule localization microscopy experiments (PALM/STORM

1039 and DNA-PAINT) on custom built microscopy systems, as previously described[59–61].

1040 The system used for PALM/STORM imaging of microtubules is based on an inverted

1041 microscope body (Nikon Ti Eclipse) equipped with a either a 60x 1.49 NA oil

1042 immersion objective (Nikon) or a 60x 1.2 NA water immersion objective (Nikon)

1043 and with the Perfect Focus System active. A 642 nm wavelength laser with 500 mW

1044 power was used to excite Alexa-647 fluorophores and an AOTF (AA optics) was used

1045 to modulate laser excitation. Sequences of diffraction limited single molecule image

1046 frames were acquired either on a sCMOS camera (Hamamatsu ORCA-Flash4.0),

1047   which can capture images of 2,042x2,042 pixels (for **Figures 4, 5a-i**), or on an

1048   EMCCD (Andor IXON ULTRA 897) with 512x512 pixels (for **Figures 3, 6d-f**). Both

1049   cameras were controlled by MicroManager software[62]. For experiments using the

1050   sCMOS camera, the effective pixel size was 108 nm and we used a 512x512 region of

1051   interest, which resulted in an imaged FoV of 55.3 µm x 55.3 µm. For experiments

1052   using the EMCCD camera, we used a 2x telescope and the effective pixel size was

1053   107 nm, resulting in a FoV of 54.8 µm x 54.8 µm. The exposure time was set to Δt

1054   =10 ms or 30 ms per frame. The number of frames acquired ranged from k=1,000

1055   (**Figure 4**) to K=60,000 (e.g. **Figure 5c**) per FoV.

1056   For the DNA-PAINT experiment (**Figure 5j,l**), we used an inverted Nikon Ti-E

1057   Eclipse microscope equipped with a 100x 1.49 NA TIRF objective and with the

1058   Perfect Focus System active. A 561 nm wavelength laser with 500 mW power was

1059   used to excite Cy3 dyes. Highly inclined laser illumination was used to reduce out-

1060   of-focus background signal. Images were acquired on an EMCCD camera as above,

1061   with a 1.5 x telescope, resulting in an effective pixel size of 106 nm and a FoV of

1062   54 µm x 54 µm. The sample was mounted in a magnetic sample holder filled with

1063   the imaging buffer provided with the Ultivue kit. Exposure time was set to Δt =30 ms

1064   and the EM gain of the EMCCD was set to 300. The laser power was increased until

1065   isolated fluorescent spots were observed. For the experiment shown in **Figure 5g,i**,

1066 K=60,000 frames were acquired.

1067   The *Xenopus* nuclear pore data (**Figure 6a-c** and **Supplementary**

1068   **Figure 14**), were acquired on a Zeiss Elyra P.S.1 microscope as described

1069   previously[36].

1070   For high-throughput imaging of microtubules (**Figure 4**), we used the Multi-

1071 Dimensional Acquisition tool in Micro-manager to define the positions of 1,089 FoVs

1072 of 55.3 μm x 55.3 μm on a 33x33 grid, with overlaps of 1 μm; the stage was

1073 automatically shifted to each of these 1,089 positions. We first acquired only

1074 widefield images, taking five frames at each of these positions (the first two were

1075 ignored because of motion blur), in a total acquisition time of 12 minutes. Then, the

1076 laser power was raised to bleach out preactivated molecules and k=1,000 frames of

1077 single molecule images were acquired at each of the 1,089 positions, in a total

1078 acquisition time of 3 hours and 8 minutes. Raw image frames were written directly

1079 to a remote storage server via Samba networking protocol.

1080

1081 **Localization microscopy image analysis**

1082 The input to ANNA-PALM reconstruction is a localization image, defined as a 2D

1083 histogram of n single molecule positions $(x_i, y_i)_{i=1..n}$. The histogram bin, i.e. the

1084 pixel size of the localization image, was set to 7.8 nm for the simulated data

1085 (**Figure 2** and **Supplementary Figures 1-6,15**) and 20 nm for the experimental

1086 data (**Figures 3-6** and **Supplementary Figures 7-14**). The positions $(x_i, y_i)_{i=1..n}$

1087 were obtained by analyzing sequences of diffraction limited frames using standard

1088 single molecule localization algorithms. For experimental microtubule images, we

1089 used the ThunderSTORM[63] plugin of ImageJ, applying wavelet filters for detection

1090 and weighted least squares Gaussian fitting for precise estimation of subpixelic

1091 positions. We used the cross-correlation feature in ThunderSTORM for drift

1092 correction, and filtered out the least certain localizations based on the fitted

1093 Gaussian's standard deviation and the χ^2 of the residual. Localizations in

1094 consecutive frames separated by less than 20 nm were assumed to originate from

1095 the same molecule and merged into a single localization. The final number n of

1096 localizations was ~7 million for the full 55 µm x 55 µm FoV of the images shown in

1097 **Figure 3** and **Supplementary Figures 8,9** (obtained from K=30,000 frames). For

1098 the high-throughput experiment (**Figure 4** and **Supplementary Figure 12**), the

1099 number of localizations per 55 µm x 55 µm FoV ranged from n=2,949 to

1100 n=1,442,048 with an average $\langle n \rangle = 610{,}983$ and standard deviation $\sigma(n)$ =273,606.

1101 The total number of localizations across all 1,089 FoVs was ≈ 665 million.

1102 ThunderSTORM analyses were performed either on high end workstations or on

1103 Institut Pasteur's high performance computer (HPC) cluster. For the high-

1104 throughput experiments, we used Python scripts to run ThunderSTORM in batch

1105 mode (without user intervention) on the HPC cluster and assembled mosaic images

1106 (**Figure 4a,b** and **Supplementary Figure 12a,b**) using a stitching plugin of

1107 ImageJ[64].

1108 The nuclear pore images were analyzed using the ZEN software from Zeiss as

1109 previously described[36]. For the DNA-PAINT experiments, we used PALMTT, a

1110 modified version of the single molecule tracking algorithm MTT[65], based on Matlab

1111 (Mathworks). This algorithm uses Gaussian smoothing and thresholding for

1112 detection, and Gaussian fitting for precise estimation of subpixelic positions. Drift

1113 correction was performed computationally by tracking fluorescent beads used as

1114 fiducial markers.

1115

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 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]

1139   prescribes a five-fold larger number of localizations to reach resolution R:

1140 $n \geq N_{5 \times \text{Nyq}} = 5 \times N_{\text{Nyq}}$, implying that the sampling limit to resolution is

1141 $R_{5 \times \text{Nyq}} = 2 \langle d(n/5) \rangle$. Accordingly, if this condition is met, the resolution is no longer

1142   limited by sampling, but by the localization precision, $R_{\text{loc}} \approx 2.3 \sigma_{\text{loc}}$, where σ_{loc} is

1143   the standard deviation of localization errors along each coordinate. In general, the

1144   resolution of a PALM image, as limited by both sampling and localization precision,

1145   can be written: $R = \max(R_{5 \times \text{Nyq}}, R_{\text{loc}}) = \max(2 \langle d(n/5) \rangle, 2.3 \sigma_{\text{loc}})$.

1146

1147   **Life Sciences Reporting Summary.**

1148   Further information is available in the Life Sciences Reporting Summary.

1149

1150   **Code availability statement**

1151   The source code of ANNA-PALM is available from https://annapalm.pasteur.fr/ .

1152

1153   **Data availability statement**

1154   The localization data used in this paper can be downloaded directly from

1155   https://annapalm.pasteur.fr/.

1156

1157   **Methods-only references**

1158   51.   Xu, L., Choy, C. S. & Li, Y. W. Deep sparse rectifier neural networks for speech
1159   denoising. in *2016 International Workshop on Acoustic Signal Enhancement,*
1160   *IWAENC 2016* (2016). doi:10.1109/IWAENC.2016.7602891
1161   52.   Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier Nonlinearities Improve Neural
1162   Network Acoustic Models. *Proc. 30 th Int. Conf. Mach. Learn.* (2013).
1163   53.   Mao, X. *et al.* Least Squares Generative Adversarial Networks. in *ICCV* 2794–
1164   2802 (2017). doi:10.1109/ICCV.2017.304

1165 54. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ICLR* 1–15
1166 (2015).
1167 doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503
1168 55. Rampasek, L. & Goldenberg, A. TensorFlow: Biology's Gateway to Deep
1169 Learning? *Cell Syst.* **2,** 12–14 (2016).
1170 56. Srinivasa Reddy, B. & Chatterji, B. N. An FFT-based technique for translation,
1171 rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5,**
1172 1266–1271 (1996).
1173 57. Simard, P. Y., Steinkraus, D. & Platt, J. C. Best practices for convolutional neural
1174 networks applied to visual document analysis. *Seventh Int. Conf. Doc. Anal.*
1175 *Recognition, 2003. Proceedings.* **1,** 958–963 (2003).
1176 58. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J.*
1177 *Comput. Phys.* **117,** 1–19 (1995).
1178 59. van de Linde, S. *et al.* Direct stochastic optical reconstruction microscopy with
1179 standard fluorescent probes. *Nat. Protoc.* **6,** 991–1009 (2011).
1180 60. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy
1181 image processing in ImageJ. *Nat. Methods* **7,** 339–340 (2010).
1182 61. Lelek, M. *et al.* Superresolution imaging of HIV in infected cells with FlAsH-
1183 PALM. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8564–9 (2012).
1184 62. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control
1185 of microscopes using manager. *Current Protocols in Molecular Biology* (2010).
1186 doi:10.1002/0471142727.mb1420s92
1187 63. Ovesny, M., Kiek, P., Borkovec, J., Vindrych, Z. & Hagen, G. M. ThunderSTORM:
1188 a comprehensive ImageJ plug-in for PALM and STORM data analysis and
1189 super-resolution imaging. *Bioinformatics* **30,** 2389–2390 (2014).
1190 64. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D
1191 microscopic image acquisitions. *Bioinformatics* **25,** 1463–1465 (2009).
1192 65. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target
1193 tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods*
1194 **5,** 687–694 (2008).
1195
1196
1197
1198
1199
1200

1   **ANNA-PALM: Deep learning accelerates super-resolution localization**

2   **microscopy by orders of magnitude**

3

4

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

6   Zimmer[1,2,3]

7

8   [1] Institut Pasteur, Unité Imagerie et Modélisation, Paris, France

9   [2] UMR 3691, CNRS

10   [3] C3BI, USR 3756, IP CNRS

11

12

13

14

15

16

17

18

19

20

21

22   Corresponding author: C. Zimmer

23 The speed of super-resolution microscopy methods based on single molecule

24 localization, e.g. PALM or STORM, is severely limited by the need to record

25 many thousands of frames with a low number of observed molecules in each.

26 Here, we present ANNA-PALM, a computational strategy that uses artificial

27 neural networks to reconstruct super-resolution views from sparse, rapidly

28 acquired localization images and/or widefield images. Simulations and

29 experimental imaging of microtubules, nuclear pores and mitochondria show

30 that high-quality super-resolution images can be reconstructed from up to two

31 orders of magnitude fewer frames than usually needed, without

32 compromising spatial resolution. Super-resolution reconstructions are even

33 possible from widefield images alone, though adding localization data

34 improves image quality. We demonstrate super-resolution imaging of >1,000

35 fields of view containing >1,000 cells in ~3 h, yielding an image spanning

36 spatial scales from ~20 nm to ~2 mm. The drastic reduction in acquisition

37 time and sample irradiation afforded by ANNA-PALM enables faster and

38 gentler high-throughput and live cell super-resolution imaging.

