nature methods

ARTICLES

https://doi.org/10.1038/s41592-019-0622-5

Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning

Yichen Wu[1,2,3,8], Yair Rivenson[1,2,3,8], Hongda Wang[1,2,3], Yilin Luo[1,2,3], Eyal Ben-David[4], Laurent A. Bentolila[3,5], Christian Pritz[6] and Aydogan Ozcan[1,2,3,7*]

We demonstrate that a deep neural network can be trained to virtually refocus a two-dimensional fluorescence image onto user-defined three-dimensional (3D) surfaces within the sample. Using this method, termed Deep-Z, we imaged the neuronal activity of a *Caenorhabditis elegans* worm in 3D using a time sequence of fluorescence images acquired at a single focal plane, digitally increasing the depth-of-field by 20-fold without any axial scanning, additional hardware or a trade-off of imaging resolution and speed. Furthermore, we demonstrate that this approach can correct for sample drift, tilt and other aberrations, all digitally performed after the acquisition of a single fluorescence image. This framework also cross-connects different imaging modalities to each other, enabling 3D refocusing of a single wide-field fluorescence image to match confocal microscopy images acquired at different sample planes. Deep-Z has the potential to improve volumetric imaging speed while reducing challenges relating to sample drift, aberration and defocusing that are associated with standard 3D fluorescence microscopy.

High-throughput volumetric fluorescence imaging is very important in various fields including, for example, biology, life sciences and engineering, and still remains a challenge in microscopy research. 3D fluorescence information is usually acquired through scanning an excitation source through the sample volume to obtain images at multiple planes, forming the basis of volumetric imaging in confocal[1], two-photon[2], light-sheet[3–5] and various super-resolution[6–11] microscopy techniques. However, scanning can limit the imaging speed and throughput, potentially introducing phototoxicity and photobleaching, even with optimized scanning strategies[1] or point-spread-function (PSF) engineering[3,12]. 3D fluorescence information of a specimen can also be acquired using non-scanning microscopy methods that simultaneously map the axial information onto a two-dimensional (2D) image, such as fluorescence light-field microscopy[13–18], Fresnel correlation holography[19–21] and others[22,23]. However, these non-scanning 3D fluorescence microscopy approaches require relatively time-consuming iterative algorithms to solve the inverse problem for reconstructing a new image, and use customized optical components and hardware, which increase the complexity of the setup.

There are emerging approaches that use deep learning to solve inverse problems in fluorescence microscopy[24], for example, to enhance the lateral[25–31] and axial[31–33] resolution using artificial neural networks trained with image data. Such deep-learning-based image reconstruction and enhancement methods take a relatively long time to train; however, this training process is a one-time effort, and after it is complete, each new sample of interest can be rapidly reconstructed through a single forward pass through the trained network, without the need for any iterations or hyperparameter tuning, which in general forms an important advantage of deep-learning-based solutions to inverse imaging problems[34].

Here we introduce a digital image refocusing framework in fluorescence microscopy by training a deep neural network using microscopic image data, enabling 3D imaging of fluorescent samples using a single 2D wide-field image, without the need for any mechanical scanning, additional hardware or parameter estimation. This framework rapidly refocuses a 2D fluorescence image onto user-defined 3D surfaces (such as tilted planes, curved surfaces and others), and can be used to digitally correct for various aberrations caused by the sample and/or the optical system. We term this deep-learning-based approach Deep-Z, and use it to computationally refocus a single 2D wide-field fluorescence image onto 3D surfaces within the sample, without sacrificing the imaging speed, resolution or field of view (FOV) of a standard microscope.

In Deep-Z, an input 2D fluorescence image is first appended with a user-defined digital propagation matrix (DPM) that represents, pixel-by-pixel, the axial distance of the target surface from the plane of the input image (Fig. 1). Deep-Z is trained using a conditional generative adversarial neural network (GAN)[35,36] with accurately matched pairs of (1) various fluorescence images axially focused at different depths and appended with different DPMs, and (2) the corresponding fluorescence images (that is, the ground-truth labels) captured at the correct (target) focus plane defined by the corresponding DPM. Through this training process that only uses experimental image data, the generator network of the GAN learns to interpret the values of each DPM pixel as an axial refocusing distance, and outputs an equivalent fluorescence image that is digitally refocused within the sample to the 3D surface defined by the user.

[1]Electrical and Computer Engineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [2]Bioengineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [3]California Nano Systems Institute (CNSI), University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [5]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA. [6]Department of Genetics, Hebrew University of Jerusalem, Jerusalem, Israel. [7]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [8]These authors contributed equally: Yichen Wu, Yair Rivenson. *e-mail: ozcan@ucla.edu

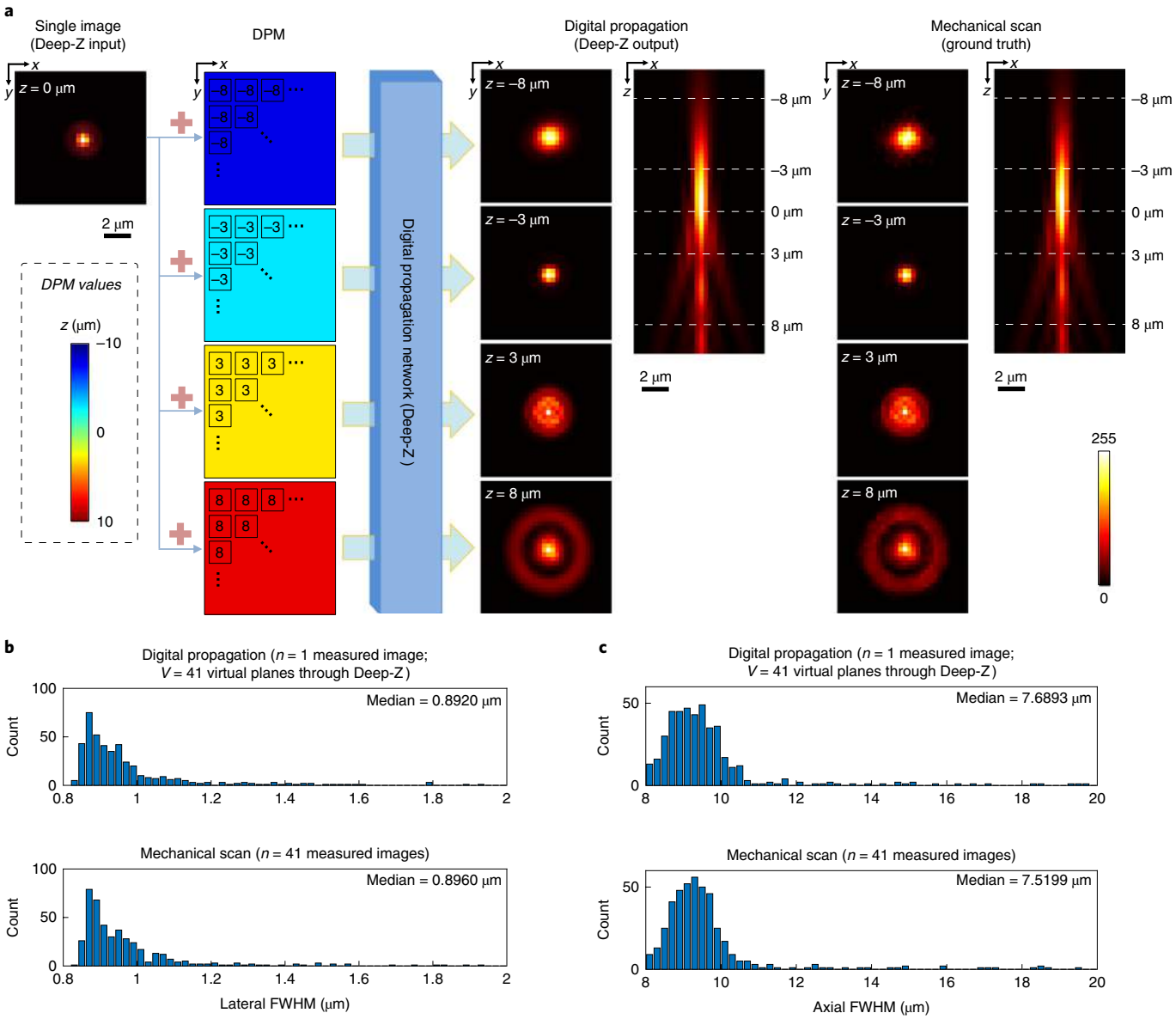


Fig. 1 | Refocusing of fluorescence images using Deep-Z. a, Steps involved in using the Deep-Z network. By appending a DPM to a single fluorescence image (left) and passing it through a trained Deep-Z network, refocused images at different planes can be virtually obtained. The PSF generated by Deep-Z (middle) and mechanical scanning (right) are shown for comparison. Color scale indicates intensity. **b**, Lateral FWHM histograms for 461 individual isolated fluorescence nanobeads (300 nm) distributed over ~500 × 500 μm², measured using Deep-Z inference ($n = 1$ captured image) and images obtained using mechanical axial scanning ($n = 41$ captured images) match each other very well. **c**, As in **b**, except using the axial FWHM measurements for the same dataset, revealing a very good match between Deep-Z inference results and the axial mechanical scanning results.

Using Deep-Z, we imaged *Caenorhabditis elegans* neurons using a standard wide-field fluorescence microscope and extended the native depth of field (DOF) by ~20-fold. Using Deep-Z, we further demonstrated 3D tracking of the neuron activity of a *C. elegans* worm over an extended DOF using a time sequence of fluorescence images acquired at a single focal plane. Furthermore, we used spatially non-uniform DPMs to refocus a 2D input fluorescence image onto user-defined 3D surfaces to computationally correct for aberrations such as sample drift, tilt and spherical aberrations, all performed after the image acquisition and without any modifications to the optical hardware of a standard fluorescence microscope.

Another important feature of Deep-Z is that it permits cross-modality digital refocusing of fluorescence images, where the GAN is trained with gold-standard label images obtained by a different fluorescence microscopy modality. We term this framework Deep-Z+. To demonstrate a proof of concept, we trained Deep-Z+ with input and label images that were acquired with a wide-field fluorescence microscope and a confocal microscope, respectively, to blindly generate the output of this cross-modality Deep-Z+: digitally refocused images of an input wide-field fluorescence image that match confocal microscopy images of the same sections.

Results

Digital refocusing of fluorescence images using Deep-Z. Figure 1a demonstrates Deep-Z-based digital refocusing of a single image of a 300-nm fluorescent bead (excitation and emission wavelengths of 538 nm and 584 nm, respectively) to multiple user-defined planes, represented by different DPMs; each one of these propagation matrices represents, pixel-by-pixel, the axial distance of the target surface from the plane of the input image (Fig. 1a). The native

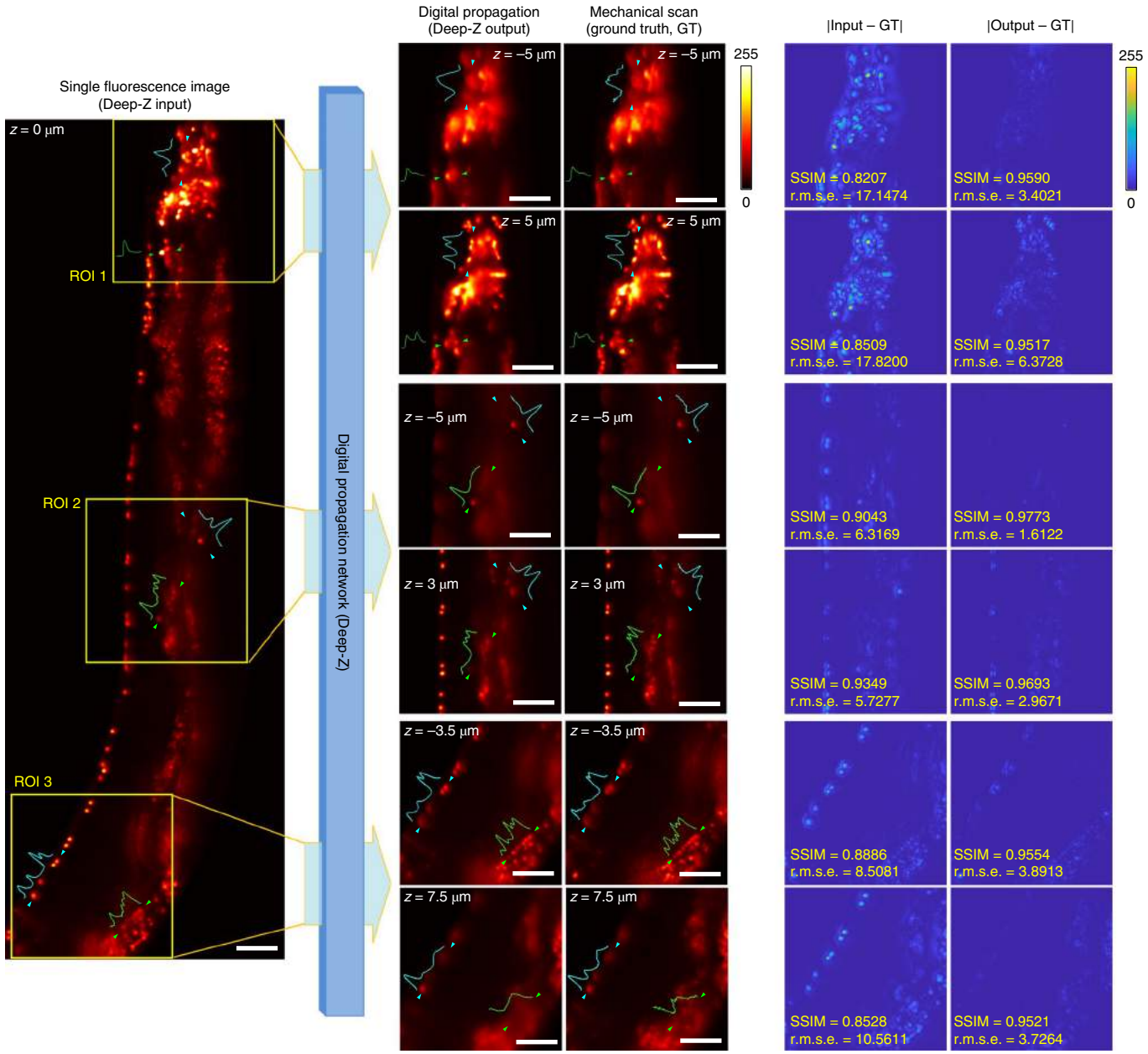


Fig. 2 | 3D imaging of *C. elegans* neuron nuclei using Deep-Z. Different ROIs are digitally refocused using Deep-Z to different planes within the sample volume; the resulting images provide a very good match to the corresponding ground-truth images, acquired using a scanning fluorescence microscope. The absolute difference images of the input and output with respect to the corresponding ground-truth image are also provided on the right, with structural similarity index (SSIM) and root mean square error (r.m.s.e.) values reported, further demonstrating the success of Deep-Z. Lines represent cross-sectional plots and arrowheads indicate where the cross-section was taken; blue and green represent two separate cross-section lateral locations. Scale bars, 25 μm. Experiments were repeated with 20 images, achieving similar results. Color scales indicate intensity.

