

# Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes

Jiji Chen (✉ [jiji.chen@nih.gov](mailto:jiji.chen@nih.gov))

Advanced Imaging and Microscopy Resource, National Institutes of Health, Bethesda, Maryland, USA  
<https://orcid.org/0000-0002-4426-3035>

**Hideki Sasaki**

DRVISION Technologies, LLC

**Hoyin Lai**

DRVISION Technologies, LLC

**Yijun Su**

National Institutes of Health

**Jiamin Liu**

Advanced Imaging and Microscopy Resource, National Institutes of Health, Bethesda, Maryland, USA

**Yicong Wu**

National Institutes of Health

**Alexander Zhovmer**

Laboratory of Molecular Cardiology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

**Christian Combs**

nih

**Ivan Rey-Suare**

Biophysics Program, University of Maryland

**Hung-Yu Chang**

DRVISION Technologies LLC <https://orcid.org/0000-0002-7666-4511>

**Xuesong Li**

Section on High Resolution Optical Imaging, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD, USA

**Min Guo**

National Institutes of Health <https://orcid.org/0000-0002-2093-8771>

**Srineil Nizambad**

National Institutes of Health

**Arpita Upadhyaya**

University of Maryland, Baltimore

**Shih-Jong Lee**

DRVISION Technologies, LLC

**Luciano Lucas**

DRVISION Technologies, LLC

**Hari Shroff**

Section on High Resolution Optical Imaging, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD, USA

---

**Article****Keywords:**

**Posted Date:** September 10th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-68002/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Methods on May 31st, 2021. See the published version at <https://doi.org/10.1038/s41592-021-01155-x>.

1 **Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy**  
2 **image volumes**

3 Jiji Chen<sup>1\*^</sup>, Hideki Sasaki<sup>2\*^</sup>, Hoyin Lai<sup>2\*</sup>, Yijun Su<sup>1, 3\*</sup>, Jiamin Liu<sup>1</sup>, Yicong Wu<sup>3</sup>, Alexander Zhovmer<sup>4</sup>,  
4 Christian A. Combs<sup>5</sup>, Ivan Rey-Suarez<sup>6,8</sup>, Hungyu Chang<sup>2</sup>, Chi Chou Huang<sup>2</sup>, Xuesong Li<sup>3</sup>, Min Guo<sup>3</sup>, Srineil  
5 Nizambad<sup>1</sup>, Arpita Upadhyaya<sup>6, 7, 8</sup>, Shih-Jong J. Lee<sup>2</sup>, Luciano A.G. Lucas<sup>2§</sup>, Hari Shroff<sup>1,3§</sup>

6 \* Equal contribution

7 § Equal contribution

8 Correspondence to [jiji.chen@nih.gov](mailto:jiji.chen@nih.gov), [sasakih@drvtechnologies.com](mailto:sasakih@drvtechnologies.com)

9

- 10 1. Advanced Imaging and Microscopy Resource, National Institutes of Health, Bethesda, Maryland,  
11 USA  
12 2. DRVISION Technologies, LLC, Bellevue, WA  
13 3. Laboratory of High Resolution Optical Imaging, National Institute of Biomedical Imaging and  
14 Bioengineering, National Institutes of Health, Bethesda, Maryland, USA  
15 4. Laboratory of Molecular Cardiology, National Heart, Lung, and Blood Institute, National  
16 Institutes of Health, Bethesda, MD 20892, USA  
17 5. NHLBI Light Microscopy Facility, National Institutes of Health, Bethesda, Maryland  
18 6. Biophysics Program, University of Maryland, College Park, MD 20742, USA  
19 7. Department of Physics, University of Maryland, College Park, MD 20742, USA  
20 8. Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742,  
21 USA

22

23 **Abstract**

24 We demonstrate residual channel attention networks (RCAN) for restoring and enhancing volumetric  
25 time-lapse (4D) fluorescence microscopy data. First, we modify RCAN to handle image volumes, showing  
26 that our network enables denoising competitive with three other state-of-the-art neural networks. We  
27 use RCAN to restore noisy 4D super-resolution data, enabling image capture over tens of thousands of  
28 images (thousands of volumes) without apparent photobleaching. Second, using simulations we show  
29 that RCAN enables class-leading resolution enhancement, superior to other networks. Third, we exploit  
30 RCAN for denoising and resolution improvement in confocal microscopy, enabling ~2.5-fold lateral  
31 resolution enhancement using stimulated emission depletion (STED) microscopy ground truth. Fourth,  
32 we develop methods to improve spatial resolution in structured illumination microscopy using  
33 expansion microscopy ground truth, achieving improvements of ~1.4-fold laterally and ~3.4-fold axially.  
34 Finally, we characterize the limits of denoising and resolution enhancement, suggesting practical  
35 benchmarks for evaluating and further enhancing network performance.

36 **Introduction**

37 All fluorescence microscopes suffer drawbacks and tradeoffs because they partition a finite  
38 signal budget in space and time. These limitations manifest when comparing different microscope types  
39 (e.g., three-dimensional structured illumination microscopy<sup>1</sup> (SIM) offers better spatial resolution than

40 high numerical aperture light sheet microscopy<sup>2</sup>, but worse photobleaching); different implementations  
41 of the same microscope type (e.g., traditional implementations of SIM offer better spatial resolution  
42 than instant SIM (iSIM)<sup>3</sup>, but worse depth penetration and lower speed<sup>4</sup>); and within the same  
43 microscope (longer exposures and bigger pixels increase signal-to-noise ratio (SNR) at the expense of  
44 speed and resolution<sup>5</sup>). Performance tradeoffs are especially severe<sup>6</sup> when considering live-cell super-  
45 resolution microscopy applications, in which the desired spatiotemporal resolution must be balanced  
46 against sample health<sup>7</sup>.

47 Deep learning<sup>8</sup>, which harnesses neural networks for data-driven statistical inference, has  
48 emerged as a promising method for alleviating drawbacks in fluorescence microscopy. Content-aware  
49 image restoration (CARE<sup>9</sup>) networks use the popular U-net<sup>10</sup> neural network architecture in conjunction  
50 with synthetic, semi-synthetic and physically acquired training data to improve resolution, resolution  
51 isotropy, and signal-to-noise ratio in fluorescence images. U-nets have also been incorporated into  
52 generative adversarial networks (GAN<sup>11</sup>) that enable cross-modality super-resolution microscopy,  
53 transforming confocal images into STED images<sup>12</sup> or transforming a series of widefield or sparse  
54 localization microscopy images into high resolution localization microscopy images<sup>13</sup>. Other recent  
55 examples include denoising confocal<sup>14</sup> or SIM<sup>15</sup> data and deconvolving light-sheet data<sup>16</sup>.

56 Here we investigate the use of an alternative network architecture, the residual channel  
57 attention network (RCAN)<sup>17</sup>, for use in super-resolution microscopy applications. RCAN has been shown  
58 to preferentially learn high spatial frequency detail within natural scene images, but this capability has  
59 not been exploited for image restoration in fluorescence microscopy applications, nor on longitudinally  
60 acquired image volumes. First, we modify RCAN for 3D applications, showing that it matches or exceeds  
61 the performance of previous networks in denoising fluorescence microscopy data. We apply this  
62 capability for super-resolution imaging over thousands of image volumes (tens of thousands of images).  
63 Second, we characterize RCAN and other networks in terms of their ability to extend resolution, finding  
64 that RCAN provides better resolution enhancement than alternatives, especially along the axial  
65 dimension. Finally, we demonstrate 4-5 fold volumetric resolution improvement in multiple fixed- and  
66 live-cell samples when using stimulated emission depletion (STED)- and expansion<sup>18</sup>- microscopy ground  
67 truth to train RCAN models.

## 68 **Results**

### 69 *RCAN enables super-resolution imaging over thousands of volumes*

70 The original RCAN was proposed specifically for resolution enhancement<sup>17</sup>. A key challenge in  
71 this task is the need to bypass abundant low-resolution information in the input image in favor of high-  
72 resolution prediction. The RCAN architecture achieves this by employing multiple skip connections  
73 between network layers to bypass low-resolution content, as well as a ‘channel-attention’ mechanism<sup>19</sup>  
74 that emphasizes the more relevant feature channels, preventing low resolution features from  
75 dominating the prediction. We modified the original RCAN architecture to handle image volumes rather  
76 than images, also improving network efficiency so that our modified 3D RCAN model fits within graphics  
77 processing unit (GPU) memory (**Fig. 1a, Methods, Supplementary Note 1**).

78 To investigate RCAN denoising performance on fluorescence data, we began by acquiring  
79 matched pairs of low- and high- SNR iSIM volumes of fixed U2OS cells transfected with mEmerald-  
80 Tomm20 (**Methods, Supplementary Table 1, 2**), labeling the outer mitochondrial membrane (**Fig. 1b**).  
81 We programmed our acousto-optic tunable filter to rapidly switch between low (4.2 W/cm<sup>2</sup>) and high  
82 (457 W/cm<sup>2</sup>) intensity illumination, rapidly acquiring 35 low SNR raw volumes and matching high SNR

83 data, which we deconvolved to yield high SNR ‘ground truth’. We then used 30 of these volumes for  
84 training and held out 5 volumes for testing network performance. Using the same training and test data,  
85 we compared four networks: RCAN, CARE, SRResNET<sup>20</sup>, and ESRGAN<sup>21</sup>. SRResNet and ESRGAN are both  
86 class-leading deep residual networks used in image super-resolution, with ESRGAN winning the 2018  
87 Perceptual Image Restoration and Manipulation challenge on perceptual image super-resolution<sup>22</sup>.

88 For the mEmerald-Tomm20 label, RCAN, CARE, ESRGAN, and SRResNET predictions all provided  
89 clear improvements in visual appearance, structural similarity index (SSIM) and peak signal-to-noise-  
90 ratio (PSNR) metrics relative to the raw input (**Fig. 1b**), also outperforming direct deconvolution on the  
91 noisy input data (**Supplementary Fig. 1**). The RCAN output provided PSNR and SSIM values competitive  
92 with the other networks (**Fig. 1b**), prompting us to investigate whether this performance held for other  
93 organelles. We thus conducted similar experiments for fixed U2OS cells with labeled actin, endoplasmic  
94 reticulum (ER), golgi, lysosomes, and microtubules (**Supplementary Fig. 2**), acquiring 15-23 volumes of  
95 training data and training independent networks for each organelle. In almost all cases, RCAN  
96 performance met or exceeded the other networks (**Supplementary Fig. 3, Supplementary Table 3**).

97 An essential consideration when using any deep learning method is understanding when  
98 network performance deteriorates. Independently training an ensemble of networks and computing  
99 measures of network disagreement can provide insight into this issue<sup>9,16</sup>, yet such measures were not  
100 generally predictive of disagreement between ground truth and RCAN output (**Supplementary Fig. 4**).  
101 Instead, we found that estimating the per-pixel SNR in the raw input (**Methods, Supplementary Fig. 4**)  
102 seemed to better correlate with network performance, with extremely noisy input generating a poor  
103 prediction, as intuitively expected. For example, for the mEmerald-Tomm20 and ERmoxGFP labels, we  
104 observed obvious artifacts when input SNR dropped below  $\sim 3$  (**Fig. 1c**). We observed similar effects  
105 when using synthetic spherical phantoms in the presence of large noise levels (**Supplementary Fig. 5**).

106 We also examined linearity and spatial resolution in the denoised RCAN predictions. We verified  
107 that the RCAN output reflected spatial variations in fluorescence intensity evident in the input data,  
108 demonstrating that linearity is preserved (**Supplementary Fig. 6**). To estimate spatial resolution, we  
109 examined the apparent full width at half maximum of 10 labeled microtubule filaments in noisy raw  
110 input; high SNR deconvolved ground truth; and the RCAN prediction (**Fig. 1d**). While lateral resolution  
111 was not recovered to the extent evident in the ground truth (170 +/- 13 nm, mean +/- standard  
112 deviation), predictions offered noticeable resolution improvement compared to the input data (194 +/-  
113 9 nm RCAN vs. 353 +/- 58 nm input).

114 Next, we tested the performance of RCAN on live cells, for extended volumetric time-lapse (4D)  
115 imaging applications. At high SNR, relatively few volumes can be obtained with iSIM, due to significant  
116 volumetric bleaching. For example, when volumetrically imaging pShooter pEF-Myc-mito-GFP (labeling  
117 the mitochondrial matrix) in live U2OS cells every 5.6 s at high intensity (360 W/cm<sup>2</sup>, **Fig. 1e**,  
118 **Supplementary Video 1**), only seven volumes could be acquired before fluorescence dropped to half its  
119 initial value. Lowering the illumination intensity to 4.2 W/cm<sup>2</sup> so that photobleaching is negligible  
120 compared to the rate of protein synthesis circumvents this problem, but the resulting low SNR usually  
121 renders the data unusable (**Fig. 1e**). To determine whether deep learning could help to address this  
122 tradeoff between SNR and imaging duration, we accumulated 36 matched low (4.2 W/cm<sup>2</sup>)/high  
123 intensity (457 W/cm<sup>2</sup>) volumes on fixed cells, and trained an RCAN model, which we then tested on our  
124 low SNR live data. This approach enabled super-resolution imaging over an extended duration, allowing  
125 capture of 2600 image volumes ( $\sim 50,000$  images, 2.2 W/cm<sup>2</sup>) acquired every 5.6 s over four hours with  
126 no detectable photobleaching and an apparent increase in fluorescence signal over the course of the  
127 recording (**Fig. 1e, Supplementary Video 2**). The restored image quality was sufficiently high that  
128 individual mitochondria could be manually segmented, a task difficult or impossible on the raw input  
129 data (**Supplementary Fig. 7**). To our knowledge, light-sheet microscopy is the only technique capable of

130 generating 4D data of similar quality and duration, but the sub-200 nm spatial resolution of our method  
131 is better than that of high-NA light-sheet microscopy<sup>23</sup>. In another application, a dual-color example, we  
132 applied the same strategy to imaging pShooter pEF-Myc-mito-GFP in conjunction with mApple-LAMP1  
133 labeled lysosomes. In this case, we obtained ~300 super-resolution volumes recorded every 5.1 s in a  
134 representative cell (**Supplementary Video 3**), allowing inspection (**Fig 1f**) of mitochondrial fission and  
135 fusion near lysosomal contacts. Manually quantifying these events from 16 cells, we found that fission  
136 occurred ~2.5x as often as fusion (**Fig. 1g**).  
137

### 138 *Estimating the resolution enhancement offered by deep learning*

139 In addition to denoising fluorescence images, deep learning can also be used for resolution  
140 enhancement<sup>9,12,13</sup>. We were curious about the extent to which RCAN (and other networks) could  
141 retrieve resolution degraded by the optical system, since this capability has not been systematically  
142 investigated. We were particularly interested in understanding when network performance breaks  
143 down, i.e., how much blurring is too much. To empirically assess the relative performance of different  
144 networks, we simulated ground truth noiseless spherical phantoms and subjected them to increasing  
145 amounts of blur (**Fig. 2, Supplementary Videos 4-6**). We trained RCAN, CARE, SRResNet, and ESRGAN  
146 networks with the same 23 matched volumes of ground truth and blurred data, and then challenged  
147 each network with 7 volumes of previously unseen test data (**Fig. 2a-c, Supplementary Figure 8**).