39
40 Fluorescence microscopy methods that overcome the diffraction limit of resolution

41 (~200-300 nm) allow imaging of biological structures with molecular specificity

42 closer to the molecular scale. Among super-resolution microscopy approaches,

43 those based on single molecule localization, such as PALM[1] or STORM[2] (hereafter

44 referred to collectively as PALM) are particularly attractive owing to their exquisite

45 spatial resolution and ease of implementation. In these methods, random subsets of

46 fluorophores are imaged in many consecutive diffraction-limited frames,

47 computationally localized to high precision, and the combined localizations are used

48 to generate a super-resolution view. In practice, typically 10^3-10^5 diffraction-limited

49 frames are needed to assemble a single super-resolution image. This requirement

50 follows from two conditions that must be simultaneously satisfied to ensure high

51 spatial resolution: (i) a low number (~10-10^2) of active fluorophores per frame, to

52 avoid overlaps between diffraction limited spots and enable precise localization of

53 individual molecules, and (ii) a large number of independent localizations to ensure

54 a sufficiently dense sampling of the underlying biological structures[3,4]. The large

55 number of required frames makes localization microscopy inherently slow, thereby

56 limiting its potential for high-throughput imaging, where many fields of view (FoVs)

57 are to be imaged, and for imaging live cell dynamics. As a result, most localization

58 microscopy studies are restricted to analyzing a small number of cells (typically less

59 than ten).

60 Multiple approaches have been explored to accelerate localization

61 microscopy. Using bright dyes with rapid switching kinetics, high power lasers and

62 fast cameras allows to minimize exposure time without losing signal to noise

63 ratio[5,6], but reaching sub-millisecond exposure remains challenging, and intense

64 irradiation exacerbates phototoxicity in live cell imaging[7,8]. Increasing the number

65 of active fluorophores per frame can reduce acquisition time, but despite algorithms

66 designed to handle overlapping fluorescent spots[9–13] this approach necessarily

67 compromises spatial resolution[14,15].

68   Here, we introduce a computational strategy that allows reducing the total

69 number of frames and independent localizations without losing spatial resolution.

70 Unlike previous approaches, our method leverages the structural redundancy of

71 most biological images to reconstruct high quality images from vastly under-

72 sampled localization microscopy data. Our method leverages deep learning, which

73 employs artificial neural networks (ANNs) to learn complex non-linear mappings

74 between numerical inputs and outputs[16]. Accordingly, we call it 'artificial neural

75 network accelerated PALM', or ANNA-PALM.

76

77 **RESULTS**

78 **A deep learning approach to super-resolution image reconstruction**

79 We aim to reconstruct a super-resolution image of approximately similar

80 information content as a standard PALM acquisition (with K frames and N

81 localizations) from a much smaller number of raw frames ($k<<K$) without changing

82 the average density of localizations, ρ, i.e. from a much smaller number of total

83 localizations ($n = \rho k \ll N = \rho K$). If PALM images are defined as 2D histograms of

84 independent localizations, this task can be formulated as restoring an image

85 corrupted by Poisson noise (and potentially additional forms of noise). Image

4

86   restoration is an ill-posed problem that has an infinity of solutions in the high-

87   dimensional space of all possible images, unless additional constraints (priors) are

88   imposed that restrict the solution to a lower dimensional subspace. Suitable

89   subspaces exist because most natural images are highly redundant, and can be

90   represented to very good approximation with a much smaller number of coefficients

91   than pixels, via appropriate functions that map feature space to pixel space[17,18]. In

92   recent years, ANNs with multiple layers (deep nets) have proven very successful at

93   learning meaningful features and non-linear mappings for image classification,

94   segmentation, restoration and many other tasks[16,18,19]. Inspired by these

95   developments, we designed ANNA-PALM, a deep learning approach for restoring

96   super-resolution views from under-sampled (sparse) localization microscopy data.

97   ANNA-PALM comprises a training stage and an inference stage (**Figure 1**).

98   For training (**Figure 1a**), a few super-resolution images representative of the

99   structure of interest (e.g. microtubules, nuclear pores, or mitochondria) are

100   obtained using standard PALM imaging, i.e. by acquiring long diffraction limited

101   image sequences (e.g. $K \sim 10^4 - 10^5$, $N \sim 10^5 - 10^7$) and processing them with

102   standard localization software[20], resulting in highly sampled (dense) PALM images.

103   In addition, a low resolution (widefield) image can also be acquired, as is commonly

104   done before single molecule imaging when bleaching out preactivated fluorophores.

105   Next, the dense PALM images are under-sampled by using a much smaller number

106   of input frames, $k \ll K$, thus yielding sparse PALM images from the same

107   localization data. Then, an ANN is trained to recover approximations of the dense

108   PALM images from these sparse PALM images (and the optional widefield image).

109   Once trained, the ANN is applied to new sparse PALM images (with or without a

110   widefield image), obtained from new image sequences with small numbers of

111   frames ($k \ll K$) -and hence in much shorter time- in order to reconstruct high

112   quality super-resolution images not previously seen (inference, **Figure 1b**).

113

114   **Neural net architecture and learning strategy**

115   Our ANN, hereafter called A-net, contains a total of 25 convolutional layers, and

116   roughly 42 million trainable parameters. A-net is adapted from the pix2pix

117   network[21], which itself builds on U-nets[22] and generative adversarial networks

118   (GANs)[23], two recent successful deep learning techniques. U-nets are special types of

119   convolutional neural networks (CNNs) that have proven effective at learning multi-

120   scale representations of images and accurate, pixel-wise mappings[22,24]. GANs can

121   generate new samples from real image distributions using a generator network that

122   outputs synthetic images, and a discriminator network that outputs the probability

123   that an input image is real or synthetic, both networks being trained simultaneously

124   to compete against each other[23]. Importantly, the generator can be conditioned on

125   input data (conditional GAN, or cGAN)[21,23], e.g. on images as in the pix2pix network.

126   We modified the pix2pix architecture to accept a computational switch as additional

127   input to handle multiple types of data, and introduced an additional network to

128   evaluate the consistency between the reconstructed image and the widefield input

129   image.

130   Training of our A-net proceeds as follows. Randomly under-sampled (i.e.

131   sparse) versions of PALM images are fed as input to the A-net, while the

132   corresponding dense PALM images are defined as the A-net's targets, i.e. desired

133   outputs (**Figure 1a**). Additional, optional inputs are widefield images, if available,

134   and the switch, which indicates the image type when multiple types of images (e.g.

135   microtubules and nuclear pores) are used during training. ANN training requires

136   defining an objective function (also called loss), which measures how well the

137   outputs match the targets. We implemented a loss function containing three terms.

138   The first term measures the difference between the A-net output and the dense

139   PALM image. Instead of the widely used mean squared error (MSE), which poorly

140   reflects visual quality[25], we used a combination of the absolute difference (L1 norm)

141   with a multi-scale version of the structural similarity index, a perceptually

142   motivated quality metric shown to improve image restoration with deep learning

143   (MS-SSIM)[26]. The second term measures the consistency between the A-net output

144   and the widefield image. Although in theory the latter should simply be a blurred

145   version of the PALM image, this is often not the case in practice[27]. Therefore, we

146   introduced another CNN (with 4-layers), called 'low resolution estimator' to predict

147   the widefield image from the super-resolution image. The corresponding loss was

148   defined as the MS-SSIM between this CNN's output and the observed widefield

149   image. (In absence of a widefield image, this loss is set to zero). The third term

150   contains a cGAN discriminator loss[21], where the discriminator is a 5-layer CNN,

151   whose inputs are the sparse PALM (and widefield) image(s) and either the dense

152   PALM image or the output of the generator above; the discriminator's output is

153   compared to 0s and 1s (for synthetic and real, respectively), respectively, via the

154   MSE. We use dropout[28] and extensive data augmentation, including random

155 rotations, translations, elastic deformations and addition of noise in the input image

156 to mimic false detections and unspecific labeling. As a result, only a few dense PALM

157 images are required for successful training without overfitting. On graphical

158 processing units (GPU), training ANNA-PALM from scratch takes on the order of

159 hours to days, but when starting from a previously trained A-net, retraining can be

160 done in an hour or less.

161   Once trained, the A-net can take sparse localization data with an optional

162 widefield image as input(s), and output a reconstructed super-resolution image in

163 less than a second (**Figure 1b**). In addition, the A-net produces an 'error map' that

164 measures the consistency of this super-resolution image with the widefield image[27]

165 (when available) and can be used to estimate the degree of reliability and highlight

166 potential reconstruction artifacts. For more details, see **Online Methods** and

167 **Supplementary Note 1**.

168

169 **Validating ANNA-PALM on simulated images**

170 We first sought to validate ANNA-PALM on synthetic data. For this, we used

171 Brownian dynamics simulations[29] to generate 200 dense PALM images of semi-

172 flexible filaments mimicking microtubules, with a resolution $R_{loc} \approx 23$ nm. These

173 represent "perfect" PALM images that would be obtained with an infinite number of

174 localizations. We applied varying levels of Poisson noise to these perfect images to

175 create sparse PALM images corresponding to finite numbers of localizations

176 (**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 $R_{5\times\text{Nyq}} \approx 85$ nm; according to this criterion,

186    $N > N_{5\text{xNyq}} = 60,000$ 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 $R_{\text{loc}} \approx 23$ nm,   as   in   the   perfect   images

199   (**Supplementary   Figure   3e,f**). These   results   indicate   that   high   quality   super-

200   resolution images can be obtained from only a small fraction of the number of

201   localizations traditionally required (here, approximately ~11% of N_{5xNyq} above; see

202   **Supplementary Figure 3d**), hence enabling a strong reduction in acquisition time.

203   Nevertheless, reconstruction errors can still occur in areas where the sparse

204   localization data are most ambiguous, e.g. where filament density is highest

205   (**Figure 2d,e,f,** white arrow). These errors can be reduced by increasing the

206   localization number n, implying a trade-off between acquisition time and

207   reconstruction quality (**Supplementary Figure 4**).

208   To quantify this trade-off, we computed the MS-SSIM between reconstructed

209   ANNA-PALM and perfect PALM images ($n = \infty$) as function of localization number,

210   from $n \sim 200$ to $n \sim 2x10^6$, in comparison with the standard PALM images

211   (**Figure 2g**). The MS-SSIM ranges from 0 to 1 and reaches 1 for perfect

212   reconstructions. For standard PALM images, the MS-SSIM increases monotonically,

213   as expected, from <0.2 to >0.95 for $n = 2x10^6$ million localizations (**Figure 2g**, black

214   curve). Using only the sparse image as input, ANNA-PALM reconstructions achieve

215   MS-SSIM that are consistently higher and increase with localization number n much

216   more rapidly than standard PALM, already exceeding 0.9 for $n \approx 10,000$

217   localizations (**Figure 2**, dashed blue curve). ANNA-PALM achieves the same MS-

218   SSIM as standard PALM at the five-fold Nyquist sampling level (≈ 0.65) with only

219     $n = 2,248$ localizations instead of $n = 58,588$, suggesting a ~26-fold speed-up. If the

220   widefield image is used as additional input, the MS-SSIM further increases, and

221   dramatically so for low localization numbers (**Figure 2g**, solid blue curve). For

222   example, with $n = 7,477$ localizations, ANNA-PALM achieves a MS-SSIM (≈ 0.95)

223 similar to standard PALM with n =644,844, implying a speed-up of roughly two

224 orders of magnitude. (Note that, if the perfect PALM image was not available for

225 these quantifications, it could be replaced by the ANNA-PALM reconstruction of a

226 dense PALM image with a large number of localizations, e.g. $n = 10^5$, with similar

227 results- see **Supplementary Figure 5**).