DOF of the input fluorescence image, as defined by the numerical aperture (NA) of the objective lens (20×/0.75 NA), is ~1 μm; using Deep-Z, we digitally refocused the image of this fluorescent bead over an axial range of approximately ±10 μm, matching the corresponding mechanically scanned images of the same region-of-interest (ROI), which form the ground-truth. Note that the PSF in Fig. 1a is asymmetric in the axial direction, which provides directional cues to the neural network regarding the digital propagation of an input image by Deep-Z. Unlike a symmetric Gaussian beam[37], such PSF asymmetry along the axial direction is ubiquitous in fluorescence microscopy systems[38].

Deep-Z also provides an improved signal-to-noise ratio (SNR) at its output as compared to a fluorescence image of the same object measured at the corresponding depth (Supplementary Fig. 1). To further quantify Deep-Z output we used PSF analysis; Fig. 1b,c illustrates the histograms of both the lateral and the axial full-width-half-maximum (FWHM) values of 461 individual isolated nanobeads. These histograms agree with each other very well (Fig. 1b,c), confirming the match between Deep-Z output images calculated from a single fluorescence image and the corresponding axially scanned ground-truth images.

Next, we tested Deep-Z by imaging the neurons of a *C. elegans* nematode expressing pan-neuronal tagRFP[39]. Figure 2 demonstrates our blind-testing results for Deep-Z-based refocusing of different parts of a *C. elegans* worm from a single wide-field fluorescence input image. Using Deep-Z, non-distinguishable fluorescent

neurons in the input image were brought into focus at different depths, while other in-focus neurons in the input image became out-of-focus and smeared into the background, according to their true axial positions in 3D; comparisons to the ground-truth mechanical scans are provided as cross-sections and image difference analyses in Fig. 2 and Supplementary Fig. 2. For optimal performance, this Deep-Z model was specifically trained using *C. elegans* samples, and the axial range of its refocusing capability is determined by the training data range (±10 µm), and fails outside of this training range (Supplementary Figure 3). Using Deep-Z, we also generated (from a single 2D fluorescence image) a virtual 3D stack (Supplementary Video 1) and 3D visualization (Supplementary Video 2) of a *C. elegans* worm, over an axial range of approximately ±10 µm. Similar results were also obtained for a *C. elegans* imaged under a 40×/1.3 NA objective lens, where Deep-Z successfully refocused the input image over an axial range of approximately ±4 µm (Supplementary Fig. 4).

Next, we captured a video of four moving *C. elegans* worms, where each frame of this fluorescence video was digitally refocused to various depths using Deep-Z. This enabled us to create simultaneously running videos of the same sample, each one focused at a different depth (Supplementary Video 3). Each one of these virtually created videos are temporally synchronized to each other (that is, the frames at different depths have identical timestamps), which is not possible with a scanning-based 3D imaging system owing to the unavoidable time delay between successive measurements of different parts of the sample. Quite importantly, Deep-Z also enables correction for sample-drift-induced defocus after the image capture. Supplementary Video 4 shows a moving *C. elegans* recorded by a fluorescence microscope, where Deep-Z digitally brought the defocused nematode into focus (also see Supplementary Note 1). In addition to 3D imaging of a nematode, Deep-Z also works well to digitally refocus the images of fluorescent samples that are spatially denser such as the mitochondria and F actin structures within bovine pulmonary artery endothelial cells (BPAEC) (Supplementary Fig. 5).

Deep-Z not only substantially boosts the imaging speed, but also reduces photobleaching on the sample. For a wide-field fluorescence microscopy experiment, where an axial image stack is acquired, the illumination excites the fluorophores through the entire specimen, and the total light exposure of a given point within the sample volume is proportional to the number of imaging planes that are acquired during a single-pass z stack. By contrast, Deep-Z only requires a single image acquisition, if its axial training range covers the sample depth. This reduction, enabled by Deep-Z, in the number of axial planes that need to be imaged within a sample directly helps to reduce the photobleaching of samples (Supplementary Fig. 6 and Supplementary Note 2). This reduced light dose is also likely to reduce the phototoxicity associated with volumetric imaging.

So far, the blindly tested samples were inferred with a network that was trained using the same type of sample and the same microscopy system. In Supplementary Notes 3–4, we evaluated the performance of Deep-Z under different scenarios, where a change in the test data distribution is introduced in comparison to the training image set, such as (1) a different type of sample is imaged, (2) a different microscopy system is used for imaging and (3) a different illumination power or SNR is used. Our results (Supplementary Figs. 7–9) reveal the robustness of Deep-Z to these changes; however, to achieve the best performance using Deep-Z, the network should be trained (from scratch or through transfer learning, which expedites the training process) using training images obtained with the same microscope system and the same types of samples as are expected to be used at the testing phase.

As illustrated in Supplementary Notes 5–6, Deep-Z is also robust to changes in the density of the fluorescent objects within the sample (up to a limit, which is a function of the axial refocusing distance), the exposure time of the input images, as well as the illumination intensity modulation (Supplementary Figs. 10–13 and Supplementary Video 5).

Table 1 | Neuron segmentation results for a *C. elegans* worm using a watershed-based segmentation algorithm

	Watershed-based neuron segmentation		
	n **(neurons)**	**Δ*z*** **(mean ± s.d.; µm)**	**|Δ*z*|** **(mean ± s.d.; µm)**
Input image (***M*** = 1 image)	95	−0.265 ± 1.437	1.156 ± 0.885
Deep-Z output stack (***M*** = 1 image)	128	−0.575 ± 1.377	0.852 ± 1.223
Merged stack (***M*** = 2 images)	148	−0.157 ± 0.983	0.639 ± 0.761
Mechanical scan stack (***M*** = 41 images)	146	Ground truth	Ground truth

The resulting segmented neuron locations were also compared against the ground truth (that is, the corresponding mechanically scanned image stack, $M = 41$ images), reporting the axial error, Δz, as well as the absolute axial error, $|\Delta z|$, as mean ± s.d. in micrometers, respectively (also see Supplementary Note 7).

C. elegans **neuron segmentation**. For *C. elegans* neuron imaging, by virtual refocusing over an extended DOF, Deep-Z helps to segment more neurons and accurately predict their depth location as compared to a single focal plane image. To demonstrate this capability, we show the segmentation results of a *C. elegans* worm (Supplementary Figs. 14d–i) calculated using a watershed segmentation algorithm[40] from a 2D input image, the corresponding Deep-Z virtual image stack and the mechanically scanned ground-truth image stack (41 depths with 0.5-µm axial spacing); the results are summarized in Table 1. In comparison to the segmentation results obtained from the 2D input image (Supplementary Fig. 14e), the segmentation obtained using the Deep-Z virtual image stack (Supplementary Fig. 14f) detected 33 additional neurons, predicting the correct 3D positions of 128 neurons in total. In comparison to the ground-truth mechanically scanned 3D image stack (Supplementary Fig. 14i), the segmentation algorithm recognized 18 fewer neurons for the Deep-Z generated virtual stack, which were mostly located within the head, where the neurons are much denser and are relatively more challenging to recover and segment. In sparser regions of the worm, the neurons were mostly correctly segmented, matching the results obtained using the mechanically scanned 3D image stack (41 axial scans). The depth locations of the segmented neurons also matched well with the corresponding depths measured using the ground-truth mechanically scanned 3D image stack, with an average depth difference of $\Delta z = -0.575 \pm 1.377$ µm (Table 1).

To further improve Deep-Z-based neuron segmentation in denser regions of the sample (such as the head of a worm), images from more than one focal plane can be used as input. In comparison to the mechanically scanned 3D image stack, this is still substantially faster, requiring fewer images to recover the volume of the specimen. For instance, in Supplementary Fig. 14h we demonstrate the segmentation results of merging two virtual image stacks created by Deep-Z (taking the per-pixel maximum), both spanning −10 µm to 10 µm but generated from two different input images at $z = 0$ µm and $z = 4$ µm, respectively. The segmentation algorithm in this case identified $n = 148$ neurons and the results match better to the ground-truth axial scanning results, $n = 146$ (Table 1). To shed more light on this comparison, we also used another segmentation algorithm (TrackMate[41]) on the same image dataset; the results of

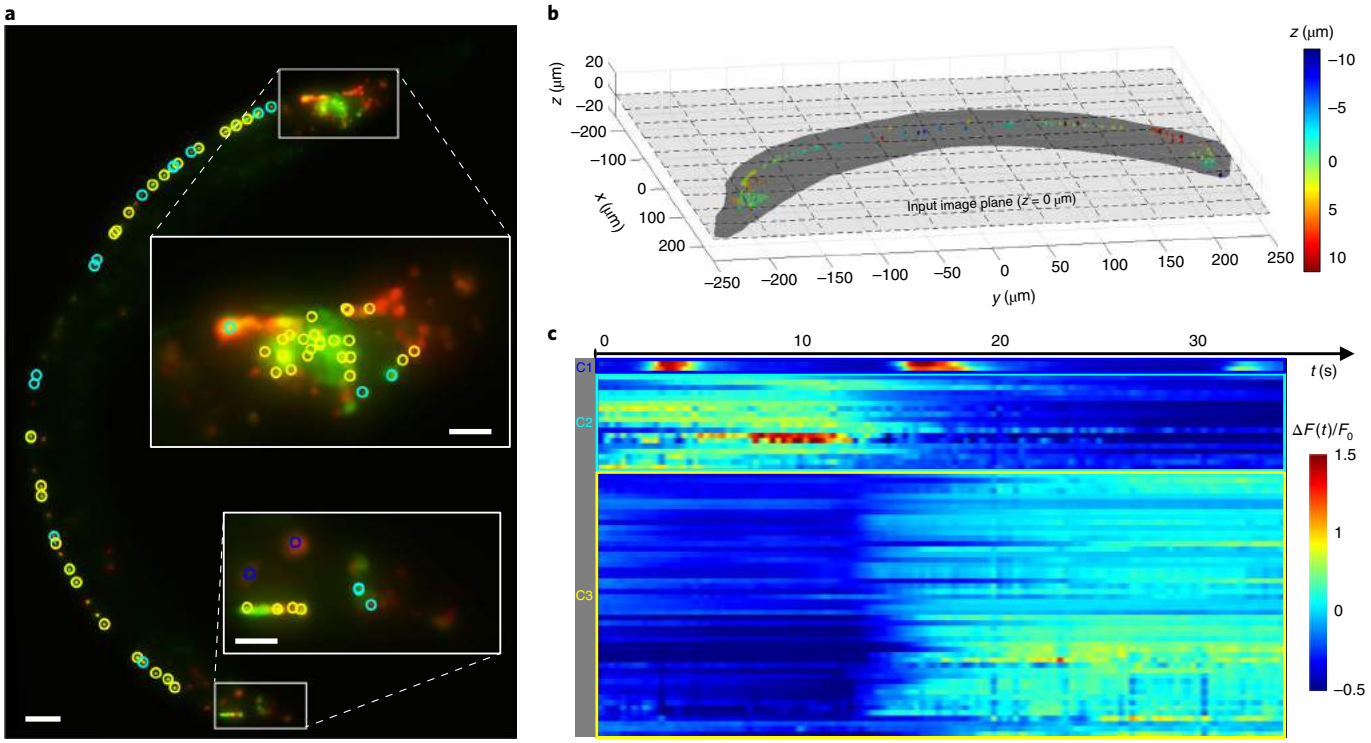


Fig. 3 | *C. elegans* neuron activity tracking in 3D using Deep-Z. a, MIP along the axial direction of the median-intensity image taken across the time sequence. The red channel (Texas Red) labels neuron nuclei. The green channel (FITC) labels neuron calcium activity. Scale bar, 25 μm. Scale bars for the expanded regions, 10 μm. **b**, All the 155 localized neurons are shown in 3D, depths are color-coded. **c**, 3D tracking of neuron calcium activity events corresponding to the 70 most active neurons. The neurons were grouped into three clusters (C1–C3) on the basis of similarities in their calcium activity patterns (Methods). The locations of these neurons are marked by the circles in **a**. The colors of the circles in **a** represents different clusters: C1 (blue), C2 (cyan) and C3 (yellow).

this analysis, summarized in Supplementary Note 7, confirmed a similar trend as that shown in Table 1.

3D functional imaging of *C. elegans* using Deep-Z. To highlight the utility of Deep-Z for tracking the activity of neurons in 3D, we recorded the fluorescence video of a *C. elegans* worm at a single focal plane ($z=0$ μm) at ~3.6 Hz for ~35 s, using a 20×/0.8 NA objective lens with two fluorescence channels, FITC for neuron activity and Texas Red for neuron locations. Each frame at each channel of the acquired video was digitally refocused using Deep-Z to a series of planes (−10 μm to 10 μm, 0.5-μm step size), generating a virtual 3D fluorescence stack for each acquired frame. Supplementary Video 6 shows a comparison of the recorded input video and a video of the maximum intensity projection (MIP) along z for these virtual stacks. As can be seen in this comparison, the neurons that are defocused in the input video can be refocused on demand at the Deep-Z output for both of the fluorescence channels. This enables accurate spatiotemporal tracking of individual neuron activity in 3D from a temporal sequence of 2D fluorescence images, captured at a single focal plane.

Next, we segmented the voxels of each neuron using the Texas Red channel, and tracked the change in the fluorescence intensity, that is, $\Delta F(t)=F(t)-F_0$, in the FITC channel (neuron activity) inside each neuron segment over time, where $F(t)$ is the neuron fluorescence emission intensity and F_0 is its time average (Methods). A total of 155 neurons in 3D were isolated using Deep-Z output images (Fig. 3b, Supplementary Video 7). For comparison, in Supplementary Fig. 14b we report the results of the same segmentation algorithm applied to just the input 2D image, in which 99 neurons were identified, without any depth information.

