

# Deep learning enables cross-modality super-resolution in fluorescence microscopy

Hongda Wang  <sup>1,2,3,9</sup>, Yair Rivenson <sup>1,2,3,9</sup>, Yiyin Jin  <sup>1</sup>, Zhensong Wei <sup>1</sup>, Ronald Gao <sup>4</sup>, Harun Günaydin <sup>1</sup>, Laurent A. Bentolila <sup>3,5</sup>, Comert Kural <sup>6,7</sup> and Aydogan Ozcan  <sup>1,2,3,8\*</sup>

We present deep-learning-enabled super-resolution across different fluorescence microscopy modalities. This data-driven approach does not require numerical modeling of the imaging process or the estimation of a point-spread-function, and is based on training a generative adversarial network (GAN) to transform diffraction-limited input images into super-resolved ones. Using this framework, we improve the resolution of wide-field images acquired with low-numerical-aperture objectives, matching the resolution that is acquired using high-numerical-aperture objectives. We also demonstrate cross-modality super-resolution, transforming confocal microscopy images to match the resolution acquired with a stimulated emission depletion (STED) microscope. We further demonstrate that total internal reflection fluorescence (TIRF) microscopy images of subcellular structures within cells and tissues can be transformed to match the results obtained with a TIRF-based structured illumination microscope. The deep network rapidly outputs these super-resolved images, without any iterations or parameter search, and could serve to democratize super-resolution imaging.

Super-resolution microscopy methods such as localization microscopy<sup>1–4</sup>, STED microscopy<sup>5</sup>, and structured illumination microscopy (SIM)<sup>6–8</sup> provide unprecedented access to the inner workings of cells and various biological processes. However, these methods often rely on relatively sophisticated optical setups, specific fluorophores and mounting media, and extensive computational post-processing of acquired image data<sup>9–11</sup>, which in and of itself may require a priori knowledge about the sample and/or its preparation, as well as a physical model of the image-formation process<sup>12–15</sup>, including, for example, the point-spread function (PSF) of the imaging system. In general, more accurate models yield higher-quality results, often with a trade-off of exhaustive parameter search and computational cost.

Here we present a deep-learning-based framework to achieve super-resolution and cross-modality image transformations in fluorescence microscopy without the need for making any assumptions about or modeling of the image-formation process. We trained a deep neural network using a generative adversarial network (GAN)<sup>16</sup> model to transform an acquired low-resolution image into a high-resolution one using matched pairs of experimentally acquired low- and higher-resolution images. The success of this super-resolution approach is a result of a highly accurate multi-stage image registration and alignment process (discussed in the Methods section) between the lower-resolution and corresponding higher-resolution images, which allows the network to solely focus on the task of improving the resolution of a previously unseen input image.

Once the deep network is trained, it remains fixed and can be used to rapidly output batches of high-resolution images in, for example, 0.4 s for an image size of 1,024 × 1,024 pixels using a single graphics processing unit (GPU). The network inference is

non-iterative and does not require a manual parameter search to optimize its performance.

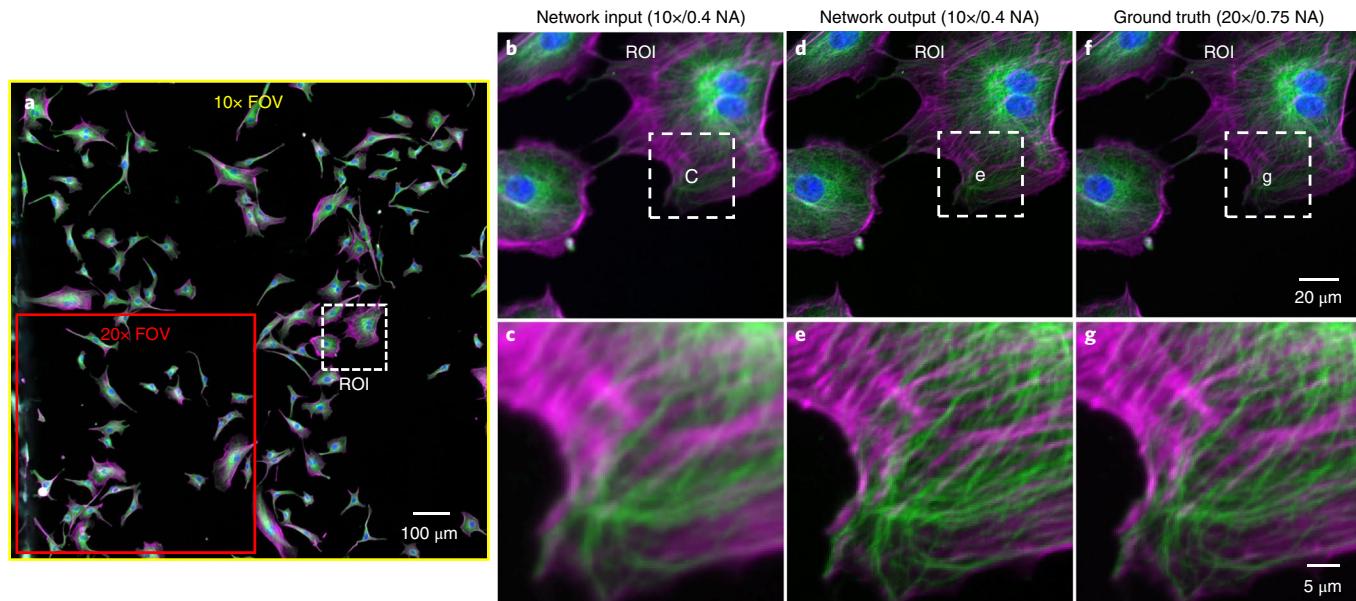
We demonstrate the success of this deep-learning-based framework by improving the resolution of raw images captured by different imaging modalities, including wide-field fluorescence, confocal, and TIRF microscopes. In the wide-field imaging case, we transformed the images acquired using a 10×/0.4-NA objective lens into resolution-enhanced images that matched the images of the same samples acquired with a 20×/0.75-NA objective. In the second case, we performed cross-modality transformation of diffraction-limited confocal microscopy<sup>17</sup> images to match the images that were acquired using a STED microscope, super-resolving Histone 3 distributions within HeLa cell nuclei and also showing a PSF width that improved from ~290 nm to ~110 nm. As another example of this GAN-based cross-modality image transformation framework, we super-resolved time-lapse TIRF microscopy images to match TIRF-SIM<sup>18</sup> images of endocytic clathrin-coated structures in SUM159 cells and *Drosophila* embryos. This deep-learning-based fluorescence super-resolution approach improves both the field of view (FOV) and imaging throughput of fluorescence microscopy and can be used to transform lower-resolution and wide-field images acquired using various imaging modalities into higher-resolution ones.

## Results

**Resolution enhancement in wide-field fluorescence microscopy.** We initially demonstrated the resolution improvement of the presented approach by imaging bovine pulmonary artery endothelial cell (BPAEC) structures. In the training stage, for each excitation line (DAPI, FITC, and TxRed) we used a multi-stage image registration process to accurately align 2,625 pairs of low- and

<sup>1</sup>Electrical and Computer Engineering Department, University of California, Los Angeles, CA, USA. <sup>2</sup>Bioengineering Department, University of California, Los Angeles, CA, USA. <sup>3</sup>California NanoSystems Institute, University of California, Los Angeles, CA, USA. <sup>4</sup>Computer Science Department, University of California, Los Angeles, CA, USA. <sup>5</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. <sup>6</sup>Department of Physics, Ohio State University, Columbus, OH, USA. <sup>7</sup>Biophysics Graduate Program, Ohio State University, Columbus, OH, USA. <sup>8</sup>Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. <sup>9</sup>These authors contributed equally: Hongda Wang and Yair Rivenson.

\*e-mail: ozcan@ucla.edu



**Fig. 1 | Deep-learning-based super-resolved images of bovine pulmonary artery endothelial cells (BPAECs).** **a**, Network input image acquired with a 10×/0.4-NA objective lens. **b–g**, Smaller ROIs are magnified and shown in **(b,c)** network input, **(d,e)** network output, and **(f,g)** ground truth (20×/0.75-NA). Experiments were repeated with >250 images, achieving similar results. Color map: magenta for F-actin, green for microtubules, blue for nuclei.