228   As any image restoration method, ANNA-PALM can make errors. The low

229 resolution error map described above (**Figure 1b**) provides a means to estimate

230 where errors are most likely to occur. When applied to ANNA-PALM reconstructions

231 of a sparse PALM image, this error map highlights regions containing the highest

232 density of filament crossings, where reconstructions tend to be least accurate

233 (**Supplementary Figure 6i,k**). If we artificially displace a small piece of filament in

234 this image to simulate a false positive and a false negative in the reconstruction

235 (**Supplementary Figure 6b,d,** white and blue arrows, respectively), the affected

236 regions also light up in the error map (**Supplementary Figure 6j,l**). Thus, the error

237 map offers a useful tool to highlight regions most likely to contain reconstruction

238 errors, and conversely, to outline regions where reconstructions are most

239 trustworthy. Thus, simulations suggest that ANNA-PALM can considerably reduce

240 acquisition time in localization microscopy and also map reconstruction reliability.

241

242 **ANNA-PALM reconstructions of immunostained microtubules**

243 We next tested our method on real images of immunolabeled microtubules

244 (**Figure 3**). We trained our A-net on seven dense PALM images (with corresponding

245 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-

269   PALM reconstruction from the sparse localization data alone (**Figure 3e,h**, blue

270   arrows). However, when combining the widefield and sparse PALM data, these

271   artifacts are largely removed and reconstructions agree very well with the dense

272   PALM image (**Figure 3f,i**). Reconstruction quality increases with the number of

273   frames k (**Figure 3j, Supplementary Figures 8-9** and **Supplementary Video 1**).

274   More quantitatively, a MS-SSIM analysis similar to that for the simulated data above

275   (with the ANNA-PALM output of the dense PALM image defined as ground truth; see

276   **Supplementary Figures 5, 9f**) suggests that ANNA-PALM allows a hundred-fold

277   reduction of acquisition time compared to standard PALM (**Supplementary**

278   **Figure 9g**). **Supplementary Figure 10** shows other examples of sparse

279   microtubule images reconstructed by ANNA-PALM.

280   As for simulations above, we used the widefield image to compute an error

281   map (**Supplementary Figures 6,11**). Bright areas in this error map highlight

282   regions where the reconstruction indeed disagrees with the dense PALM image;

283   conversely, reconstructions are of high quality in the majority of regions where the

284   error map is dim (**Supplementary Figure 11d-f**). These results demonstrate

285   experimentally that ANNA-PALM can restore high quality approximations of super-

286   resolution images from much shorter acquisition time than typical for PALM

287   imaging, and also predict where reconstruction errors are most likely.

288

289   **ANNA-PALM enables high-throughput super-resolution imaging**

290   The drastic improvement in imaging efficiency afforded by ANNA-PALM permits

291   super-resolution imaging of orders of magnitude more cells and FoVs per unit time.

292   To demonstrate this, we used an automated acquisition protocol to image >1,000

293   cells with immunolabeled microtubules in 1,089 (33x33), partly overlapping, FoVs

294   of 55.3 μm x 55.3 μm each (**Figure 4**, **Supplementary Figure 12**). We first acquired

295   widefield images at each of these positions, in a total of ~12 minutes, mostly

296   consisting of stage stabilization delays (**Supplementary Figure 12a**). Next, we

297   obtained 1,089 sparse PALM images using only 10 s of imaging time per FoV

298   (k=1,000 frames, Δt = 10 ms), in a total of only ~3.1 hours (**Figure 4a**). Neither the

299   widefield nor the sparse PALM images provided much small scale information

300   (**Figure 4c, Supplementary Figure 12 c,e**). However, ANNA-PALM reconstructions

301   led to high quality super-resolution images, allowing to visualize the microtubule

302   network with clarity and to distinguish microtubule filaments in dense areas that

303   appeared as unstructured regions in the sparse PALM image (**Figure 4b,d**). The

304   FWHM across filaments in the reconstructed image was ~51 nm (**Figure 4d**), within

305   the range measured for the training data (**Supplementary Figure 7a**). Similar

306   images can be obtained by ANNA-PALM using the widefield images alone

307   (**Supplementary Figure 12**), although of lower quality, as discussed above.

308   Stitching the reconstructed images together yielded a single super-resolution image

309   that contained approximately seven billion 20x20 nm pixels and covered an area of

310   1.8 mm x 1.8 mm, thereby spanning almost five orders of magnitude in spatial scales

311   (**Figure 4b,d, Supplementary Figure 12b** and **Supplementary Video 2**).

312

313   **ANNA-PALM is robust to experimental perturbations**

314   ANNA-PALM can reconstruct accurate super-resolution images from sparse data

315   because it was trained on similar images before. This raises the question of whether

316   an ANN trained in one experimental condition can be successfully applied to

317   another condition. To test this, we used the A-net as trained above to analyze

318   microtubule images of cells subjected to drugs affecting the cytoskeletal network.

319   We first treated U373 cells with 1 µM of Taxol, an antimitotic agent that inhibits the

320   depolymerization of microtubules and increases their bending rigidity[30,31].

321   **Figure 5d** shows a sparse PALM image of these cells (k=800, kΔt=8 s), in which the

322   microtubule network is barely recognizable. By contrast, the ANNA-PALM

323   reconstructions clearly display a complex microtubule network and agree well with

324   the dense PALM image obtained from K=60,000 frames ($K\Delta$t=10 min)

325   (**Figure 5d,e,f**). These images display a larger density of straighter and more

326   parallel filaments with less frequent crossings than in the untreated cells

327   (**Figure 5a-c**), consistent with microtubule stabilization and increased rigidity[31].

328   Next, we treated cells with 1 µM of Nocodazole, a drug that, on the opposite,

329   promotes microtubule depolymerization and is expected to more dramatically alter

330   the cytoskeletal network[32]. Again, whereas the sparse image obtained from k=300

331   frames ($k\Delta$t=3 s) contained little exploitable information (**Figure 5g**), the ANNA-

332   PALM reconstruction offered clear and detailed views of the disrupted microtubule

333   network (**Figure 5h**), exhibiting a much smaller number of filaments, with higher

334   curvature, than in untreated cells. These reconstructions were in good (though not

335   perfect) agreement with dense PALM images obtained from K=60,000 frames

336   ($K\Delta$t= 10 min) (**Figure 5i**). Thus, ANNA-PALM, when trained on microtubules in

337   one experimental condition, may be successfully applied to new experimental

338   conditions without retraining, thereby highlighting the method's robustness to

339   biologically relevant structural alterations.

340   We further asked if ANNA-PALM is robust to changes in technical imaging

341   conditions. To address this, we performed localization microscopy on microtubules

342   by simultaneously changing multiple important imaging parameters relative to the

343   training data. Instead of PALM/STORM, we used DNA-PAINT a technique where

344   single molecule detection relies on transient binding of fluorophore-conjugated

345   short DNA strands to complementary, antibody-conjugated, DNA strands[33], rather

346   than on fluorophore blinking. The continuously emitting freely diffusing dyes lead to

347   higher background noise in DNA-PAINT compared to PALM/STORM. Moreover, we

348   used primary mouse antibodies instead of rat antibodies, Cy3 dyes instead of Alexa-

349   647 dyes, and an EMCCD instead of a sCMOS camera. Despite all these differences,

350   when ANNA-PALM was applied without retraining on a sparse microtubule image

351   (k=400 frames, kΔt=12 s) , the reconstructed image still agreed very well with the

352   dense DNA-PAINT image obtained from K=60,000 frames ($K\Delta$t=30 min) (**Figure 5j-**

353   **l**). These data demonstrate the high robustness of ANNA-PALM to changes in

354   experimental imaging conditions.

355

356   **ANNA-PALM adapts to different biological structures**

357   To demonstrate that ANNA-PALM is not restricted to filamentary structure, we

358   turned to nuclear pores, a very different biological structure, and another popular

359   target of super-resolution imaging studies[34–36]. We retrained A-net simultaneously

360   on microtubule images and on a single PALM image of the nucleoporin gp210 in

361   immunolabeled nuclear membranes of *Xenopus* frog eggs[34,36] (*K*=30,000). With the

362   switch (**Figure 1**) set to microtubules ('MT'), this newly trained A-net can still

363   reconstruct sparse images of microtubules as when trained exclusively on

364   microtubule data (**Supplementary Figure 13a-c**). We then applied the same A-net

365   with the switch set to nuclear pores ('NPC') to reconstruct a new sparse PALM

366   image of gp210 obtained from the first *k*=3,000 frames (**Figure 6a**). The sparsity of

367   this image makes it difficult to clearly distinguish individual nuclear pores. ANNA-

368   PALM, however, reconstructs a much clearer image, containing many easily

369   identifiable ring-like structures, as expected for nuclear pores[34] (**Figure 6b**), and in

370   good agreement with the dense PALM image obtained from *K*=30,000 frames (even

371   though the latter shows mostly incomplete, open rings, presumably due to

372   suboptimal labeling) (**Figure 6c**). An automated procedure based on cross-

373   correlation with a ring template indeed identified ~2.7 times more putative nuclear

374   pores from the ANNA-PALM image than the sparse image (**Supplementary Figure**

375   **14a-c**). Moreover, computed pore locations were in good agreement with a PALM

376   image of wheat germ agglutinin (WGA), a lectin that concentrates in the inner

377   nuclear pore channel[36] (**Supplementary Figure 14 d-f**).   These results show that

378   ANNA-PALM can successfully analyze non-filamentary structures, when properly

379   retrained, and that a single ANN, with a simple computational switch, can

380   reconstruct very different types of structures.

381   Finally, we imaged TOM22, a protein of the mitochondrial outer membrane[37].

382   Whereas, at the resolution of our experiments, microtubules and nucleoporins are

383   essentially one-dimensional and zero-dimensional structures, mitochondrial

384   membranes are two-dimensional surfaces. Furthermore, their complex 3D

385   morphology might seem less predictable than filaments or nuclear pores, potentially

386   hampering ANNA-PALM reconstruction. Despite these differences, after being

387   trained on nine PALM images of TOM22 (with frame numbers ranging from

388   K=24,000 to K=40,000), ANNA-PALM reconstructions of distinct sparse PALM

389   images (k=400 frames) displayed mitochondrial morphologies in good agreement

390   with the dense PALM images (**Figure 6d-f**) - although the protein's localization

391   along the membrane was less well reproduced. Taken together, our results illustrate

392   the versatility of ANNA-PALM and its applicability to images of very different

393   structural content.