Figure 3c plots the activities of the 70 most active neurons, which were grouped into clusters C1–C3 on the basis of similarities in their calcium activity patterns (Methods). The activity of all of the 155 neurons inferred using Deep-Z are provided in Supplementary Fig. 15. Figure 3c reports that cluster C3 calcium activities increased at $t=14$ s, whereas the activities of cluster C2 decreased at a similar time point. These neurons very likely correspond to the type A and B motor neurons that promote backward and forward motion, respectively, which typically anticorrelate with each other[42]. Cluster C1 features two cells that were comparatively larger in size, located in the middle of the worm. These cells had three synchronized short spikes at $t=4$, 17 and 32 s. Their 3D positions and the regularity of their calcium activity pattern suggest that they are either neuronal or muscle cells of the defecation system that initiates defecation in regular intervals in coordination with the locomotion system[43]. We emphasize that all this 3D-tracked neuron activity was in fact embedded in the input 2D fluorescence image sequence, which was acquired at a single focal plane. Through Deep-Z, the neuron locations and activities were accurately tracked using a 2D microscopic time sequence, without the need for mechanical scanning, additional hardware or a trade-off of resolution or imaging speed.

As Deep-Z generates temporally synchronized virtual image stacks through digital refocusing, it can be used to match the imaging speed to the limit of the camera framerate. To highlight this opportunity, we used the stream mode of the camera of our microscope (Methods) and captured two videos at 100 frames per second to monitor the neuron nuclei (Texas Red) and the neuron calcium activity (FITC) of a moving *C. elegans* over a period of 10 s, and used Deep-Z to generate virtually refocused videos over an axial range of ±10 μm (Supplementary Videos 8 and 9).

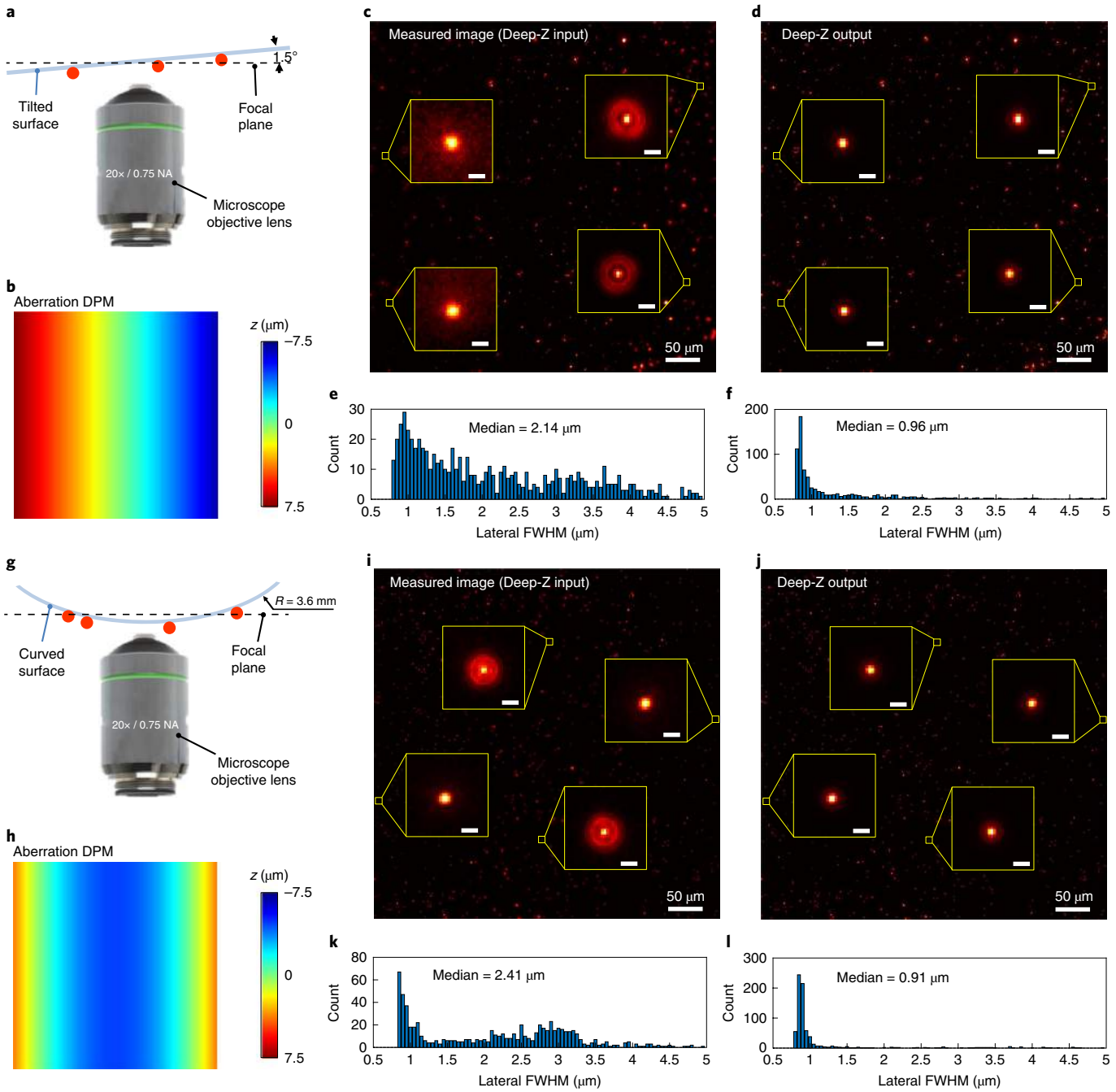


Fig. 4 | Non-uniform DPMs enable digital refocusing of a single fluorescence image onto user-defined 3D surfaces using Deep-Z. a, Measurement of a tilted fluorescent sample (300-nm beads). **b**, The corresponding DPM for this tilted plane. **c**, Measured raw fluorescence image; the left and right parts are out of focus in different directions, owing to the sample tilt. **d**, The Deep-Z output rapidly brings all the regions into correct focus. **e**,**f**, Lateral FWHM values of the nanobeads shown in **c**,**d**, respectively, clearly demonstrating that Deep-Z with the non-uniform DPM brought the out-of-focus particles into focus. **g**, Measurement of a cylindrical surface with fluorescent beads (300-nm beads). **h**, The corresponding DPM for this curved surface. **i**, Measured raw fluorescence image; the middle region and the edges are out-of-focus owing to the curvature of the sample. **j**, The Deep-Z output rapidly brings all the regions into correct focus. **k**,**l**, report the lateral FWHM values of the nanobeads shown in **i**,**j**, respectively, clearly demonstrating that Deep-Z with the non-uniform DPM brought the out-of-focus particles into focus, generating monodispersed distributions with a median of ~0.96 μm and ~0.91 μm, respectively. Experiments were repeated with 32 images, achieving similar results. Scale bars for the expanded regions, 2 μm.

Deep-Z-based aberration correction using spatially non-uniform DPMs. Even though Deep-Z is trained with uniform DPMs, during testing one can also use spatially non-uniform entries as part of a DPM to refocus an input fluorescence image onto user-defined 3D surfaces. Such a unique capability can be useful, among many applications, for simultaneous autofocusing of different parts of a fluorescence image after image capture and measurement or assessment of the aberrations introduced by the optical system (and/or the sample), as well as for correction of such aberrations by applying a desired non-uniform DPM. To exemplify this opportunity, Fig. 4 demonstrates the correction of the planar tilting and cylindrical curvature of two different samples, after the acquisition of a single

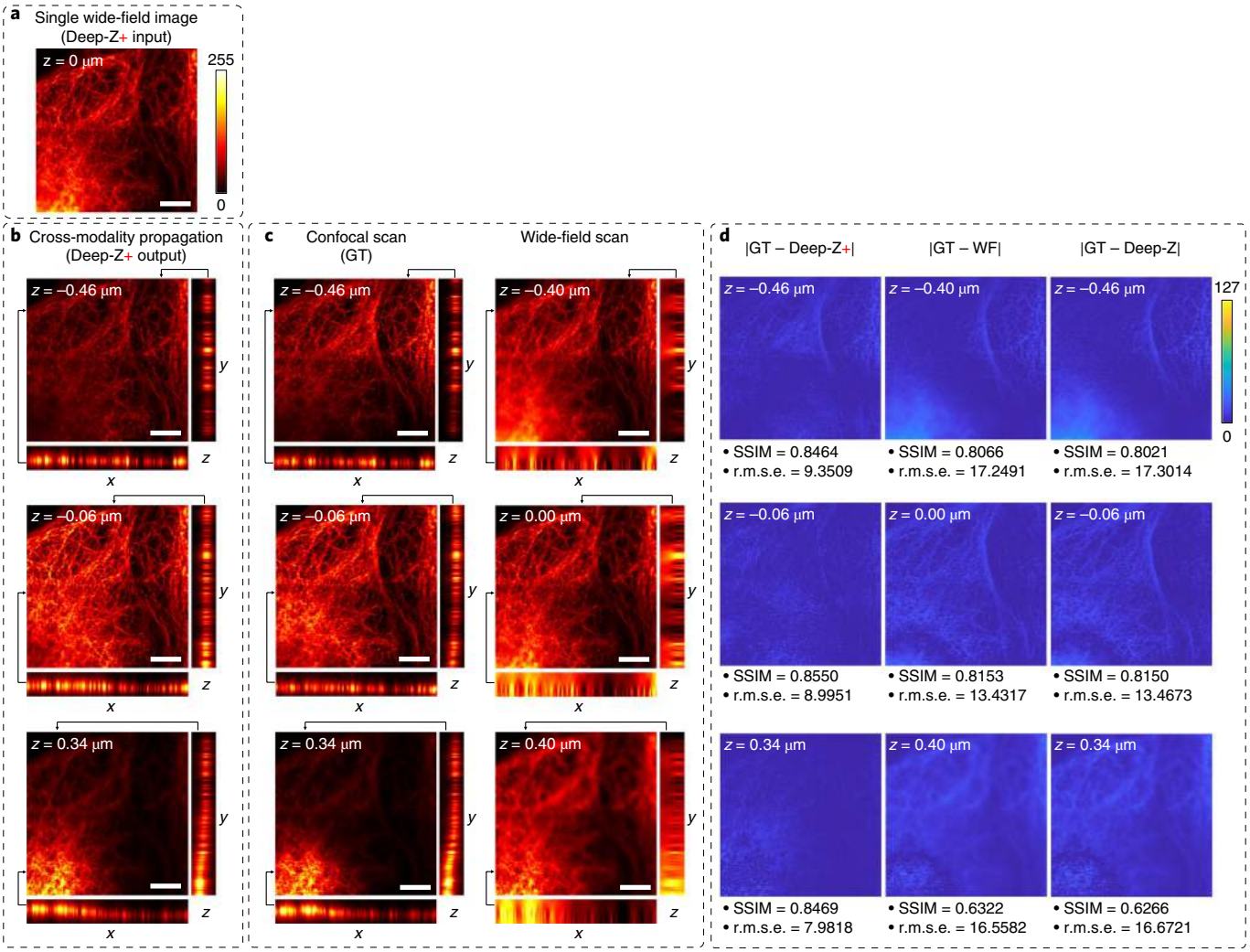


Fig. 5 | Deep-Z+, cross-modality digital refocusing of fluorescence images. a–c, A single wide-field fluorescence image (63×/1.4 NA objective lens) of BPAEC microtubule structures (**a**) was digitally refocused using Deep-Z+ to different planes in 3D (**b**), matching the images captured by a confocal microscope at the corresponding planes (**c**), retrieving volumetric information from a single input image and performing axial sectioning at the same time. Wide-field (WF) images are also shown in **c** for comparison. The cross-sections (*x*–*z* and *y*–*z*) of refocused images are shown to demonstrate the match between Deep-Z+ inference and the ground-truth (GT) confocal microscope images of the same planes; the same cross sections (*x*–*z* and *y*–*z*) are also shown for a wide-field scanning fluorescence microscope, reporting a substantial axial blur in each case. Each cross-sectional expanded image spans 1.6 μm in the *z* direction (with an axial step size of 0.2 μm) and the dotted arrows mark the locations at which the *x*–*z* and *y*–*z* cross sections were taken. **d**, The absolute difference images of the Deep-Z+ output with respect to the corresponding confocal images, with SSIM and r.m.s.e. values, further quantifying the performance of Deep-Z+. For comparison, we also show the absolute difference images of the 'standard' Deep-Z output images and the scanning wide-field fluorescence microscope images with respect to the corresponding confocal images, both of which report increased error and weaker SSIM as compared to |GT – Deep-Z+|. The quantitative match between |GT – WF| and |GT – Deep-Z| also suggests that the impact of 60-nm axial offset between the confocal and wide-field image stacks is negligible. Scale bars, 10 μm. Experiments were repeated with 42 images, achieving similar results. Color scales indicate intensity.

2D fluorescence image per object. Figure 4a illustrates the first measurement, in which the plane of a fluorescent nanobead sample was tilted with respect to the focal plane of the objective lens (Methods). By using a non-uniform DPM (see Fig. 4b), which represents the sample tilt, Deep-Z can act on the blurred input image (Fig. 4c) and accurately bring all the nanobeads into focus (Fig. 4d), even though it was only trained using uniform DPMs. The lateral FWHM values calculated at the network output image became monodispersed, with a median of ~0.96 μm (Fig. 4f), as compared to a median of ~2.14 μm at the input image (Fig. 4e). Similarly, Fig. 4g illustrates the second measurement, where the nanobeads were distributed on a cylindrical surface with a diameter of ~7.2 mm. Using a non-uniform DPM that defines this cylindrical surface (Fig. 4h), the aberration in Fig. 4i was corrected using Deep-Z (Fig. 4j); the lateral FWHM values calculated at the network output once again became monodispersed (Fig. 4l). Supplementary Note 8 details an analysis on the 3D surface curvature that a DPM can have without generating artifacts.

Cross-modality digital refocusing of fluorescence images. Deep-Z can also be used to perform cross-modality digital refocusing of an input image, where the generator can be trained using pairs of input and label images captured by two different fluorescence imaging modalities, which we term as Deep-Z+. To demonstrate this capability, we trained a Deep-Z+ network using pairs of wide-field microscopy images (inputs) and confocal microscopy

images at the corresponding planes (ground-truth labels) to perform cross-modality digital refocusing (Methods). Figure 5 demonstrates our blind-testing results for imaging microtubule structures of BPAEC using Deep-Z+. The trained Deep-Z+ network digitally refocused the input wide-field fluorescence image onto different axial distances, while at the same time rejecting some of the defocused spatial features at the refocused planes, matching the confocal images of the corresponding planes, which serve as our ground-truth (Fig. 5). For example, the microtubule structure at the lower left corner of Fig. 5b, which was prominent at a refocusing distance of $z = 0.34\,\mu m$, was digitally rejected by Deep-Z+ at a refocusing distance of $z = -0.46\,\mu m$ as it became out of focus at this axial distance, matching the corresponding image of the confocal microscope at the same depth (Fig. 5c). Wide-field images are also shown in Fig. 5c for comparison. These scanning wide-field images report the closest heights to the corresponding confocal images, and have an axial offest of 60 nm, as the two image stacks are discretely scanned and digitally aligned to each other. Figure 5 reports x–z and y–z cross sections of Deep-Z+ output images, in which the axial distributions of the microtubule structures are substantially sharper as compared to the axial scanning images of a wide-field fluorescence microscope, providing a very good match to the cross sections obtained with a confocal microscope, matching the aim of its training.