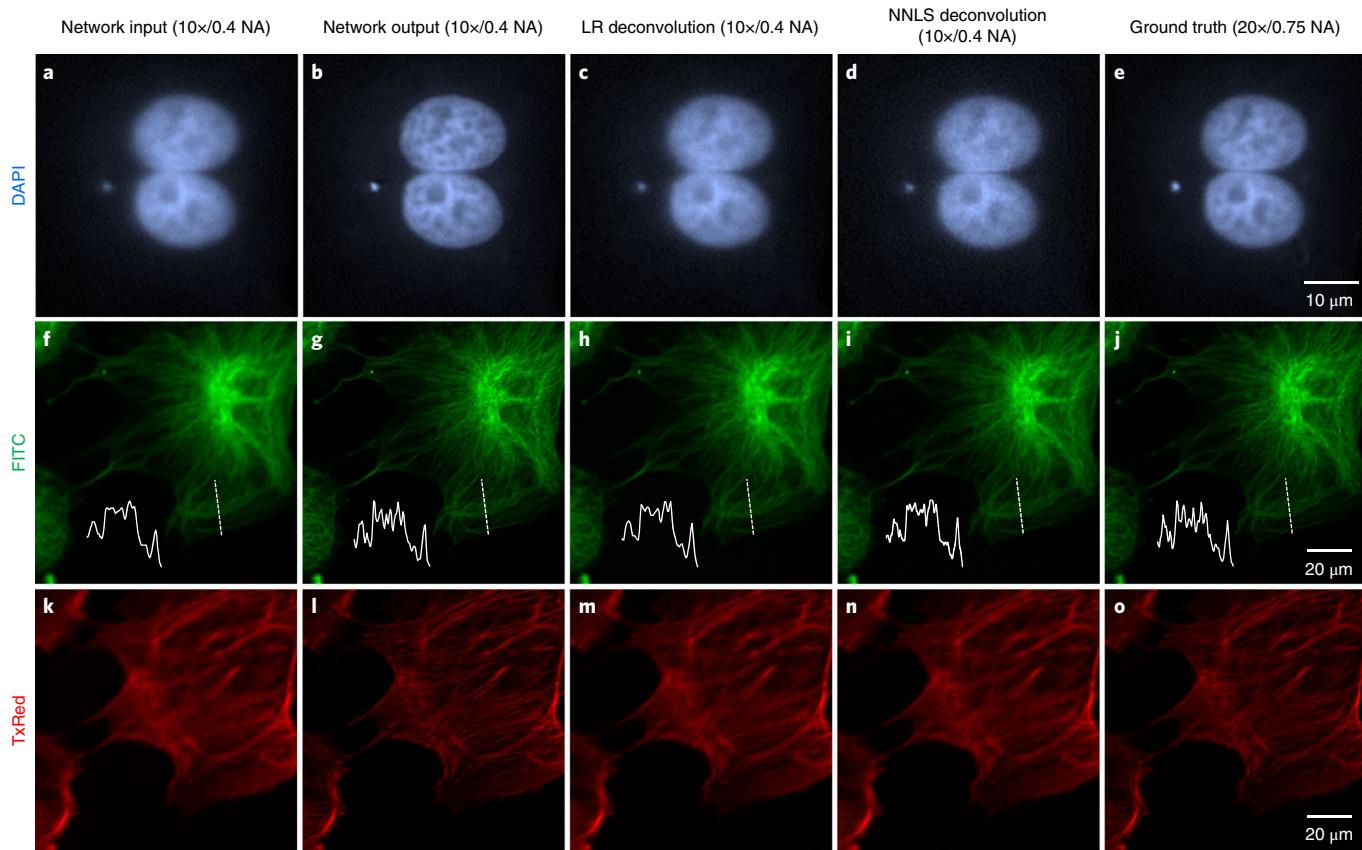
high-resolution image patches to each other, and we trained a separate model for each filter set to achieve optimal results (Methods). Each image patch had a size of  $1,024 \times 1,024$  pixels, and the raw input images to the network were acquired using a 10×/0.4-NA objective. The results of the network were compared against the ground truth images, which were captured using a 20×/0.75-NA objective. An example of the network input image is shown in Fig. 1a, where the FOV of the 10× and 20× objectives are also labeled. Figure 1b,c shows some zoomed-in regions of interest (ROIs) revealing further details of a cell's F-actin and microtubules. A pretrained deep neural network was applied to each color channel of these input images (10×/0.4-NA), outputting the resolution-enhanced images shown in Fig. 1d,e, where various features of F-actin, microtubules, and nuclei are clearly resolved at the network output, providing very good agreement with the ground truth images (20×/0.75-NA) shown in Fig. 1f,g. Note that all the network output images shown in this article were blindly generated by the deep network, that is, the input images were not previously seen by the network.

Next, we compared the results of deep-learning-based super-resolution against widely used image deconvolution methods, specifically, the Lucy–Richardson (LR) deconvolution and the non-negative least square (NNLS) algorithm<sup>19–21</sup>. For this, we used an estimated model of the PSF of the imaging system, which is required by these deconvolution algorithms to approximate the forward model. Following its parameter optimization (Methods), the LR deconvolution algorithm, as expected, demonstrated resolution improvements compared to the input images (Fig. 2a,f,k); however, compared to our deep learning results (Fig. 2b,g,l), the improvements observed with LR deconvolution (Fig. 2c,h,m) were modest, despite the fact that it used parameter search, optimization, and a priori knowledge on the PSF of the imaging system. The NNLS algorithm, in contrast, yielded slightly sharper features (Fig. 2d,i,n) compared to LR deconvolution results, at the cost of having additional artifacts as shown in Supplementary Fig. 1; regardless, both of these deconvolution methods are inferior to our deep learning results reported in Fig. 2, exhibiting a shallower modulation depth in comparison to the deep learning results and the ground truth images.

We also noticed that the deep network output image shows sharper details compared to the ground truth image, especially for

the F-actin structures. This result is in line with the fact that all the images were captured by finding the autofocusing plane within the sample using the FITC channel (see, for example, Fig. 2f–j), and therefore the Texas-Red channel (for example, Fig. 2k–o) can remain slightly out of focus owing to the thickness of the cells. This means the shallow depth of field (DOF) of a 20×/0.75-NA objective (~1.4 μm) might have caused some blurring in the F-actin structures (Fig. 2o). This out-of-focus imaging of different color channels did not affect the network output as much because the input image to the network was captured with a much larger DOF (~5.1 μm), using a 10×/0.4-NA objective. Therefore, in addition to an increased FOV resulting from a low-NA input image, the network output image is also benefiting from an increased DOF, helping to reveal some finer features that might be out of focus in different color channels with a high-NA objective.

Next, we tested the generalization of our trained network model in improving image resolution on new types of samples that were not present in the training phase; Supplementary Note 1 summarizes the success of our results. Here, we emphasize that a new network model should be trained for optimal super-resolution performance on input images corresponding to different types of samples, or captured with a new experimental setup. However, in case such training image pairs are not available to follow our super-resolution image transformation framework, one can attempt to use an existing trained model, although this might not produce ideal results in all cases. To exemplify such a scenario where training image pairs are not available, we used the network model trained with only the images of F-actin captured with the Texas Red (TxRed) filter set to blindly super-resolve the images captured with DAPI and FITC filter sets (Supplementary Fig. 2a–h). Compared with the optimal network models trained with the images acquired with the right filter sets, the model that was trained using a different filter set (TxRed) was still able to infer almost identical images, although it was applied on input images that were captured using a different filter set. In fact, even if the imaging modalities and sample scales are different, the wide-field TxRed model might still be used to improve the images of other microscopy modalities, for example, TIRF and confocal microscopy as shown in Supplementary Fig. 2i–p; however, the image inference performance in these cases cannot match



**Fig. 2 | Comparison of deep learning results against Lucy-Richardson (LR) and non-negative least square (NNLS) image deconvolution algorithms.**

Also see Supplementary Fig. 10. Experiments were repeated with >250 images, achieving similar results.

the results obtained with the optimal model, which is trained on the same imaging platform and same type of samples.

We also quantified the deep network inference results using spatial frequency spectrum analysis and successfully demonstrated the frequency extrapolation feature of our deep learning framework, as detailed in Supplementary Note 2. To further quantify the improvement achieved using our approach, we imaged 20-nm fluorescent beads, and using a model trained only with F-actin images, we extracted the PSFs from individual nano-beads to demonstrate the resolution improvement and the enhanced DOF of our network output images (Supplementary Note 3).

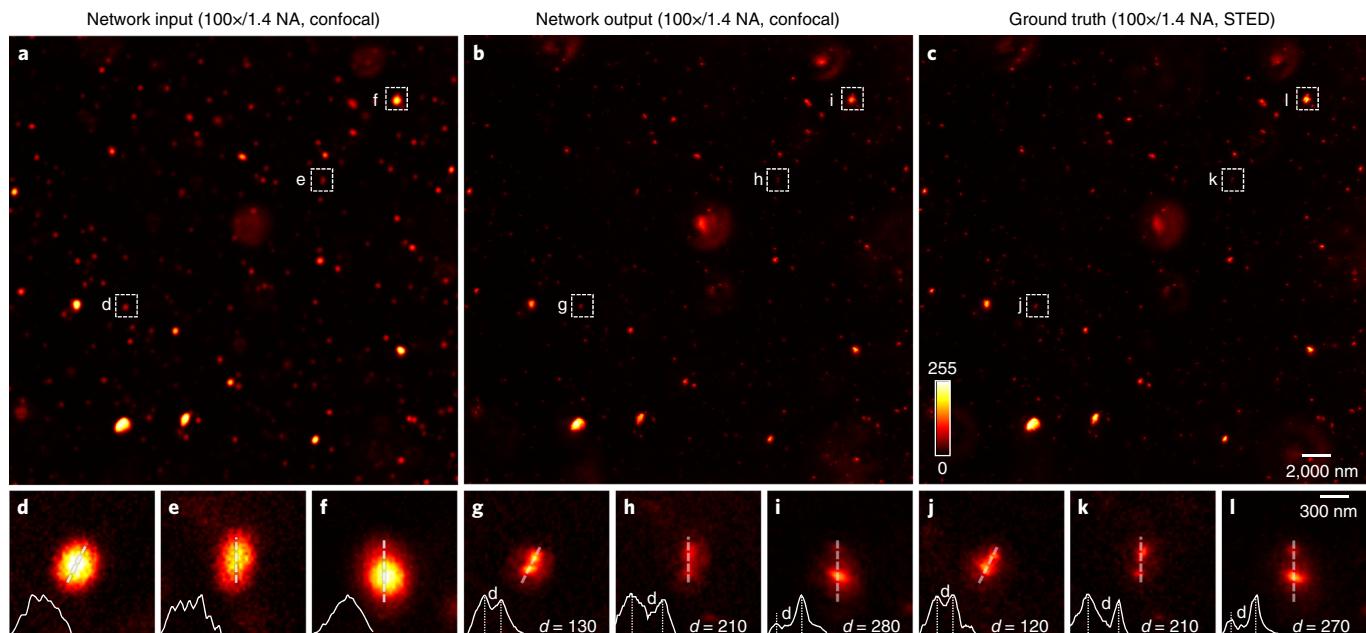
**Cross-modality imaging from confocal to STED.** We also applied the presented framework to transform confocal microscopy images into images that match those obtained by STED microscopy (Figs. 3 and 4; Supplementary Note 4). Training data were acquired using 20-nm fluorescent beads (645-nm emission) imaged on the same instrument using both confocal microscopy and STED modes. After the training phase, the neural network, as before, blindly takes an input image (confocal) and outputs a super-resolved image that matches the STED image of the same sample. Some of the nano-beads in our samples were spaced close to each other, within the classical diffraction limit, that is, under ~290 nm, as shown in, for example, Fig. 3d–f, and therefore could not be resolved in the raw confocal microscopy images. The neural network resolved these closely spaced nano-particles, providing a good match to STED images of the same regions of the sample (see Fig. 3g–i vs. Fig. 3j–l).