394

395   **Discussion**

396   We introduced ANNA-PALM, a computational method based on deep learning, that

397   reconstructs high quality super-resolution images from sparse, rapidly acquired,

398   single molecule localization data (and/or widefield images). Our method enables

399   considerable gains in acquisition time compared to standard localization

400   microscopy without increasing active fluorophore density, thereby preserving

401   spatial resolution. In fact, ANNA-PALM even improves spatial resolution when

402   applied to images of lower resolution than the training data

403   (**Supplementary Figures 7a, 10**), and greatly diminishes the detrimental effect of

404   drift. The improvement in imaging efficiency afforded by ANNA-PALM alleviates the

405   incompatibility between high resolution and high-throughput microscopy by

406   enabling super-resolution imaging of thousands of cells within a few hours or even

407   less (**Figure 4, Supplementary Figure 12**). This will facilitate super-resolution

408   studies of rare events, cellular heterogeneity and of partly stochastic structures such

409   as cytoskeletal polymers or chromosomes, whose characterization requires

410   statistics on many configurations[38,39]. ANNA-PALM may also be beneficial for high-

411   throughput imaging screens, e.g. of drug treatments or gene knock-outs[40–42]. In

412   addition, we envision applications to super-resolution imaging of large samples by

413   stitching together multiple images of spatially adjacent fields. The ability to generate

414   images spanning many orders of magnitude in scale could be well adapted to

415   expansion microscopy, a super-resolution technique that physically increases

416   sample size, but often requires tiling many fields of view to image even a single

417   cell[43,44]. With correlative microscopy[45], it might also be possible to train ANNA-

418   PALM to reconstruct electron microscopy (EM) images from fluorescence images,

419   potentially extending the method to molecular resolutions currently out of reach of

420   localization microscopy. Adaptation of ANNA-PALM to 3D[15,46] and multi-color[47,48]

421   localization microscopy should be relatively straightforward. Localization

422   microscopy of cellular dynamics remains very challenging[3,49]. By using much fewer

423   frames (or even only widefield images), ANNA-PALM could dramatically improve

424   the temporal resolution of live cell localization microscopy without sacrificing

425   spatial resolution or increasing phototoxicity and photobleaching. Thus, ANNA-

426   PALM provides multiple novel avenues for multi-scale imaging beyond standard

427   spatio-temporal resolution limits.

428   Neverthelesss, important caveats should be stressed. First, although ANNA-

429   PALM can be applied successfully to very different types of images (**Figure 6**), the

430   method fails in absence of statistical redundancies between molecular localizations,

431   e.g. for entirely random distributions of molecules. Second, ANNA-PALM requires

432   prior training on dense PALM images with structures similar to those in the images

433   to be reconstructed. We showed that ANNA-PALM is robust, i.e. does not require

434   retraining, for some experimentally induced changes in structures and variations in

435   imaging parameters (**Figure 5**). Nevertheless, indiscriminate application of ANNA-

436   PALM to very different structures without retraining, or incorrect setting of the

437   switch, may result in artifacts (**Supplementary Figure 13h**). Third, even when

438   applied to data similar to the training images, ANNA-PALM can produce errors -as

439   any reconstruction method in a context of information scarcity. The frequency of

440   errors can be reduced by increasing the number of recorded frames, at the cost of

441   reduced acceleration (see **Figures 2,3** and **Supplementary Figures 4, 8, 9**). In

442   addition, ANNA-PALM can use widefield images to estimate the reliability of

443   reconstructions, thereby helping their interpretation, providing some protection

444   against artifacts and indicating when retraining may be needed (**Supplementary**

445   **Figures 11, 15**). Future work, e.g. using Bayesian deep learning[50], may provide

446   additional assessments of reconstruction uncertainty and shed more light on the

447   scope and limitations of our approach.

448   As a purely computational technique, ANNA-PALM does not necessitate any

449   changes to existing microscopy systems, but only requires one or a few standard

450   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.

481
482

Author contributions

484   W.O.: conceived method, developed ANNA-PALM software and web application,

485   performed experiments and analyses. A.A., M.L., X.H.: performed experiments. C.Z.:

486   conceived method, supervised project and wrote manuscript.

487

488

Competing financial interests statement

490   W.O. and C.Z. are listed as inventors on European patent application EP17306022

491   filed by Institut Pasteur.

492

493

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).

540 21. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with
541   Conditional Adversarial Networks. *http://arxiv.org/abs/1611.07004* (2016).
542 22. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for
543   biomedical image segmentation. in *Medical Image Computing and Computer-*
544   *Assisted Intervention − MICCAI 2015. Lecture Notes in Computer Science,* **9351,**
545   234–241 (Springer International Publishing, 2015).
546 23. Goodfellow, I. *et al.* Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.*
547   *27* 2672–2680 (2014). doi:10.1017/CBO9781139058452
548 24. Ciresan, D., Giusti Alessandro, Luca, G. & Schmidhuber, J. Mitosis Detection in
549   Breast Cancer Histology Images with Deep Neural Networks. *Med. Image*
550   *Comput. Comput. Interv. − MICCAI 2013* **8150,** (2013).
551 25. Wang, Z. W. Z. & Bovik, A. C. A. C. Mean squared error: Love it or leave it? A
552   new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **26,** 98–117
553   (2009).
554 26. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss Functions for Image Restoration
555   with Neural Networks. *IEEE Trans. Comput. IMAGING* **3,** 47–57 (2017).
556 27. Culley, S. *et al.* NanoJ-SQUIRREL: quantitative mapping and minimisation of
557   super-resolution optical imaging artefacts. *bioRxiv* 158279 (2017).
558   doi:10.1101/158279
559 28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R.
560   Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach.*
561   *Learn. Res.* **15,** (2014).
562 29. Arbona, J.-M., Herbert, S., Fabre, E. & Zimmer, C. Inferring the physical
563   properties of yeast chromatin through Bayesian analysis of whole nucleus
564   simulations. *Genome Biol.* **18,** 81 (2017).
565 30. Arnal, I. & Wade, R. H. How does taxol stabilize microtubules? *Curr. Biol.* **5,**
566   900–908 (1995).
567 31. Wu, S. *et al.* Microtubule motors regulate ISOC activation necessary to
568   increase endothelial cell permeability. *J. Biol. Chem.* **282,** 34801–8 (2007).
569 32. De Brabander, M., De May, J., Joniau, M. & Geuens, G. Ultrastructural
570   immunocytochemical distribution of tubulin in cultured cells treated with
571   microtubule inhibitors. *Cell Biol. Int. Rep.* **1,** 177–83 (1977).
572 33. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R.
573   Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12,** 1198–1228
574   (2017).
575 34. Löschberger, A. *et al.* Super-resolution imaging visualizes the eightfold
576   symmetry of gp210 proteins around the nuclear pore complex and resolves
577   the central channel with nanometer resolution. *J. Cell Sci.* **125,** 570–5 (2012).
578 35. Szymborska, A. *et al.* Nuclear pore scaffold structure analyzed by super-
579   resolution microscopy and particle averaging. *Science* **341,** 655–8 (2013).
580 36. Sellés, J. *et al.* Nuclear pore complex plasticity during developmental process
581   as revealed by super-resolution microscopy. *Sci. Rep.* **7,** 14732 (2017).
582 37. Bellot, G. *et al.* TOM22, a core component of the mitochondria outer
583   membrane protein translocation pore, is a mitochondrial receptor for the
584   proapoptotic protein Bax. *Cell Death Differ.* **14,** 785–794 (2007).
585 38. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin

586   folding for different epigenetic states. *Nature* **529,** 418–422 (2016).

587   39.   Zhang, Z., Nishimura, Y. & Kanchanawong, P. Extracting microtubule networks
588   from superresolution single-molecule localization microscopy data. *Mol. Biol.*
589   *Cell* **28,** 333–345 (2017).

590   40.   Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse
591   microscopy reveals cell division genes. *Nature* **464,** 721–727 (2010).

592   41.   Beghin, A. *et al.* Localization-based super-resolution imaging meets high-
593   content screening. *Nat. Methods* **14,** 1184–1190 (2017).

594   42.   Ouyang, W. & Zimmer, C. The imaging tsunami: Computational opportunities
595   and challenges. *Curr. Opin. Syst. Biol.* **4,** 105–113 (2017).

596   43.   Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science (80-.).*
597   **347,** 543–548 (2015).

598   44.   Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* (2017).
599   doi:10.1038/nmeth.4261

600   45.   de Boer, P., Hoogenboom, J. P. & Giepmans, B. N. G. Correlated light and
601   electron microscopy: ultrastructure lights up! *Nat. Methods* **12,** 503–513
602   (2015).

603   46.   Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-
604   resolution imaging by stochastic optical reconstruction microscopy. *Science*
605   *(80-.).* **319,** 810–813 (2008).

606   47.   Bates, M., Huang, B., Dempsey, G. T. & Zhuang, X. Multicolor Super-Resolution
607   Imaging with Photo-Switchable Fluorescent Probes. *Science (80-.).*
608   1146598v1 (2007).

609   48.   Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with
610   DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11,** 313–318 (2014).

611   49.   Shroff, H., Galbraith, C. G., Galbraith, J. A. & Betzig, E. Live-cell photoactivated
612   localization microscopy of nanoscale adhesion dynamics. *Nat Methods* **5,** 417–
613   423 (2008).

614   50.   Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep
615   Learning for Computer Vision? *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017)*
616   *pre-proceedings* 5580–5590 (2017).

617

FIGURE LEGENDS

619

620   **FIGURE 1: Overview of ANNA-PALM**

621   ANNA-PALM consists of two main stages: (**a**) acquisition of training images using

622   standard localization microscopy (PALM) followed by artificial neural network

623   (ANN) training, and (**b**) reconstruction of super-resolution views and low resolution

624   error maps from new sparse PALM and/or widefield images (inference).

625   (**a**) Training images are obtained by acquiring one or a few long sequences, of

626    $K \sim 10^3$-10^5 diffraction limited, single molecule image frames, as in standard PALM

627   experiments; optionally, a widefield image W can also be acquired (top). The

628   acquisition time for each image sequence is $K\Delta t$, where Δt is the single frame

629   exposure time. Standard localization microscopy algorithms (grey 'localization'

630   boxes) are used to generate super-resolution images. For each experiment, a highly

631   sampled (dense) super-resolution image (PK) is generated using all (or in practice

632   ≥95%) available K frames. Sparse PALM images (Pk) from the same experiment are

633   obtained by using only $k<<K$ frames. A switch (light blue) can be set to distinguish

634   different types of structures, e.g. nuclear pore complexes ('NPC'), mitochondria

635   ('Mito') or microtubules ('MT'). An ANN (labeled $\mathcal{G}$ for 'generator', orange) is trained

636   by using the sparse PALM images Pk (plus an upsampled version L of the widefield

637   image W and the switch setting) as inputs and the corresponding dense PALM image

638   PK as target output. During training, the output of the generator $\mathcal{G}$ (Ak) is compared

639   to the target image PK and the widefield image L (if available) via three loss, or error

640   functions (gray bordered boxes): (i) the 'super-resolution reconstruction error'

641 measures the difference between the reconstructed image Ak and the target PK

642 using a combination of the L1 norm and the MS-SSIM; (ii) the 'low resolution

643 reconstruction error' measures the MS-SSIM between the low resolution image WA

644 predicted from the reconstruction Ak and the low resolution image WP predicted

645 from the target image PK. Images WA and WP are predicted using a second ANN,

646 called low-resolution estimator (labeled Q, blue) that is trained to produce an

647 approximation of the actual widefield image W based on the MS-SSIM metric; (iii)

648 the 'conditional GAN error' uses a third ANN (labeled D for 'discriminator', red) that

649 attempts to distinguish between real dense PALM images PK and the generator's

650 output Ak. The combined loss functions are iteratively optimized using stochastic

651 gradient descent. (**b**) A short sequence of diffraction limited single molecule images

652 (with $k<<K$ frames, i.e. acquisition time $k\Delta t$), and an optional widefield image (W'),

653 are acquired. Standard localization algorithms generate a sparse (under-sampled)

654 PALM image (P'k). This sparse image P'k (and the upscaled widefield image L' and

655 switch setting) are fed as inputs to the trained generator G, which outputs a

656 reconstructed ANNA-PALM image (A'k). In addition, the low resolution estimator Q

657 predicts a low resolution image WA', which can be compared to the input widefield

658 image W' via the MS-SSIM to produce a low resolution error map (top).