148 The RCAN generated plausible reconstructions even with blurring 3-fold greater (in all spatial  
149 dimensions) than the iSIM PSF (**Fig. 2b**), largely preserving the size of the smallest particles (**Fig. 2b,c**).  
150 However, RCAN performance degraded with increasingly blurry input, with SSIM and PSNR decaying  
151 from 0.98 to 0.93 and 38 dB to 32 dB for two- to three-fold blur, with other networks also showing  
152 worsened performance at increasing blur (**Fig. 2d, Supplementary Table 4**). Compared to the other  
153 networks, RCAN predictions offered improved resolution along the axial dimension (**Fig. 2b, c,**  
154 **Supplementary Fig. 8**), and superior SSIM and PSNR (**Fig. 2d, Supplementary Table 4**). We noticed  
155 obvious artifacts in all networks at 4x blur, suggesting an effective limit for deblurring with deep learning  
156 (**Supplementary Fig. 9, Supplementary Video 6**).  
157

### 158 *Using RCAN for confocal to STED resolution enhancement in fixed and live cells*

159  
160 Since the noiseless spherical phantoms suggested that RCAN provides class-leading performance  
161 for resolution enhancement, we sought to benchmark RCAN performance using noisy experimental  
162 data. As a first test, we studied the ability to ‘transform’ confocal volumes into volumes with STED-like  
163 spatial resolution (**Fig. 3**), which is attractive because confocal imaging provides gentler, higher SNR  
164 imaging than STED microscopy but worse spatial resolution. Such ‘cross-modality’ super-resolution has  
165 been demonstrated before with GANs, but only with 2D images obtained from fixed cells<sup>12</sup>.

166 We collected training data (22-26 volumes, **Supplementary Table 2**) on fixed, fluorescently  
167 labeled mouse embryonic fibroblast cells using a commercial Leica SP8 3X STED microscope (**Fig. 3a-c**).  
168 This system was particularly convenient as the STED images could be acquired immediately after the  
169 confocal images, on the same instrument. We imaged fixed mouse embryonic fibroblasts,  
170 immunostained with ATTO647-secondary antibodies against anti- $\alpha$ -tubulin primary antibodies for  
171 marking microtubules (**Fig. 3a**); and with Alexa Fluor 594-secondary antibodies against anti-NPC primary  
172 antibodies, marking nuclear pores (**Fig. 3b**). Next, we trained RCAN models and applied them to unseen  
173 data (**Supplementary Fig. 10**), using a modified decorrelation analysis<sup>24</sup> (**Methods, Supplementary Fig.**  
174 **11**) to estimate average spatial resolution. Confocal spatial resolution was 273 +/- 9 nm ( $N = 18$  images  
175 used for these measurements) in the microtubule dataset and 313 +/- 14 nm in the pore dataset, with

176 STED microscopy providing ~2-fold improvement in resolution (129 +/- 6 nm for microtubules, 144 +/- 9  
177 nm for the pores) and the RCAN prediction providing similar gains (121 +/- 4 nm microtubules, 123 +/-  
178 14 nm nuclear pores, **Fig. 3d**) that could not be matched by deconvolving the confocal data  
179 (**Supplementary Fig. 12**). We suspect that the slight improvement in spatial resolution in RCAN output  
180 relative to the STED ground truth is because the RCAN denoised the data as well as improved resolution,  
181 resulting in higher SNR than the STED ground truth. Close examination of the RCAN prediction for  
182 nuclear pores revealed slight differences in pore placement relative to the STED microscopy ground  
183 truth. We suspect that this result is due to slight differences in image registration between the confocal  
184 and STED data (**Supplementary Fig. 13**), perhaps due to sample drift between acquisitions or slight  
185 instrument misalignment. Applying an affine registration between the confocal and STED training data  
186 improved agreement between the confocal and STED data, improving network output (**Supplementary**  
187 **Fig. 13**). However, small deviations in nuclear pore placement between the ground truth STED and RCAN  
188 predictions were still evident.

189 We also examined a third label, SiR-DNA, a DNA stain well suited for labeling live and fixed cells  
190 in both confocal and STED microscopy<sup>25</sup>. Collecting matched confocal and STED volumes on fixed nuclei  
191 in a variety of mitotic stages enabled us to train a robust RCAN model that produced predictions on  
192 different nuclear morphologies (**Fig. 3c, Supplementary Fig. 14**) that were sharper and less noisy than  
193 confocal input. Improvement relative to the confocal data was particularly striking in the axial dimension  
194 (**Fig. 3c**). Given the quality of these reconstructions, we wondered whether the same RCAN model could  
195 be adapted for transfer learning on live samples.

196 Point-scanning confocal imaging can produce time-lapse volumetric recordings of living cells at  
197 SNR much higher than STED microscopy, given that more signal is collected per pixel. Nevertheless, even  
198 confocal microscopy recordings are quite noisy if high speed acquisitions are acquired. To demonstrate  
199 that our RCAN model trained on fixed cells could simultaneously denoise and improve resolution in live  
200 cells, we acquired noisy resonant confocal recordings of dividing cells labeled with SiR-DNA (**Fig. 3e**). Our  
201 illumination conditions were sufficiently gentle and rapid that we could acquire tens of imaging volumes  
202 without obvious bleaching or motion blur (**Supplementary Video 7**). Although the raw resonant confocal  
203 data poorly defined nuclei and chromosomes, these structures were clearly resolved in the RCAN  
204 predictions (**Fig. 3e, Supplementary Video 7**). The RCAN also better captured chromosome  
205 decondensation and the return to interphase DNA structure (**Fig. 3f**, see also additional interphase cell  
206 comparisons in **Supplementary Video 8**).

207

### 208 *Using expansion microscopy to improve iSIM resolution in fixed and live cells*

209 Our success in using fixed STED training data to improve the spatial resolution of confocal  
210 microscopy made us wonder whether a similar strategy could be used to improve spatial resolution in  
211 iSIM. Since our iSIM did not inherently possess a means to image specimens at higher resolution than  
212 that of the base microscope, we used expansion microscopy (ExM<sup>18</sup>) to provide higher-resolution  
213 training data (**Fig. 4a**). ExM physically expands fixed tissue using a hydrogel and can improve resolution  
214 near-isotropically up to a factor given by the gel expansion. We used ultrastructure expansion  
215 microscopy (U-ExM<sup>26</sup>, a variant of the original ExM protocol) to expand mitochondria (immunolabeled  
216 with Rabbit- $\alpha$ -Tomm20 primary and Donkey- $\alpha$ -Rabbit Biotin secondary antibodies and Alexa Fluor 488  
217 Streptavidin) and microtubules (labeled with Mouse- $\alpha$ -Tubulin primary and Donkey-  $\alpha$ -Mouse Biotin  
218 secondary antibodies and Alexa Fluor 488 Streptavidin) in fixed U2OS cells by 3.2- and 4-fold,  
219 respectively (**Methods, Supplementary Fig. 15**), also developing protocols to locate and image the same  
220 region before- and after ExM with iSIM (**Supplementary Fig. 16, Methods**).

221 We first attempted to directly register pre-ExM iSIM data to post-ExM data to build a training  
222 dataset suitable RCAN. Unfortunately, local distortions in the post-ExM data prevented the subpixel  
223 registration needed for accurate correspondence between pre- and post-ExM data, even when using  
224 landmark-based non-affine based registration methods (**Supplementary Fig. 17**). Instead, we digitally  
225 degraded the post-ExM data so that it resembled the lower resolution, pre-ExM iSIM data (**Fig. 4a**).  
226 Simply blurring the post-ExM data is insufficient, as blurring also oversmooths the background to the  
227 point that the images are noticeably smoother and less noisy than acquired pre-ExM iSIM data  
228 (**Supplementary Fig. 18**). Instead, we developed methods to match noise and background signal so that  
229 the digitally degraded post-ExM iSIM data better resembled deconvolved, pre-ExM iSIM data  
230 (**Supplementary Fig. 19, Methods**). This approach allowed us to register image pairs perfectly and to  
231 train RCAN models for microtubule and mitochondrial labels (**Methods, Supplementary Video 9,**  
232 **Supplementary Fig. 20**).

233 On fixed samples, the trained networks provided modest lateral resolution enhancement on  
234 synthetic data derived from ground truth images of expanded immunostained mitochondria and  
235 microtubules from fixed U2OS cells (**Fig. 4b**), allowing us to occasionally resolve closely spaced filaments  
236 otherwise blurred in the synthetic images (red arrows, **Fig. 4b**). However, the axial resolution  
237 enhancement offered by RCAN was more dramatic, showing clear improvement similar to the ground  
238 truth images. Using decorrelation analysis to estimate the degree of resolution enhancement on the  
239 microtubule data, we found that RCAN offered 1.5-fold increase laterally and 2.8-fold increase axially  
240 relative to the synthetic deconvolved data, compared to 2.2-fold improvement (lateral) and 3.5-fold  
241 (axial) offered by the ground truth data (**Supplementary Fig. 21**). We observed similar enhancements on  
242 experimentally acquired pre-expansion data: 1.4- and 3.4- fold improvement laterally and axially by the  
243 RCAN, versus 2.7-fold and 3.7-fold improvement in the ground truth data (**Fig. 4c**).

244 The improvements in fixed cells prompted us to apply our ExM-trained RCAN models to living  
245 cells imaged with iSIM in volumetric time-lapse sequences (**Fig. 4d, e, Supplementary Videos 10-13**). In  
246 a first example, we applied the RCAN to mitochondria labeled with EGFP-Tomm20 in live U2OS cells (**Fig.**  
247 **4d, Supplementary Video 10**). Modest improvements in lateral resolution and contrast with RCAN  
248 offered better definition of individual mitochondria, including the void regions contained within the  
249 outer-mitochondrial space (**Fig. 4d**, red arrows). As with the fixed cells, improvements in axial views of  
250 the specimen were more dramatic (**Supplementary Video 11**), allowing us to discern closely packed  
251 mitochondria that were otherwise blurred in the deconvolved iSIM data (**Fig. 4d**, yellow arrows).

252 In a second, transfer-learning example, we applied our expansion-RCAN model derived from  
253 immunostained U2OS cells to live Jurkat T cells transiently expressing EMTB-3xGFP<sup>27</sup>, a protein that  
254 labels microtubule filaments. Jurkat T cells settled onto anti-CD3 coated activating coverslips (**Fig. 4e,**  
255 **Supplementary Videos 12-14**), which mimic antigen presenting cells and enable investigation of the  
256 early stages of immune synapse formation<sup>28</sup>. Dynamics and organization of the actin and microtubule  
257 cytoskeleton during cell spreading are important regulators of this phenomenon. The RCAN output  
258 offered clear views of the microtubule cytoskeleton during the initial stages of this dynamic process,  
259 including the deformation of microtubule bundles surrounding the nucleus. We observed pronounced  
260 deformation of the central microtubule bundles at the dorsal cell surface as spreading initiated (blue  
261 arrowheads), suggesting that these bundles may be anchored to the actin cortex. Anchoring of  
262 microtubules to the actin cortex allows the repositioning of the centrosome, a hallmark of immune  
263 synapse maturation<sup>29</sup>. Interestingly, we observed a higher deformation of the microtubule bundles on  
264 the right side of the cell shown in **Fig. 4e**, likely due to the forces that push and pull the centrosome  
265 towards the substrate (initially also located on the right side of the cell, red arrow at 228 s). RCAN  
266 output offered views with better resolution and contrast than the deconvolved iSIM input, particularly  
267 axially and towards the top of the cell. In some cases, dim or axially blurred filaments barely discerned in



268 the input data were clearly resolved in the RCAN view (yellow arrows in **Fig. 4e, Supplementary Video**  
269 **12, 14**).

270

## 271 **Discussion**

272 Here we focused on 4D imaging applications, because sustained volumetric imaging over  
273 extended duration at diffraction-limited or better spatial resolution remains a major challenge in  
274 fluorescence microscopy. We have shown that RCAN denoises and deconvolves fluorescence  
275 microscopy image volumes with performance competitive to state-of-the-art neural networks (**Fig. 1**). In  
276 live 4D super-resolution applications, which typically exhibit pronounced bleaching that limits  
277 experiment duration, RCAN restoration allows the illumination to be turned down to a level where the  
278 rate of photobleaching is drastically reduced or even negligible. Unacceptably noisy images can be  
279 restored, allowing for extended volumetric imaging similar to that attained with light-sheet microscopy,  
280 but with better spatial resolution. We suspect that RCAN carefully combined with high-resolution, but  
281 noisy, confocal microscopy may thus challenge the current primacy of light-sheet microscopy,  
282 particularly when imaging thin samples. At the same time, we expect RCAN denoising to synergize with  
283 light-sheet microscopy, allowing even greater gains in experiment duration (or speed) with that  
284 technique. RCAN also deblurs images, with better performance than the other networks we've tested  
285 (**Fig. 2**). We used this feature to improve spatial resolution in confocal microscopy (**Fig. 3**), achieving 2.5-  
286 fold improvement in lateral resolution and iSIM data (**Fig. 4**), achieving 1.4-fold improvement laterally  
287 and ~3-fold improvement axially.