To further quantify this resolution improvement achieved by the network, we measured the PSFs arising from the images of single/isolated nano-beads across the sample FOV<sup>22</sup>, repeated for >400 individual nanoparticles that were tracked in the images of the

confocal microscope and STED microscope, as well as the network output image (in response to the confocal image). The results are summarized in Fig. 4, where the FWHM of the confocal microscope PSF is centered at ~290 nm, roughly corresponding to the lateral resolution of a diffraction-limited imaging system at an emission wavelength of 645 nm. As shown in Fig. 4, the PSF FWHM distribution of the network output provides a very good match to the PSF results of the STED system, with a mean FWHM of ~110 nm versus ~120 nm, respectively. Also see Supplementary Notes 4 and 5 for related discussions, revealing the spatially varying PSF information that is indirectly learned at the end of the training phase of this confocal-to-STED cross-modality network, without the need for prior information on, for example, the image formation model or sensor-specific noise patterns, which are typically required for standard deconvolution and localization methods.

An additional benefit of using our deep learning approach is improved SNR, for which we conducted a comparative analysis using the confocal-to-STED transformation results to quantify this improvement. Supplementary Note 6 further details that the deep neural network suppresses noise and improves the SNR compared to the input (confocal) and the ground truth (STED) images.

Next, we applied this confocal-to-STED image transformation framework to super-resolve Histone 3 distributions within fixed HeLa cell nuclei (see Fig. 5). Because nanoparticles do not accurately represent the spatial feature diversity observed in biological specimens, direct application of a network that is trained only with nano-beads would not be ideal to image complex biological systems (see Supplementary Fig. 3b). Therefore, we made use of a concept known as transfer learning<sup>23</sup>, in which a learned neural network (trained, for example, with nanoparticles; Figs. 3 and 4) was used to initialize a model to super-resolve cell nuclei using

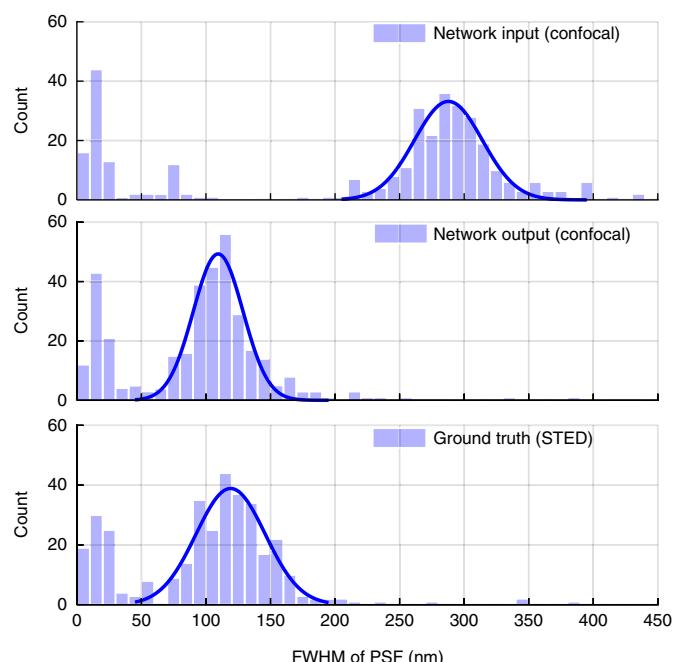


**Fig. 3 | Image resolution improvement beyond the diffraction limit: from confocal microscopy to STED.** **a–c**, A diffraction-limited confocal microscope image is used as input to the network and is super-resolved to blindly yield (**b**) the network output, which is comparable to (**c**) a STED image of the same FOV, used as the ground truth. **d–f**, Examples of closely spaced nano-beads that cannot be resolved by confocal microscopy. **g–l**, The trained neural network takes **d–f** as input and resolves the individual beads (**g–i**), very well agreeing with STED microscopy images (**j–l**). The cross-sectional profiles reported in **d–l** are extracted from the original images. Peak-to-peak distance ( $d$ ) in these cross-sectional profiles is reported in nanometers. Also see Fig. 5 for further quantification of the performance of the deep network on confocal images, and its comparison to STED. Experiments were repeated with 75 images, achieving similar results.

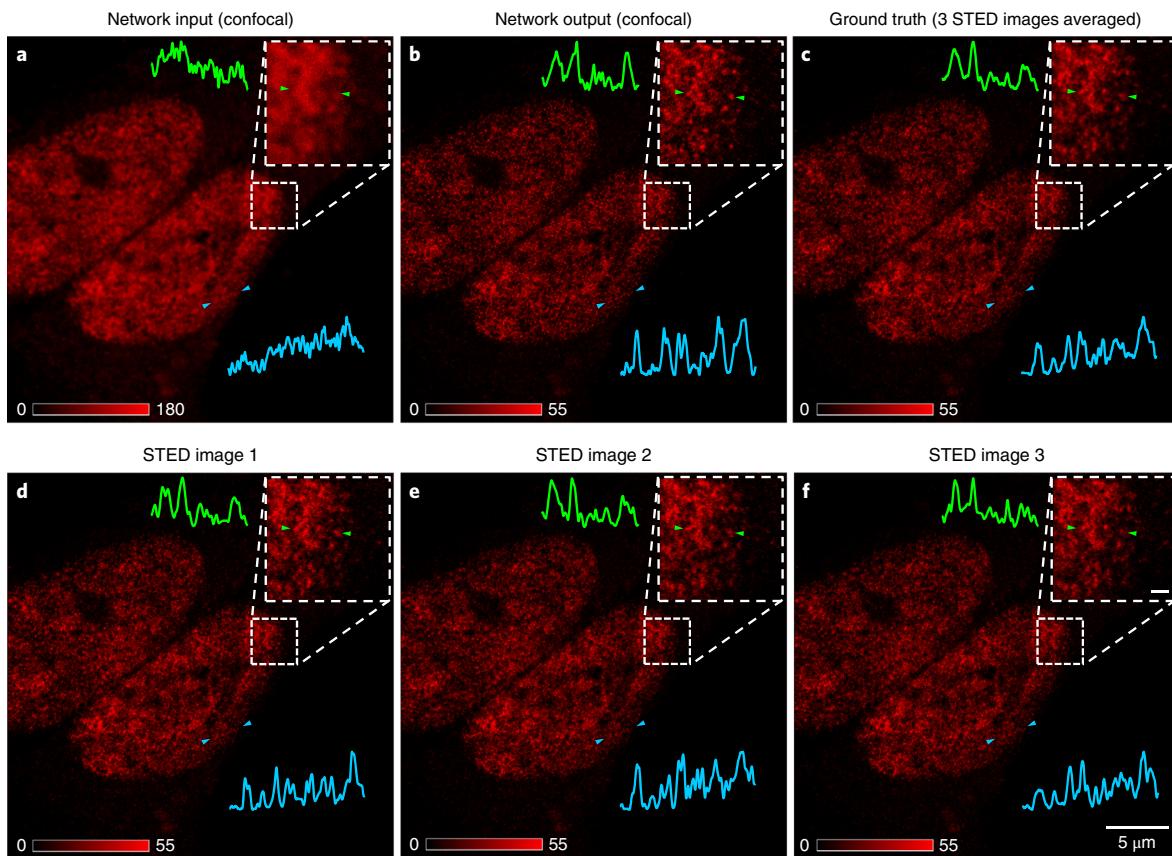
confocal-to-STED transformation; this transfer learning approach also significantly speeds up the training process, as detailed in the Methods section. Despite some challenges associated with STED imaging of densely labeled specimens and sample drift, after transfer learning, the neural network successfully improved the resolution of a confocal microscope image (input), matching the STED image of the same nuclei (Fig. 5). Some of the discrepancies between the network output and the STED image can be related to the fluctuations observed in STED imaging, as shown in Fig. 5d–f, where three consecutive STED scans of the same FOV show frame-to-frame variations due to fluorophore state changes and sample drift. In this case, the network's output image better correlates with the average of three STED images that are drift-corrected (see Fig. 5b,c). Using the same confocal-STED experimental data, Supplementary Fig. 4 further illustrates the advantages of the presented GAN-based super-resolution approach over a standard CNN (convolutional neural network) without the discriminative loss, which results in a lower-resolution image compared to GAN-based inference.