659

660 **FIGURE 2: Validation of ANNA-PALM on simulated images**

661 (**a**) Simulated widefield image of microtubules. (**b**) Simulated sparse PALM image of

662 microtubules with $n=$ 6,834 localizations. (**c**) ANNA-PALM reconstruction using only

663 the widefield image **a** as input. (**d**) ANNA-PALM reconstruction using both the

664 widefield image **a** and the sparse PALM **b** image as inputs. (**e**) Simulated "perfect"

665 PALM image, equivalent to a PALM image with an infinite number of localizations

666 ($n=\infty$) and a resolution of 23 nm. This image was used to generate **a** (by convolution

667 with a Gaussian kernel approximating the microscope point spread function) and **b**

668 (by application of Poisson noise). (**f**) Merged image showing the perfect PALM image

669 **e** in green and the ANNA-PALM reconstruction **d** in red. Note that the ANNA-PALM

670 images **c,d** provide many high resolution details that are absent from the widefield

671 image **a** and the sparse PALM image **b** and that are in good (**c**) or very good (**d**)

672 agreement with the perfect PALM image **e**. Some reconstruction errors are

673 highlighted by arrows. Blue arrows in panel **c** point to errors of ANNA-PALM

674 reconstruction from the widefield image only, the white arrow in panel **d** points to

675 an error of ANNA-PALM reconstruction from both widefield and sparse PALM

676 images combined. Reconstruction errors diminish for larger numbers of

677 localizations, *n* (**Supplementary Figure 4**). (**g**) Reconstruction quality of PALM and

678 ANNA-PALM images, measured by the MS-SSIM with the perfect PALM image **e**, as

679 function of localization number *n*. Black curve: reconstruction quality of the

680 standard PALM images. Dashed blue curve: reconstruction quality of ANNA-PALM

681 using the sparse PALM images as input. Solid blue curve: reconstruction quality of

682 ANNA-PALM using both the sparse PALM and widefield images as inputs. Red

683 dashed line: reconstruction quality of ANNA-PALM using the widefield images as

684 input only. Dots are averages from 10 simulations; error bars show standard

685 deviations. The vertical dashed orange line indicates the minimum number of

686 localizations needed to achieve a resolution of $R_{5\times\text{Nyq}} =23$ nm according to the five-

687 fold Nyquist criterion[3]. The dashed grey line indicates the minimum number of

688 localizations needed to achieve a double mean nearest neighbor distance less than

689 23 nm. ANNA-PALM reconstructions from sparse PALM images only (i.e. without

690 widefield images) achieve the same average MS-SSIM as standard PALM at the five-

691 fold Nyquist sampling limit with 26 times less localizations (blue double arrow).

692 ANNA-PALM reconstruction quality is highest when using both widefield and sparse

693 PALM images as inputs.

694

695 **FIGURE 3: ANNA-PALM imaging of microtubules**

696 ANNA-PALM reconstructions of a localization microscopy image of immunostained

697 microtubules. (**a**) Widefield image. (**b**) Sparse PALM image obtained from the first

698 9 s of acquisition (k=300 frames, n=11,740 localizations). (**c**) Dense PALM image

699 obtained from a 15 min long acquisition (K=30,000 frames, N=409,364

700 localizations). (**d**) ANNA-PALM reconstruction from the widefield image **a** only.

701 (**e**) ANNA-PALM reconstruction from the sparse PALM image **b** only. (**f**) ANNA-

702 PALM reconstruction from the widefield image **a** and sparse PLAM image **b**

703 combined. In panels **b-f**, pixel values are linearly mapped to colors from the look-up

704 table shown below. Black and white correspond to values V_{min} and V_{max},

705 respectively, with V_{min}=0 for all panels, V_{max}=3, 24, 102, 102 and 102 for panels **b**, **c**,

706 **d**, **e** and **f**, respectively. (**g-i**) Merged images comparing ANNA-PALM

707 reconstructions from panels **d-f** to the dense PALM image **c**. ANNA-PALM

708 reconstructions are shown in red, the dense PALM image in green. (**j**) Gradual

709 improvement of image quality for increasing acquisition time $k\Delta$t, shown for the

710　area highlighted in the insets of panels **a-i**. Top row: sparse PALM images. Middle

711　row: ANNA-PALM reconstructions from the sparse PALM images only (without

712　widefield). Bottom row: ANNA-PALM reconstructions from the widefield and sparse

713　PALM images combined. **Supplementary Video 1** shows the gradual increase in

714　quality of PALM and ANNA-PALM images with increased acquisition time for the

715　larger region of interest shown in panels **a-i.**

716

717　**FIGURE 4: High-throughput imaging with ANNA-PALM**

718　Application of ANNA-PALM to high-throughput imaging of a 1.8 mm x 1.8 mm area

719　containing more than 1,000 cells. (**a**) Sparse PALM image of this area, constructed

720　by assembling a mosaic of 33x33=1,089 sparse PALM images of individual fields of

721　view, obtained from k=1,000 raw frames each (with Δt=10 ms exposure time per

722　frame, i.e. in $k\Delta$t =10 s). Total image acquisition time was 1,089x10s, i.e. ~3.1 hours.

723　The sparsity of the image is not apparent at this large scale. (**b**) ANNA-PALM

724　reconstruction of the image in **a**, obtained by assembling a mosaic of 1,089

725　individual reconstructions (one per field of view). (**c**) Magnified view of the green

726　boxed region in **a**. The inset shows a further magnified view of the yellow boxed

727　region, highlighting the sparsity of the image. (**d**) Same as **c**, but for the ANNA-PALM

728　reconstruction. A line profile across a microtubule is shown, with a FWHM≈51 nm.

729　Non-linear contrast adjustment was applied manually for panels **a** and **b,** with black

730　corresponding to values of zero in both panels. In panels **c** and **d**, pixel values were

731　linearly mapped to colors from the look-up table in **Figure 3;** Black and white

732　correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, and

733   V_{max}=3 and 51 for panels **c** and **d**, respectively. See also **Supplementary Video 2** for

734   an animated 'zoom-in' highlighting the spatial scales covered by the assembled

735   image. See also **Supplementary Figure 12** for ANNA-PALM reconstructions of the

736   same area from the widefield images only.

737

738   **FIGURE 5: Robustness of ANNA-PALM to experimental perturbations**

739   This figure shows ANNA-PALM reconstructions using an ANN trained on PALM

740   images of microtubules in untreated cells and applied without retraining to sparse

741   localization images of microtubules in different experimental conditions: untreated

742   control cells (**a-c**); cells treated with 1 μM of Taxol (**d-f**); cells treated with 1 μM of

743   Nocodazole (**g-i**); untreated cells imaged with DNA-PAINT (**j-l**). (**a,d,g,j**) Sparse

744   localization images obtained from the first k frames of the acquired image sequence,

745   with k=500, 800, 300, and 400 for **a, d, g,** and **j**, respectively. (**b,e,h,k**) ANNA-PALM

746   reconstructions using the sparse localization images immediately to the left as

747   input. (**c,f,i,l**) Dense localization images obtained from K=60,000 frames. Pixel

748   values are linearly mapped to colors from the look-up table in **Figure 3**. Black and

749   white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels,

750   and V_{max}=10, 120, 90, 25, 150, 40, 18, 150, 50, 18, 120, and 200 for panels **a**, **b**, **c**, **d**,

751   **e**, **f, g, h**, **i**, **j**, **k**, and **l**, respectively.

752

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

754   PALM and ANNA-PALM images of nuclear pores (**a-c**) and mitochondria (**d-f**).

755   (**a**) Sparse PALM image of the immunolabeled *Xenopus* nucleoporin gp210 obtained

756   from the first k=3,000 frames. Note that individual nuclear pores are hard to

757   identify. (**b**) ANNA-PALM reconstruction of image **a**. (**c**) Dense PALM image obtained

758   from all K=30,000 frames. (**d**) Sparse PALM image of the immunolabeled

759   mitochondrial outer membrane protein TOM22, obtained from the first k=400

760   frames. (**e**) ANNA-PALM reconstruction of image **d**. (**f**) Dense PALM image obtained

761   from all K=30,000 frames. Pixel values are linearly mapped to colors from the look-

762   up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max},

763   respectively, with V_{min}=0 for all panels, and V_{max}=3, 51, 3, 3, 128, and 18 for panels **a**,

764   **b**, **c**, **d**, **e**, and **f**, respectively

ONLINE METHODS

767 **Artificial neural network**

768 Our ANN, called 'A-net', is based on the pix2pix architecture[21], which is a special

769 conditional generative adversarial network (cGAN)[23] for image to image

770 "translation", i.e. mapping from one type of image to another. The A-net consists of

771 three distinct neural networks: (i) a generator network G that produces the

772 reconstructed super-resolution image, (ii) a network Q called 'low resolution

773 estimator' that produces the low resolution error map, (iii) a cGAN discriminator

774 network D that provides the adversarial loss (**Figure 1a**). The generator network G

775 builds on the U-net architecture, and consists of an encoder-decoder network with

776 skip connections[22] and 16 convolutional layers. Its inputs and outputs are image

777 patches containing $(256m)$x$(256m)$ pixels, where m is an integer (we used m=1 or 2

778 but this can be adjusted for different sizes of CPU/GPU memory or input images).

779 The input is a sparse PALM image, a widefield image upscaled to the same size (see

780 below), and a computational switch number that allows the network to switch

781 between different types of image structures (e.g. nuclear pores or microtubules).

782 The switch setting is encoded numerically and coupled by convolutional operations

783 into the A-net encoder. The output of the generator G is a reconstructed image

784 (called ANNA-PALM reconstruction or ANNA-PALM image elsewhere) of the same

785 size as the input images. The low resolution estimator Q has four convolutional

786 layers. It takes the $(256m)$x$(256m)$ dense PALM image patch or the ANNA-PALM

787 image patch as input and outputs a low resolution image with $(64m)$x$(64m)$ pixels.

788   The cGAN discriminator network D has five convolutional layers. Its inputs are three

789    $(256m)$x$(256m)$ pixel image patches (the sparse PALM image, the widefield image,

790   and either the ANNA-PALM reconstruction or the corresponding dense PALM

791   image), plus the upscaled widefield image, and its output is a $(30m)$x$(30m)$ image

792   whose pixel values indicate whether the corresponding input patch is real (i.e. an

793   experimentally obtained conventional PALM image) or produced by the generator $\mathcal{G}$.

794   All convolutional layers are followed by batch normalization[51]. Dropout layers[28]

795   (with dropout probability p=0.5) are introduced in the central layers of the A-net

796   generator and turned on during training, but switched off during inference.

797   Activation functions are rectified linear units (ReLU) $x \to \sup(x, 0)$ or "leaky" ReLUs

798    $x \to \sup(x, 0) + \inf(\varepsilon x, 0)$ with ε =0.2[52], except for the last layer of $\mathcal{G}$, which uses the

799   hyperbolic tangent $x \to \tanh(x)$ and the last layer of Q, which uses a sigmoid

800   function $x \to (1 + \exp(-x))^{-1}$. The A-net architecture is fully described in

801   **Supplementary Note 1** and **Supplementary Tables 1-4**.

802

803   **Training objectives and error map**

804   Each of the three networks mentioned above ($\mathcal{G}, Q,$ and D) is associated to a distinct

805   objective function - also called loss- and hereafter noted $\mathcal{L}_G$, $\mathcal{L}_Q$, and $\mathcal{L}_D$ respectively.

806   These loss functions are specified in detail below. In the following equations, for

807   notational simplicity, we designate the sparse input image as S, the low resolution

808   (widefield) input image as W, the corresponding dense PALM image (i.e. the target)

809   as T, and the A-net's output as $A = \mathcal{G}(S, W)$ (in **Figure 1a**, images S, T and A are

810   labeled Pk, PK and Ak, respectively); the computational switch indicating the image

811   type is noted $\boldsymbol{M}$. Low resolution images produced by the low resolution estimator

812   network Q from $\boldsymbol{A}$ and $\boldsymbol{T}$ are designated as $\boldsymbol{W_A} = Q(\boldsymbol{A})$ and $\boldsymbol{W_T} = Q(\boldsymbol{T})$,

813   respectively.