288 Our findings highlight limitations of current neural networks and workflows and point the way to  
289 further improvements. First, on denoising applications we found 'breaking points' of the RCAN network  
290 at low input SNR. Estimating such input SNR may be useful in addition to computing measures of  
291 network disagreement<sup>9</sup>, especially given that the latter were not especially predictive of differences  
292 between ground truth and denoised data (**Supplementary Fig. 4**). Second, for resolution enhancement  
293 applications, our simulations on noiseless data revealed that all networks suffer noticeable deterioration  
294 when attempting to deblur at blur levels greater than 2-fold. Perhaps this explains why attempts to  
295 restore blurry microscopy images with neural networks have enabled only relatively modest levels of  
296 deblurring<sup>9,14</sup>. The fact that RCAN yielded better reconstructions than other networks even at 3-fold  
297 blurring suggests that network architecture itself may have substantial impact on deblurring  
298 performance. Our simulations also show that increased degradation in network output correlates with  
299 increased blur (**Fig. 2d**), implying caution is prudent when attempting extreme levels of deblurring.  
300 Exploring the fundamental limits of deblurring with neural networks would be an interesting avenue of  
301 further research. Third, practical factors still limit the performance of network output, suggesting that  
302 further improvement is possible. For the confocal-to-STED restorations, local deviations in spatial  
303 alignment between the training data pairs likely contribute to error in nuclear pore placement  
304 (**Supplementary Fig. 13**), suggesting that a local registration step during training would boost the quality  
305 of the restorations. For the expansion microscopy data, although we bypassed the need to finely  
306 register input and ground truth data by simulating pre-expansion data, improved registration schemes  
307 may enable direct use of experimentally derived pre- and post- expansion pairs. We suspect this would  
308 further improve the degree of resolution enhancement as complex noise and background variations in  
309 the data could be incorporated into the training procedure. We also expect that increasing label density  
310 would further improve the quality of our training data, as at the ~65 nm resolution we achieved in the  
311 ground truth expansion data, stochastic variations in labelling were evident (**Supplementary Fig. 22**) and

312 likely contribute an additional source of noise. Such improvements would probably also increase the  
313 SSIM and PSNR in the expansion predictions (**Supplementary Fig. 20**), which were markedly lower than  
314 in the confocal to STED predictions (**Supplementary Fig. 10**). Finally, achieving better spatial resolution  
315 in live samples usually demands corresponding improvements in temporal resolution, lest motion blur  
316 defeat gains in spatial resolution. We did not attempt to further increase the speed of our live  
317 recordings to account for this effect but doing so may result in sharper images.

318 Despite these caveats, the RCAN in its current form improves noisy super-resolution  
319 acquisitions, enabling image capture over tens of thousands of images; quantification, segmentation,  
320 and tracking of organelles and organelle dynamics; and prediction and inspection of fine details in  
321 confocal and iSIM data otherwise hidden by blur. We hope that our work inspires further advances in  
322 the rapidly developing field of image restoration.

323

### 324 **Author Contributions**

325

326 Conceived project: S-J.J.L., L.A.G.L., H.Shroff. Designed experiments: J.C., Y.S., A.Z., C.C., I.R-S., A.U.,  
327 H.Shroff. Performed experiments: J.C., Y.S., A.Z., C.C., I.R-S. Developed and tested 3D RCAN: H.Sasaki,  
328 H.L., H.C., C.C.H., S-J J.L., L.A.G.L. Adapted and tested CARE, SRResNet, ESRGAN: J.C. and J.L. Wrote  
329 software for hardware control: X.L. Conceived, developed, and tested expansion microscopy pipeline:  
330 J.C., H.Sasaki, H.L., Y.S., J.L., Y.W., X.L., M.G., H.Shroff. Performed and analyzed simulations: J.C., J.L.,  
331 Y.W., M.G., S.N., H.S. All authors analyzed data. Wrote paper: H.Shroff with input from all authors.  
332 Supervised research: A.U., S-J.J.L., L.A.G.L., H.Shroff.

333

### 334 **Acknowledgements**

335 J.C., Y.W., and H.Shroff thank the Marine Biological Laboratory at Woods Hole for providing the Deep  
336 Learning for Microscopy Image Analysis Course as well as Florian Jug, Jenny Folkesson, and Patrick La  
337 Riviere for providing superb instruction. We thank William Bement for the gift of the EMTB-3XEGFP  
338 plasmid (also available as Addgene plasmid #26741), Justin Taraska for the LAMP1-GFP plasmid, George  
339 Patterson for the GalT-GFP plasmid, and Panagiotis Chandris for the pShooter pEF-Myc-mito-GFP  
340 plasmid. We also thank Patrick La Riviere and Hank Eden for their careful read and comments on this  
341 work. This research was supported by the intramural research programs of the National Institute of  
342 Biomedical Imaging and Bioengineering and the National Institute of Heart, Lung, and Blood within the  
343 National Institutes of Health and by an SBIR cooperative agreement of the National Institute of General  
344 Medical Sciences: 1U44GM136091-01. A.U. and I.R. acknowledge support from NIH R01 GM131054 and  
345 NSF PHY 1607645 grants.

### 346 **Conflicts of Interest**

347 H.Sasaki, H.L., H.C., C.C.H., S-J J.L., L.A.G.L. are employees of DRVISION, LLC, a machine vision company.  
348 They have developed Aivia (a commercial software platform) that offers the 3D RCAN developed here.

### 349 **Disclaimer**

350 The NIH, its staff, and officers do not recommend or endorse any company, product, or service.

### 351 **Methods**

352 *Neural networks used for image restoration*

353 *3D RCAN*

354 The RCAN consists of multiple residual groups which themselves contain residual structure. Such  
355 'residual in residual' structure forms a very deep network consisting of multiple residual groups with  
356 long skip connections (**Fig. 1a**). Each residual group also contains residual channel attention blocks  
357 (RCAB) with short skip connections. The long and short skip connections, as well as shortcuts within the  
358 residual blocks, allow low spatial frequency information to be bypassed, facilitating the prediction of  
359 high spatial frequency information. Additionally, a channel attention mechanism<sup>19</sup> within the RCAB is  
360 used to adaptively rescale channel-wise features by considering interdependencies among channels,  
361 further improving the capability of the network to achieve higher resolution.

362 We extended the original RCAN<sup>17</sup> to handle image volumes. Since 3D models with a large patch  
363 size may consume prohibitive GPU memory, we also changed various network parameters to ensure  
364 that our modified RCAN fits within GPU memory. These changes relative to the original RCAN model  
365 include: (1) we set the number of residual groups (RG) to  $G = 5$  in the RIR structure; (2) in each RG, the  
366 RCAB number is set to 3; (3) the number of convolutional (Conv) layers in the shallow feature extraction  
367 and RIR structure is  $C = 32$ ; (4) the Conv layer in channel-downscaling has  $C/r = 4$  filters, where the  
368 reduction ratio  $r$  is set to 8; (5) all 2D Conv layers are replaced with 3D conv layers; (6) the upscaling  
369 module at the end of the network is omitted because network input and output have the same size in  
370 our case. In the original RCAN paper<sup>17</sup>, a small patch with size  $48 \times 48$  is used for training. By contrast, we  
371 used a much larger patch size ( $256 \times 256 \times 16$ ). We tried using a smaller patch size, but the training process  
372 was unstable and the results were poor. We suspect this is because microscopy images may show less  
373 high spatial frequency content than natural images, so a larger patch is necessary to extract enough  
374 gradient information for back-propagation.

375 The percentile-based image normalization proposed in the CARE manuscript<sup>9</sup> is applied as a pre-  
376 processing step prior to training. In microscopy images, foreground objects of interest may be  
377 distributed sparsely. In such cases the model may overfit the background, failing to learn the structure  
378 of foreground objects if the entire image is used indiscriminately for training. To avoid overfitting,  
379 patches of the background were automatically rejected in favor of foreground patches during training.  
380 Background patch rejection is performed on the fly during data augmentation. We implemented training  
381 in a 3D version of RCAN using Keras<sup>30</sup> with a TensorFlow<sup>31</sup> backend. Each model was trained on two  
382 NVIDIA GeForce GTX 1080 Ti GPUs for 400 epochs, which took 1 day. Applying the denoising model on a  
383  $1920 \times 1550 \times 12$  dataset using a desktop with a single GTX 1080 Ti GPU took  $\sim 63.3$  s per volume. This  
384 time also includes the time it takes to save the volume (with 32-bit output). On similar datasets with the  
385 same XY dimensions (but different number of Z-slices), applying the model took  $\sim 3.9$  s - 5.2 s per Z-slice.  
386 Further details are provided in **Supplementary Note 1** and **Supplementary Software**.

387  
388 *SRResNet and ESRGAN*

389  
390 SRResNet is a deep residual network for image super-resolution, which obtained state-of-the-art  
391 results in 2017<sup>20</sup>. Building on ResNet<sup>32</sup>, the SRResNet has 16 Residual Blocks (RB) with identical layout.  
392 Within each RB, there are two convolutional layers with small  $3 \times 3$  kernels and 64 feature maps,  
393 followed by batch-normalization layers and a parametric rectified linear unit (ReLU) as activation  
394 function.

395 Generative adversarial networks<sup>11</sup> (GAN) provide a powerful framework for generating  
396 plausible-looking natural images with high perceptual quality in computer vision applications. GANs are

397 used in image super-resolution applications to favor solutions that resemble natural images<sup>20</sup>. Among  
398 such methods, enhanced super-resolution generative adversarial networks (ESRGAN<sup>21</sup>) won the first  
399 place in the Perceptual Image Restoration and Manipulation (PIRM) challenge on perceptual super-  
400 resolution in 2018<sup>22</sup>. Thus, we selected ESRGAN as an additional reference method to evaluate  
401 performance on fluorescence microscopy images.

402 The key concept underlying ESRGAN is to train a Generator  $G$  with the goal of fooling a  
403 Discriminator  $D$  that is trained to distinguish predicted high-resolution images from real high-resolution  
404 images. The Generator network  $G$  has 16 Residual in Residual Dense Blocks<sup>21</sup> (RRDB) with identical  
405 layouts, which improves the RB design in SRResNet. RRDB has a residual-in-residual structure, where  
406 multi-level residual learning is used. In addition, RRDB contains dense blocks<sup>33</sup>, which increase network  
407 capacity due to the dense connections contained within each dense block.

408 The Discriminator network  $D$  is based on Relativistic GAN<sup>34</sup>. It has 8 convolutional layers with  
409 small 3×3 kernels as in the VGG network<sup>35</sup> and the resulting feature maps are followed by two dense  
410 layers. A Relativistic average Discriminator<sup>20</sup> (RaD) is used as the final activation function to predict the  
411 probability that a real high-resolution image is relatively more realistic than a fake high-resolution  
412 image.

413 In this work, we used the published SRResNet and ESRGAN (PyTorch implementation,  
414 <https://github.com/xinntao/BasicSR>) to process image volumes in a slice-by-slice manner. Before  
415 training, we normalized low-resolution (LR) and high-resolution (HR) images by percentile-based image  
416 normalization<sup>9</sup> to reduce the effect of hot and dead pixels in the camera. Then we linearly rescaled the  
417 range of LR and HR images to [0,1]. SRResNet and ESRGAN networks were trained on an NVIDIA Quadro  
418 P6000 GPU. In all experiments (except the spherical phantoms), for each mini-batch, we cropped 16  
419 random 480×480 overlapping image patches for training. Patches of background were not used for  
420 training. To determine whether a patch pair was from the background, we simply compared the mean  
421 intensity of the patch versus the whole image. If the mean intensity of the patch was less than 20% of  
422 the mean intensity of the whole image, the patch pair was not used for training. In spherical phantom  
423 experiments, we selected 16 random 2D image slices (256×256) for each mini-batch. For SRResNet,  
424 Adam optimization were used for all experiments with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.99$ , a learning rate of  $2 \times 10^{-4}$ , and  
425  $10^5$  update iterations. During testing, batch-normalization update was turned off to obtain an output HR  
426 image that depended only on the input LR image. For ESRGAN, we used Adam optimization for all  
427 experiments with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.99$ . The Generator  $G$  and Discriminator  $D$  were alternately updated  
428 with learning rate initialized as  $10^{-4}$  and decayed by a factor of 2 every  $10^4$  updates. Training time was  $\sim 8$   
429 hours for SRResNet and  $\sim 12$  hours for ESRGAN. Application usually took  $\sim 60$  s (SRResNet) to 120 s  
430 (ESRGAN) for the image volumes shown here.

#### 431 CARE

432 The content aware restoration (CARE) framework has been described in detail.<sup>9</sup> We  
433 implemented CARE through Keras and TensorFlow via GitHub (<https://github.com/CSBDeep/CSBDeep>).  
434 CARE networks were trained on an NVIDIA Titan RTX GPU card in a local workstation. Typically for each  
435 image volume, 2048 patches of size 128×128×8 were randomly cropped and used to train a CARE  
436 network with a learning rate of  $2 \times 10^{-4}$ . From the extracted patches, 10% were used as validation data.  
437 The number of epochs for training is 200 and the mean absolute error (mae) was used as loss function.  
438 Training time for a given model was 8-12 hours, application of the model on a 1920×1550×28 sized  
439 image volume took  $\sim 90$  s.

440 For all networks, we evaluated the peak-signal-to-noise-ratio (PSNR) and the structural similarity  
441 index<sup>29</sup> (SSIM) on normalized input, network output, and ground truth with built-in MATLAB  
442 (Mathworks) functions.

443

444 *Instant structured illumination microscopy (iSIM)*

445 *U2OS Cell Culture and transfection*

446 U2OS cells were cultured and maintained at 37 C and 5% CO<sub>2</sub> on glass bottom dishes (MatTek,  
447 P35G-1.5-14-C) in 1 mL of DMEM medium (Lonza, 12-604F) containing 10% FBS. At 40-60% confluency,  
448 cells were transfected with 100 μL of 1X PBS containing 2 μL of X-tremeGENE HP DNA transfection  
449 reagent (Sigma,6366244001) and 2 μL plasmid DNA (300-400 ng/μL, see **Supplementary Table 1** for  
450 plasmid information) and maintained at 37C, 5 % CO<sub>2</sub> for 1-2 days.

451 *Immunofluorescence labeling*

452 U2OS cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences, 15710) and  
453 0.25% Glutaraldehyde (Sigma, G5882) in 1X PBS at room temperature (RT) for 15 minutes. Cells were  
454 rinsed 3 times with 1X PBS, and permeabilized by 0.1% Triton X-100 (Sigma, 93443) in 1X PBS for 1  
455 minute. Cells were treated with 300μL Image-iT FX Signal enhancer (Thermofisher, R37107) for 30  
456 minutes at RT followed by 30-minute blocking with 1% BSA/PBS (Thermofisher, 37525) at RT. Cells were  
457 then labeled with fluorescent antibodies and/or fluorescent streptavidin (see **Supplementary Table 1**)  
458 in 0.1% Triton X-100/PBS for 1 hour at RT. After antibody labeling, cells were washed 3 times with 0.1%  
459 Triton X-100 and stained by 4',6-diamidino-2-phenylindole (DAPI, Sigma, D9542) (1μg/mL) in 1X PBS  
460 for 5 minutes at RT. DAPI stain was used for expansion factor estimation and rapid cell or region  
461 localization throughout the Expansion Microscopy (ExM) process.