We also emphasize that in the experiments reported in Figs. 3–5, the required excitation power for STED was threefold to ten fold stronger than that of confocal microscopy (Methods). Furthermore, the depletion beam of STED is typically orders of magnitude higher than its excitation beam<sup>24–26</sup>, which highlights an important advantage of our deep-learning-based super-resolution approach for imaging biological objects that are vulnerable to photo-bleaching or photo-toxicity<sup>24,27</sup>.

**Cross-modality imaging from TIRF to TIRF-SIM.** We further demonstrated the cross-modality image transformation capability of our method by transforming diffraction-limited TIRF images to match TIRF-SIM reconstructions (Fig. 6 and Supplementary Fig. 5). In these experiments, the sample was exposed to nine different structured illumination patterns following a reconstruction method



**Fig. 4 | PSF characterization, before and after the network, and its comparison to STED.** We extracted more than 400 bright spots from the same locations of the network input (confocal), network output (confocal), and the corresponding ground truth (STED) images. Each one of these spots was fit to a 2D Gaussian function, and the corresponding FWHM distributions are shown in each histogram. These results show that the resolution of the network output images is significantly improved from ~290 nm (top row: network input using a confocal microscope) to ~110 nm (middle row: network output), which provides a very good fit to the ground truth STED images of the same nano-particles, summarized in the bottom row.



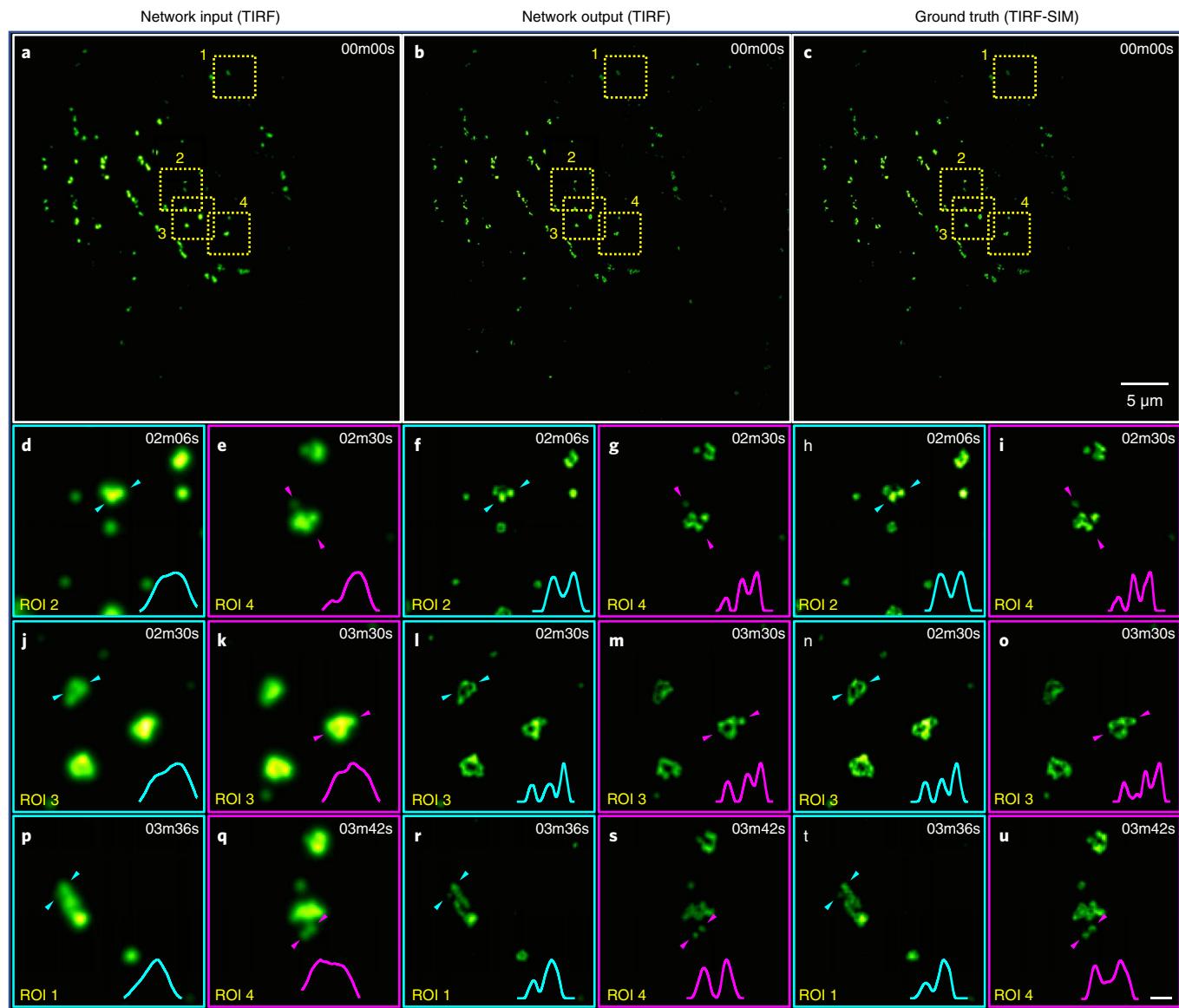
**Fig. 5 | Deep-learning enabled cross-modality image transformation from confocal to STED.** **a–c**, A diffraction-limited confocal microscope image (**a**) of Histone 3 distributions within HeLa cell nuclei is used as input to the neural network to blindly yield (**b**) the network output image, which is comparable to (**c**) a STED image of the same FOV. **d,e,f**, Three individual STED scans of the same FOV, averaged to create panel **c**. Scale bar in the inset in **f** is 500 nm and applies to all insets. Arrows in each image refer to the line of the shown cross-section. Experiments were repeated with 30 images, achieving similar results.

used in SIM<sup>18</sup>, whereas the low-resolution (diffraction-limited) TIRF images were obtained using a simple average of these nine exposures<sup>28</sup>. We trained our neural network model using images of gene-edited SUM159 cells expressing eGFP-labeled clathrin adaptor AP2, and blindly tested its inference (Fig. 6 and Supplementary Video 1). To highlight some examples, the neural network was able to detect the dissociation of clathrin-coated pits from larger clathrin patches (i.e., plaques<sup>18,29</sup>) as shown in Fig. 6r,t, as well as the development of curvature-bearing clathrin cages<sup>18,30</sup>, which appear as doughnuts under SIM (Fig. 6l–o). Next, to provide another demonstration of the network's generalization, we blindly applied it to amnioserosa tissues of *Drosophila* embryos (never seen by the network) expressing clathrin-mEmerald (Supplementary Fig. 5). Highly motile clathrin-coated structures<sup>31</sup> within the embryo that cannot be resolved in the original TIRF image can be clearly distinguished as separate objects in the network output (Supplementary Fig. 5). These results demonstrate that our network model can super-resolve individual clathrin-coated structures within cultured cells and tissues of a developing metazoan embryo.

We note that the aberrations or artifacts potentially observed in some of the ground truth training images can couple back into the network's inference and result in some residual artifacts in the network output. If the ground truth training image set is not dominated with such artifacts, the impact of this will be negligible, close to the noise floor of the output image, as illustrated in our Supplementary Protocol. Such residual artifacts can be further reduced by pre-selection of the training ground truth images to be free from major artifacts (if possible) or through an additional loss term applied to suppress such features during the training process.

**Depth-of-field enhancement.** Another important feature of the deep network-based image transformation approach is that it can resolve features over an extended DOF because of the lower NA of the input image (Fig. 2, Supplementary Figs. 6–8, and Supplementary Note 3). We further illustrated this phenomenon by acquiring a depth-resolved image set (composed of 34 images, axially separated by 0.3 μm) corresponding to the blood-vessel sample using a 20×/0.75-NA objective, and synthesized an extended-DOF image using the ImageJ plugin EDF<sup>32</sup>, which provides a significantly improved ground truth image compared to a single high-resolution image. These results and the comparison reported in Supplementary Fig. 9 clearly demonstrate the extended-DOF capabilities of our super-resolution method. This extended DOF is also favorable in terms of photo-damage to the sample, by eliminating the need for a fine axial scan within the sample volume, which might reduce the overall light delivered to the sample, while also making the imaging process more efficient. Although some thicker samples will ultimately require axial scanning, the presented approach will still reduce the number of scans required by inferring high-resolution images from parts of the sample that would have been defocused with higher-NA imaging systems (Supplementary Figs. 6 and 7).

**Artifact analysis.** A common concern for computational approaches that enhance image resolution is the potential emergence of spatial artifacts that may degrade the image quality, such as the Gibbs phenomenon in LR deconvolution<sup>33</sup>. To explore this, we randomly selected an example in the test image dataset, and quantified the artifacts of the network output using the NanoJ-Squirrel Plugin;<sup>13</sup> this analysis (Supplementary Note 7) revealed that the network



**Fig. 6 | Deep-learning enabled cross-modality image transformation from TIRF to TIRF-SIM.** **a**, TIRF image of a gene-edited SUM159 cell expressing AP2-eGFP. **b,c**, The network model super-resolves the diffraction-limited TIRF image (input; **b**) and matches TIRF-SIM reconstruction results (**c**). **d–u**, Zoomed-in regions of **a–c** at the labeled ROIs and time points. Supplementary Video 1 shows the same FOV to further highlight the success of the network's inference as a function of temporal dynamics within the cell. Also see Supplementary Fig. 5 for additional results, demonstrating the generalizability of the inference of this network on a new type of sample (amnioserosa tissues of a *Drosophila* embryo) that it has never seen before. Scale bar in **u** represents 500 nm and applies to **d–u**. Arrows in each image refer to the line of the shown cross-section. Also see Supplementary Fig. 11. Experiments were repeated with >1,000 images, achieving similar results.

output does not generate noticeable super-resolution artifacts and in fact has the same level of spatial mismatch error that the ground truth image has with respect to the input image of the same sample (see Supplementary Fig. 1 and Supplementary Note 7). This conclusion is further confirmed by Supplementary Fig. 1, which overlays the network output image and the ground truth image in different colors, revealing no obvious feature mismatch between the two. The same conclusion remained consistent for other test images as well.