814   The generator loss function $\mathscr{L}_G$ is the sum of three terms. The first term of $\mathscr{L}_G$

815   is the super-resolution reconstruction error, hereafter called $\mathscr{L}_{\mathrm{SuperRes}}(\mathcal{G})$. This term

816   penalizes the difference between the generator output $\boldsymbol{A}$ and the target image $\boldsymbol{T}$.

817   Based on a recent analysis of loss functions for image restoration with neural

818   networks[26], we defined this difference as a weighted average of two quantities:

819   (i) the multi-scale structural similarity index (MS-SSIM) between $\boldsymbol{A}$ and $\boldsymbol{T}$ and (ii) a

820   modification of the L1 norm, where the absolute difference between $\boldsymbol{A}$ and $\boldsymbol{T}$ is

821   smoothed by a Gaussian kernel:

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

822   where $\mathbb{E}$ denotes expectation; $p_{\mathrm{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})$ is the joint probability density of the

823   sparse PALM images $\boldsymbol{S}$, dense PALM images $\boldsymbol{T}$, widefield images $\boldsymbol{W}$ and switch

824   settings $\boldsymbol{M}$ from the training data set; $MS\_SSIM(\boldsymbol{A},\boldsymbol{T})$ is the multi-scale structural

825   similarity index between $\boldsymbol{A}$ and $\boldsymbol{T}$; G_σ is a Gaussian smoothing kernel; $*$ denotes

826   convolution; $|\boldsymbol{A} - \boldsymbol{T}|$ is the absolute difference image (i.e. pixel (i,j) has value

827    $|\boldsymbol{A}(\mathrm{i,j}) - \boldsymbol{T}(\mathrm{i,j})|$ and $\rho \in [0,1]$ is a scalar weight that balances the relative

828   contributions of MS-SSIM and the modified L1 norm and is set to $\rho = 0.84$ as in ref.

829   [26].

830   The second term of $\mathscr{L}_{\mathcal{G}}$ is called $\mathscr{L}_{\text{LowRes}}(\mathcal{G}, Q)$ and measures the consistency

831   between the low resolution images $\boldsymbol{W}_A$ and $\boldsymbol{W}_T$ predicted by the low resolution

832   estimator network Q:

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

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

833   Alternatively, in the above objective function, $\boldsymbol{W}_T$ can be replaced by the actually

834   observed widefield image $\boldsymbol{W}$, although with our data this led to slightly lower

835   reconstruction quality. The low resolution estimator network Q is trained

836   simultaneously with the generator $\mathcal{G}$ to produce a low resolution image from the

837   dense PALM image $\boldsymbol{T}$ that is consistent with the observed low resolution image $\boldsymbol{W}$.

838   This training is done based on the following objective function:

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

839   Note that the reconstructed low resolution image $Q(\boldsymbol{T})$ is four times smaller than

840   the dense PALM image $\boldsymbol{T}$, as described in **Supplementary Note 1**. Because the input

841   widefield image $\boldsymbol{W}$ can have a different size, we use bilinear interpolation to resize

842    $\boldsymbol{W}$ to the same size as $Q(\boldsymbol{T})$. (If needed, a scaling factor different from four can be

843   obtained by adding or removing downsample layers in network Q). At inference, the

844   low resolution estimator Q is also used to produce the error map, as shown in

845   **Figure 1b**, and **Supplementary Figures 6, 11, 15**. This error map is defined as:

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

846   High (respectively low) values of the error map indicate large (respectively small)

847   inconsistencies, between the reconstructed super-resolution image $\boldsymbol{A}$ and the

848   observed widefield image $\boldsymbol{W}$.

849   The third term of $\mathscr{L}_G$ draws from recent work on generative adversarial

850   networks (GAN)[21,23,53] and is noted $\mathscr{L}_{\text{cGAN}}(\mathcal{G}, \mathcal{D})$. In a GAN, a generator network $\mathcal{G}$

851   learns to transform random input vectors z (drawn from a probability density

852    $p_z(z)$) into new samples of a data probability density $p_{\text{data}}(x)$. In our case, the data

853   samples x are the dense PALM images T. The generator $\mathcal{G}$ learns by working against

854   a discriminator network $\mathcal{D}$ that simultaneously learns to discriminate between

855   original data samples and samples generated by $\mathcal{G}$. Adversarial training thus

856   consists in playing a minmax game such that $(\mathcal{G}^*, \mathcal{D}^*) = \arg\min_{\mathcal{G}} \max_{\mathcal{D}} \mathscr{L}_{\text{GAN}}(\mathcal{G}, \mathcal{D})$,

857   with an objective function of the form[23]: $\mathcal{L}_{\text{GAN}}(\mathcal{G}, \mathcal{D}) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log \mathcal{D}(x)] +$

858    $\mathbb{E}_{z \sim p_z(z)} \log[1 - \mathcal{D}(\mathcal{G}(z))]$, or equivalently by simultaneous optimization of two

859   coupled loss functions:

$$\begin{cases} \mathcal{D}^* = \arg \max_{\mathcal{D}} \left(\mathbb{E}_{x \sim p_{\text{data}}(x)}[\log \mathcal{D}(x)] + \mathbb{E}_{z \sim p_z(z)} \log[1 - \mathcal{D}(\mathcal{G}(z))] \right) \\ \mathcal{G}^* = \arg \min_{G} \left(\mathbb{E}_{z \sim p_z(z)} \log[1 - \mathcal{D}(\mathcal{G}(z))] \right) \end{cases}$$

860   In a conditional GAN (cGAN), the generator and the discriminator have an extra

861   input vector c and the first objective function above becomes: $\mathcal{L}_{\mathcal{D}}(\mathcal{G}, \mathcal{D}) =$

862    $\mathbb{E}_{(c,x) \sim p_{\text{data}}(c,x)}[\log \mathcal{D}(c, x)] + \mathbb{E}_{c \sim p_{\text{data}}(c), z \sim p_z(z)} \log[1 - \mathcal{D}(c, \mathcal{G}(c, z))]$, such that the

863   generator learns a conditional probability density $p_{\text{data}}(x|c)$; and the second

864   objective function likewise becomes $\mathcal{L}_{\text{cGAN}}(\mathcal{G}, \mathcal{D}) = \mathbb{E}_{c \sim p_{\text{data}}(c), z \sim p_z(z)} \log[1 -$

865    $\mathcal{D}(\mathcal{G}(z))]$. In our A-net, we replaced the logarithmic losses above by least square

866   losses[53], as they empirically yielded better results. Thus, we used the objective

867   functions:

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

869     $\mathcal{L}_{\text{cGAN}}(\mathcal{G}, \mathcal{D}) = \mathbb{E}_{c \sim p_{\text{data}}(c), z \sim p_z(z)}\left[1 - \mathcal{D}\big(c, \mathcal{G}(c, z)\big)\right]^2$. In our case, the input c is the

870   sparse PALM image S combined with the upsampled version $L = \mathcal{B}(W)$ of the

871   widefield image W, where $\mathcal{B}$ denotes bilinear interpolation. Note that in practice the

872   noise z in our ANN was introduced only through the use of dropout layers, as in the

873   pix2pix implementation[21]. Thus, the objective functions are:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{D}) = \mathbb{E}_{S,T,W,M \sim p_{\text{data}}(S,T,W,M)}(\mathcal{D}(\mathcal{B}(W), S, T) - 1)^2$$

$$+ \mathbb{E}_{z \sim p_z(z), S \sim p_{\text{data}}(S)}\left(\mathcal{D}\big(\mathcal{B}(W), S, \mathcal{G}(S, W, M)\big)\right)^2$$

874   and:

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

875   In the end, combining the three loss terms described above, we implemented

876   the following optimization problem :

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

877   The weights α, β and γ are hyperparameters, which we set manually to $\alpha = 50$,

878     $\beta = 25$ and $\gamma = 1$ for most experiments. In absence of widefield images W, β was

879   simply set to zero. The reported results are not very sensitive to these parameters.

880   We trained the A-net end-to-end using stochastic gradient descent (SGD)

881   with Adam[54] and a batch size of 1 with 200,000 or more iterations (backpropagation

882   steps). Our implementation was adapted from affinelayer's TensorFlow[55]

883   implementation, which is ported from the Torch implementation of pix2pix[21]. Both

884   network training and inference were performed on Tesla P100, Tesla M40, Tesla

885   K80 or GTX TitanXP graphical processing units (GPUs) from Nvidia. A-net training

886   from scratch typically takes from hours to days on a single GPU. Once trained, the A-

887   net takes only ~1 second or less to reconstruct a super-resolution image of

888   2560x2560 pixels (corresponding to an entire FoV). Training time could be further

889   reduced by pretraining (or transfer learning), use of GPU clusters, or optimized data

890   augmentation.

891

892   **Experimental training images and data augmentation**

893   Experimental training data are obtained from standard localization microscopy data

894   (dense PALM images). To achieve good performance, ANNs generally necessitate

895   large amounts of training data. However, ANNA-PALM typically requires PALM

896   images from no more than 10 FoVs (of 55 μm x 55 μm each) and can even be trained

897   with a single FoV. This is possible thanks to an extensive on-the-fly data

898   augmentation strategy, as described below. Each of the dense PALM training images

899   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

900   the diffraction limited frame from which localization x_i, y_i originates, and K is the

901   total frame number. PALM images are obtained as plain 2D histograms of these

902   localizations with typical pixel sizes of 10-20 nm. From each list of localizations

903   (corresponding to a dense PALM image of a single FoV), we generate 10-30 pairs of

904   input and target images (S,T) for training. To define the target image T, we take a

905   random consecutive subset $[k_T, k_T + 0.95K]$ of 95% of all available K frames (k_T is

906   chosen randomly between 0 and $0.05K$) and create the 2D histogram image based

907   on localizations from those frames only, i.e. from all (x_j, y_j) such that $f_j \in$

908　$[k_T, k_T + 0.95K]$. To define the sparse input image S, we take random subsets of

909　300-500 consecutive frames from the first half of the image sequence and similarly

910　create a 2D histogram of the localizations from those frames only. When a widefield

911　image W is available, this image must first be aligned with the corresponding dense

912　PALM image T. This is done using an FFT-based phase correlation algorithm[56] after

913　histogram equalization of image T, smoothing by convolution with a Gaussian

914　Kernel of standard deviation 6 pixels, and resizing with bilinear interpolation to the

915　same size as image W. The registered widefield image W is then scaled up using

916　bilinear interpolation to an image $L = \mathcal{B}(W)$ with the same size as image T.

917　　　　During training, for each iteration of SGD, we crop the images S, L and T with

918　a randomly placed 712x712 pixel sized region $\mathcal{R} = [x_{min}, x_{min} + 712] \times$

919　$[y_{min}, y_{min} + 712]$. We then use random geometric transformations and apply them

920　identically to the three images. Specifically, we rotate the images by a random angle

921　between 0 and 360 degrees, apply elastic transformations [57], and then crop the

922　center region of size 512x512 pixels. In addition to geometric transformations, we

923　also introduce realistic noise from experimental background images. This is done by

924　manually outlining regions of background in selected PALM training images,

925　splitting these regions into small patches of 40x40 pixels, grouping them according

926　to their summed pixel values, then assembling them into a larger image with the

927　same size as T. During training, these semi-synthetic noise images are randomly

928　selected and added to the input image S without altering T or W. Finally, we

929　normalize the input image S by subtracting its mean and dividing by standard

930　deviation. If a widefield image W is provided, its pixel values are scaled to a

931   minimum of 0 and a maximum of 1. Otherwise, W is replaced by an image containing

932   zeros only. The target image T is truncated at a maximum value of 255 and then

933   scaled to have a minimum of 0 and a maximum of 1. For the switch M, we used an

934   integer number to define the type of training images, e.g. 0 for microtubules, 1 for

935   nucleoporins, 2 for mitochondria. When training on different types of images, e.g.