462 *iSIM imaging for denoising*

463 iSIM data was obtained on our previously reported home-built system<sup>3</sup>. A 60x NA 1.42 oil  
464 objective (Olympus) was used for all imaging, except the training data acquired for the iSIM to  
465 expansion microscopy cross-modality experiments (which used a 1.2 NA water immersion lens,  
466 described below in more detail). To obtain high and low SNR image pairs for training, high (usually 33  
467 mW for 488 nm, 72 mW for 561 nm) and low powers (0.3 mW for 488 nm, 0.6 mW for 561 nm) were  
468 rapidly switched via an AOTF. Green and red fluorescence images were acquired with a filter wheel  
469 (Sutter, FG-LB10-BIQ and FG-LB10-NW) and notch filters (Semrock, NF03-488E-25 and NF03-561E-25).  
470 Samples were deposited on 35-mm-diameter high-precision 1.5 dish (Matek; P35G-0.170-14-C). For live  
471 cell imaging, the dishes were mounted within an incubation chamber (Okolab; H301-MINI) to maintain  
472 temperature at 37°C.

473 *Estimating illumination intensity*

474 A power meter (Thorlabs, PM100D) was used to measure the excitation laser power  
475 immediately prior to the objective. The average intensity was calculated using the measured intensity  
476 divided by the field of view (FOV, 106 μm by 68 μm).

477 *Jurkat T Cell Culture, substrate preparation, and iSIM imaging*

478 E6-1 Wild Type Jurkat cells were cultured in RPMI 1640 supplemented with 10% fetal bovine  
479 serum and 1% Penn-Strep antibiotics. Cells were transiently transfected with EMTB-3XGFP plasmid using  
480 the Neon (ThermoFisher Scientific) electroporation system two days before imaging, using the  
481 manufacturer's protocol.

482 Coverslips attached to 8 well Labtek chambers were incubated in Poly-L-Lysine (PLL) at 0.01%  
483 W/V (Sigma Aldrich, St. Louis, MO) for 10 min. PLL was aspirated and the slide was left to dry for 1 hour  
484 at 37 °C. T cell activating antibody coating is performed by incubating the slides in a 10 µg/ml solution of  
485 anti-CD3 antibody (Hit-3a, eBiosciences, San Diego, CA) for 2 hours at 37 °C or overnight at 4 °C. Excess  
486 anti-CD3 was removed by washing with L-15 imaging media immediately prior to the experiment.

487 Imaging of live EMTB-3XGFP expressing Jurkat cells was performed at 37 °C using iSIM, with a  
488 1.42 numerical aperture 60× lens (Olympus) and 488 nm laser for excitation using the same home-built  
489 system as above<sup>3</sup>. For volumetric live imaging, the exposure was set to 100 ms per slice, the spacing  
490 between slices to 250 nm, and the inter-volume temporal spacing to 12.3 s.

#### 491 *Linearity estimate*

492 Linearity was assessed by measuring the intensity in different regions in maximum intensity  
493 projections (MIP) of raw images of fixed cells expressing U2OS cells expressing the mEmerald-Tomm20  
494 label, and the corresponding RCAN predictions (**Supplementary Fig. 6**). Small regions of interest (ROIs, 8  
495 by 8 pixels) were selected and the average intensity value in each region used in comparisons between  
496 raw input and RCAN predictions.

497

#### 498 *Expansion Microscopy (ExM)*

499 Expansion microscopy was performed as described<sup>26</sup>. Immunolabeled U2OS cells were post-  
500 fixed in 0.25% Glutaraldehyde/1X PBS for 10 minutes at RT and rinsed three times in 1X PBS. Fixed cells  
501 were incubated with 200 µL of monomer solution (19% (wt/vol) Sodium acrylate (Sigma, 408220),  
502 10% (wt/vol) Acrylamide (Sigma, A3553), 0.1% (wt/vol) N,N-Methylenebis(acrylamide) (Sigma, 146072)  
503 in 1X PBS) for 1 minute at RT. To start gelation, the monomer solution was replaced by fresh monomer  
504 solution containing 0.2% (vol/vol) Ammonium Persulfate (Thermofisher, 17874) and 0.2% (vol/vol)  
505 Tetramethylethylenediamine (Thermofisher, 17919). Gelation was allowed to proceed 40 minutes at RT,  
506 and the resulting gel was digested in 1 mL of digestion buffer (0.8M guanidine hydrochloride and 0.5%  
507 Triton X-100 in 1X TAE buffer) by Proteinase K (0.2mg/mL, Thermofisher, AM2548) for 1 hour at 45°C.  
508 After digestion, gels were expanded in 5mL of pure water (Millipore, Direct-Q 5UV, ZRQSVR5WW), and  
509 fresh water exchanged 3-4 times every 15 minutes.

#### 510 *Pre-ExM and post-ExM on the same cell*

511 To compare images between pre- and post-ExM, the same group of cells needs to be located  
512 and imaged before and after ExM (**Supplementary Fig. 16**). After initial antibody (**Supplementary Table**  
513 **1**) and DAPI staining, the pre-ExM cells were imaged under a wide field microscope with a 20X air  
514 objective (Olympus, UPlanFL N, 0.5 NA). Based on the DAPI signal, the nuclear shape, diameter, and  
515 distribution pattern of selected cells can be recorded, a useful aid in finding the same cells again if post-  
516 ExM images are acquired on the wide field microscope. The coarse location of a group of cells was  
517 marked by drawing a square with a Sharpie marker underneath the coverslip. The marked cells are then  
518 imaged on our home-built instant structured illumination microscope (iSIM<sup>3</sup>) before and after ExM in  
519 later steps. Before expansion, the marked region was imaged on iSIM with a 60X, NA 1.2 water objective  
520 (Olympus, PSF grade) to acquire pre-ExM data. The correction collar was adjusted to the 0.19 setting,  
521 which was empirically found to minimize spherical aberration. After ExM, a square portion of expanded  
522 gel was cut out, based on the marked region drawn underneath the cover glass, then remounted on a  
523 poly-L-lysine coated glass bottom dish (MatTek, P35G-1.5-14-C) and secured by depositing 0.1% low

524 melt agarose around the periphery of the gel. To create the coated glass bottom dish, we applied poly-L-  
525 lysine (0.1% in water, Sigma, P892) for 30 minutes at room temperature, rinsed three times with pure  
526 water, and air dried. The same group of cells was then found on the wide field microscope using the  
527 DAPI stain and the 20X air objective. By comparing to the wide field DAPI image acquired before  
528 expansion, coarse estimation of the expansion factor as well as potential cell distortion/damage can be  
529 assayed. Finally, another square was drawn underneath the coverslip to locate the expanded cells,  
530 which were then imaged on the iSIM with the same objective and correction collar settings for post-ExM  
531 image acquisition.

#### 532 *Attempting to register pre- and post-expansion data*

533 Pre- and post-expansion images were registered using the landmark registration module in 3D  
534 Slicer<sup>36</sup> (<http://www.slicer.org/>). Landmark-based registration in 3D Slicer is an interactive registration  
535 method that allows the user to view registration results and manipulate landmarks in real time. We first  
536 rescaled the pre-expansion images according to the estimated expansion factor in the X, Y, and Z axes.  
537 During the registration process, pre-expansion images were used as fixed volumes and post-expansion  
538 images were used as moving volumes. Pre- and post-expansion images were coarsely aligned by affine  
539 registration based on 2-3 manually selected landmarks. Image registration was further refined using thin  
540 plate spline registration by interactively manipulating the landmarks. Finally, a transformation grid was  
541 generated to transform the post-expansion images to the pre-expansion images (**Supplementary Fig.**  
542 **17**).

#### 543 544 *Estimating expansion factor*

545 Pre- and post-expansion mitochondrial and microtubule data were inspected in 3D Slicer and  
546 registered with landmark-based registration as described above. Apparent distances between feature  
547 points were manually measured and ratioed to obtain the local expansion factor, which varied between  
548 3.1-3.4 for mitochondria and 3.9-4.1 for microtubules (**Supplementary Fig. 15**). Based on this analysis  
549 we used a value 3.2 for mitochondria and 4.0 for microtubules in all downstream processing.

#### 550 *Stage scanning with iSIM*

551 To rapidly tile multiple iSIM image fields to capture large expanded samples, we added a stage  
552 scan function into our control software, available on request from the authors. In the software, a step  
553 size of 0 to 150  $\mu\text{m}$  can be selected for both horizontal (X) and vertical (Y) directions. We set this step  
554 size to be  $\sim 70 \mu\text{m}$ , a value smaller than the field of view to ensure that each image had at least 20%  
555 overlap with adjacent images for stitching. We used up to 100 steps in both directions. The stage scan  
556 experiment was performed in a “zigzag” format (adjacent rows were scanned in opposite directions) to  
557 avoid large movements and maintain sample stability. At each stage position, 3D stacks were acquired.  
558 Stacks were stitched in Imaris Stitcher (Bitplane).

#### 559 *Generating synthetic pre-expansion data*

560 To first order, we can interpret the post-expansion image as enlarging the objects by an expansion  
561 factor  $M$  and blurring with the system PSF,  $sPSF$ :

$$562 \quad g_{POST} = S_M * sPSF$$

563 where  $s_M$  is the expanded object,  $g_{POST}$  is the post-expansion image of the expanded object, and  $*$  is the  
564 convolution operation. Similarly, if we upsample the pre-expansion image by a factor  $M$  we can  
565 approximate it as

$$566 \quad g_{PRE} = s_M * ePSF$$

567 where  $ePSF$  is  $sPSF$  enlarged  $M$  times. We seek to express  $g_{PRE}$  in terms of  $g_{POST}$ , thus obtaining an estimate  
568 of  $g_{PRE}$  in terms of the measured post-expansion image.

569 Fourier transforming (FT) both equations, dividing to eliminate the object spectrum, and  
570 rearranging terms, we obtain

$$571 \quad G_{PRE} = (G_{POST})(mOTF)$$

572 where  $mOTF$  is a modified OTF equivalent to the ratio of the OTFs corresponding to  $ePSF$  and  $sPSF$ , i.e.  
573  $mOTF = FT(ePSF) / FT(sPSF)$ . To avoid zero or near zero division in this calculation, we set the amplitude  
574 of  $FT(sPSF)$  to 1 beyond the cut-off frequency of  $sPSF$ . Finally, inverse Fourier transforming yields a  
575 synthetic estimate of  $g_{PRE}$ .

576 We improved this estimate by also modifying the background and noise levels to better match  
577 experimental pre-expansion images, computing the SSIM between the synthetic image and the  
578 experimental pre-expansion image as a measure of similarity. We tried to maximize the SSIM by (1)  
579 laterally and axially modifying the modelled  $sPSF$  so that the FWHM value is equal to the FWHM measured  
580 with 100 nm beads and resolution-limited structures in the experimental images; (2) modifying the  
581 background level, i.e., adding or subtracting a constant value; and (3) adding Gaussian and Poisson noise.  
582 We optimized these parameters in a range +/- 15 % of the values derived from experimental pre-  
583 expansion data (2-3 pre-expansion images that could be reasonably well registered to corresponding post-  
584 expansion data), and then applied these optimized parameters for all synthetic data. Finally, we  
585 performed a visual check before deconvolving the synthetic data and post-expansion data in preparation  
586 for RCAN training. 15 iterations of Richardson-Lucy deconvolution were applied, using  $sPSF$  for the  
587 expanded images and the modified  $ePSF$  for the synthetic data. These steps are shown in **Supplementary**  
588 **Fig. 19**.

#### 589 *Estimating signal-to-noise (SNR) ratio in experiments*

590 We assumed a simple model for per-pixel SNR, accounting for Poisson noise arising from the  
591 signal and read noise from the camera. After subtracting a constant background offset (100 counts) and  
592 converting the digital signal in each pixel to photons using the manufacturer-supplied conversion factor  
593 (0.46 photoelectrons/digital count), we used

$$594 \quad SNR = S / (S + N_r^2)^{0.5}$$

595  
596 where  $S$  is the observed, background-corrected signal in photoelectrons, and  $N_r$  the read noise (1.3  
597 electrons from the manufacturer).

#### 598 599 *Spherical simulations*

600  
601 For the images in **Fig. 2** and images and analysis in **Supplementary Fig. 5, 8, 9** the simulated  
602 ground-truth images consisted of spheres seeded at random locations and with random size and



603 intensity, generated with `ImgLib2`<sup>37</sup>. The maximum radius of the spheres was set at 3 pixels and the  
604 intensity range set to 1000 to 20000. We generated a set of 30 such images with size of 256x256x256.  
605 Ground truth (GT) images were generated by blurring this set of 30 images with the iSIM PSF (simulated  
606 as the product of the excitation and emission PSFs, generated in PSF generator  
607 (<http://bigwww.epfl.ch/algorithms/psfgenerator/>) with an NA of 1.42 and wavelengths 488 nm and 561  
608 nm, respectively). Noisy phantom images were obtained by adding Gaussian noise (simulating the  
609 background noise of the camera in the absence of fluorescence) and Poisson noise (proportional to the  
610 square root of the signal) to the GT images. The 2x, 3x and 4x blurred noiseless phantom images were  
611 obtained by blurring the initial 30 images with a kernel 2x, 3x and 4x the size of the iSIM PSF.

612

### 613 *Estimating spatial resolution*

614 The resolution measures in **Fig. 1d** were estimated by computing the FWHM as a measure of  
615 apparent size of a subdiffractive object (microtubule width). However, all other resolution estimates  
616 were based on decorrelation analysis<sup>24</sup>. This method estimates average image resolution from the local  
617 maxima of a series of decorrelation functions, providing an estimated resolution that corresponds to the  
618 highest spatial frequency with sufficient SNR, rather the Abbe resolution limit.