Furthermore, we also calculated the difference of the network inference and the ground truth images for all the modalities used in our manuscript (Figs. 2–6), to demonstrate the high degree of spatial agreement between the two (Supplementary Note 8 and Supplementary Figs. 10 and 11); these results also indicate that the minor differences between the network output and

ground truth images are partially due to the extended DOF of our output images.

As an additional inquiry of potential artifacts, Supplementary Note 2 reports spatial frequency spectrum analysis to demonstrate the agreement between the spatial frequencies of the network output and the ground truth images, which further supports the success of our inference results.

## Discussion

Our deep learning approach allows for the generation of super-resolution images directly from images acquired on conventional, diffraction-limited microscopes without a priori knowledge about the sample and/or the image formation process. In addition to democratizing super-resolution microscopy, our approach offers the benefits of rapidly imaging larger FOVs and DOFs, creating

higher-resolution images with fewer frames and/or lower light doses, which enables new opportunities for imaging objects with reduced photo-bleaching and photo-toxicity<sup>24,27</sup>.

An essential step of the presented super-resolution framework is the accurate alignment and registration between the lower-resolution and the higher-resolution label images. This multi-stage image registration process (Methods) allows the network to learn a pixel-to-pixel transformation and is used as a regularization for the network to learn the resolution enhancement, while avoiding warping of the input images, which in turn significantly reduces potential artifacts. This data-driven cross-modality transformation framework is further discussed in Supplementary Note 5 with an emphasis on the fact that the input and output distributions share a high degree of mutual information, with an output probability distribution that is conditional upon the input data distribution.

As illustrated in Supplementary Note 6, our deep learning approach also improves the image SNR. In fact, the resolution limit of a microscopy modality is fundamentally limited by its SNR<sup>34</sup>; stated differently, the lack of some spatial frequencies at the image plane (for example, carried by evanescent waves) does not pose a fundamental limit for the achievable resolution of a computational microscope. These missing spatial frequencies (although not detected at the image) can in principle be extrapolated based on the measured or known spatial frequencies of an object<sup>34</sup>. For example, the full spatial frequency spectrum of an object function that has a limited spatial extent with finite energy can in theory be recovered from the partial knowledge of its spectrum using the analytical continuation principle, as its Fourier transform defines an entire function<sup>35</sup>. In practice, however, this is a challenging task and the success of such a frequency extrapolation method is strongly dependent on the SNR of the measured image information and a priori information regarding the object. Although the presented neural-network-based super-resolution approach does not include any such analytical continuation models or any a priori assumptions about the known frequency bands or support information of the object, through image data it learns to statistically separate out noise patterns from the structural information of the object, helping us achieve effectively much improved frequency extrapolation (Supplementary Note 2) and resolution enhancement compared to the state-of-the-art methods as reported in our Results.

To practice our approach on new types of samples or new imaging systems that were not part of the training process, fresh application of our presented framework is recommended for getting optimal results, starting with the image registration between the input images (lower resolution) and the desired labels (higher resolution), followed by the training of a GAN, as detailed in the Methods section. Transfer learning from a previously trained network for another type of sample might speed up the convergence of this learning process; however, this is neither a required step nor a replacement for the entire image registration and GAN training processes performed on new sample types of interest. After a sufficiently large number of training iterations (for example, >10,000), the optimal network model can be selected when the validation loss value no longer decreases.

Taken together, our work represents an important step forward for the fields of computational microscopy and super-resolution imaging, and should help us democratize high-resolution imaging systems, potentially enabling new biological observations beyond what can be achieved in well-resourced institutions and laboratory settings. Our ability to close the gap between lower-resolution and higher-resolution imaging systems using a deep learning framework is fundamentally tied to image SNR in both the training and blind testing phases, and in this sense the presented image transformation framework is limited in its performance by noise, very much like all the other super-resolution imaging modalities.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41592-018-0239-0>.

Received: 23 April 2018; Accepted: 5 November 2018;

Published online: 17 December 2018

## References

- Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
- Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
- van de Linde, S. et al. Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat. Protoc.* **6**, 991–1009 (2011).
- Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
- Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, 82–87 (2000).
- Cox, S. Super-resolution imaging in live cells. *Dev. Biol.* **401**, 175–181 (2015).
- Gustafsson, M. G. L. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc. Natl. Acad. Sci. USA* **102**, 13081–13086 (2005).
- Henriques, R. et al. QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat. Methods* **7**, 339–340 (2010).
- Small, A. & Stahlheber, S. Fluorophore localization algorithms for super-resolution microscopy. *Nat. Methods* **11**, 267–279 (2014).
- Abraham, A. V., Ram, S., Chao, J., Ward, E. S. & Ober, R. J. Quantitative study of single molecule location estimation techniques. *Opt. Express* **17**, 23352–23373 (2009).
- Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011).
- Culley, S. et al. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat. Methods* **15**, 263–266 (2018).
- Sage, D. et al. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12**, 717–724 (2015).
- Almada, P., Culley, S. & Henriques, R. PALM and STORM: into large fields and high-throughput microscopy with sCMOS detectors. *Methods* **88**, 109–121 (2015).
- Goodfellow, I. J. et al. Generative adversarial networks. *arXiv Preprint at https://arxiv.org/abs/1406.2661* (2014).
- Wilson, T. & Masters, B. R. Confocal microscopy. *Appl. Opt.* **33**, 565–566 (1994).
- Li, D. et al. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).
- Richardson, W. H. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* **62**, 55 (1972).
- Lucy, L. B. An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745 (1974).
- Landweber, L. An iteration formula for Fredholm integral equations of the first kind. *Am. J. Math.* **73**, 615–624 (1951).
- Farahani, J. N., Schibler, M. J. & Bentolila, L. A. Stimulated emission depletion (STED)microscopy: from theory to practice. *Microsc. Sci. Technol. Appl. Educ.* **2**, 1539–1547 (2010).
- Hamel, P., Davies, M. E. P., Yoshii, K. & Goto, M. Transfer learning in MIR: sharing learned latent representations for music audio classification and similarity. *Google AI* <https://ai.google/research/pubs/pub41530> (2013).
- Wälchen, S., Lehmann, J., Klein, T., van de Linde, S. & Sauer, M. Light-induced cell damage in live-cell super-resolution microscopy. *Sci. Rep.* **5**, 15348 (2015).
- Hein, B., Willig, K. I. & Hell, S. W. Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell. *Proc. Natl. Acad. Sci. USA* **105**, 14271–14276 (2008).
- Hein, B. et al. Stimulated emission depletion nanoscopy of living cells using SNAP-tag fusion proteins. *Biophys. J.* **98**, 158–163 (2010).
- Dyba, M. & Hell, S. W. Photostability of a fluorescent marker under pulsed excited-state depletion through stimulated emission. *Appl. Opt.* **42**, 5123–5129 (2003).
- Kner, P., Chhun, B. B., Griffis, E. R., Winoto, L. & Gustafsson, M. G. L. Super-resolution video microscopy of live cells by structured illumination. *Nat. Methods* **6**, 339–342 (2009).

29. Leyton-Puig, D. et al. Flat clathrin lattices are dynamic actin-controlled hubs for clathrin-mediated endocytosis and signalling of specific receptors. *Nat. Commun.* **8**, 16068 (2017).
30. Fiolka, R., Shao, L., Rego, E. H., Davidson, M. W. & Gustafsson, M. G. L. Time-lapse two-color 3D imaging of live cells with doubled resolution using structured illumination. *Proc. Natl. Acad. Sci. USA* **109**, 5311–5315 (2012).
31. Ferguson, J. P. et al. Deciphering dynamics of clathrin-mediated endocytosis in a living organism. *J. Cell. Biol.* **214**, 347–358 (2016).
32. Forster, B. et al. M. Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images. *Microsc. Res. Tech.* **65**, 33–42 (2004).
33. Liu, R. & Jia, J. Reducing boundary artifacts in image deconvolution. in *2008 15th IEEE International Conference on Image Processing* 505–508 (IEEE, New York, 2008).
34. Cox, I. J. & Sheppard, C. J. R. Information capacity and resolution in an optical system. *J. Opt. Soc. Am. A* **3**, 1152–1158 (1986).
35. Katznelson, Y. *An Introduction to Harmonic Analysis* (Dover Publications, New York, 1976).

## Acknowledgements

The Ozcan Research Group at UCLA acknowledges the support of NSF Engineering Research Center (ERC, PATHS-UP), the Army Research Office (ARO; W911NF-13-1-0419 and W911NF-13-1-0197), the ARO Life Sciences Division, the National Science Foundation (NSF) CBET Division Biophotonics Program, the NSF Emerging Frontiers in Research and Innovation (EFRI) Award, the NSF INSPIRE Award, NSF Partnerships for Innovation: Building Innovation Capacity (PFI:BIC) Program, the National Institutes of Health (NIH, R21EB023115), the Howard Hughes Medical Institute (HHMI), Vodafone Americas Foundation, the Mary Kay Foundation, and Steven & Alexandra Cohen Foundation. Yair Rivenson is partially supported by the European Union's Horizon 2020 research and innovation programme under the Marie

Sklodowska-Curie grant agreement No H2020-MSCA-IF-2014-659595 (MCMQCT). Confocal and STED laser scanning microscopy was performed at the California NanoSystems Institute (CNSI) Advanced Light Microscopy/Spectroscopy Shared Resource Facility at UCLA. We also thank the Advanced Imaging Center (AIC) at Janelia Research Campus for access to their TIRF-SIM microscope. The AIC is jointly supported by the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation. Finally, we thank H. Chang (Purdue University, West Lafayette, IN, USA) for sharing the CLC-mEmerald fly strain.