936   microtubules and nucleoporins (**Supplementary Figure 13**), we assign the

937   corresponding switch value to M and use it as additional input to the A-net together

938   with images S and W, as described in **Supplementary Note 1.**

939

940   **Image simulations**

941   Our procedure to simulate localization microscopy (PALM) images of microtubules

942   is illustrated in **Supplementary Figure 1**. To simulate microtubule filaments, we

943   used a Langevin dynamics simulation[29] that generates random configurations of

944   semiflexible curves with a specified rigidity (persistence length), starting from a

945   random initial configuration (**Supplementary Figure 1a,b**). The initial

946   configurations were generated with a Python library named cpolymer and the

947   Langevin dynamics was implemented using the molecular dynamics code

948   LAMMPS[58]. Although the simulation generates 3D polymer chains, we only

949   considered their 2D projections, consisting of N_p connected positions $(x_k, y_k)_{k=1..N_p}$.

950   To obtain smooth filaments we further interpolated these connected segments using

951   spline functions with the Scipy function scipy.interpolate.splev. Next, we turned

952   these 2D curves into a grey scale image of 800x800 pixels, with an assumed pixel

953   size of 7.8 nm, using the Python library Matplotlib. This image was further

954   convolved with a Gaussian kernel of standard deviation 1.5 pixels, resulting in a

955   smooth image $I(i,j)$ as shown in **Supplementary Figure 1c,** and normalized to a

956   probability density ($\sum\sum_{i,j} I(i,j) = 1$, with all $I(i,j) \geq 0$). This image was used to

957   mimic a "perfect" PALM image of filaments corresponding to an infinite number of

958   localizations ($n = \infty$). Such perfect images were used as targets during ANN

959   training for simulated data and defined as ground truth for the quantification of

960   reconstruction quality by MS-SSIM (**Figure 2g**). During training, we applied the

961   same rotations and elastic transformations described for experimental data in the

962   previous section.

963   Localization microscopy images obtained from a finite number of

964   localizations $n < \infty$ (sparse PALM images $\boldsymbol{S}$), can be considered as a sampling of the

965   probability density $I(i,j)$ with n samples. These images can therefore be simulated

966   by applying Poisson noise to a rescaled version of the perfect PALM image, i.e.:

967     $\boldsymbol{S}(\lambda, I) = \mathcal{P}(\lambda I/I_{max})$, where I_{max} is the maximum value of I, $\mathcal{P}(\mu)$ denotes the

968   Poisson probability distribution of mean μ and where the peak parameter λ controls

969   the level of sampling. In order to simulate sparse PALM images for various levels of

970   sampling, we varied the peak value λ following a log-normal distribution where

971     $\ln(\lambda)$ has mean -0.5 and standard deviation 0.001 and applied Poisson noise using

972   the numpy library function random.poisson. An example of a simulated sparse

973   PALM image is shown in **Supplementary Figure 1d**. Besides finite sampling,

974   localization microscopy images are corrupted by additional noise sources such as

975   false detections from background noise due to out-of-focus light or unspecific

976   binding of antibodies. To mimic this, we first created a probability density

977 $I_b = I * G_{\sigma_b}$ for the background noise by convolving I with a Gaussian kernel G_{σ_b} of

978 large standard deviation σ_b=25 pixels, and applied Poisson noise with $\lambda = 0.06$. To

979 create training images, we added this background noise image to the sparse PALM

980 image S above (**Supplementary Figure 2**). We did not add background noise to the

981 test images used during inference (**Figure 2**). To simulate the widefield images

982 (**Figure 2a**), we first blurred the perfect PALM image by convolution with a

983 Gaussian kernel of standard deviation 8 pixels, then added Gaussian noise with zero

984 mean and standard deviation chosen randomly between 0.5 and 1.5.

985   For simulations of nuclear pore images (**Supplementary Figure 15**), we

986 applied a similar procedure, except that the perfect PALM images were obtained by

987 randomly distributing circles of diameter 150 nm in the plane (avoiding overlaps)

988 and placing eight Gaussian spots (of standard deviation 1.7 pixels) at equal distance

989 from each other on each circle to mimic the octogonal shape of nuclear pores.

990

991 **Sample preparation**

992 For microtubule imaging experiments (**Figures 3-5** and **Supplementary Figures 8-**

993 **10**), except those using DNA-PAINT (**Figure 5j,l**), U-373 MG (Uppsala) cells were

994 cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12;

995 Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v)

996 penicillin-streptomycin (Gibco), in a 5% CO_2 environment at 37°C on 18-mm

997 cleaned coverslips in 12-well plates. 24 hours after plating, cells were pre-extracted

998 for 10 s in 0.25% (v/v) Triton X-100 (Triton) in BRB80 (80 mM PIPES, 1 mM MgCl2,

999 1 mM EGTA, adjusted to pH 6.8 with KOH) supplemented with 4 mM EGTA, and

1000 immediately fixed for 10 min with 0.25% (v/v) Triton + 0.5% Glutaraldehyde in

1001 BRB80, followed by reduction for 7 min with 0.1% NaBH4 solution in PBS and

1002 another washing step in PBS. Cells were directly incubated for 1h at room

1003 temperature in PBS with 1:500 rat alpha-tubulin antibodies (Bio-Rad MCA77G),

1004 followed by 3 washing steps with PBS, and then incubated for 45min in PBS with

1005 1:500 anti-rat Alexa-647 conjugated secondary antibodies from donkey (Jackson

1006 ImmunoResearch Laboratories, ref. 712-605-153).

1007 For the DNA-PAINT experiment on microtubules (**Figure 5k-m**), U-373 cells

1008 stuck on 18 mm diameter coverslips were fixed at 37°C with 4% PFA in PHEM

1009 buffer and permeabilized in 0.2% glutaldehyde. Next, cells were incubated for 1 h

1010 with 1:500 primary mouse antibodies against alpha-tubulin. The sample was

1011 washed 3 times in PBS, then incubated with 1:100 anti-mouse oligo-conjugated

1012 antibodies from Ultivue Kit 2 for DNA-PAINT imaging[33]. After washing the sample 3

1013 times in PBS, and just before imaging, 2nM of complementary oligos coupled to Cy3

1014 fluorophores were added to the sample.

1015 Nuclear pore imaging data of gp210 and WGA (**Figure 6a-c** and

1016 **Supplementary Figure 14**) were kindly provided by J. Sellés and O. Falklaris and

1017 obtained from nuclear membranes of *Xenopus* frog eggs prepared as described

1018 previously[36].

1019 For mitochondria imaging experiments (**Figure 6e-f**), COS7 cells were

1020 cultured under the same conditions as U-373 cells above using phenol-red free

1021 DMEM medium and fixed with 4% PFA in PBS for 10 min. The sample was blocked

1022 with 3% BSA in PBS for 20 min and immunostained with 1:500 mouse antibodies

1023   against TOM22 (Sigma, ref. T6319) in wash buffer (PBS with 0.5% BSA) for 1 h.

1024   After extensive washing with wash buffer, the sample was incubated with 1:500

1025   anti-mouse secondary antibodies from donkey conjugated to Alexa-647 dyes

1026   (Jackson ImmunoResearch Laboratories, ref. 715-605-151) in wash buffer for 30

1027   min. After washing 5 times with wash buffer and 2 times with PBS, samples were

1028   post-fixed with 2% PFA in PBS for 10 min and washed 5 times with PBS.

1029   For all localization microscopy experiments except DNA-PAINT, we used a

1030   photoswitching buffer[59] composed of 50 mM Tris-HCl + 10 mM NaCl + 10% (w/v)

1031   glucose + 168 AU/mL Glucose-Oxidase + 1404 AU/mL Catalase + 1% 2-

1032   Mercaptoethanol. For microtubule imaging experiments, we used this buffer to fill a

1033   square hole that was manually cut in a parafilm sheet, which was deposited on a

1034   rectangular coverglass. The round coverslips were sealed with nail polish.

1035

1036   **Image acquisition in localization and high-throughput microscopy**

1037   We performed single molecule localization microscopy experiments (PALM/STORM

1038   and DNA-PAINT) on custom built microscopy systems, as previously described[59–61].

1039   The system used for PALM/STORM imaging of microtubules is based on an inverted

1040   microscope body (Nikon Ti Eclipse) equipped with a either a 60x 1.49 NA oil

1041   immersion objective (Nikon) or a 60x 1.2 NA water immersion objective (Nikon)

1042   and with the Perfect Focus System active. A 642 nm wavelength laser with 500 mW

1043   power was used to excite Alexa-647 fluorophores and an AOTF (AA optics) was used

1044   to modulate laser excitation. Sequences of diffraction limited single molecule image

1045   frames were acquired either on a sCMOS camera (Hamamatsu ORCA-Flash4.0),

1046   which can capture images of 2,042x2,042 pixels (for **Figures 4, 5a-i**), or on an

1047   EMCCD (Andor IXON ULTRA 897) with 512x512 pixels (for **Figures 3, 6d-f**). Both

1048   cameras were controlled by MicroManager software[62]. For experiments using the

1049   sCMOS camera, the effective pixel size was 108 nm and we used a 512x512 region of

1050   interest, which resulted in an imaged FoV of 55.3 μm x 55.3 μm. For experiments

1051   using the EMCCD camera, we used a 2x telescope and the effective pixel size was

1052   107 nm, resulting in a FoV of 54.8 μm x 54.8 μm. The exposure time was set to Δt

1053   =10 ms or 30 ms per frame. The number of frames acquired ranged from k=1,000

1054   (**Figure 4**) to K=60,000 (e.g. **Figure 5c**) per FoV.

1055   For the DNA-PAINT experiment (**Figure 5j,l**), we used an inverted Nikon Ti-E

1056   Eclipse microscope equipped with a 100x 1.49 NA TIRF objective and with the

1057   Perfect Focus System active. A 561 nm wavelength laser with 500 mW power was

1058   used to excite Cy3 dyes. Highly inclined laser illumination was used to reduce out-

1059   of-focus background signal. Images were acquired on an EMCCD camera as above,

1060   with a 1.5 x telescope, resulting in an effective pixel size of 106 nm and a FoV of

1061   54 μm x 54 μm. The sample was mounted in a magnetic sample holder filled with

1062   the imaging buffer provided with the Ultivue kit. Exposure time was set to Δt =30 ms

1063   and the EM gain of the EMCCD was set to 300. The laser power was increased until

1064   isolated fluorescent spots were observed. For the experiment shown in **Figure 5g,i**,

1065 K=60,000 frames were acquired.

1066   The *Xenopus* nuclear pore data (**Figure 6a-c** and **Supplementary**

1067   **Figure 14**), were acquired on a Zeiss Elyra P.S.1 microscope as described

1068   previously[36].