619 There are four main steps in the algorithm. First, the Fourier transform of the input image  $I(k)$   
620 and its normalized version  $I_n(k)$  are cross-correlated using Pearson correlation, producing a single value  
621 between 0 and 1, denoted  $d$ . Second, the normalized Fourier transform  $I_n(k)$  is repeatedly filtered by a  
622 binary circular mask with different radius  $r \in [0,1]$  (here  $r$  is expressed as a normalized spatial  
623 frequency), and the cross-correlation between  $I(k)$  and each filtered  $I_n(k)$  is recalculated, yielding a  
624 decorrelation function  $d(r)$ . This decorrelation function exhibits a local maximum of amplitude  $A_0$  that  
625 indicates the spatial frequency  $r_0$  of best noise rejection and signal preservation ratio. Third, the input  
626 image is repeatedly filtered with different Gaussian high-pass filters to attenuate the energy of low  
627 frequencies. For each filtered image, another decorrelation function is computed, generating a set of  $[r_i,$   
628  $A_i]$  pairs, where  $r_i$  and  $A_i$  are the position and amplitude of the local maximum, respectively. Last, the  
629 most suitable peak position (i.e., selected from the  $r_i$ ) is selected as the estimate of resolution. In the  
630 original algorithm<sup>24</sup>, two choices are used and validated in many applications - (1) the peak  
631 corresponding to the highest frequency (i.e., the maximum  $r_i$  value); (2) the peak corresponding to the  
632 highest geometric mean of  $r_i$  and  $A_i$ .

633 However, we found that both criteria often failed when using them on our images, i.e., the  
634 estimated resolution was often a value much beyond the theoretic resolution limit. Plotting  $[r_i, A_i]$  pairs  
635 shows three phases:  $A_i$  first increases in phase I, then gradually decreases in phase II, and finally  
636 increases again in phase III (**Supplementary Fig. 11a**). Resolution values in phase III exist due to digital  
637 upsampling of the pixel size, but are not reliable, as they extend past the Abbe limit. We thus modified  
638 the algorithm by (1) setting a theoretical resolution limit in computing the decorrelation functions; and  
639 (2) adopting a new criterion to determine the resolution estimate. Our new criterion finds the local  
640 minimum of  $A_i$  to locate  $r_i$  at the transition between phase II and phase III, which provides a reliable  
641 resolution estimate that is robust to changes in pixel size. We validated this strategy on a microtubule  
642 image with 1x, 1.5x and 3x digital upsampling (**Supplementary Fig. 11b**), finding that our criterion gave  
643 identical estimates of spatial resolution in each case.

644 For estimating the lateral and axial resolution in our data (input, ground truth, and deep  
645 learning outputs), we first interpolated the stacks along the axial dimension to achieve isotropic pixel  
646 size. Then we performed our modified decorrelation analysis on a series of xy slices to obtain lateral  
647 resolution estimate (with mean and standard deviations derived from the slices). For axial resolution, we  
648 implemented sectorial resolution estimate<sup>24</sup> on a series of xz slices, where the binary circular mask was  
649 replaced with a sectorial mask (22.5 degree opening angle, **Supplementary Fig. 11c**) that captured

650 spatial frequencies predominantly along the z dimension.

651

## 652 *Confocal and STED microscopy*

### 653 *Sample preparation*

654 Mouse embryonic fibroblasts (MEF) were grown in #1.5 glass-bottom dishes (MatTek, P35G-1.5-  
655 20-C) using DMEM (Gibco, 10564011) supplemented with 10% FBS (Quality Biological, 110-001-101HI).

656 For microtubule and nuclear pore samples, we fixed and permeabilized cells with -20°C  
657 methanol (Sigma-Aldrich, 322415) for 10 min at -20°C. Samples were rinsed and blocked for 1 hour with  
658 1x Blocker BSA (ThermoFisher Scientific, 37525) and incubated overnight with 1:500 dilution of primary  
659 rabbit anti-alpha tubulin (Abcam, ab18251) and mouse anti-nuclear pore complex (Abcam, ab24609)  
660 antibodies in 1x Blocker BSA at 4°C. Samples were washed three times for 5 minutes with 1x Blocker  
661 BSA. After the last washing step, we fluorescently labeled samples by incubation with 1:500 dilution of  
662 secondary Alexa Fluor 594 goat anti-mouse (ThermoFisher Scientific, A-11005) and ATTO 647N goat anti-  
663 rabbit (Sigma-Aldrich, 40839) antibodies in 500 µL of 1x Blocker BSA for 4 hours at room temperature.  
664 Samples were washed four times for 5 minutes with 1x Blocker BSA. After final washing, samples were  
665 mounted in glass-bottom dishes using 90% Glycerol (Sigma-Aldrich, G2025) in PBS (KD Medical, RGF-  
666 3210).

667 For SiR-DNA imaging, we used live MEF cells, grown as before, and MEF cells fixed with 4°C 4%  
668 formaldehyde (Sigma-Aldrich, 252549) in PBS for 20 minutes at room temperature. Sample labelling was  
669 performed with the SiR-DNA kit (Spirochrome, SC007) following the manufacturers protocol:

670 [https://spirochrome.com/documents/202003/datasheet\\_SPY650-DNA\\_202003.pdf](https://spirochrome.com/documents/202003/datasheet_SPY650-DNA_202003.pdf). Fixed samples were  
671 mounted as before.

672

### 673 *Imaging*

674 We acquired 33 matched sets of confocal/STED volumes for microtubule- and nuclear pore  
675 complex-labeled samples. For these experiments all images were acquired using a Leica SP8 3X STED  
676 microscope, a white-light laser for fluorescence excitation (470-670nm), a Leica HyD SMD time-gated  
677 PMT, and a Leica 100x (1.4 N.A.) STED White objective (Leica Microsystems, Inc.). ATTO 647 was excited  
678 at 647 nm and emission was collected over a bandwidth of 657-700 nm. Alexa Fluor 594 was imaged with  
679 580 nm excitation, and emission was collected over a bandwidth of 590-650 nm. All images (both confocal  
680 and STED) were acquired with a pinhole size of 0.7 A.U., a scan speed of 600 Hz, a pixel format of 1024 x  
681 1024 (pixel sizes of 25 nm), a 6-slice z-stack acquired at an interslice distance of 0.16 µm, and time gating  
682 on the HyD SMD set to a time range of 0.7-6.5 ns. The STED images for both labels were acquired with  
683 depletion at 775nm laser (pulsed at 80 MHz) at a power of 105 mW at the back aperture for ATTO 647-  
684 labeled microtubules (25% of full power) and 85 mW at the back aperture for Alexa Fluor 594 labeled  
685 nuclear pore complexes (20% of full laser power). Fluorescence excitation for STED imaging was set to 4x  
686 and 1.5x the confocal excitation power levels for ATTO 647 and Alexa Fluor 594 respectively. For ATTO  
687 647, HyD SMD gain was set to 100% for confocal and STED imaging. For Alexa Fluor 594, HyD SMD gain  
688 was set to 64% for confocal imaging and 100% for STED imaging. For both colors, confocal images were  
689 acquired with a 2-frame line average and STED images were acquired with a 2-frame line average  
690 combined with 2-frame integration.

691 SiR-DNA labelled MEF cells were imaged both in the fixed (confocal and STED) and live-cell  
692 (confocal only) mode. Low SNR confocal and high-quality STED image replicates were taken on similar

693 fixed samples (35 data sets) to train a deep-learning model to apply to the live cell confocal data. Low-  
694 excitation level (thus low SNR) live cell confocal images were followed over time to capture cell division.  
695 For these experiments, the same microscope hardware listed above was used but scanned in the resonant  
696 mode (to afford more rapid imaging capable of capturing cell division). For live cell confocal stacks of 25  
697 or more slices (interslice distance of 0.16  $\mu\text{m}$ ) were taken approximately every minute (2s/frame)  
698 continuously for a period of ~30-45 minutes. Images were taken with a scan rate of 8000 Hz, 8 line-  
699 average, a pinhole set to 1 A.U, 647nm excitation (5% of total laser power), an emission bandwidth of  
700 657-637nm, and at a pixel size of 25 nm at a format of 2048 x 2048. For the fixed cell experiments the  
701 confocal settings were the same except that line averaging was set to 16, the frame rate was 6s/frame,  
702 the excitation power at 647nm was set to 0.1% total laser power (to approximately match the SNR in the  
703 live cell data), and only one z-stack was taken. STED experiments were the same, except that 647 nm  
704 excitation was set to 1.5% and the depletion power at 775 nm was 7.5% (approximately 35 mW at the  
705 back aperture). Time gating windows on the HyD SMD was set to 0.3 to 6.5 ns or 0.7 to 6.5ns for the  
706 confocal and STED experiments, respectively. For live experiments, temperature was set to 37°C using a  
707 culture dish heater and temperature control unit (DH-35 and TC-344B, Warner Instruments, Hamden, CT)  
708 and an objective heater (Bioprotechs, Butler, PA).

709

#### 710 *Deconvolution*

711 Huygens Professional (version 19.1, Scientific Volume Imaging, Hilversum, The Netherlands) was  
712 used to deconvolve some confocal images. All deconvolution was based on idealized point spread  
713 functions, using the classic maximum likelihood estimation (CMLE) deconvolution algorithm. In some  
714 cases, the object stabilizer module was used to compensate for drift and minor mechanical instabilities.

#### 715 *Data availability*

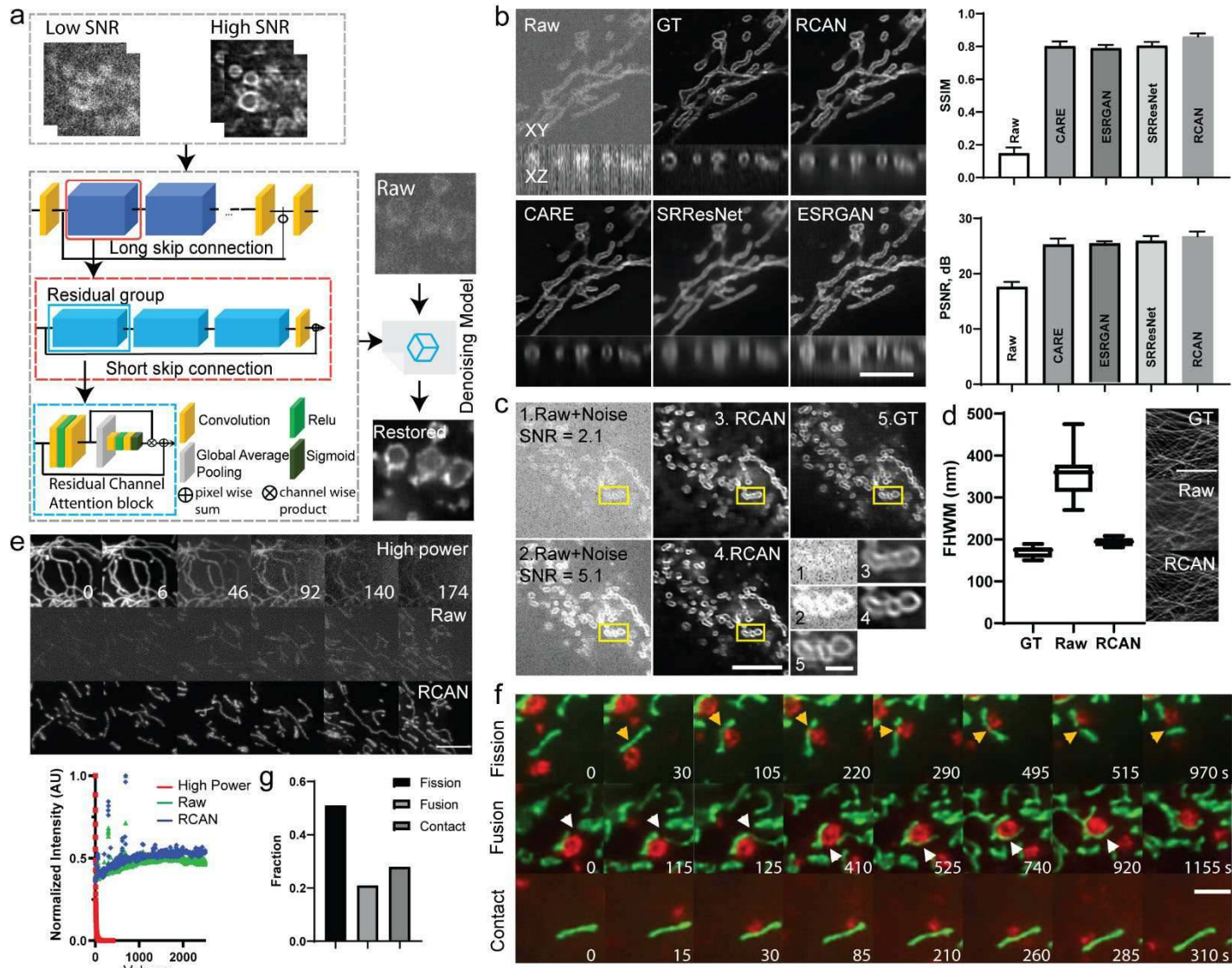
716 The data that support the findings of this study are available from the corresponding author  
717 upon reasonable request.

#### 718 *Code availability*

719 The code used in this study will be available as Supplementary Software. We plan to upload  
720 software and supplementary data to GitHub in a finalized version of this manuscript.