## Author contributions

H.W., Y.R., and A.O. conceived the research. H.W., Y.R., L.B., and C.K. contributed to the experiments. H.W., Y.J., Z.W., and H.G. processed the data. H.W. and Y.J. prepared the figures. H.W., Y.R., and A.O. prepared the manuscript, and all the authors contributed to the manuscript. H.W., Y.R., and R.G. developed the Fiji/ImageJ plugin. A.O. supervised the research.

## Competing interests

A.O., Y.R., and H.W. have a pending patent application on the contents of the presented results.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41592-018-0239-0>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to A.O.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

## Methods

**Wide-field fluorescence microscopy image acquisition.** The fluorescence microscopy images (Figs. 1 and 2) were captured by scanning a microscope slide containing multi-labeled bovine pulmonary artery endothelial cells (BPAECs) (FluoCells Prepared Slide #2, Thermo Fisher Scientific) on a standard inverted microscope equipped with a motorized stage (IX83, Olympus Life Science). The low-resolution (LR) and high-resolution (HR) images were acquired using 10 $\times$ /0.4-NA (UPLSAPO10X2, Olympus Life Science) and 20 $\times$ /0.75-NA (UPLSAPO20X, Olympus Life Science) objective lenses, respectively. Three bandpass optical filter sets were used to image the three different labeled cell structures and organelles: Texas Red for F-actin (OSFI3-TXRED-4040C, EX562/40, EM624/40, DM593, Semrock), FITC for microtubules (OSFI3-FITC-2024B, EX485/20, EM522/24, DM506, Semrock), and DAPI for cell nuclei (OSFI3-DAPI-5060C, EX377/50, EM447/60, DM409, Semrock). The imaging experiments were controlled by MetaMorph microscope automation software (Molecular Devices), which performed translational scanning and auto-focusing at each position of the stage. The auto-focusing was performed on the FITC channel, and the DAPI and Texas Red channels were both exposed at the same plane as FITC. With a 130-W fluorescence light source set to 25% output power (U-HGLGPS, Olympus Life Science), the exposure time for each channel was set as follows: Texas Red, 350 ms (10 $\times$ ) and 150 ms (20 $\times$ ); FITC, 800 ms (10 $\times$ ) and 400 ms (20 $\times$ ); DAPI, 60 ms (10 $\times$ ) and 50 ms (20 $\times$ ). The images were recorded by a monochrome sCMOS camera (ORCA-flash4.0 v2, Hamamatsu Photonics K.K.) and saved as 16-bit grayscale images with regard to each optical filter set. The additional test images (Supplementary Figs. 6 and 7) are captured using the same setup with FluoCells Prepared Slide #1 (Thermo Fisher Scientific), with the filter setting of Texas Red for mitochondria and FITC for F-actin, and FluoCells Prepared Slide #3 (Thermo Fisher Scientific), with the filter setting of Texas Red for actin and FITC for glomeruli and convoluted tubules. The mouse brain tumor sample was prepared with mouse brains perfused with Dylight-594-conjugated Tomato Lectin (1 mg/ml) (Vector Laboratories, CA), fixed in 4% paraformaldehyde for 24 h and incubated in 30% sucrose in phosphate-buffered saline, then cut in 50- $\mu$ m-thick sections as detailed in ref.<sup>36</sup>, and imaged using Texas Red filter set for blood vessels, and FITC filter set for tumor cells.

**Confocal and STED image acquisition.** For the Histone 3 imaging experiments, the HeLa cells were grown as a monolayer on high-performance coverslips (170  $\mu$ m  $\pm$  10  $\mu$ m) (Carl Zeiss Microscopy) and fixed with methanol. Nuclei were labeled with a primary Rabbit anti-Histone H3 trimethyl Lys4 (H3K4me3) antibody (Active motif #39159) and a secondary Atto-647N Goat anti-rabbit IgG antibody (Active Motif # 15048) using the reagents of the MAXpack Immunostaining Media Kit (Active Motif #15251). The labeled cells were then embedded with Mowiol 4-8-8 and mounted on a standard microscope slide.

The nano-bead samples for confocal and STED experiments (Figs. 3 and 4) were prepared with 20-nm fluorescent nano-beads (FluoSpheres Carboxylate-Modified Microspheres, crimson fluorescent (625/645), 2% solids, Thermo Fisher Scientific) that were diluted 100 times with methanol and sonicated for 3  $\times$  10 min, and then mounted with antifade reagents (ProLong Diamond, Thermo Fisher Scientific) on a standard glass slide, followed by placement on high-performance coverslips.

Samples were imaged on a Leica TCS SP8 STED confocal microscopy using a Leica HC PL APO 100 $\times$ /1.40-NA Oil STED White objective. The scanning for each FOV was performed by a resonant scanner working at 8,000 Hz with 16 times line average and 30 times frame average for nano-beads, and 8 times line average and 6 times frame average for cell nuclei. The fluorescent nano-beads were excited with a laser beam at 633-nm wavelength. The emission signal was captured with a hybrid photodetector (HyD SMD, Leica Microsystems) through a 645–752-nm bandpass filter. The excitation laser power was set to 5% for confocal imaging and 50% for STED imaging, so that the signal intensities remained similar while the same scanning speed and gain voltage were maintained. A depletion beam of 775 nm was also applied when capturing STED images with 100% power. The confocal pinhole was set to 1 Airy unit (for example, 168.6  $\mu$ m for 645-nm emission wavelength and 100 $\times$ magnification) for both the confocal and STED imaging experiments. The cell nuclei samples were excited with a laser beam at 635 nm and captured with the same photodetector, which was set to 1 $\times$ gain for confocal and 1.9 $\times$ gain for STED with a 650–720-nm bandpass filter. The confocal pinhole was set to 75.8  $\mu$ m (for example, 0.457 Airy unit for 650-nm emission wavelength and 100 $\times$ magnification) for both the confocal and STED imaging experiments. The excitation laser power was set to 3% and 10% for confocal and STED experiments, respectively. The scanning step size (i.e., the effective pixel size) for both experiments was  $\sim$ 30 nm to ensure sufficient sampling rate. All the images were exported and saved as 8-bit grayscale images.

**TIRF-SIM image acquisition.** Gene-edited SUM159 cells expressing AP2-eGFP<sup>37</sup> were grown in F-12 medium containing hydrocortisone, penicillin-streptomycin and 5% FBS. Transient expression of mRuby-CLTB (Addgene; Plasmid #55852) was carried out with the Gene Pulser Xcell electroporation system (Bio-Rad Laboratories, CA, USA) according to the manufacturer's instructions, and imaging was performed 24–48 h after transfection. Cells were imaged in

phenol-red-free L15 (Thermo Fisher Scientific) supplemented with 5% FBS at 37 °C ambient temperature. Clathrin dynamics were monitored in lateral epidermis and amniocerous tissues of *Drosophila* embryos using the UAS/GAL4 system as described in ref.<sup>38</sup>. *Drosophila* embryos were gently pressed against the coverslip to position the apical surface of the lateral epidermis and amniocerous cells within the evanescence field of the TIRF system. Arm-GAL4 strain was provided by the Bloomington *Drosophila* Stock Center; CLC-mEmerald strain was provided by Dr. Henry Chang (Purdue University, USA). TIRF-SIM images were acquired with a 100 $\times$ /1.49-NA objective lens (Olympus Life Science, CA, USA) fitted on an inverted microscope (Axio Observer; ZEISS) equipped with an sCMOS camera (ORCA-Flash4.0; Hamamatsu). Structured illumination was provided by a spatial light modulator as described in ref.<sup>18</sup>.

**Image pre-processing.** For wide-field images (Figs. 1 and 2, and Supplementary Figs 1, 2a–h, and 6–9), a low intensity threshold was applied to subtract background noise and auto-fluorescence, as a common practice in fluorescence microscopy. The threshold value was estimated from the mean intensity value of a region without objects, which is  $\sim$ 300 out of 65,535 in our 16-bit images. The LR images are then linearly interpolated two times to match the effective pixel size of the HR images. Accurate registration of the corresponding LR and HR training image pairs is of crucial importance because the objective function of our network consists of adversarial loss and pixel-wise loss. We employed a two-step registration workflow to achieve the needed registration with sub-pixel-level accuracy. First, the FOVs of LR and HR images are digitally stitched in a MATLAB script interfaced with the Fiji<sup>39</sup> Grid/Collection stitching plugin<sup>40</sup> through MIJ<sup>41</sup>, and matched by fitting of their normalized cross-correlation map to a 2D Gaussian function and identification of the peak location (Supplementary Note 9). However, because of optical distortions and color aberrations of different objective lenses, the local features might still not be exactly matched. To address this, the globally matched images are fed into a pyramidal elastic registration algorithm to achieve sub-pixel-level matching accuracy, which is an iterative version of the registration module in Fiji Plugin NanoJ, with a shrinking block size (Supplementary Fig. 12)<sup>13,39,42,43</sup>. This registration step starts with a block size of 256  $\times$  256 and stops at a block size of 64  $\times$  64, with the block size shrunk by 1.2 times every 5 iterations with a shift tolerance of 0.2 pixels. Because of the slightly different placement and the distortion of the optical filter sets, we performed the pyramidal elastic registration for each fluorescence channel independently. At the last step, the precisely registered images were cropped 10 pixels on each side to avoid registration artifacts, and converted to single-precision floating data type and scaled to a dynamic range of 0–255. This scaling step is not mandatory but creates convenience for fine tuning of hyper-parameters when working with images from different microscopes/sources.