1069   For high-throughput imaging of microtubules (**Figure 4**), we used the Multi-

1070   Dimensional Acquisition tool in Micro-manager to define the positions of 1,089 FoVs

1071   of 55.3 µm x 55.3 µm on a 33x33 grid, with overlaps of 1 µm; the stage was

1072   automatically shifted to each of these 1,089 positions. We first acquired only

1073   widefield images, taking five frames at each of these positions (the first two were

1074   ignored because of motion blur), in a total acquisition time of 12 minutes. Then, the

1075   laser power was raised to bleach out preactivated molecules and k=1,000 frames of

1076   single molecule images were acquired at each of the 1,089 positions, in a total

1077   acquisition time of 3 hours and 8 minutes. Raw image frames were written directly

1078   to a remote storage server via Samba networking protocol.

1079

1080   **Localization microscopy image analysis**

1081   The input to ANNA-PALM reconstruction is a localization image, defined as a 2D

1082   histogram of n single molecule positions $(x_i, y_i)_{i=1..n}$. The histogram bin, i.e. the

1083   pixel size of the localization image, was set to 7.8 nm for the simulated data

1084   (**Figure 2** and **Supplementary Figures 1-6,15**) and 20 nm for the experimental

1085   data (**Figures 3-6** and **Supplementary Figures 7-14**). The positions $(x_i, y_i)_{i=1..n}$

1086   were obtained by analyzing sequences of diffraction limited frames using standard

1087   single molecule localization algorithms. For experimental microtubule images, we

1088   used the ThunderSTORM[63] plugin of ImageJ, applying wavelet filters for detection

1089   and weighted least squares Gaussian fitting for precise estimation of subpixelic

1090   positions. We used the cross-correlation feature in ThunderSTORM for drift

1091   correction, and filtered out the least certain localizations based on the fitted

1092   Gaussian's standard deviation and the χ^2 of the residual. Localizations in

1093   consecutive frames separated by less than 20 nm were assumed to originate from

1094   the same molecule and merged into a single localization. The final number n of

1095   localizations was ~7 million for the full 55 µm x 55 µm FoV of the images shown in

1096   **Figure 3** and **Supplementary Figures 8,9** (obtained from K=30,000 frames). For

1097   the high-throughput experiment (**Figure 4** and **Supplementary Figure 12**), the

1098   number of localizations per 55 µm x 55 µm FoV ranged from n=2,949 to

1099 n=1,442,048 with an average $\langle n \rangle = 610,983$ and standard deviation $\sigma(n)$ =273,606.

1100   The total number of localizations across all 1,089 FoVs was ≈ 665 million.

1101   ThunderSTORM analyses were performed either on high end workstations or on

1102   Institut Pasteur's high performance computer (HPC) cluster. For the high-

1103   throughput experiments, we used Python scripts to run ThunderSTORM in batch

1104   mode (without user intervention) on the HPC cluster and assembled mosaic images

1105   (**Figure 4a,b** and **Supplementary Figure 12a,b**) using a stitching plugin of

1106   ImageJ[64].

1107   The nuclear pore images were analyzed using the ZEN software from Zeiss as

1108   previously described[36]. For the DNA-PAINT experiments, we used PALMTT, a

1109   modified version of the single molecule tracking algorithm MTT[65], based on Matlab

1110   (Mathworks). This algorithm uses Gaussian smoothing and thresholding for

1111   detection, and Gaussian fitting for precise estimation of subpixelic positions. Drift

1112   correction was performed computationally by tracking fluorescent beads used as

1113   fiducial markers.

1114

Quality metrics and sampling resolution

1115

1116   In order to quantitatively assess the quality of PALM images and ANNA-PALM

1117   reconstructions, we calculated the multi-scale structural similarity index (MS-SSIM)

1118   between either image and the ground truth (**Figure 2g** and **Supplementary**

1119   **Figures 5,9**). For the simulated data, the ground truth was simply defined as the

1120   "perfect" PALM image, corresponding to an infinite number of localizations (see

1121   'Image simulations' above, **Figure 2e** and **Supplementary Figure 1c**). For the

1122   experimental data, the ground truth was defined as the ANNA-PALM reconstruction

1123   of a dense PALM image obtained from all available frames (e.g. **Supplementary**

1124   **Figure 9c**). Before calculation of the MS-SSIM, all simulated images were linearly

1125   normalized without clipping to a maximum value of 255.

1126   In order to evaluate the effect of sampling on the resolution of PALM images,

1127   we computed the double mean distance, $R_{\text{Nyq}} = 2\langle d \rangle$ between nearest neighbors in

1128   the underlying sets of localizations[49] as function of localization number

1129   (**Supplementary Figure 3d**). For the simulated sparse PALM data, sets of

1130   localizations were obtained by interpreting each image S as a 2D histogram of

1131   localizations, and creating a random subpixelic position $(x_k, y_k)_{k=1..m}$ within each

1132   pixel (i,j), as many times as given by the pixel value $m = S(i,j)$ (therefore resulting

1133   in a set of $n = \sum_{i,j} S(i,j)$ localizations). The quantity $R_{\text{Nyq}}(n) = 2\langle d(n) \rangle$ decreases

1134   towards zero with increasing number of localizations, n. A Nyquist criterion

1135   introduced in ref.[49] suggests that the resolution is limited by sampling to no less

1136   than R_{Nyq}, implying that at least $n \geq N_{\text{Nyq}} = R_{\text{Nyq}}^{-1}(R)$ localizations are needed to

1137   achieve a given resolution R. However, a more stringent and realistic criterion[3]

1138   prescribes a five-fold larger number of localizations to reach resolution R:

1139 $n \geq N_{5\times\text{Nyq}} = 5 \times N_{\text{Nyq}}$, implying that the sampling limit to resolution is

1140 $R_{5\times\text{Nyq}} = 2\langle d(n/5) \rangle$. Accordingly, if this condition is met, the resolution is no longer

1141   limited by sampling, but by the localization precision, $R_{\text{loc}} \approx 2.3\sigma_{\text{loc}}$, where σ_{loc} is

1142   the standard deviation of localization errors along each coordinate. In general, the

1143   resolution of a PALM image, as limited by both sampling and localization precision,

1144   can be written: $R = \max\left(R_{5\times\text{Nyq}}, R_{\text{loc}}\right) = \max(2\langle d(n/5)\rangle, 2.3\sigma_{\text{loc}})$.

1145

1146   **Life Sciences Reporting Summary.**

1147   Further information is available in the Life Sciences Reporting Summary.

1148

1149   **Code availability statement**

1150   The source code of ANNA-PALM is available from https://annapalm.pasteur.fr/ .

1151

1152   **Data availability statement**

1153   The localization data used in this paper can be downloaded directly from

1154   https://annapalm.pasteur.fr/.

1155

1156   **Methods-only references**

1157   51.   Xu, L., Choy, C. S. & Li, Y. W. Deep sparse rectifier neural networks for speech
1158   denoising. in *2016 International Workshop on Acoustic Signal Enhancement,*
1159   *IWAENC 2016* (2016). doi:10.1109/IWAENC.2016.7602891
1160   52.   Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier Nonlinearities Improve Neural
1161   Network Acoustic Models. *Proc. 30 th Int. Conf. Mach. Learn.* (2013).
1162   53.   Mao, X. *et al.* Least Squares Generative Adversarial Networks. in *ICCV* 2794–
1163   2802 (2017). doi:10.1109/ICCV.2017.304

1164 54. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ICLR* 1–15
1165 (2015).
1166 doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503
1167 55. Rampasek, L. & Goldenberg, A. TensorFlow: Biology's Gateway to Deep
1168 Learning? *Cell Syst.* **2,** 12–14 (2016).
1169 56. Srinivasa Reddy, B. & Chatterji, B. N. An FFT-based technique for translation,
1170 rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5,**
1171 1266–1271 (1996).
1172 57. Simard, P. Y., Steinkraus, D. & Platt, J. C. Best practices for convolutional neural
1173 networks applied to visual document analysis. *Seventh Int. Conf. Doc. Anal.*
1174 *Recognition, 2003. Proceedings.* **1,** 958–963 (2003).
1175 58. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J.*
1176 *Comput. Phys.* **117,** 1–19 (1995).
1177 59. van de Linde, S. *et al.* Direct stochastic optical reconstruction microscopy with
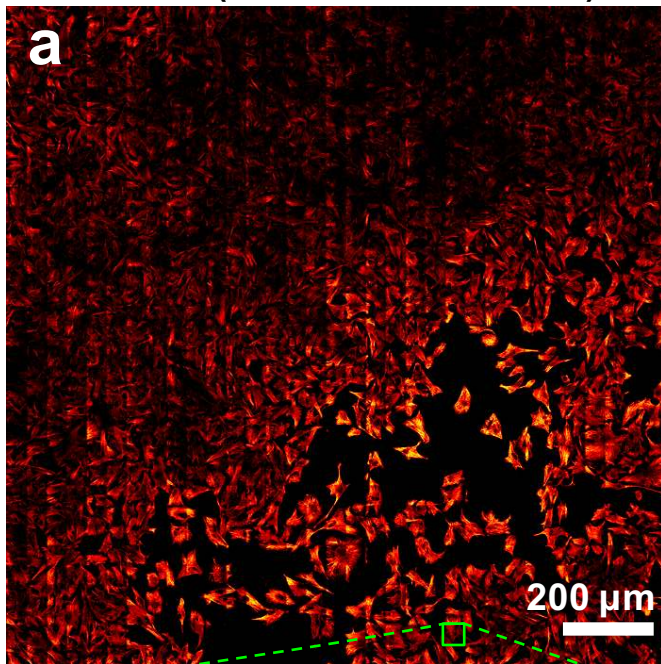1178 standard fluorescent probes. *Nat. Protoc.* **6,** 991–1009 (2011).
1179 60. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy
1180 image processing in ImageJ. *Nat. Methods* **7,** 339–340 (2010).
1181 61. Lelek, M. *et al.* Superresolution imaging of HIV in infected cells with FlAsH-
1182 PALM. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8564–9 (2012).
1183 62. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control
1184 of microscopes using manager. *Current Protocols in Molecular Biology* (2010).
1185 doi:10.1002/0471142727.mb1420s92
1186 63. Ovesny, M., Kiek, P., Borkovec, J., Vindrych, Z. & Hagen, G. M. ThunderSTORM:
1187 a comprehensive ImageJ plug-in for PALM and STORM data analysis and
1188 super-resolution imaging. *Bioinformatics* **30,** 2389–2390 (2014).
1189 64. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D
1190 microscopic image acquisitions. *Bioinformatics* **25,** 1463–1465 (2009).
1191 65. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target
1192 tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods*
1193 **5,** 687–694 (2008).
1194
1195

a

widefield

0   kΔt   KΔt   time

k<<K

localization   localization

W

L

PALM(k)

Pk

+

PALM(K)

PK

NPC, Mito,......MT

𝒢

ANNA-PALM(k)

Ak

Super-res reconstruction error

Ak   PK

L1 + MS-SSIM

Low-res reconstruction error

Ak

PK

Q

WA

WP

MS-SSIM

MS-SSIM

W

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

k<<K

W'

localization

L'

PALM(k)

P'k

+

NPC, Mito,......MT

𝒢

ANNA-PALM(k)

A'k

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

PALM
ANNA-PALM (sparse+widefield)
ANNA-PALM (sparse)
ANNA-PALM (widefield)

Reconstruction quality (MS-SSIM)

Localization counts, n

a Widefield

b PALM (k=300)
n=11,740

c PALM (K=30,000)
n=409,364

1 µm

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**

V$_{min}$ V$_{max}$

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

200 μm

b

200 μm

c

10 μm

1 μm

d

10 μm

50.7nm

0 50 100 150 200 250

1 μm

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) **b** ANNA-PALM (k=3,000) **c** PALM (K=30,000)

500 nm

d PALM (k=400) **e** ANNA-PALM (k=400) **f** PALM (K=30,000)

5 μm