721

722

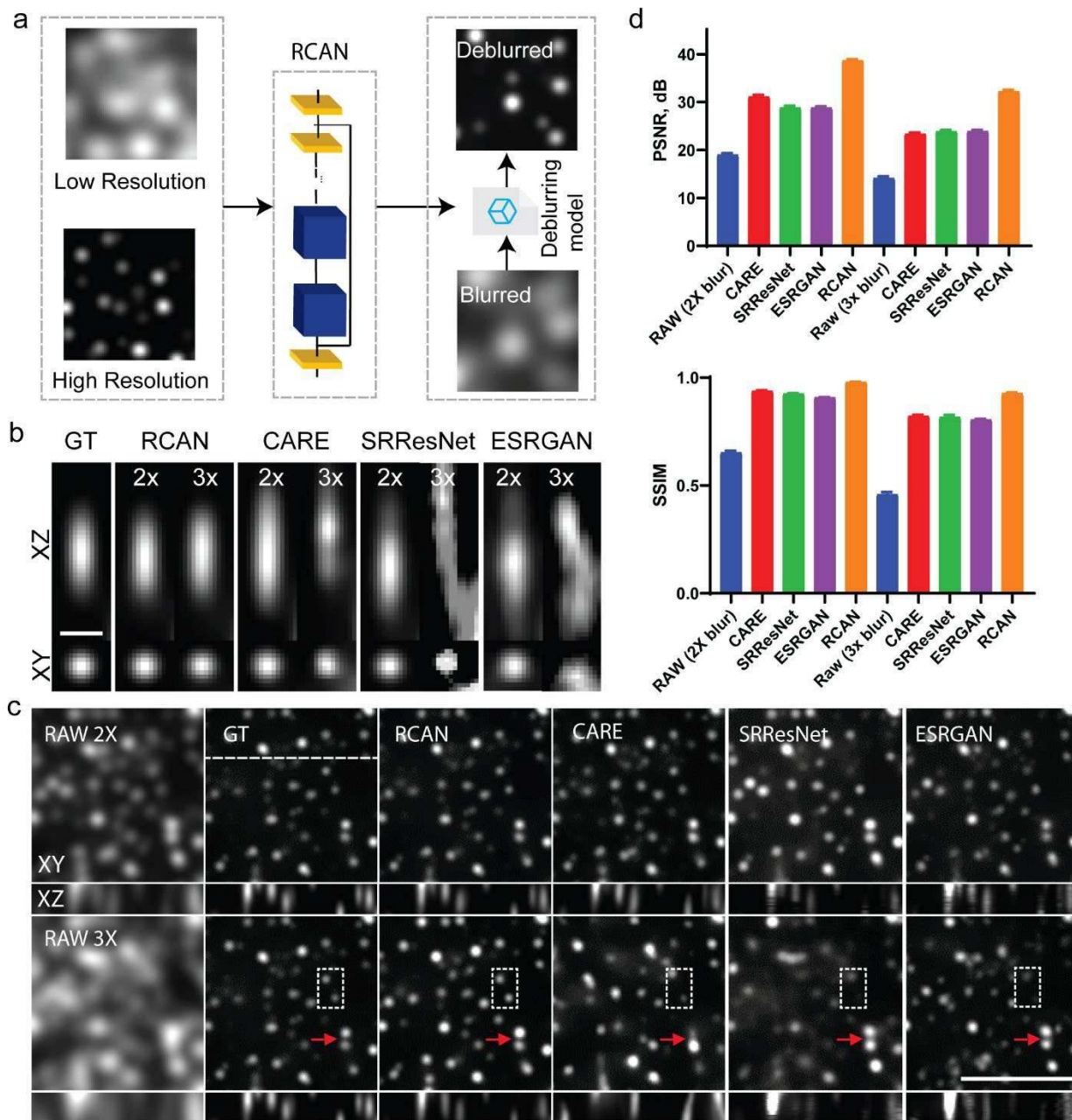


723

724 **Fig. 1, Residual channel attention networks denoise super-resolution data.**

725 **a)** Residual channel attention network (RCAN) architecture used throughout this work. Matched low  
 726 and high SNR image volumes are used to train our RCAN, a residual in residual structure which consists  
 727 of several residual groups (dark blue, red outline) with long skip connections. Each residual group itself  
 728 contains additional residual channel attention blocks (RCAB, light blue, blue outline) with short skip  
 729 connections, convolution, rectified linear unit (ReLU), sigmoid, and pooling operations. Long and short  
 730 skip connections, as well as short-cuts within the residual blocks, allow abundant low-frequency  
 731 information to be bypassed through such identity-based skip connections, facilitating the learning of  
 732 high frequency information. A channel attention mechanism within the RCAB further aids the  
 733 representational ability of the network in learning high-resolution information. **b) Left:** noisy raw  
 734 instant SIM (iSIM) data acquired with low-intensity illumination, low-noise deconvolved ground truth  
 735 (GT) data acquired with high-intensity illumination, RCAN, CARE, SRResNet, and ESRGAN output. Lateral  
 736 (upper) and axial (lower) cross sections are shown. Samples are fixed U2OS cells expressing mEmerald-  
 737 Tomm20, imaged via iSIM. **Right:** Comparison of network output using structural similarity index (SSIM)  
 738 and peak signal-to-noise-ratio (PSNR). Means and standard deviations are reported, obtained from  $N =$

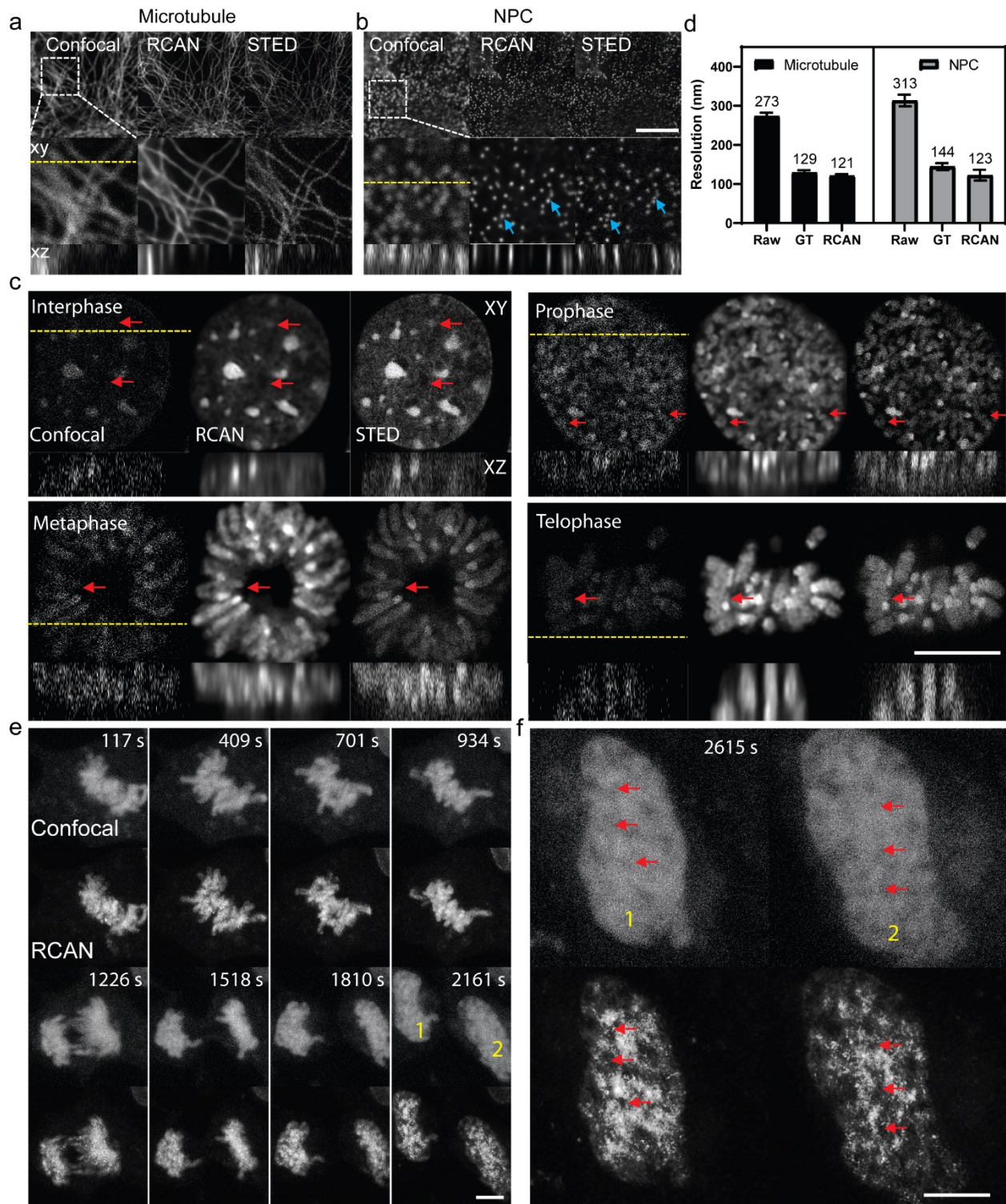
739 10 planes from one volume. See also **Supplementary Figs. 1, 2. c)** RCAN performance at different input  
740 SNR levels, simulated by adding Gaussian and Poisson noise to raw input. Noisy raw input data at SNR  
741 2.1 (top row) and 5.1 (bottom row) were used to generate predictions, which were then compared to  
742 ground truth. SNR values are calculated as mean of values within the yellow rectangular regions. Higher  
743 magnification views of mitochondria marked in yellow rectangular regions are shown at lower right. See  
744 also **Supplementary Fig. 5. d)** Full width at half maximum values (mean +/- standard deviations) from 10  
745 microtubule filaments for deconvolved, high SNR ground truth (GT); noisy iSIM input ('Raw'); and  
746 network output ('RCAN'). **e)** RCAN denoising enables collection of thousands of iSIM volumes without  
747 photobleaching. Mitochondria in live U2OS cells were labeled with pShooter pEF-Myc-mito-GFP and  
748 imaged with high (360 W/cm<sup>2</sup>) and low (4.2 W/cm<sup>2</sup>) intensity illumination. Top row: selected examples  
749 at high illumination power, illustrating severe photobleaching. Middle row: selected examples from a  
750 different cell, imaged at low illumination power, illustrating low SNR ('Raw'). Bottom row: RCAN output  
751 given low SNR input. Numbers in top row indicate volume #. Graph quantifies normalized signal in each  
752 case, 'jumps' in Raw and RCAN signal correspond to manual refocusing during acquisition. Maximum  
753 intensity projections are shown. See also **Supplementary Videos 1, 2, Supplementary Fig. 6, 7. f)** Dual-  
754 color imaging of mitochondria (green, pShooter pEF-Myc-mito-GFP) and lysosomes (mApple-Lamp1) in  
755 live U2OS cells. RCAN output illustrating mitochondrial fission (orange arrowheads), mitochondrial  
756 fusion (white arrowheads), and mitochondrial-lysosomal contacts. Single lateral planes are shown. See  
757 also **Supplementary Video 3. g)** Graph shows quantification of fission, fusion, and contact events  
758 quantified from 16 cells. All scale bars: 5 μm, except 1 μm for higher magnification views shown in **c)**.  
759



760

761 **Fig. 2, RCAN resolution enhancement assayed with simulated spherical phantoms. a)** Noiseless mages  
 762 of simulated spherical phantoms were created (High Resolution) and blurred (Low Resolution),  
 763 generating matched volumes for RCAN training. Blurred volumes unseen by the trained network were  
 764 then tested to evaluate deblurring performance. **b)** Examples of RCAN, CARE, SRResNet and ESRGAN  
 765 performance on increasingly blurred data (blurred with a kernel 2x and 3x larger than the iSIM PSF used  
 766 for ground truth (GT) data). Axial (top row) and lateral (bottom row) cross sections are shown. Networks  
 767 are compared on the same test object, a sub-resolution sphere that approximates the iSIM PSF after  
 768 blurring (GT, shown in leftmost column). Scale bar: 40 pixels. See also **Supplementary Fig. 8. c)**  
 769 Additional examples of input data after progressively more severe blur (RAW, left column, with blurring  
 770 kernels 2x and 3x the size of the iSIM PSF indicated in successive rows). Ground truth and different

771 network outputs (right column) are also shown. Scale bar: 100 pixels, lateral (XY, top images) and axial  
772 slices (XZ, bottom images) along the dotted horizontal line are shown. Dotted rectangles and red arrows  
773 highlight features for comparison across the different networks. See also **Supplementary Fig. 9,**  
774 **Supplementary Videos 4-6. d)** SSIM (top) and PSNR (bottom) for data shown in **c)**. Means and standard  
775 deviations from 8 measurements are shown, see also **Supplementary Table 4.**



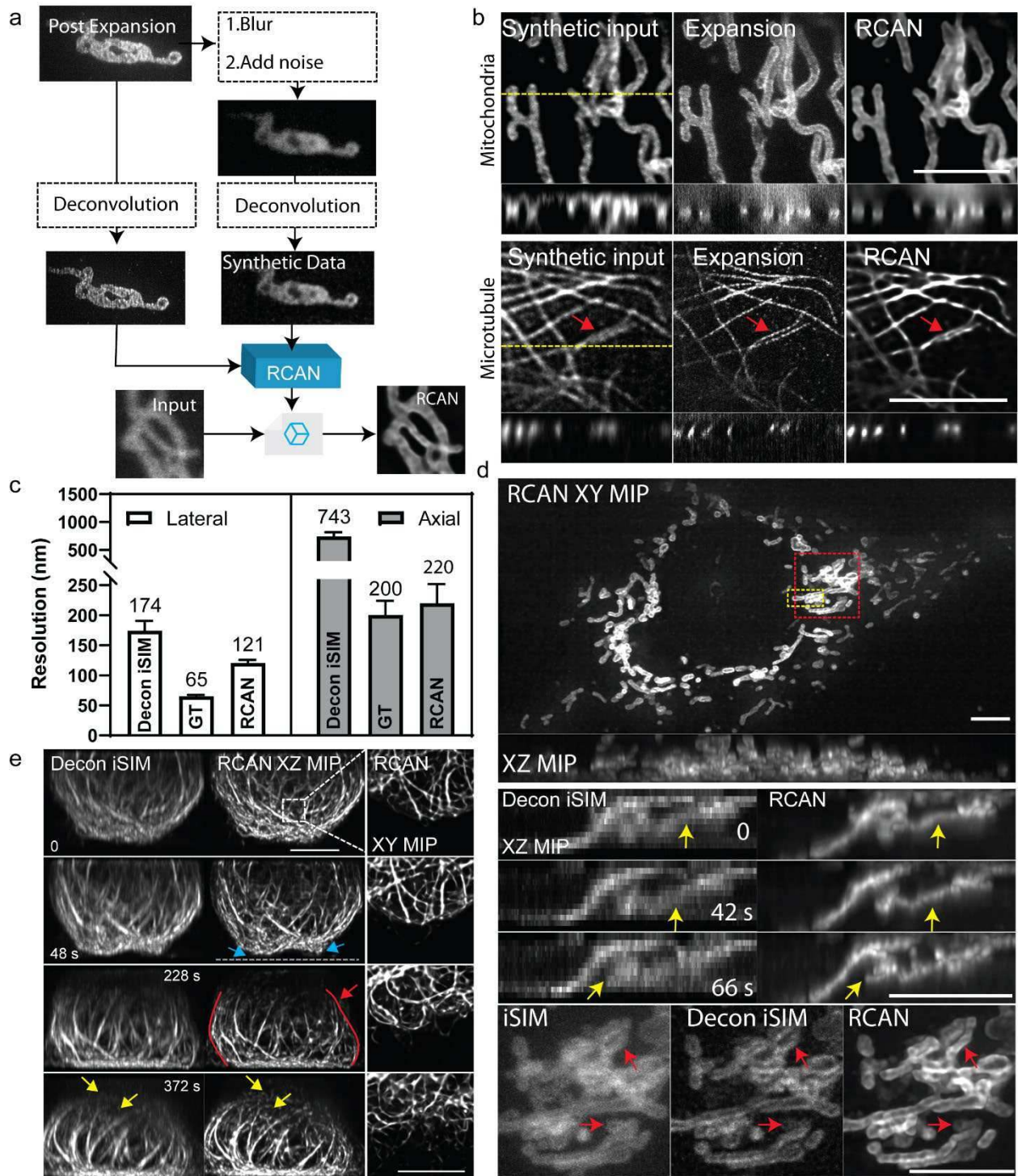
776

777 **Fig. 3, Confocal- to STED- microscopy restoration with RCAN.** **a**) Example confocal input (left), RCAN  
778 prediction (middle) and ground truth STED (right) images for fixed mouse embryonic fibroblast (MEF)  
779 cells with microtubules stained with ATTO647-secondary antibodies against anti- $\alpha$ -tubulin primary  
780 antibodies (**a**), nuclear pore complexes (NPCs) stained with Alexa Fluor 594-secondary antibodies against  
781 anti-NPC primary antibodies (**b**) and nuclei stained with SiR-DNA (**c**). Higher magnification views of the