Discussion

We developed a unique framework, termed Deep-Z, powered by deep neural networks, that enables rapid 3D refocusing within a sample using a single 2D fluorescence image as input. This framework is non-iterative and does not require hyperparameter tuning after its training stage. Even though the network is only trained using uniform DPMs, one can still apply various non-uniform DPMs during the inference stage to enable, for example, correction of sample drift, tilt, curvature or other optical aberrations, which might prove useful for longitudinal imaging experiments in biology and life sciences, by digitally recovering valuable information that might otherwise be lost owing to, for example, the sample becoming out of focus or tilted over time. On the basis of these unique features, Deep-Z also has the potential to reduce the photobleaching of samples that is associated with volumetric fluorescence imaging.

Yet another unique feature of this Deep-Z framework is that it permits cross-modality virtual refocusing of fluorescence images, where the network is trained with gold-standard label images obtained by a different fluorescence microscopy modality (for example, confocal) to teach the generator network to digitally refocus an input image (for example, an image acquired by wide-field microscopy) onto another plane within the sample volume, but this time to match the image of the same plane acquired by a different fluorescence imaging modality as compared to the input image. Figure 5 contains an example of wide-field to confocal transformation results.

We also demonstrated the efficacy of Deep-Z for structural and functional imaging of neurons in *C. elegans* nematodes. For neuron segmentation applications, we observed that Deep-Z output images in denser regions of a sample (for example, the head of a *C. elegans*) resulted in under-counting of the segmented neurons, which was improved by merging the Deep-Z output images resulting from two different focal planes as input. In fact, neuron segmentation is in general a challenging task, and not all the neurons in the body of a worm can be accurately identified in each experiment, even using a mechanically scanned image stack with a high NA objective and state-of-the-art neuron-segmentation algorithms[44]. Although not demonstrated here, Deep-Z can potentially be used as a front-end module to jointly optimize future deep-learning-based neuron-segmentation algorithms, which can make use of Deep-Z to reduce the number of images required to accurately and efficiently track neural activity of model organisms. This could also benefit from new generator architectures that utilize more than one input images, for example, from different focal planes, to more effectively combine additional 3D information acquired at different planes.

Finally, we should note that the retrievable axial range in Deep-Z depends on the SNR of the recorded image, that is, if the depth information carried by the PSF falls below the noise floor, accurate inference will be challenging. To validate the performance of a pre-trained Deep-Z network under variable SNR, we tested the inference of Deep-Z at different exposure conditions (Supplementary Figure 7), revealing the robustness of its inference over a broad range of exposure times that were not included in the training data (Supplementary Note 3). Our results demonstrated an enhancement of ~20× in the DOF of a wide-field fluorescence image using Deep-Z. This refocusing range is in fact not an absolute limit but rather a practical choice for our training data, and it may be further improved through hardware modifications to the optical setup by engineering the PSF in the axial direction[3,12,45–47]. Supplementary Video 10 shows an experimental demonstration of Deep-Z blind inference for a double-helix PSF.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41592-019-0622-5.

Received: 25 March 2019; Accepted: 30 September 2019;
Published online: 04 November 2019

References

1. Nguyen, J. P. et al. Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **113**, E1074–E1081 (2016).
2. Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M. & Vaziri, A. Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat. Methods* **10**, 1013–1020 (2013).
3. Tomer, R. et al. SPED light sheet microscopy: fast mapping of biological system structure and function. *Cell* **163**, 1796–1806 (2015).
4. Siedentopf, H. & Zsigmondy, R. Uber sichtbarmachung und größenbestimmung ultramikoskopischer teilchen, mit besonderer anwendung auf goldrubingläser. *Ann. Phys.* **315**, 1–39 (1902).
5. Lerner, T. N. et al. Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* **162**, 635–647 (2015).
6. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
7. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
8. Henriques, R. et al. QuickPALM: 3D real-time photoactivation nanoscopy image processing in Image. *J. Nat. Methods* **7**, 339–340 (2010).
9. 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).
10. 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).
11. Juette, M. F. et al. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nat. Methods* **5**, 527–529 (2008).
12. Pavani, S. R. P. et al. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Natl Acad. Sci. USA* **106**, 2995–2999 (2009).
13. Prevedel, R. et al. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* **11**, 727–730 (2014).
14. Levoy, M., Ng, R., Adams, A., Footer, M. & Horowitz, M. Light Field Microscopy. In *ACM SIGGRAPH 2006 Papers* 924–934 (ACM, 2006).
15. Pégard, N. C. et al. Compressive light-field microscopy for 3D neural activity recording. *Optica* **3**, 517–524 (2016).
16. Broxton, M. et al. Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express* **21**, 25418–25439 (2013).
17. Cohen, N. et al. Enhancing the performance of the light field microscope using wavefront coding. *Opt. Express* **22**, 24817–24839 (2014).
18. Wagner, N. et al. Instantaneous isotropic volumetric imaging of fast biological processes. *Nat. Methods* **16**, 497–500 (2019).
19. Rosen, J. & Brooker, G. Non-scanning motionless fluorescence three-dimensional holographic microscopy. *Nat. Photonics* **2**, 190–195 (2008).

20. Brooker, G. et al. In-line FINCH super resolution digital holographic fluorescence microscopy using a high efficiency transmission liquid crystal GRIN lens. *Opt. Lett.* **38**, 5264–5267 (2013).
21. Siegel, N., Lupashin, V., Storrie, B. & Brooker, G. High-magnification super-resolution FINCH microscopy using birefringent crystal lens interferometers. *Nat. Photonics* **10**, 802–808 (2016).
22. Abrahamsson, S. et al. Fast multicolor 3D imaging using aberration-corrected multifocus microscopy. *Nat. Methods* **10**, 60–63 (2013).
23. Abrahamsson, S. et al. MultiFocus polarization microscope (MF-PolScope) for 3D polarization imaging of up to 25 focal planes simultaneously. *Opt. Express* **23**, 7734–7754 (2015).
24. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* https://doi.org/10.1038/s41592-019-0458-z (2019).
25. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
26. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).
27. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018).
28. Rivenson, Y. et al. Deep learning enhanced mobile-phone microscopy. *ACS Photonics* **5**, 2354–2364 (2018).
29. Haan, K., de, Ballard, Z. S., Rivenson, Y., Wu, Y. & Ozcan, A. Resolution enhancement in scanning electron microscopy using deep learning. *Sci. Rep.* **9**, 1–7 (2019).
30. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103– (2019).
31. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090– (2018).
32. Zhang, X. et al. Deep learning optical-sectioning method. *Opt. Express* **26**, 30762–30772 (2018).
33. Wu, Y. et al. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl.* **8**, 25 (2019).
34. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
35. Goodfellow, I. et al. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014).
36. Mirza, M. & Osindero, S. Conditional Generative Adversarial Nets. Preprint at *arXiv* https://arxiv.org/abs/1411.1784 (2014).
37. Shaw, P. J. & Rawlins, D. J. The point-spread function of a confocal microscope: its measurement and use in deconvolution of 3-D data. *J. Microsc.* **163**, 151–165 (1991).
38. 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).
39. Nguyen, J. P., Linder, A. N., Plummer, G. S., Shaevitz, J. W. & Leifer, A. M. Automatically tracking neurons in a moving and deforming brain. *PLoS Comput. Biol.* **13**, e1005517 (2017).
40. Gonzalez, R. C., Woods, R. E. & Eddins, S. L. *Digital Image Processing Using MATLAB* (McGraw-Hill, 2004).
41. Tinevez, J.-Y. et al. TrackMate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
42. Kato, S. et al. Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell* **163**, 656–669 (2015).
43. Nagy, S., Huang, Y.-C., Alkema, M. J. & Biron, D. *Caenorhabditis elegans* exhibit a coupling between the defecation motor program and directed locomotion. *Sci. Rep.* **5**, 17174 (2015).
44. Toyoshima, Y. et al. Accurate automatic detection of densely distributed cell nuclei in 3D space. *PLoS Comput. Biol.* **12**, e1004970 (2016).
45. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
46. Antipa, N. et al. DiffuserCam: lensless single-exposure 3D imaging. *Optica* **5**, 1–9 (2018).
47. Shechtman, Y., Sahl, S. J., Backer, A. S. & Moerner, W. E. Optimal point spread function design for 3D imaging. *Phys. Rev. Lett.* **113**, 133902 (2014).

Methods

Sample preparation. The 300-nm red fluorescence nanobeads were purchased from MagSphere (PSF-300NM 0.3 UM RED), diluted 5,000 times with methanol and ultrasonicated for 15 min before and after dilution to break down the clusters. For the fluorescent bead samples on a flat surface and a tilted surface, a number 1 coverslip (22 × 22 mm^2, thickness of ~150 μm) was thoroughly cleaned and plasma treated. Then, a 2.5-μl droplet of the diluted bead sample was pipetted onto the coverslip and dried. For the fluorescent bead sample on a curved (cylindrical) surface, a glass tube (diameter of ~7.2 mm) was thoroughly cleaned and plasma treated. Then, a 2.5-μl droplet of the diluted bead sample was pipetted onto the outer surface of the glass tube and dried.

Structural imaging of *C. elegans* neurons was carried out in strain AML18. AML18 carries the genotype wtfIs3 [rab-3p::NLS::GFP+rab-3p::NLS::tagRFP] and expresses GFP and tagRFP in the nuclei of all the neurons[39]. For functional imaging, we used the strain AML32, carrying wtfIs5 [rab-3p::NLS::GCaMP6s+rab-3p::NLS::tagRFP]. The strains were acquired from the *Caenorhabditis* Genetics Center. Worms were cultured on nematode growth medium seeded with OP50 bacteria using standard conditions[48]. For imaging, worms were washed off the plates with M9 and anesthetized with 3 mM levamisole[49]. Anesthetized worms were then mounted on slides seeded with 3% agarose. To image moving worms, the levamisole was omitted.

Two slides of multilabeled BPAEC were acquired from Thermo Fisher: FluoCells Prepared Slide 1 and FluoCells Prepared Slide 2. These cells were labeled to express different cell structures and organelles. The first slide uses Texas Red for mitochondria and FITC for F-actin structures. The second slide uses FITC for microtubules.

Fluorescence image acquisition. The fluorescence images of nanobeads, *C. elegans* structure and BPAEC samples were captured by an inverted scanning microscope (IX83, Olympus Life Science) using a 20×/0.75 NA objective lens (UPLSAPO20X, Olympus Life Science). A 130-W fluorescence light source (U-HGLGPS, Olympus Life Science) was used at 100% output power. Two bandpass optical filter sets were used, Texas Red and FITC. The bead samples were captured by placing the coverslip with beads directly on the microscope sample mount. The tilted surface sample was captured by placing the coverslip with beads on a 3D-printed holder, which created a 1.5° tilt with respect to the focal plane. The cylindrical tube surface with fluorescent beads was placed directly on the microscope sample mount. These fluorescent bead samples were imaged using a Texas Red filter set. The *C. elegans* sample slide was placed on the microscope sample mount and imaged using a Texas Red filter set. The BPAEC slide was placed on the microscope sample mount and imaged using Texas Red and FITC filter sets. For all the samples, the scanning microscope had a motorized stage (ProScan *XY* stage kit for IX73/83) that moved the samples to different FOVs and performed image-contrast-based autofocus at each location. The motorized stage was controlled using MetaMorph microscope automation software (Molecular Devices). At each location, the control software autofocused the sample on the basis of the s.d. of the image, and a *z*-stack was taken from −20 μm to 20 μm with a step size of 0.5 μm. The image stack was captured by a monochrome scientific CMOS camera (ORCA-flash4.0 v2, Hamamatsu Photonics K.K), and saved in uncompressed tiff format with 81 planes and 2,048 × 2,048 pixels in each plane.

The images of *C. elegans* neuron activity were captured by another scanning wide-field fluorescence microscope (TCS SP8, Leica Microsystems) using a 20×/0.8 NA objective lens (HCPLAPO20x/0.80DRY, Leica Microsystems) and a 40×/1.3 NA objective lens (HC PL APO 40×/1.30 oil, Leica Microsystems). Two bandpass optical filter sets were used, Texas Red and FITC. The images were captured by a monochrome scientific CMOS camera (Leica DFC9000GTC-VSC08298). For capturing image stacks of anesthetized worms, the motorized stage controlled by a control software (LAS X, Leica Microsystems) moved the sample slide to different FOVs. At each FOV, the control software took a *z* stack from −20 μm to 20 μm with a step size of 0.5 μm for the 20×/0.8 NA objective lens images or a step size of 0.27 μm for the 40×/1.3 NA objective lens images, with respect to a middle plane ($z = 0$ μm). Two images were taken at each *z* plane, for the Texas Red channel and FITC channel, respectively. For capturing 2D videos of dynamic worms, the control software took a time-lapse video that also time-multiplexed the Texas Red and FITC channels at the maximum speed of the system. This resulted in an average framerate of ~3.6 frames per second for a maximum camera framerate of 10 frames per second, for imaging both channels.

The BPAEC wide-field and confocal fluorescence images were captured by another inverted scanning microscope (TCS SP5, Leica Microsystems). The images were acquired using a 63×/1.4 NA objective lens (HC PL APO 63×/1.40 oil CS2, Leica Microsystems) and a FITC filter set was used. The wide-field images were recorded by a charge-coupled device with 1,380 × 1,040 pixels and a 12-bit dynamic range, whereas the confocal images were recorded by a photo-multiplier tube with 1,024 × 1,024 pixels and an 8-bit dynamic range. The scanning microscope had a motorized stage that moved the sample to different FOVs and depths. For each location, a stack of 12 images with 0.2-μm axial spacing was recorded.

Image preprocessing and preparation of training data. Each captured image stack was first axially aligned using the ImageJ plugin 'StackReg'[50], which corrects the rigid shift and rotation caused by the microscope stage inaccuracy. Then an extended depth of field (EDF) image was generated using the ImageJ plugin 'Extended Depth of Field'[51]. This EDF image was used as a reference image to normalize the whole image stack in the following steps: (1) a triangular threshold[52] was used on the image to separate the background and foreground pixels; (2) the mean intensity of the background pixels of the EDF image was determined to be the background noise and subtracted; (3) the EDF image intensity was scaled to 0–1, where the scale factor was determined such that 1% of the foreground pixels above the background were greater than one (that is, saturated); and (4) the background level was subtracted from each image in the stack and each image was normalized by the intensity scaling factor. For testing data without an image stack, steps 1–3 were applied on the input image instead of the EDF image.