For confocal and STED images (Figs. 3–5) that were scanned in sequence on the same platform, only a drift correction step was required, which was calculated from the 2D Gaussian fit of the cross-correlation map. The drift was found to be  $\sim$ 10 nm for each scanning FOV between the confocal and STED images. We did not perform thresholding to the nano-bead dataset for the network training. However, after the test images were enhanced by the network, we subtracted a constant value (calculated by taking the mean value of an empty region) from the confocal (network input), the super-resolved (network output), and the STED (ground truth) images, respectively, for better visualization and comparison of the images. The total number of images used for training, validation and blind testing of each network are summarized in Supplementary Table 1.

**Generative adversarial network structure and training.** In this work, our deep neural network was trained following the generative adversarial network (GAN) framework<sup>46</sup>, which has two sub-networks being trained simultaneously, a generative model which enhances the input LR image, and a discriminative model which returns an adversarial loss to the resolution-enhanced image, as illustrated in Supplementary Fig. 13. We designed our objective function as the combination of the adversarial loss with two regularization terms: the mean square error (MSE), and the structural similarity (SSIM) index<sup>44</sup>. Specifically, we aim to minimize

$$\begin{aligned} \mathcal{L}(G; D) &= -\log D(G(x)) + \lambda \times \text{MSE}(G(x), y) \\ &\quad - \nu \times \log [(1 + \text{SSIM}(G(x), y))/2] \\ \mathcal{L}(D; G) &= -\log D(y) - \log [1 - D(G(x))] \end{aligned} \quad (1)$$

where  $x$  is the LR input,  $G(x)$  is the generative model output,  $D(\cdot)$  is the discriminative model prediction of an image (network output or ground truth image), and  $y$  is the HR image used as ground truth. The structural similarity index is defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

where  $\mu_x, \mu_y$  are the averages of  $x, y$ ;  $\sigma_x^2, \sigma_y^2$  are the variances of  $x, y$ ;  $\sigma_{x,y}$  is the covariance of  $x$  and  $y$ ; and  $c_1, c_2$  are the variables used to stabilize the division with a

small denominator. An SSIM value of 1.0 refers to identical images. When training with the wide-field fluorescence images, the regularization constants  $\lambda$  and  $v$  were set to accommodate the MSE loss and the SSIM loss to be  $\sim 1\text{--}10\%$  of the combined generative model loss  $\mathcal{L}(G; D)$ , depending on the noise level of the image dataset. When training with the confocal-STED image datasets, we kept  $\lambda$  the same and set  $v$  to 0. While the adversarial loss guides the generative model to map the LR images into HR, the two regularization terms assure that the generator output image is established on the input image with matched intensity profile and structural features. These two regularization terms also help us stabilize the training schedule and smooth out the spikes on the training loss curve before it reaches equilibrium. For the sub-network models, we employed a similar network structure as described in ref.<sup>43</sup>. The relatively low weight that is given to the MSE and SSIM terms is due to the fact that these values already represent a high degree of agreement between the low-resolution input and the gold standard label (for example,  $\sim 0.87\text{--}0.94$  for the wide-field microscopy experiments). Hence, a large weight given to these loss terms will drive the network to converge to a local minimum that will strongly resemble the low-resolution input and not learn the desired (super-resolved) output distribution. Therefore, it might be beneficial for some other applications to increase the weights of these terms, for example, for low SNR images, where the task of denoising might be of main interest for automated segmentation and related image processing tasks.

**Generative model.** U-net is a CNN architecture that was first proposed for medical image segmentation, yielding high performance with very few training datasets<sup>45</sup>. A similar network architecture has also been successfully applied in recent image reconstruction and virtual staining applications<sup>43,46</sup>. The structure of the generative network used in this work is illustrated in Supplementary Fig. 13, which consists of four downsampling blocks and four upsampling blocks. Each downsampling block consists of three residual convolutional blocks, within which it performs

$$\begin{aligned} x_k &= x_{k-1} + \text{LReLU}[\text{Conv}\{\text{LReLU}[\text{Conv}\{\text{LReLU}[\text{Conv}\{x_{k-1}\}]\}]\}], \\ k &= 1, 2, 3, 4 \end{aligned} \quad (3)$$

where  $x_k$  represents the output of the  $k$ th downsampling block, and  $x_0$  is the LR input image.  $\text{Conv}\{\cdot\}$  is the convolution operation,  $\text{LReLU}[\cdot]$  is the leaky rectified linear unit activation function with a slope of  $\alpha=0.1$ , that is,

$$\text{LReLU}(x; \alpha) = \max(0, x) - \alpha \times \max(0, -x) \quad (4)$$

The input of each downsampling block is zero-padded and added to the output of the same block. The spatial downsampling is achieved by an average pooling layer after each downsampling block. A convolutional layer lies at the bottom of this U-shape structure that connects the downsampling and upsampling blocks.

Each upsampling block also consists of three convolutional blocks, within which it performs

$$\begin{aligned} y_k &= \text{LReLU} \\ &\quad [\text{Conv}\{\text{LReLU}[\text{Conv}\{\text{LReLU}[\text{Conv}\{\text{Concat}(x_{5-k}, y_{k-1})\}\}]\}]\}], \\ k &= 1, 2, 3, 4 \end{aligned} \quad (5)$$

where  $y_k$  represents the output of the  $k$ th upsampling block, and  $y_0$  is the input of the first upsampling block.  $\text{Concat}(\cdot)$  is the concatenation operation of the downsampling block output and the upsampling block input on the same level in the U-shape structure. The last layer is another convolutional layer that maps the 32 channels into 1 channel that corresponds to a monochrome grayscale image.

**Discriminative Model.** As shown in Supplementary Fig. 13, the structure of the discriminative model begins with a convolutional layer, which is followed by 5 convolutional blocks, each of which performs the following operation:

$$z_k = \text{LReLU}[\text{Conv}\{\text{LReLU}[\text{Conv}\{z_{k-1}\}]\}], k = 1, 2, 3, 4, 5 \quad (6)$$

where  $z_k$  represents the output of the  $k$ th convolutional block, and  $z_0$  is the input of the first convolutional block. The output of the last convolutional block is fed into an average pooling layer whose filter shape is the same as the patch size, that is,  $H \times W$ . This layer is followed by two fully connected layers for dimension reduction. The last layer is a sigmoid activation function whose output is the probability of an input image being ground truth, defined as

$$D(z) = \frac{1}{1 + \exp(-z)} \quad (7)$$

**Network training schedule.** During our training the patch size is set as  $64 \times 64$ , with a batch size of 12 on each of the two GPUs. Within each iteration, the generative model and the discriminative model are each updated once while the other is kept unchanged. Both the generative model and the discriminative model were randomly initialized and optimized using the adaptive moment estimation

(Adam) optimizer<sup>47</sup> with a starting learning rate of  $1 \times 10^{-4}$  and  $1 \times 10^{-5}$ , respectively. This framework was implemented with TensorFlow framework version 1.7.0<sup>48</sup> and Python version 3.6.4 in the Microsoft Windows 10 operating system. The training was performed on a consumer-grade laptop (EON17-SLX, Origin PC) equipped with dual GeForce GTX1080 graphic cards (NVIDIA) and a Core i7-8700K CPU @ 3.7 GHz (Intel). The final models for wide-field images were selected with the smallest validation loss at around the 50,000th iteration, which took  $\sim 10$  h to train. The final model for confocal-STED transformation (Figs. 3 and 4) is selected with the smallest validation loss at around the 500,000th iteration, which took  $\sim 90$  h to train. The transfer learning for the confocal-STED transformation network (Fig. 5) was implemented with the same framework on a desktop computer with dual GTX1080Ti graphic cards, with the patch size set as  $256 \times 256$  with 4 patches on each GPU. It was first initialized with the confocal-STED model trained with nano-beads, and then refined with cell nuclei image data with  $\sim 20,000$  iterations, which took  $\sim 24$  h. The training of the TIRF to TIRF-SIM transformation network was also implemented with dual GTX1080Ti graphic cards, with the patch size set as  $64 \times 64$ , and 64 patches on each GPU. The final model was trained for  $\sim 20,000$  iterations, which took  $\sim 18$  h. A typical plot of the loss functions during the GAN training is shown in Supplementary Fig. 14, where the generative and discriminative models compete in an equilibrium state for  $\sim 60,000$  iterations before they start to diverge. The iteration time is also dependent on the patch and batch size. We also demonstrate in Supplementary Fig. 4 that the role of the discriminative model of GAN is critical to achieving super-resolution, as it provides an adaptive loss function and helps the generative model to jump out of local minima. Training without the discriminative loss can result in over-smoothed images (see, for example, Supplementary Fig. 4), as the generative model optimizes only a specific group of statistical metrics. A step-by-step training instruction and guideline, with several critical steps discussed and emphasized, are provided in Supplementary Note 10.