782 dotted rectangular regions are shown below **a, b**, and axial reslices along yellow dotted lines marked in  
783 the lateral images are shown for **a-c**. Blue arrows in **b**) highlight areas of discrepancy between RCAN  
784 output and ground truth data while red arrows in **c**) are intended to highlight areas that are predicted  
785 well by RCAN but barely visible in the raw data. See also **Supplementary Fig. 10, 12, 13**. Phases of the  
786 cell cycle are also indicated in **c**). See also **Supplementary Fig. 14**. **(d)** Average image resolution in  
787 microtubule (left) and NPC (right) images obtained from decorrelation analysis. Means (also shown  
788 above each column) and standard deviations (from  $N = 18$  image planes) are shown for raw confocal  
789 input, ground truth STED, and RCAN output. **(e)** Live MEF cells stained with SiR-DNA were imaged in  
790 resonant confocal mode (top) and the RCAN model trained on fixed datasets similar to those shown in  
791 **(c)** was applied to yield predictions (bottom). Single planes from volumetric time series are shown See  
792 also **Supplementary Video 7, 8**. **f**) Higher magnification view from series in **e**) 2615 s after the start of  
793 imaging, corresponding to nuclei marked 1, 2 in **e**). Red arrows highlight areas absent SiR-DNA signal  
794 that are more easily defined in RCAN prediction vs. confocal data. All scale bars: 5  $\mu\text{m}$ .

795



797

798 **Fig. 4, Using expansion microscopy to improve spatial resolution in fixed and live instant structured**  
 799 **illumination microscopy (iSIM). a)** Simplified schematic showing generation of synthetic data used for  
 800 training RCAN network. Post-expansion data are acquired and deconvolved, generating ground truth  
 801 data (left). Post-expansion data are also blurred, noise is added, and the resulting images are

802 deconvolved to generate synthetic pre-expansion data (right). Ground truth and synthetic data are then  
803 used to train RCAN models for resolution enhancement on blurry input data (bottom). See also  
804 **Supplementary Fig. 19. b)** Example input data (either synthetic or experimental) not seen by network,  
805 mimicking deconvolved iSIM (left); expansion ground truth (middle); and RCAN predictions. Lateral and  
806 axial (taken along dotted line in lateral view) slices are shown for mitochondria (top, labeled with EGFP-  
807 Tomm20 in fixed, expanded U2OS cells) and microtubules (bottom, immunolabeled with Alexa Fluor 488  
808 secondary against anti- $\alpha$ -tubulin primary antibody in fixed, expanded U2OS cells). See also  
809 **Supplementary Fig. 20 and Supplementary Video 9. c)** Average resolution quantification from  
810 decorrelation analysis on microtubule samples. Lateral (left) and axial (right) values are shown for  
811 experimentally acquired deconvolved iSIM (left columns, 174 +/- 16 nm and 743 +/- 73 nm), ground  
812 truth expanded data (middle columns, 65 +/- 2 nm and 200 +/- 24 nm), and RCAN predictions (right  
813 columns, 120 +/- 5 nm and 220 +/- 31 nm). Note discontinuous representation of ordinate axis. Mean  
814 (shown also above each column) +/- standard deviations derived from  $N = 12$  images are shown. See  
815 also **Supplementary Fig. 21. d)** Images from live U2OS cells expressing EGFP-Tomm20 were imaged with  
816 iSIM, deconvolved, and input into the trained RCAN model. Top: Overview lateral and axial maximum  
817 intensity projections (MIP) of first volume in time-series from RCAN prediction. Middle: higher  
818 magnification views of axial slice corresponding to yellow rectangular region in overview, comparing  
819 deconvolved iSIM input (left) and RCAN output. Yellow arrows highlight mitochondria that are better  
820 resolved with RCAN output than input data. Bottom: higher magnification views of red rectangular  
821 region in overview, comparing raw iSIM, deconvolved iSIM, and RCAN prediction. Red arrows highlight  
822 mitochondria better resolved with RCAN than iSIM. See also **Supplementary Videos 10, 11. e)** Images  
823 from live Jurkat T cells expressing EMTB-3XGFP were deconvolved and used as input to trained RCAN  
824 model. Left: selected axial MIPs at indicated time points, comparing deconvolved iSIM vs. RCAN output.  
825 Right: lateral MIPs, corresponding to dashed rectangular region in lefthand images. Blue arrowheads  
826 indicates deformation of lower cell cortex prior to T cell spreading, red arrow indicates approximate  
827 location of centrosome, red lines indicate asymmetric deformation of microtubule bundles surrounding  
828 the nucleus, and yellow arrows indicate microtubule filaments at the top of the cell better defined with  
829 RCAN vs iSIM. See also **Supplementary Videos 12, 13.** All scale bars: 5  $\mu$ m.

830

- 831 1 Gustafsson, M. G. L. *et al.* Three-Dimensional Resolution Doubling in Wide-Field Fluorescence  
832 Microscopy by Structured Illumination. *Biophys. J.* **94**, 4957-4970 (2008).
- 833 2 Galland, R. *et al.* 3D high- and super-resolution imaging using single-objective SPIM. *Nat*  
834 *Methods* **12**, 641-644 (2015).
- 835 3 York, A. G. *et al.* Instant super-resolution imaging in live cells and embryos via analog image  
836 processing. *Nat Methods* **10**, 1122-1126 (2013).
- 837 4 Wu, Y. & Shroff, H. Faster, sharper, and deeper: structured illumination microscopy for biological  
838 imaging. *Nature Methods* **15**, 1011-1019 (2018).
- 839 5 Winter, P. W. & Shroff, H. Faster fluorescence microscopy: advances in high speed biological  
840 imaging. *Curr Opin Chem Biol.* **20**, 46-53 (2014).
- 841 6 Li, D. *et al.* Extended-resolution structured illumination imaging of endocytic and cytoskeletal  
842 dynamics. *Science* **349**, aab3500 (2015).
- 843 7 Laissue, P. P., Alghamdi, R. A., Tomancak, P. & Reynaud, E. G. S., H. Assessing phototoxicity in  
844 live fluorescence imaging. *Nat Methods* **14**, 657-661 (2017).
- 845 8 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).

846 9 Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence  
847 microscopy. *Nat Methods* **15**, 1090-1097 (2018).

848 10 Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image  
849 segmentation. *arXiv*, 1505.04597 (2015).

850 11 Goodfellow, I. J. *et al.* Generative Adversarial Networks. *arXiv*, 1406.2661 (2014).

851 12 Wang, H. *et al.* Deep learning enables cross-modality super-resolution in fluorescence  
852 microscopy. *Nat Methods* **16**, 103-110 (2019).

853 13 Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep Learning Massively Accelerates  
854 Super-Resolution Localization Microscopy. *Nature Biotechnol.* **36**, 460-468 (2018).

855 14 Fang, L. *et al.* Deep Learning-Based Point-Scanning Super-Resolution Imaging. *bioRxiv*, doi:  
856 <http://dx.doi.org/10.1101/740548> (2019).

857 15 Jin, L. *et al.* Deep learning enables structured illumination microscopy with low light levels and  
858 enhanced speed. *Nat Commun.* **11**, 1934 (2020).

859 16 Guo, M. *et al.* Rapid image deconvolution and multiview fusion for optical microscopy. *Nature*  
860 *Biotechnol.*, in press, doi:<https://doi.org/10.1101/647370> (2020).

861 17 Zhang, Y. *et al.* Image Super-Resolution Using Very Deep Residual Channel Attention Networks.  
862 *European Conference on Computer Vision*, 286-301 (2018).

863 18 Chen, F., Tillberg, P. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543-548 (2015).

864 19 Hu, J., Shen, L. & Sun, G. Squeeze-and-Excitation Networks. *IEEE/CVF Conference on Computer*  
865 *Vision and Pattern Recognition*, 7132-7141 (2018).

866 20 Ledig, C. *et al.* Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial  
867 Network. *IEEE Conference on Computer Vision and Pattern Recognition*, 4681-4690 (2017).

868 21 Wang, X. *et al.* ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks *arXiv*,  
869 1809.00219v00212 (2018).

870 22 Blau, Y., Mechrez, R., Timofte, R., Michaeli, T. & Zelnik-Manor, L. The 2018 PIRM Challenge on  
871 Perceptual Image Super-resolution. *arXiv*, 1809.07517 (2018).

872 23 Fadero, T. C. *et al.* LITE microscopy: Tilted light-sheet excitation of model organisms offers high  
873 resolution and low photobleaching. *J Cell Biol.* **217**, 1869-1882 (2018).

874 24 Descloux, A., Großmayer, K. S. & Radenovic, A. Parameter-free image resolution estimation  
875 based on decorrelation analysis. *Nature Methods* **16**, 918-924 (2019).

876 25 Lukinavičius, G. *et al.* SiR-Hoechst is a far-red DNA stain for live-cell nanoscopy. *Nat Commun.* **6**,  
877 8497 (2015).

878 26 Gambarotto, D. *et al.* Imaging cellular ultrastructures using expansion microscopy (U-ExM). *Nat*  
879 *Methods* **16**, 71-74 (2019).

880 27 Miller, A. L. & Bement, W. M. Regulation of cytokinesis by Rho GTPase flux *Nat Cell Biol* **11**, 71-  
881 77 (2009).

882 28 Bunnell, S. C., Kapoor, V., Triple, R. P., Zhang, W. & Samelson, L. E. Dynamic actin polymerization  
883 drives T cell receptor-induced spreading: a role for the signal transduction adaptor LAT.  
884 *Immunity* **14**, 315-329 (2001).

885 29 Yi, J. *et al.* Centrosome repositioning in T cells is biphasic and driven by microtubule end-on  
886 capture-shrinkage. *J Cell Biol.* **202**, 779-792. (2013).

887 30 Chollet, F. & others. *keras*, 2015).

888 31 Abadi, A. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogenous Distributed  
889 Systems. *arXiv*, 1603.04467 (2016).

890 32 He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *IEEE*  
891 *Conference on Computer Vision and Pattern Recognition*, doi:10.1109/CVPR.2016.90 (2016).

892 33 Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional  
893 Networks. *arXiv*, 1608.06993 (2016).

894 34 Jolicoeur-Martineau, A. The relativistic discriminator: a key element missing from standard GAN.  
895 *arXiv*, 1807.00734 (2018).

896 35 Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image  
897 Recognition. *ArXiv*, 1409.1556 (2014).

898 36 Jolesz, F. in *Intraoperative Imaging and Image-Guided Therapy*. (Springer, New York, NY,  
899 2014).

900 37 Pietzch, T., Preibisch, S., Tomancak, P. & Saalfeld, S. ImgLib2-generic image processing in Java.  
901 *Bioinformatics* **28**, 3009-3011 (2012).

902

# Figures

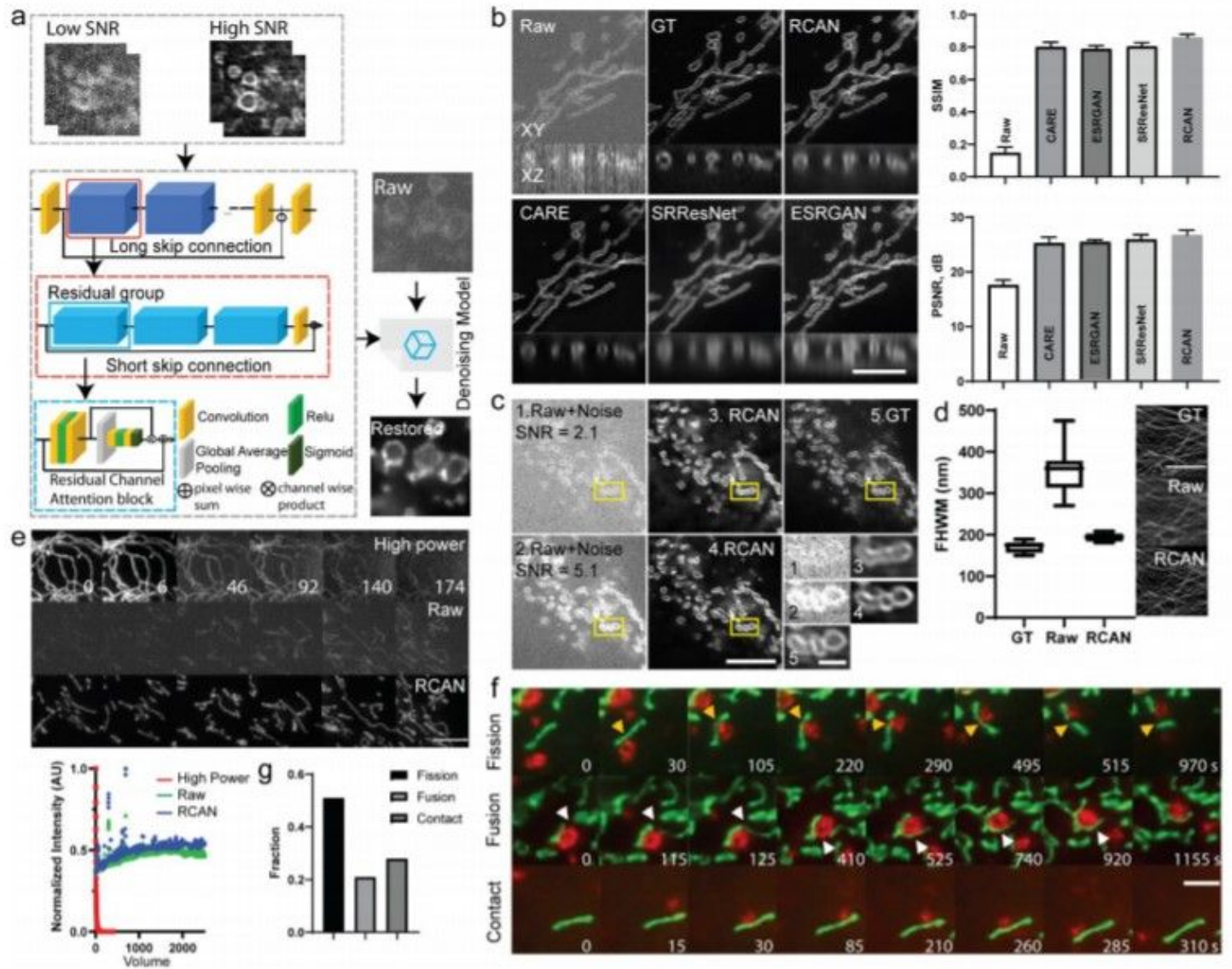


Figure 1

Residual channel attention networks denoise super-resolution data.