To prepare the training and validation datasets, on each FOV, a geodesic dilation[40] with fixed thresholds was applied on fluorescence EDF images to generate a mask that represents the regions containing the sample fluorescence signal above the background. Then, a customized greedy algorithm was used to determine a minimal set of regions with 256 × 256 pixels that covered this mask, with ~5% area overlap between these training regions. The lateral locations of these regions were used to crop images on each height of the image stack, where the middle plane for each region was set to be the one with the highest s.d. Then, 20 planes above and 20 planes below this middle plane were set to be the range of the stack and an input image plane was generated from each of these 41 planes. Depending on the size of the dataset, around 5–10 of these 41 planes were randomly selected as the corresponding target plane, forming around 150–300 image pairs. For each one of these image pairs, the refocusing distance was determined on the basis of the location of the plane (that is, 0.5 μm multiplied by the difference from the input plane to the target plane). By repeating this number, a uniform DPM was generated and appended to the input fluorescence image. The final dataset typically contained ~100,000 image pairs. This was randomly divided into a training dataset and a validation dataset, which took 85% and 15% of the data, respectively. During the training process, each data point was further augmented five times by flipping or rotating the images by a random multiple of 90°. The validation dataset was not augmented. The testing dataset was cropped from separate measurements with sample FOVs that did not overlap with the FOVs of the training and validation datasets.

Deep-Z network architecture. The Deep-Z network is formed by a least square GAN framework[53], and it is composed of two parts: a generator and a discriminator (Supplementary Note 9). The generator is a convolutional neural network inspired by the U-Net[54], and follows a similar structure as that seen in refs. [33,55]. The generator network consists of a downsampling path and a symmetric upsampling path. In the downsampling path, there are five downsampling blocks. Each block contains two convolutional layers that map the input tensor x_k to the output tensor x_{k+1}

$$x_{k+1} = x_k + \mathrm{ReLU}[\mathrm{CONV}_{k_2}\{\mathrm{ReLU}[\mathrm{CONV}_{k_1}\{x_k\}]\}] \quad (1)$$

where ReLU[.] stands for the rectified linear unit operation and CONV{.} stands for the convolution operator (including the bias terms). The subscript of CONV denotes the number of channels in the convolutional layer; along the downsampling path we have: $k_1 = 25$, 72, 144, 288, 576 and $k_2 = 48$, 96, 192, 384, 768 for levels $k = 1$, 2, 3, 4, 5, respectively. The '+' sign in equation (1) represents a residual connection. Zero padding was used on the input tensor x_k to compensate for the channel number mismatch between the input and output tensors. The connection between two consecutive downsampling blocks is a 2 × 2 max-pooling layer with a stride of 2 × 2 pixels to perform a 2× downsampling. The fifth downsampling block connects to the upsampling path, which will be detailed next.

In the upsampling path, there are four corresponding upsampling blocks, each of which contains two convolutional layers that map the input tensor y_{k+1} to the output tensor y_k using:

$$y_k = \mathrm{ReLU}[\mathrm{CONV}_{k_4}\{\mathrm{ReLU}[\mathrm{CONV}_{k_3}\{\mathrm{CAT}(x_{k+1}, y_{k+1})\}]\}] \quad (2)$$

where the CAT(·) operator represents the concatenation of the tensors along the channel direction, that is, CAT(x_{k+1},y_{k+1}) appends tensor x_{k+1} from the downsampling path to the tensor y_{k+1} in the upsampling path at the corresponding level $k+1$. The number of channels in the convolutional layers, denoted by k_3 and k_4, are $k_3 = 72$, 144, 288, 576 and $k_4 = 48$, 96, 192, 384 along the upsampling path for $k = 1$, 2, 3, 4, respectively. The connection between consecutive upsampling blocks is an up-convolution (convolution transpose) block that upsamples the image pixels by 2×. The last block is a convolutional layer that maps the 48 channels to one output channel.

The discriminator is a convolutional neural network that consists of six consecutive convolutional blocks, each of which maps the input tensor z_i to the output tensor z_{i+1}, for a given level i

$$z_{i+1} = \mathrm{LReLU}[\mathrm{CONV}_{i_2}\{\mathrm{LReLU}[\mathrm{CONV}_{i_1}\{z_i\}]\}] \quad (3)$$

where the LReLU stands for a leaky ReLU operator with a slope of 0.01. The subscript of the convolutional operator represents its number of channels, which

are $i_1 = 48, 96, 192, 384, 768, 1{,}536$ and $i_2 = 96, 192, 384, 768, 1{,}536, 3{,}072$, for the convolution block $i = 1, 2, 3, 4, 5, 6$, respectively.

After the last convolutional block, an average pooling layer flattens the output and reduces the number of parameters to 3,072. Subsequently there are fully connected layers of size $3{,}072 \times 3{,}072$ with LReLU activation functions, and another fully connected layer of size $3{,}072 \times 1$ with a sigmoid activation function. The final output represents the discriminator score, which falls within (0, 1), where 0 represents a false and 1 represents a true label.

All the convolutional blocks use a convolutional kernel size of 3×3 pixels and replicate padding of one pixel unless mentioned otherwise. All the convolutions have a stride of 1×1 pixel, except the second convolutions in equation (3), which has a stride of 2×2 pixels to perform a 2× downsampling in the discriminator path. The weights are initialized using the Xavier initializer[56] and the biases are initialized to 0.1.

Training and testing of the Deep-Z network. The Deep-Z network learns to use the information given by the appended DPM to digitally refocus the input image to a user-defined plane. In the training phase, the input data of the generator G(.) have the dimensions $256 \times 256 \times 2$, where the first channel is the fluorescence image and the second channel is the user-defined DPM. The target data of G(.) have the dimensions of 256×256, which represent the corresponding fluorescence image at a surface specified by the DPM. The input data of the discriminator D(.) have the dimensions of 256×256, which can be either the generator output or the corresponding target $z^{(i)}$. During the training phase, the network iteratively minimizes the generator loss L_G and discriminator loss L_D, defined as:

$$L_G = \frac{1}{2N}\sum_{i=1}^{N}\left[D\left(G\left(x^{(i)}\right)\right) - 1\right]^2 + \alpha\frac{1}{2N}\sum_{i=1}^{N}\mathrm{MAE}\left(x^{(i)}, z^{(i)}\right) \quad (4)$$

$$L_D = \frac{1}{2N}\sum_{i=1}^{N}\left[D\left(G\left(x^{(i)}\right)\right)\right]^2 + \frac{1}{2N}\sum_{i=1}^{N}\left[D\left(z^{(i)}\right) - 1\right]^2 \quad (5)$$

where N is the number of images used in each batch (for example, $N = 20$), $G(x^{(i)})$ is the generator output for the input $x^{(i)}$, $z^{(i)}$ is the corresponding target label, D(.) is the discriminator, and MAE(.) stands for mean absolute error. α is a regularization parameter for the GAN loss and the MAE loss in L_G. In the training phase, it was chosen as $\alpha = 0.02$. For training stability and optimal performance, an adaptive momentum optimizer was used to minimize both L_G and L_D, with a learning rate of 10^{-4} and 3×10^{-5} for L_G and L_D, respectively. In each iteration, six updates of the generator loss and three updates of the discriminator loss were performed. The validation set was tested every 50 iterations and the best network (to be blindly tested) was chosen to be the one with the smallest MAE loss on the validation set.

In the testing phase, once the training is complete, only the generator network is active. Limited by the graphical memory of our GPU, the largest image FOV that we tested was $1{,}536 \times 1{,}536$ pixels. Because the image was normalized to be in the range 0–1, whereas the refocusing distance was on the scale of around −10 to 10 (in micrometers), the DPM entries were divided by 10 to be in the range of −1 to 1 before the training and testing of the Deep-Z network to keep the dynamic range of the image and DPM matrices similar to each other.

The network was implemented using TensorFlow[57], performed on a PC with Intel Core i7-8700K six-core 3.7 GHz CPU and 32 GB RAM, using an Nvidia GeForce 1080Ti GPU. On average, the training takes ~70 h for ~400,000 iterations (equivalent to ~50 epochs). After the training, the network inference time was ~0.2 s for an image with 512×512 pixels and ~1 s for an image with $1{,}536 \times 1{,}536$ pixels on the same PC.

Measurement of the lateral and axial FWHM values of the fluorescent beads samples. For characterizing the lateral FWHM of the fluorescent beads samples, a threshold was performed on the image to extract the connected components. Then, individual regions of 30×30 pixels were cropped around the centroid of these connected components. A 2D Gaussian fit was performed on each of these individual regions, which was done using lsqcurvefit (https://www.mathworks.com/help/optim/ug/lsqcurvefit.html) in Matlab (MathWorks) to match the function:

$$\mathrm{I}(x, y) = A\exp\left[\frac{(x - x_c)^2}{2\sigma_x^2} + \frac{(y - y_c)^2}{2\sigma_y^2}\right] \quad (6)$$

The lateral FWHM was then calculated as the mean FWHM of x and y directions

$$\mathrm{FWHM}_{\mathrm{lateral}} = 2\sqrt{2\ln 2}\frac{\sigma_x\Delta_x + \sigma_y\Delta_y}{2} \quad (7)$$

where $\Delta_x = \Delta_y = 0.325\,\mu\mathrm{m}$ was the effective pixel size of the fluorescence image on the object plane. A histogram was subsequently generated for the lateral FWHM values for all the thresholded beads (for example, $n = 461$ for Fig. 1 and $n > 750$ for Fig. 4).

To characterize the axial FWHM values for the bead samples, slices along the x–z direction with 81 steps were cropped at $y = y_c$ for each bead, from either the digitally refocused or the mechanically scanned axial image stack. Another 2D Gaussian fit was performed on each cropped slice, to match the function

$$\mathrm{I}(x, z) = A\exp\left[\frac{(x - x_c)^2}{2\sigma_x^2} + \frac{(z - z_c)^2}{2\sigma_z^2}\right] \quad (8)$$

The axial FWHM was then calculated as

$$\mathrm{FWHM}_{\mathrm{axial}} = 2\sqrt{2\ln 2}\sigma_z\Delta_z \quad (9)$$

where $\Delta_z = 0.5\,\mu\mathrm{m}$ was the axial step size. A histogram was subsequently generated for the axial FWHM values.

Image quality evaluation. The network output images $\mathrm{I}^{\mathrm{out}}$ were evaluated with reference to the corresponding ground-truth images I^{GT} using the following five criteria: (1) mean square error (MSE), (2) r.m.s.e., (3) MAE, (4) correlation coefficient and (5) SSIM[58]. The MSE is one of the most widely used error metrics, defined as

$$\mathrm{MSE}\left(\mathrm{I}^{\mathrm{out}}, \mathrm{I}^{\mathrm{GT}}\right) = \frac{1}{N_x N_y}\left\|\mathrm{I}^{\mathrm{out}} - \mathrm{I}^{\mathrm{GT}}\right\|_2^2 \quad (10)$$

where N_x and N_y represent the number of pixels in the x and y directions, respectively. The square root of MSE results in r.m.s.e. In comparison to MSE, MAE uses 1-norm difference (absolute difference) instead of 2-norm difference, which is less sensitive to outlier pixels

$$\mathrm{MAE}\left(\mathrm{I}^{\mathrm{out}}, \mathrm{I}^{\mathrm{GT}}\right) = \frac{1}{N_x N_y}\left\|\mathrm{I}^{\mathrm{out}} - \mathrm{I}^{\mathrm{GT}}\right\|_1 \quad (11)$$

The correlation coefficient is defined as

$$\mathrm{corr}\left(\mathrm{I}^{\mathrm{out}}, \mathrm{I}^{\mathrm{GT}}\right) = \frac{\sum_x\sum_y\left(\mathrm{I}_{xy}^{\mathrm{out}} - \mu_{\mathrm{out}}\right)\left(\mathrm{I}_{xy}^{\mathrm{GT}} - \mu_{\mathrm{GT}}\right)}{\sqrt{\left(\sum_x\sum_y\left(\mathrm{I}_{xy}^{\mathrm{out}} - \mu_{\mathrm{out}}\right)^2\right)\left(\sum_x\sum_y\left(\mathrm{I}_{xy}^{\mathrm{GT}} - \mu_{\mathrm{GT}}\right)^2\right)}} \quad (12)$$

where μ_{out} *and* μ_{GT} are the mean values of the images $\mathrm{I}^{\mathrm{out}}$ *and* I^{GT} respectively.

While these criteria listed above can be used to quantify errors in the network output as compared to the GT, they are not strong indicators of the perceived similarity between two images. SSIM aims to address this shortcoming by evaluating the structural similarity in the images, defined as

$$\mathrm{SSIM}\left(\mathrm{I}^{\mathrm{out}}, \mathrm{I}^{\mathrm{GT}}\right) = \frac{(2\mu_{\mathrm{out}}\mu_{\mathrm{GT}} + C_1)(2\sigma_{\mathrm{out,GT}} + C_2)}{(\mu_{\mathrm{out}}^2 + \mu_{\mathrm{GT}}^2 + C_1)(\sigma_{\mathrm{out}}^2 + \sigma_{\mathrm{GT}}^2 + C_2)} \quad (13)$$

where σ_{out} and σ_{GT} are the standard deviations of $\mathrm{I}^{\mathrm{out}}$ and I^{GT} respectively, and $\sigma_{\mathrm{out,GT}}$ is the cross-variance between the two images; C_1 and C_2 are constants, used to avoid division by a small denominator.

Tracking and quantification of *C. elegans* neuron activity. The *C. elegans* neuron activity tracking video was captured by time multiplexing the two fluorescence channels (FITC, followed by Texas Red and then FITC and so on). The adjacent frames were combined so that the green color channel was FITC (neuron activity) and the red color channel was Texas Red (neuron nuclei). Subsequent frames were aligned using a feature-based registration toolbox with projective transformation in Matlab (MathWorks) to correct for slight body motion of the worms. Each input video frame was appended with DPMs representing propagation distances from −10 µm to 10 µm with a step size of 0.5 µm, and then tested through a Deep-Z network (specifically trained for this imaging system), which generated a virtual axial image stack for each frame in the video.

To localize individual neurons, the red channel stacks (Texas Red, neuron nuclei) were projected by median-intensity through the time sequence. Local maxima in this projected median-intensity stack marked the centroid of each neuron and the voxels of each neuron was segmented from these centroids by watershed segmentation[40], which generated a 3D spatial voxel mask for each neuron. A total of 155 neurons were isolated. Then, the average of the 100 brightest voxels in the green channel (FITC, neuron activity) inside each neuron spatial mask was calculated as the calcium activity intensity $F_i(t)$, for each time frame t and each neuron $i = 1, 2, \ldots, 155$. The differential activity was then calculated, $\Delta F(t) = F(t) - F_0$, for each neuron, where F_0 is the time average of $F(t)$.