**Implementation of LR and NNLS deconvolution.** For a fair comparison, the lower-resolution images were upsampled 2 times by bilinear interpolation before being deconvolved. We used the Born and Wolf PSF model<sup>49,50</sup>, with parameters set to match our experimental setup, that is, NA = 0.4, immersion refractive index = 1.0, pixel size = 325 nm. The PSF is generated by a Fiji PSF Generator Plugin<sup>39,51</sup>. We performed an exhaustive parameter search by running the LR algorithm with 1–100 iterations and damping threshold 0–10%. The results were visually assessed, with the best one obtained at 10 iterations and 0.1% damping threshold (Fig. 2, third column). The NNLS deconvolution was performed with Fiji Plugin DeconvolutionLab2<sup>52</sup> with 100 iterations and a step size of 0.5. The deconvolutions for Texas Red, FITC, and DAPI channels were performed separately, assuming the central emission wavelengths to be 630 nm, 532 nm, and 450 nm, respectively.

**Characterization of the lateral resolution by PSF fitting.** We characterized the resolution differences among the network input (confocal), the network output (confocal), and the ground truth (STED) images by fitting their PSFs to a 2D Gaussian profile, as shown in Fig. 4. For this, more than 400 independent bright spots were selected from the ground truth STED images and cropped out with the surrounding  $19 \times 19$ -pixel regions, that is,  $\sim 577 \times 577$  nm<sup>2</sup>. The same locations were also projected to the network input and output images, followed by cropping of the same image regions as in the ground truth STED images. Each cropped region was then fitted to a 2D Gaussian profile. The FWHM values of all these 2D profiles were plotted as histograms, shown in Fig. 4. For each category of images, the histogram profile within the main peak region was fitted to a 1D Gaussian function (Fig. 4). A similar process was repeated for the results reported in Supplementary Fig. 8d.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

We declare that all the data supporting the findings of this work are available within the manuscript and Supplementary Information files. Raw images can be requested from the corresponding author. Deep learning models reported in this work used standard libraries and scripts that are publicly available in TensorFlow. The instruction manual for our Fiji/ImageJ plugin and trained models (available online as Supplementary Software 1–7) is provided as a Supplementary Protocol.

## References

36. Bentolila, L. A. et al. Imaging of angiogenesis/vascular co-option in a murine model of brain melanoma: implications for melanoma progression along extravascular pathways. *Sci. Rep.* **6**, 23834 (2016).
37. Aguet, F. et al. Membrane dynamics of dividing cells imaged by lattice light-sheet microscopy. *Mol. Biol. Cell* **27**, 3418–3435 (2016).
38. Willy, N. M. et al. Membrane mechanics govern spatiotemporal heterogeneity of endocytic clathrin coat dynamics. *Mol. Biol. Cell* **28**, 3480–3488 (2017).

39. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
40. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).
41. Sage, D., Prodanov, D., Tinevez, J.-Y. & Schindelin, J. MIJ: making interoperability between ImageJ and Matlab possible. Poster presented at the ImageJ User & Developer Conference, Mondorf-les-Bains, Luxembourg, 24–26 October, 2012.
42. Rivenson, Y. et al. Deep learning enhanced mobile-phone microscopy. *ACS Photonics* **5**, 2354–2364 (2018).
43. Rivenson, Y. et al. Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue. *arXiv* Preprint at <https://arxiv.org/abs/1803.11293> (2018).
44. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
45. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. *arXiv* Preprint at <https://arxiv.org/abs/1505.04597> (2015).
46. Wu, Y. et al. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* **5**, 704–710 (2018).
47. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *arXiv* Preprint at <https://arxiv.org/abs/1412.6980> (2014).
48. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. *arXiv* Preprint at <https://arxiv.org/abs/1605.08695> (2016).
49. Aguet, F., Van De Ville, D. & Unser, M. Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy. *IEEE Trans. Image Process.* **17**, 1144–1153 (2008).
50. Born, M., Wolf, E. & Bhatia, A. B. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Cambridge University Press, 1999).
51. Kirshner, H., Aguet, F., Sage, D. & Unser, M. 3-D PSF fitting for fluorescence microscopy: implementation and localization application. *J. Microsc.* **249**, 13–25 (2013).
52. Sage, D. et al. DeconvolutionLab2: an open-source software for deconvolution microscopy. *Methods* **115**, 28–41 (2017).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

For the widefield fluorescence image data collection, we used a conventional fluorescence microscope (IX83, Olympus Corporation, Tokyo, Japan) equipped with a motorized stage, where the image acquisition process was controlled by MetaMorph® microscope automation software (Molecular Devices, LLC).  
For the confocal and stimulated emission depletion (STED) microscopy, we have used TCS SP8 (Leica Microsystems), which was controlled by the LAS X (Leica Microsystems) software, version 3.5.0.  
For total-internal reflection microscopy (TIRF), we used an inverted microscope (Axio Observer, ZEISS) equipped with a sCMOS camera (ORCA-Flash4.0; Hamamatsu).

#### Data analysis

Deep learning models reported in this work used standard libraries and scripts that are publicly available in TensorFlow v1.7.0 (Google Inc.). The custom codes were written in Python v3.6.4 (Anaconda modified version). Image matching procedures were all performed using custom Matlab vR2017a codes (The MathWorks Inc.). For the global patch matching, the Matlab script interfaced to a Fiji (ImageJ) plugin "Stitching"->"Grid/Collection stitching" (version 1.2, Preibisch et. al, Bioinformatics 2009), through Miji version 1.3.9 (available at <https://imagej.net/Miji>) java package.  
All the software was developed and executed on a Windows 10 (Home Edition) operating system.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

We declare that all the data supporting the findings of this work are available within the manuscript and Supplementary Information files. Raw images can be requested from the corresponding author.

The instruction manual on our Fiji/ImageJ plugin and trained models (available online as Supplementary Software) is provided as a Supplementary Protocol.

## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the widefield fluorescence image data, we have trained and tested the approach on 2 different fluorescently labeled samples with 2x2657 separate images, each with 1024x1024 pixels. For the confocal to STED cross modality transformation we have used 2x757 images of nano-beads, each with 1024x1024 pixels, and 2x1230 images for transfer learning to resolve cell nuclei, each with 1024x1024 pixels. For the TIRF to TIRF-SIM cross modality transformation we have used 2x4473 images, each with 1024x1024 pixels
Data exclusions	None.
Replication	For the widefield deep network, following its training, it was blindly tested with a total of 280 images (of size 1024x1024 pixels) and for the confocal to STED cross modality, we have tested the trained network on 138 images (of size 1024x1024 pixels). For the TIRF to TIRF-SIM cross modality transformation we have tested the network with 1100 images (of size 1024x1024 pixels).
Randomization	Training, validation and testing images were randomly selected.
Blinding	All the performance testing of the deep neural network results was blindly performed on images that were not included in the training or validation phase of the deep learning method, or were performed on new structures that the network was not trained for.

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

## Non-participation

*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

## Randomization

*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

## Study description

*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*

## Research sample

*Describe the research sample (e.g. a group of tagged *Passer domesticus*, all *Stenocereus thurberi* within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*

## Sampling strategy

*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*

## Data collection

*Describe the data collection procedure, including who recorded the data and how.*

## Timing and spatial scale

*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*

## Data exclusions

*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*

## Reproducibility

*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*

## Randomization

*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*

## Blinding

*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work?

Yes     No

## Field work, collection and transport

## Field conditions

*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*

## Location

*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*

## Access and import/export

*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*

## Disturbance

*Describe any disturbance caused by the study and how it was minimized.*

## Reporting for specific materials, systems and methods

**Materials & experimental systems**

n/a	Involved in the study
<input checked="" type="checkbox"/>	Unique biological materials
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants

**Methods**

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

**Unique biological materials**Policy information about [availability of materials](#)Obtaining unique materials **Antibodies**Antibodies used Validation **Eukaryotic cell lines**Policy information about [cell lines](#)Cell line source(s) Authentication Mycoplasma contamination Commonly misidentified lines  
(See [ICLAC](#) register) **Palaeontology**Specimen provenance Specimen deposition Dating methods  Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.**Animals and other organisms**Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal researchLaboratory animals Wild animals Field-collected samples **Human research participants**Policy information about [studies involving human research participants](#)Population characteristics   
  
Recruitment

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

#### Data access links

*May remain private before publication.*

N/A

#### Files in database submission

N/A

#### Genome browser session (e.g. [UCSC](#))

N/A

### Methodology

#### Replicates

N/A

#### Sequencing depth

N/A

#### Antibodies

N/A

#### Peak calling parameters

N/A

#### Data quality

N/A

#### Software

N/A

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

N/A

#### Instrument

N/A

#### Software

N/A

#### Cell population abundance

N/A

#### Gating strategy

N/A

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

## Magnetic resonance imaging

### Experimental design

#### Design type

N/A

#### Design specifications

N/A

#### Behavioral performance measures

N/A

**Acquisition**

Imaging type(s)	N/A	
Field strength	N/A	
Sequence & imaging parameters	N/A	
Area of acquisition	N/A	
Diffusion MRI	<input type="checkbox"/> Used	<input type="checkbox"/> Not used

**Preprocessing**

Preprocessing software	N/A	
Normalization	N/A	
Normalization template	N/A	
Noise and artifact removal	N/A	
Volume censoring	N/A	

**Statistical modeling & inference**

Model type and settings	N/A		
Effect(s) tested	N/A		
Specify type of analysis:	<input type="checkbox"/> Whole brain	<input type="checkbox"/> ROI-based	<input type="checkbox"/> Both
Statistic type for inference (See <a href="#">Eklund et al. 2016</a> )	N/A		
Correction	N/A		

**Models & analysis**

n/a	<input type="checkbox"/> Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input checked="" type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis

Functional and/or effective connectivity Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.