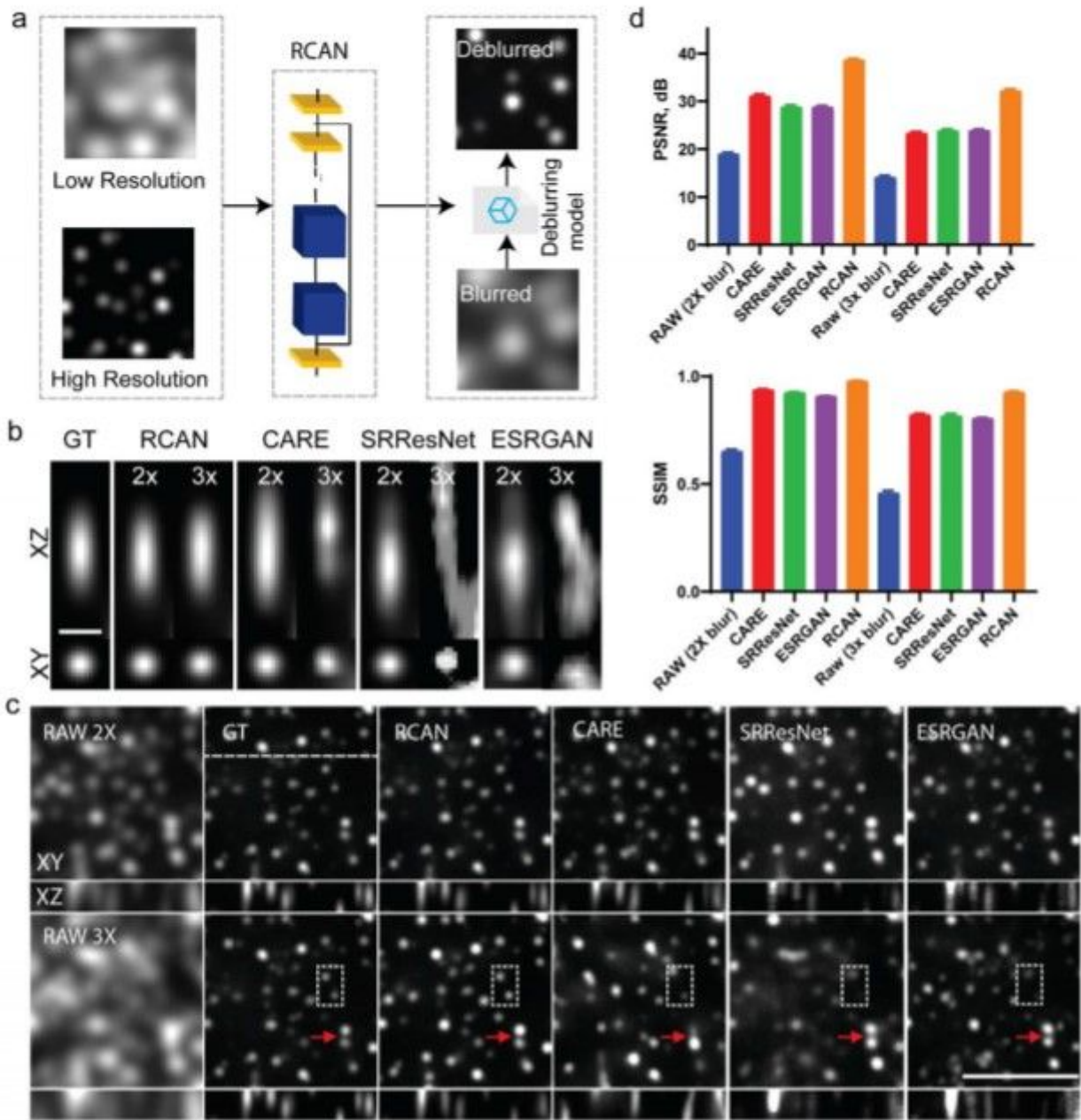
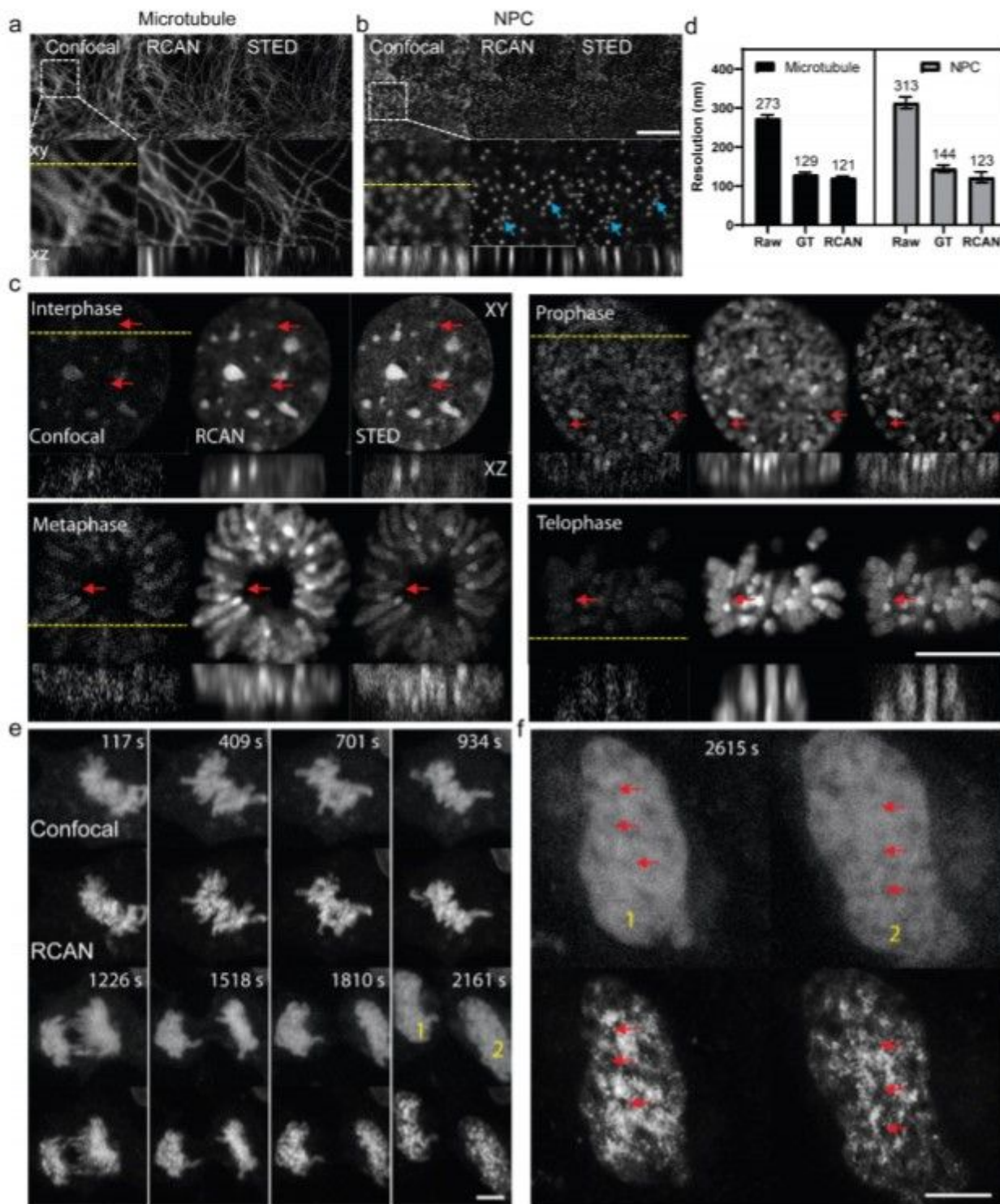


Figure 2

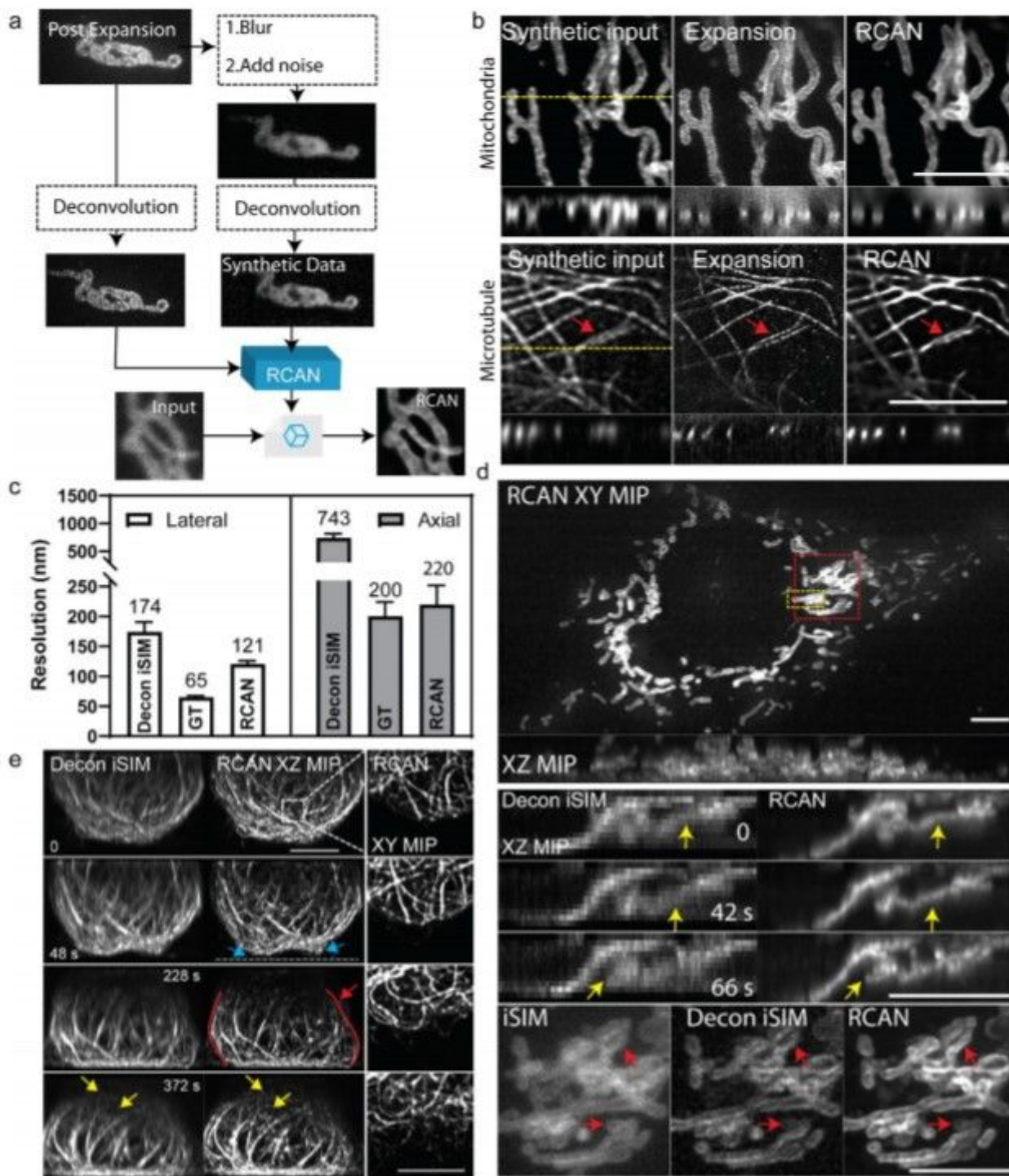
RCAN resolution enhancement assayed with simulated spherical phantoms



**Figure 3**

Confocal- to STED- microscopy restoration with RCAN





**Figure 4**

Using expansion microscopy to improve spatial resolution in fixed and live instant structured 799 illumination microscopy (iSIM).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SV1SupplementaryVideoHighpowerphotobleaching.avi
- SV2SupplementaryVideoMitoLongterm.avi
- SV3SupplementaryVideotwocolorMitoLysosome.avi
- SV4SupplementaryVideo2xdeblurXZ.avi
- SV5SupplementaryVideo3xdeblurXZ.avi
- SV6SupplementaryVideo4xdeblurXZ.avi
- SV7SupplementaryVideoconfocal2STEDDNA.avi
- SV8SupplementaryVideoconfocal2STEDDNAXYZ.avi
- SV9SupplementaryVideozstackcompareMTExpansion.avi
- SV10SupplementaryVideoFigure4dMitoexpansionxyz.avi
- SV11SupplementaryVideoMitochondriaExpansionlivecellZoomedXZ.avi
- SV12SupplementaryVideoFigure4eMTExpansionlivecellXZ.avi
- SV13SupplementaryVideoFigure4eMTExpansionlivecellXY.avi
- SV14SupplementaryVideoMTExpansionlivecellRCANiSIM1.avi
- SupMat20200828.pdf