By thresholding on the s.d. of each $\Delta F(t)$, we selected the 70 most active cells and performed further clustering on them on the basis of similarities in their calcium activity pattern (Supplementary Figure 15b) using a spectral-clustering algorithm[59,60]. The calcium activity pattern similarity was defined as

$$\mathbf{S}_{ij} = \exp\left(-\frac{\left\|\frac{\Delta F_i(t)}{F_{i0}} - \frac{\Delta F_j(t)^2}{F_{j0}}\right\|}{\sigma^2}\right) \quad (14)$$

for neurons i and j, which results in a similarity matrix $\boldsymbol{S}$ (Supplementary Figure 15c). The s.d. of this Gaussian similarity function is $\sigma = 1.5$, which controls the width of the neighbors in the similarity graph. The spectral clustering solves an eigenvalue problem on the graph Laplacian L generated from the similarity matrix $\boldsymbol{S}$, defined as the difference of weight matrix $\boldsymbol{W}$ and degree matrix $\boldsymbol{D}$

$$\mathbf{L} = \boldsymbol{D} - \boldsymbol{W} \tag{15}$$

where

$$\boldsymbol{W}_{ij} = \begin{cases} \boldsymbol{S}_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{16}$$

$$\boldsymbol{D}_{ij} = \begin{cases} \sum_j \boldsymbol{W}_{ij} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{17}$$

The number of clusters was chosen using eigengap heuristics[59], which was the index of the largest general eigenvalue (by solving the general eigenvalue problem $Lv = \lambda Dv$) before the eigengap, where the eigenvalues jump up, which was determined to be $k = 3$ (Supplementary Figure 15d). Then the corresponding first $k = 3$ eigenvectors were combined as a matrix, whose rows were clustered using standard k-means clustering[59], which resulted in the three clusters of calcium activity patterns shown in Supplementary Figure 15e and the rearranged similarity matrix shown in Supplementary Figure 15f.

Cross-modality alignment of wide-field and confocal fluorescence images. After its training, the Deep-Z+ network learns to digitally refocus a single input fluorescence image acquired by a fluorescence microscope to a user-defined target surface in 3D, but the output will match an image of the same sample captured by a different fluorescence imaging modality at the corresponding height (plane). To demonstrate this capability, we trained a Deep-Z+ network using pairs of wide-field microscopy images (inputs) and confocal microscopy images at the corresponding planes, where each stack of the wide field–confocal pair was first self-aligned and normalized using the method described above. Then the individual FOVs were stitched together using the 'Image Stitching' plugin in ImageJ (https://imagej.net/Image_Stitching). The stitched wide-field and confocal EDF images were then simultaneously registered using a feature-based registration with projective transformation performed in Matlab (MathWorks; https://www.mathworks.com/help/vision/ref/detectsurffeatures.html). Then the stitched confocal EDF images, as well as the stitched stacks, were warped using this estimated transformation to match their wide-field counterparts. The non-overlapping regions of the wide-field and warped confocal images were subsequently deleted. Then the greedy algorithm described above was used to crop non-empty regions of 256×256 pixels from the remaining stitched wide-field images and their corresponding warped confocal images. The same feature-based registration was applied on each pair of cropped regions for fine alignment. This step provides good correspondence between the wide-field image and the corresponding confocal image in the lateral directions (Supplementary Note 10).

Although the axial scanning step size was fixed to be 0.2 µm, the reference zero-point in the axial direction for the wide-field and the confocal stacks needed to be matched. To determine this reference zero-point in the axial direction, the images at each depth were compared with the EDF image of the same region using the SSIM[57], providing a focus curve. A second-order polynomial fit was performed on four points in this focus curve with highest SSIM values, and the reference zero-point was determined to be the peak of the fit. The heights of wide-field and confocal stacks were then centered by their corresponding reference zero-points in the axial direction. For each wide-field image used as input, four confocal images were randomly selected from the stack as the target, and their DPMs were calculated on the basis of the axial difference of the centered height values of the confocal and the corresponding wide-field images (Supplementary Note 10).

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 its supplementary information.

Code availability

Deep learning models reported in this work used standard libraries and scripts that are publicly available in TensorFlow. Through a custom-written Fiji-based plugin, we provided our trained network models (together with some sample test images) for the following objective lenses: Leica HC PL APO 20×/0.80 NA DRY (two different network models trained on TxRd and FITC channels), Leica HC PL APO 40×/1.30 NA oil (trained on TxRd channel), Olympus UPLSAPO20X 0.75 NA (trained on TxRd channel). We made this custom-written plugin and our models publicly available through the following links: http://bit.ly/deep-z-git and http://bit.ly/deep-z.

References

48. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
49. Strange, K. (Ed.) *C. elegans: Methods and Applications* (Humana Press, 2006).
50. Thevenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
51. Forster, B., Van de Ville, D., Berent, J., Sage, D. & Unser, 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).
52. Zack, G. W., Rogers, W. E. & Latt, S. A. Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **25**, 741–753 (1977).
53. Mao, X. et al. Least squares generative adversarial networks. In *Proc. 2017 IEEE International Conference on Computer Vision* 2813–2821 (IEEE, 2017).
54. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention 2015* (eds Navab, N. et al.) 234–241 (Springer, 2015).
55. Wu, Y. et al. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* **5**, 704–710 (2018).
56. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proc. Thirteenth International Conference on Artificial Intelligence and Statistics* 249–256 (2010).
57. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *Proc. 12th USENIX Symposium on Operating Systems Design and Implementation* 265–283 (USENIX, 2016).
58. 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).
59. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000).
60. von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007).

Acknowledgements

The authors acknowledge Y. Luo, X. Tong, T. Liu, H. C. Koydemir and Z. S. Ballard of University of California, Los Angeles (UCLA), as well as Leica Microsystems for their help with some of the experiments. The Ozcan group at UCLA acknowledges the support of Koc Group, National Science Foundation and the Howard Hughes Medical Institute. Y.W. also acknowledges the support of a SPIE John Kiel scholarship. Some of the reported optical microscopy experiments were performed at the Advanced Light Microscopy/Spectroscopy Laboratory and the Leica Microsystems Center of Excellence at the California NanoSystems Institute at UCLA with funding support from National Institutes of Health Shared Instrumentation grant S10OD025017 and National Science Foundation Major Research Instrumentation grant CHE-0722519. We also thank Double Helix Optics for providing their SPINDLE system and DH-PSF phase mask, which was used for engineered PSF data capture, and acknowledge X. Yang and M.P. Lake for their assistance with these engineered PSF experiments and related analysis.

Author contributions

A.O., Y.W. and Y.R. initiated the research. Y.W., Y.R. and H.W. performed the experiments. E.B.-D. cultured and prepared *C. elegans* samples. Y.W. and Y.L. processed the data. L.A.B. and C.P. helped with the experiments and analysis. A.O., Y.W. and Y.R. prepared the manuscript. A.O. supervised the research.

Competing interests

A.O., Y.W. and Y.R. have a pending patent application on the presented framework.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0622-5.

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

Peer review information Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Aydogan Ozcan

Last updated by author(s): Sep 21, 2019

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.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒	A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Deep-Z virtual propagation network was trained on wide-field fluorescence images collected from two standard fluorescence microscopes: IX83 (Olympus Corporation) controlled using MetaMorph® microscope automation software (Molecular Devices, LLC), and TCS SP8 controlled using LAS X software (Leica Microsystems) without modifications. Deep-Z+ virtual cross-modality propagation network was trained on wide-field and confocal fluorescence images of BPAEC collected from an inverted scanning microscope TCS SP5 controlled using LAS X software (Leica Microsystems) without modifications.
Data analysis	Image registrations were all performed using Fiji (ImageJ) plugins "StackReg", "Extended Depth of Field" and "Image Stitching", as well as Matlab R2018a feature-based registration toolbox. Deep learning models reported in this work used standard libraries and scripts that are publicly available in TensorFlow v1.10 (Google Inc.) interfaced with Python v3.6. Through a custom-written Fiji based plugin, we provided our trained network models (together with sample test images) for the following objective lenses: Leica HC PL APO 20x/0.80 DRY (two different network models trained on TxRd and FITC channels), Leica HC PL APO 40x/1.30 OIL (trained on TxRd channel), Olympus UPLSAPO20X - 0.75 NA (trained on TxRd channel). We made this custom-written plugin and our models publicly available through the following links: http://bit.ly/deep-z-git http://bit.ly/deep-z

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. 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.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Four Deep-Z virtual propagation networks with same architecture were trained on 86,000, 110,000, 105,000 and 77,000 pairs of images for four different objective lenses, where each image had 256×256 pixels. One Deep-Z+ cross-modality propagation network with same architecture was trained on 40,000 pairs of wide-field and confocal images, each with 256×256 pixels. Four additional Deep-Z virtual propagation networks with same architecture were trained using transfer learning on 25,000, 35,000, 18,000 and 110,000 pairs of images, each image with 256×256 pixels. A validation dataset (separate and independent from the training and testing datasets) of ~ 20% of the size of the training dataset is used in each training and transfer learning phase; the best network (blindly tested) is chosen to be the one with the smallest MAE loss on the validation dataset.
Data exclusions	During the training of the Deep-Z network for C. elegans, the regions of the image stacks that contained a moving worm or a dead worm were manually identified and removed from the training and validation data sets.
Replication	Deep-Z networks were first successfully tested on a total of 120 images, each with 1536×1536 pixels, where each image was virtually refocused to 20 to 40 planes at different depths within the training axial range, including the following: -- Tested with 60 images containing 300 nm red fluorescence beads captured under a microscope with a 20× objective and Texas Red filter set, achieving similar results as in Figure 1. -- Tested with 20 images containing C. elegans worms captured under a microscope with a 20× objective and Texas Red filter set through difference experiments, achieving similar results as in Figure 2 and Supplementary Figure 2. -- Tested with 18 images containing C. elegans worms captured under a microscope with a 20× objective and FITC filter set through difference experiments, achieving similar results as in Figure 2 and Supplementary Figure 2. -- Tested with 12 images containing C. elegans worms captured under a microscope with a 40× objective and Texas Red filter set, achieving similar results as in Supplementary Figure 4. -- Tested with 5 imaging FOVs (cropped into 180 images with 240×240 pixels) containing BPAEC captured under a microscope with a 20× objective and Texas Red filter set, achieving similar results as in Supplementary Figure 5 -- Tested with 5 imaging FOVs (cropped into 180 images with 240×240 pixels) containing BPAEC captured under a microscope with a 20× objective and FITC filter set, achieving similar results as in Supplementary Figure 5 The networks were also successfully tested on another 16 time-sequences of fluorescence images, each containing 120 - 1000 frames, with 1536×1536 pixels in each frame, achieving similar results as in Supplementary Videos 3-5, 7, 8. These networks were further successfully tested on 32 images virtually mapped to 3D surfaces that were not part of the training phase (training only included refocusing to planes), achieving similar results as in Figure 4. Deep-Z+ network was successfully tested on 42 image patches, each with 512×512 pixels, where each image was virtually refocused to 7 planes at different depths within the training axial range, achieving similar results as in Figure 5.
Randomization	The training, validation and testing data sets were randomly selected.
Blinding	Performances of deep neural networks were blindly tested and reported on images that were not included in the training or validation phases of the network training.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Caenorhabditis elegans (C. elegans) worms [strain AML18 and AML32] were acquired from Caenorhabditis Genetics Center (CGC) and cultured for fluorescence structural and functional imaging of their neurons.
Wild animals	None
Field-collected samples	None
Ethics oversight	No ethical approval or guidance was required because there is no requirement for ethics in worms.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

nature methods

SUPPLEMENTARY INFORMATION

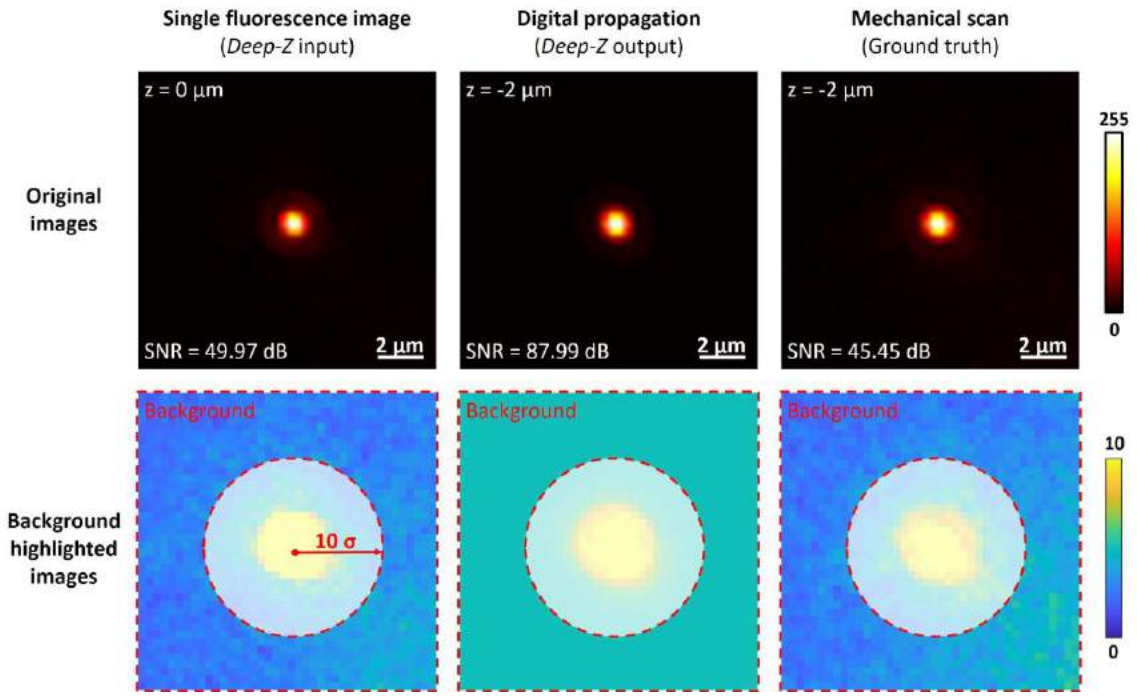
https://doi.org/10.1038/s41592-019-0622-5

In the format provided by the authors and unedited.

Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning

Yichen Wu[1,2,3,8], Yair Rivenson[1,2,3,8], Hongda Wang[1,2,3], Yilin Luo[1,2,3], Eyal Ben-David[4], Laurent A. Bentolila[3,5], Christian Pritz[6] and Aydogan Ozcan[1,2,3,7*]

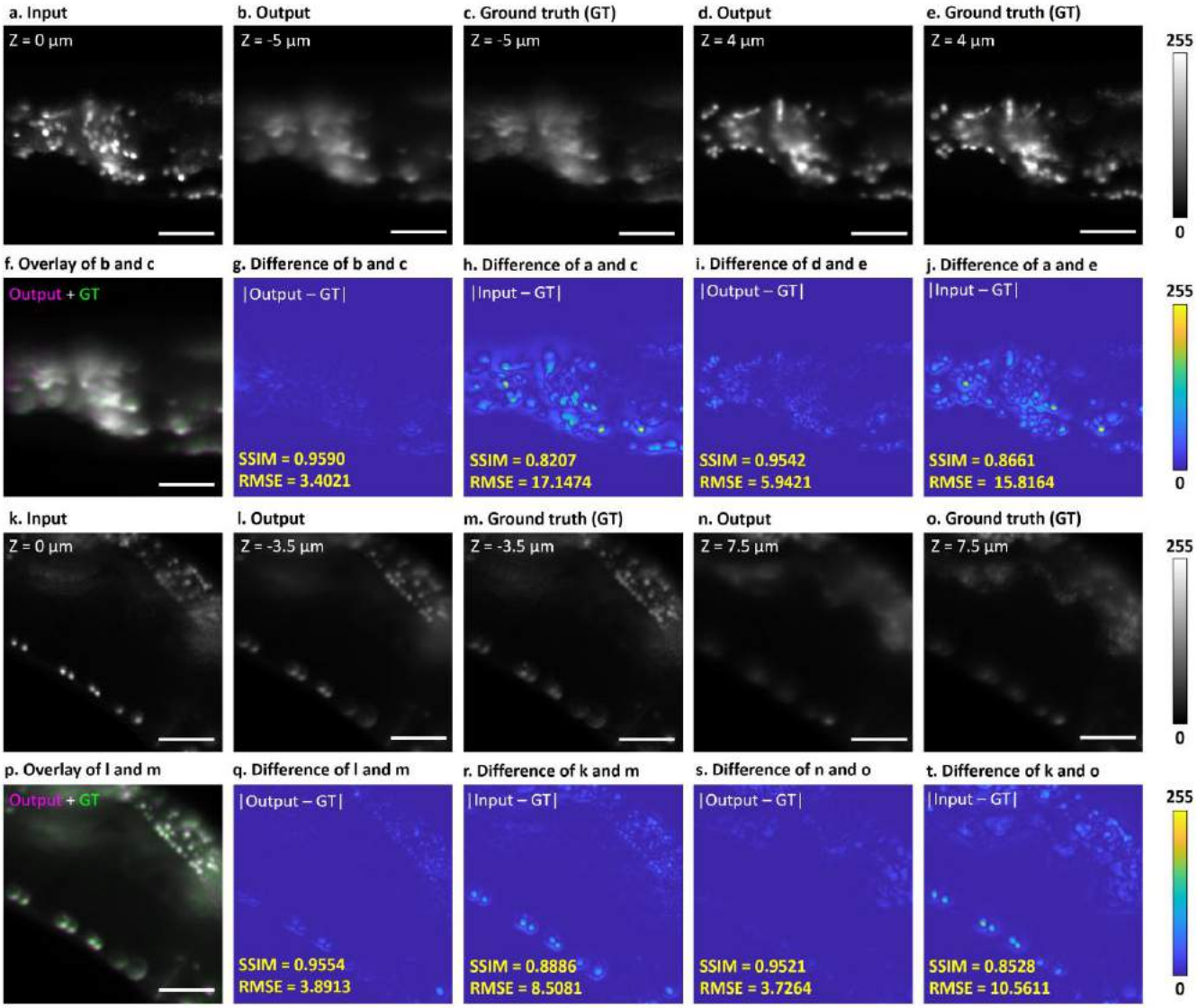
[1]Electrical and Computer Engineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [2]Bioengineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [3]California Nano Systems Institute (CNSI), University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [5]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA. [6]Department of Genetics, Hebrew University of Jerusalem, Jerusalem, Israel. [7]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [8]These authors contributed equally: Yichen Wu, Yair Rivenson. *e-mail: ozcan@ucla.edu



Supplementary Figure 1

Quantification of SNR improvement through *Deep-Z*.

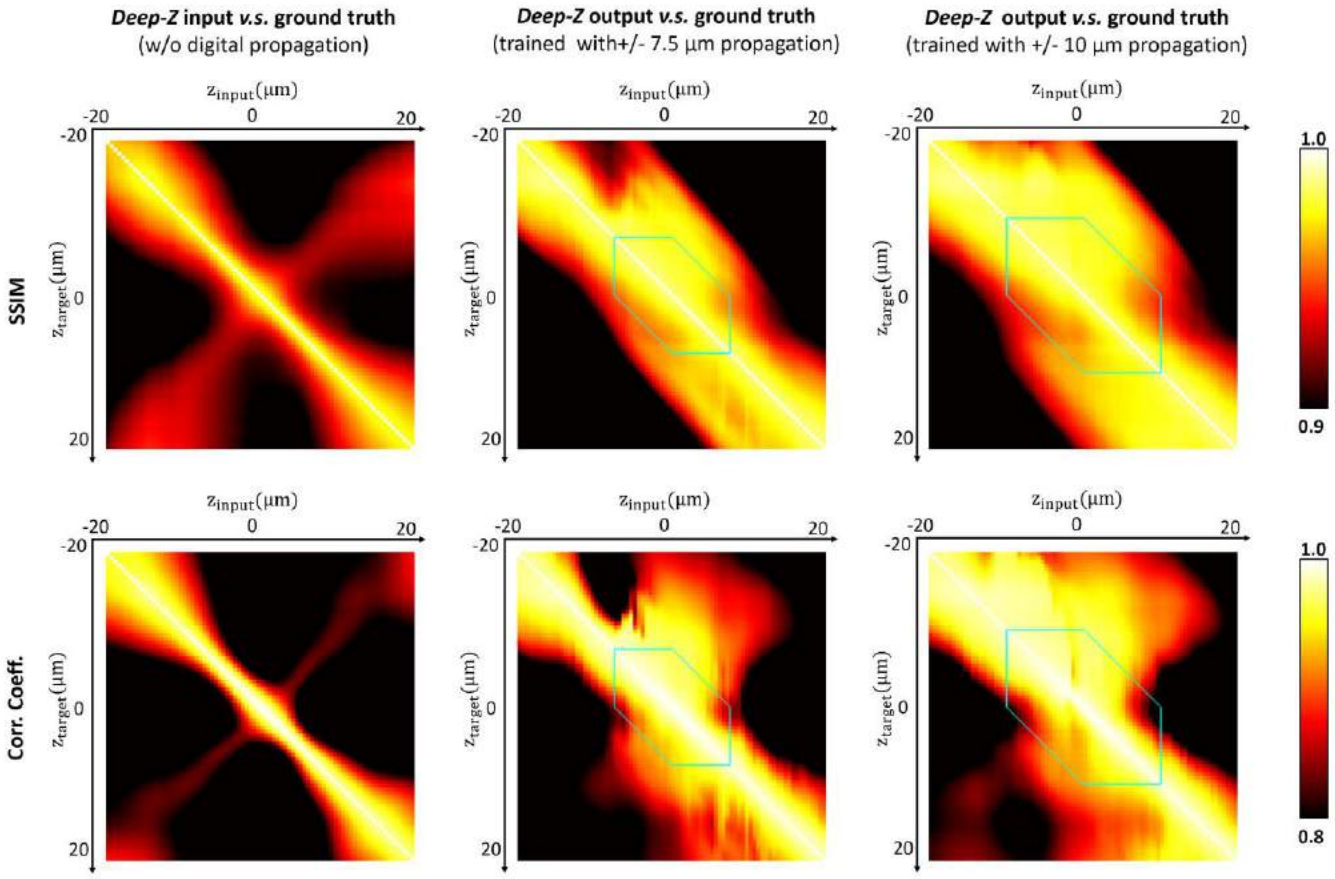
Top row: an input image of a 300 nm fluorescent bead was digitally refocused to a plane 2 μm above it using *Deep-Z*, where the ground truth was the mechanically scanned fluorescence image acquired at this plane. Bottom row: same images as the first row, but saturated to a dynamic range of [0, 10] to highlight the background. The SNR values were calculated by first taking a Gaussian fit (see the Materials and Methods section) on the pixel values of each image to find the peak signal strength. Then the pixels in the region of interest (ROI) that were 10σ away (where σ^2 is the variance of the fitted Gaussian) were regarded as the background (marked by the region outside the red dotted circle in each image) and the standard deviation of these pixel values was calculated as the background noise. The *Deep-Z* network rejects background noise and improves the output image SNR by ~ 40 dB, compared to the mechanical scan ground truth image. Analysis was performed on a randomly selected particle from a group of 96 images with similar results.



Supplementary Figure 2

Digital refocusing of fluorescence images of *C. elegans* worms.

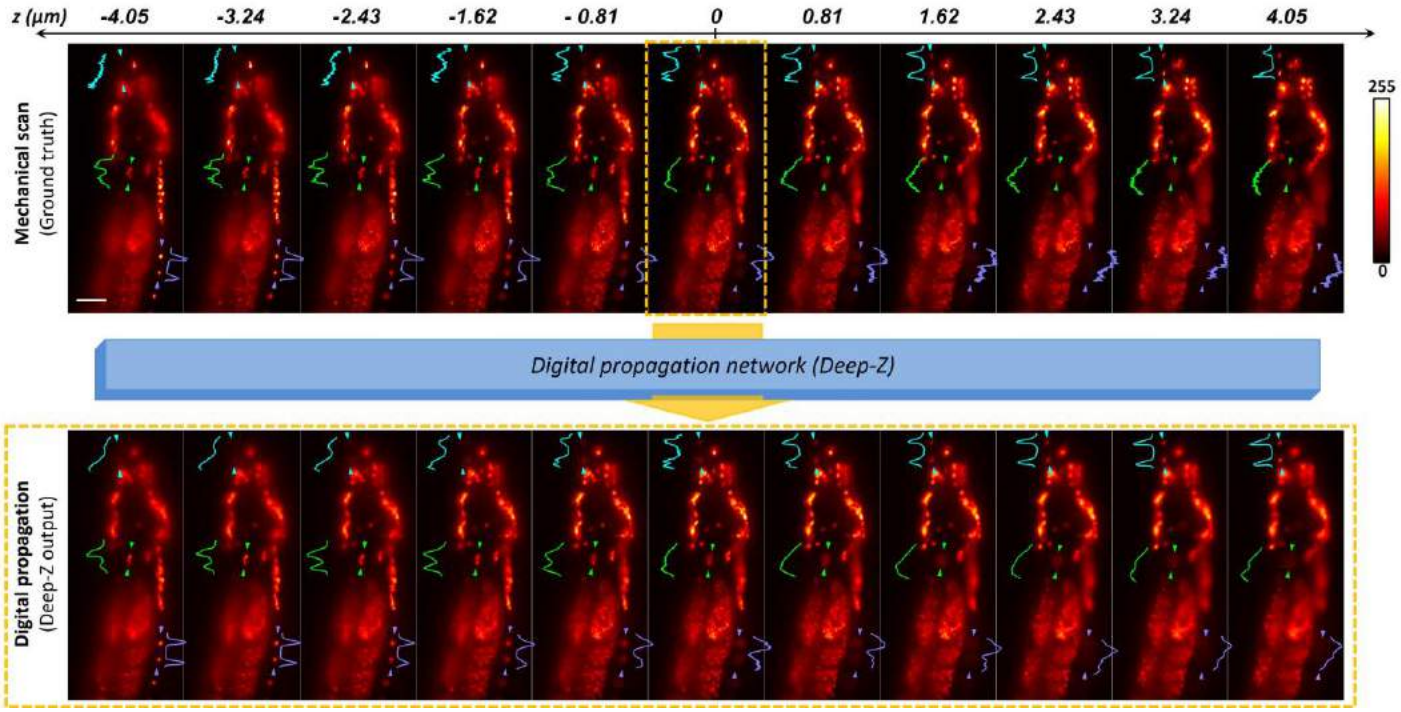
(a, k) Measured fluorescence images (*Deep-Z* input). (b, d, l, n) *Deep-Z* output at different target heights (z). (c, e, m, o) Ground truth (GT) images, captured using a mechanical axial scanning microscope at the same heights as the *Deep-Z* outputs. (f, p) overlay of *Deep-Z* output images in magenta and GT images in green. (g, i, q, s) Absolute difference images of *Deep-Z* output images and the corresponding GT images at the same heights. (h, j, r, t) Absolute difference images of *Deep-Z* input and the corresponding GT images. Structural similarity index (SSIM) and root mean square error (RMSE) were calculated for the output vs. GT and the input vs. GT for each region, displayed in (g, i, q, s) and (h, j, r, t), respectively. Scale bar: 25 µm. Experiments were repeated with 20 images, achieving similar results.



Supplementary Figure 3

Structural similarity (SSIM) index and correlation coefficient (Corr. Coeff.) analysis for digital refocusing of fluorescence images from an input plane at z_{input} to a target plane at z_{target}.

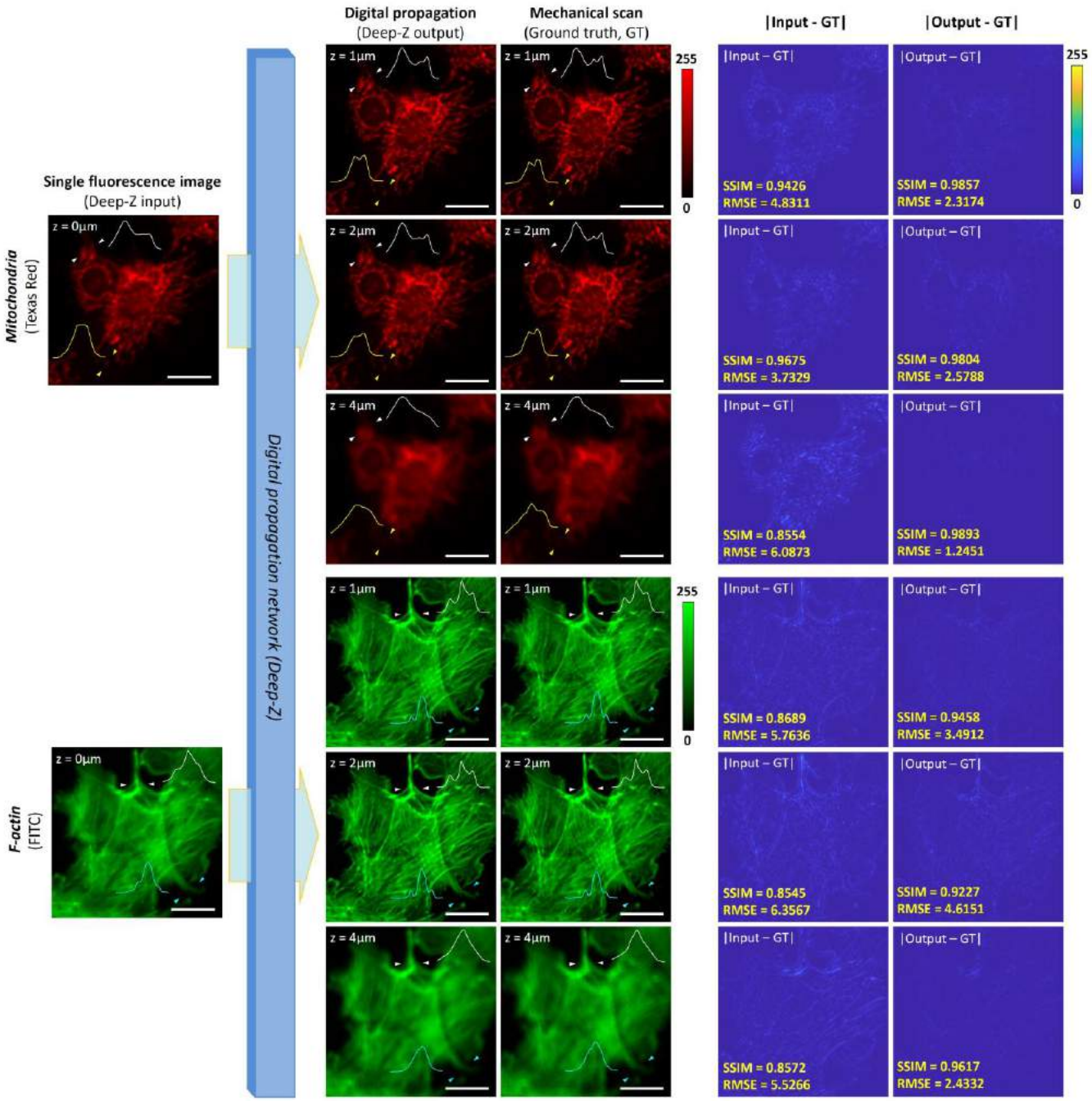
We created a scanned fluorescence z-stack of a *C. elegans* sample, within an axial range of -20 µm to 20 µm, with 1 µm spacing. *First column*: each scanned image at z_{input} in this stack was compared against the image at z_{target}, forming cross-correlated SSIM and Corr. Coeff. matrices. Both the SSIM and Corr. Coeff. fall rapidly off the diagonal entries. *Second column*: A *Deep-Z* network trained with fluorescence image data corresponding to +/- 7.5 µm propagation range (marked by the *cyan diamond* in each panel) was used to digitally refocus images from z_{input} to z_{target}. The output images were compared against the ground truth images at z_{target} using SSIM and Corr. Coeff. *Third column*: same as the second column, except the training fluorescence image data included up to +/- 10 µm axial propagation (marked by the *cyan diamond* that is now enlarged compared to the second column). These results confirm that *Deep-Z* learned the digital propagation of fluorescence, but it is limited to the axial range that it was trained for (determined by the training image dataset). Outside the training range (defined by the cyan diamonds), both the SSIM and Corr. Coeff. values considerably decrease.



Supplementary Figure 4

3D imaging of *C. elegans* head neuron nuclei using *Deep-Z*.

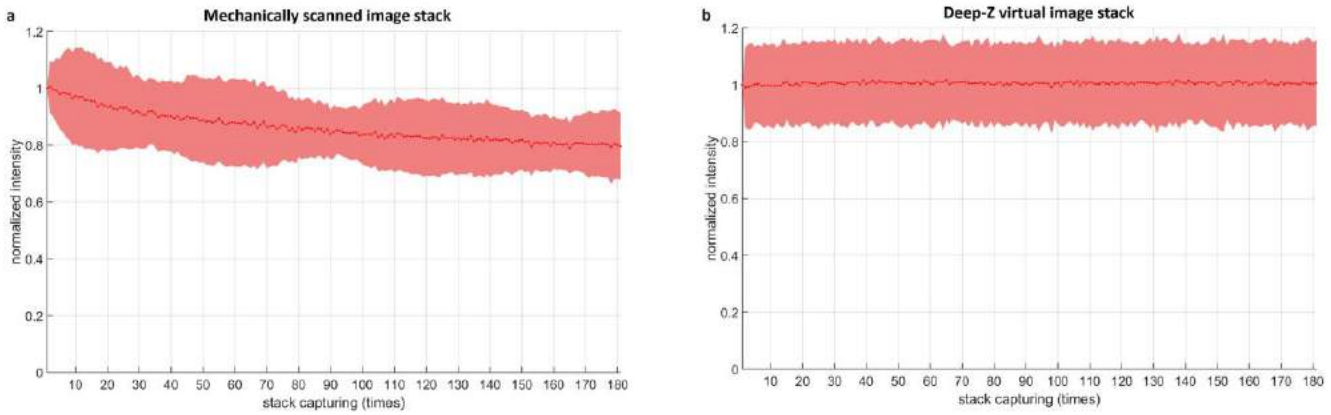
The input and ground truth images were acquired by a scanning fluorescence microscope with a 40×/1.4NA objective. A single fluorescence image acquired at z=0 μm focal plane (marked by dashed yellow rectangle) was used as the input image to *Deep-Z* and was digitally refocused to different planes within the sample volume, spanning around -4 to 4 μm; the resulting images provide a good match to the corresponding ground truth images. Scale bar: 25 μm. Experiments were repeated with 12 images, achieving similar results.



Supplementary Figure 5

Digital refocusing of fluorescence microscopy images of BPAEC using *Deep-Z*.

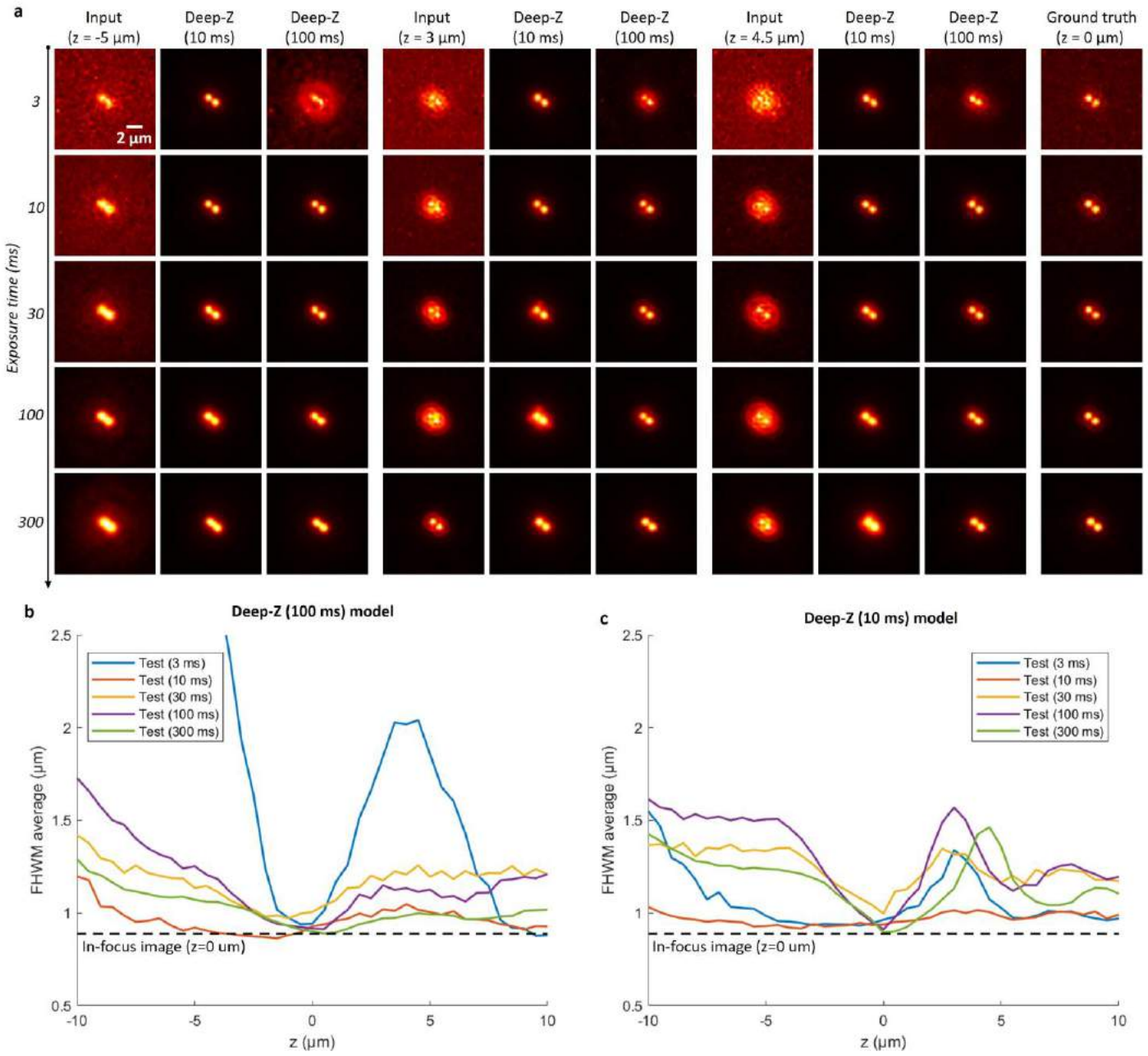
The input image was captured using a 20×/0.75 NA objective lens, using the Texas Red and FITC filter sets, occupying the red and green channels of the image, for the mitochondria and F-actin structures, respectively. Using *Deep-Z*, the input image was digitally refocused to 1 µm above the focal plane, where the mitochondrial structures in the green channel are in focus, matching the features on the mechanically-scanned image (obtained directly at this depth). The same conclusion applies for the *Deep-Z* output at z = 2 µm, where the F-actin structures in the red channel come into focus. After 3 µm above the image plane, the details of the image content get blurred. The absolute difference images of the input and output with respect to the corresponding ground truth images are also provided, with SSIM and RMSE values, quantifying the performance of *Deep-Z*. Scale bar: 20 µm. Experiments were repeated with 180 images, achieving similar results.



Supplementary Figure 6

Photobleaching analysis.

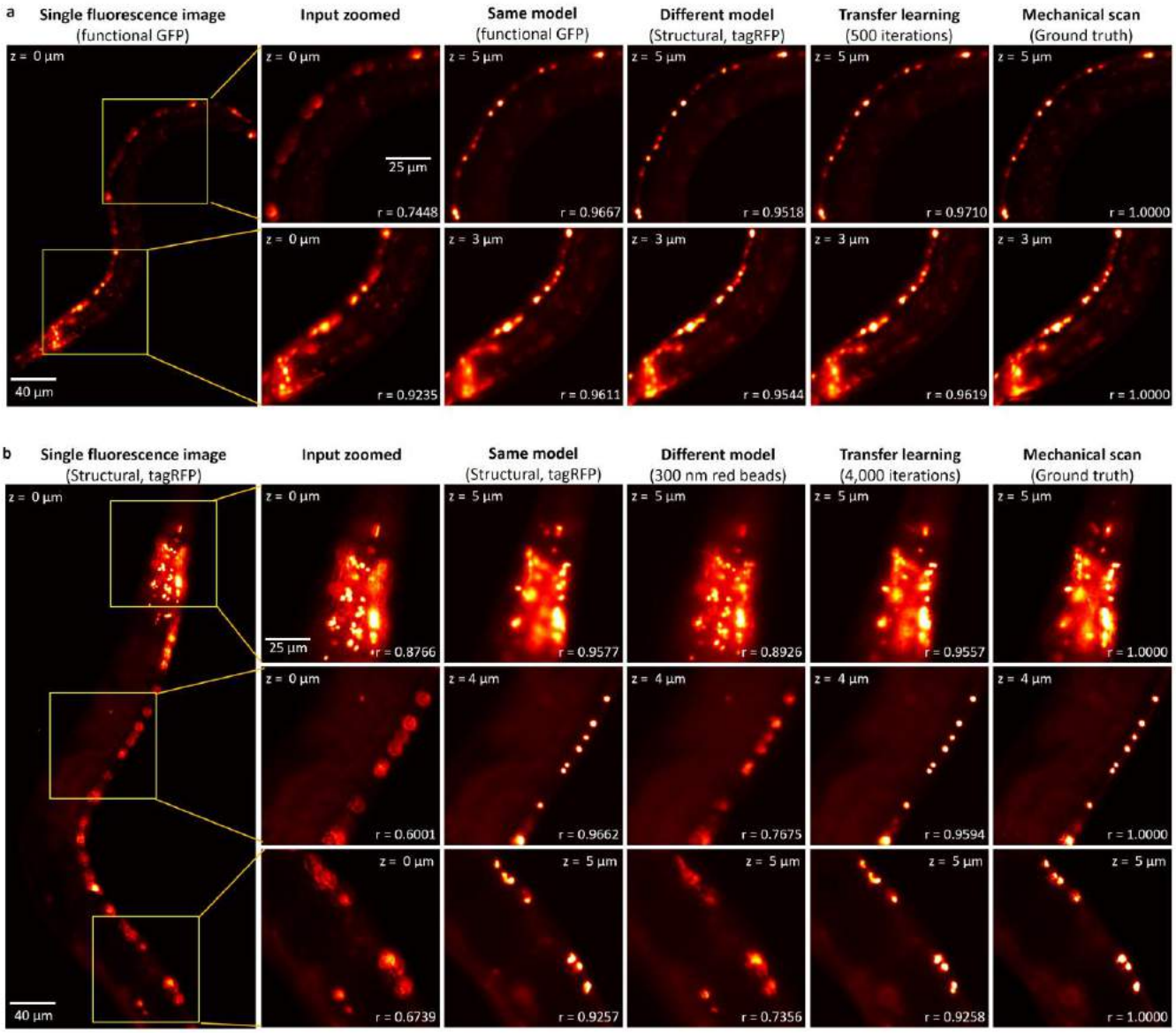
(a) Fluorescence signal of nanobeads imaged in 3D, for 180 times of repeated axial scans, containing 41 planes, spanning +/- 10 µm with a step size 0.5 µm. The accumulated scanning time is ~30 min. (b) The corresponding scan for a single plane, which is us ed by *Deep-Z* to generate a virtual image stack, spanning the same axial depth within the sample (+/- 10 µm). The accumulated scanning time for *Deep-Z* is ~ 15 seconds. The center line represents the mean and the shaded region represents the standard deviation of the normalized intensity for 681 and 597 individual nanobeads (for (a) and (b), respectively) inside the sample volume.



Supplementary Figure 7

Refocusing capability of *Deep-Z* under lower image exposure.

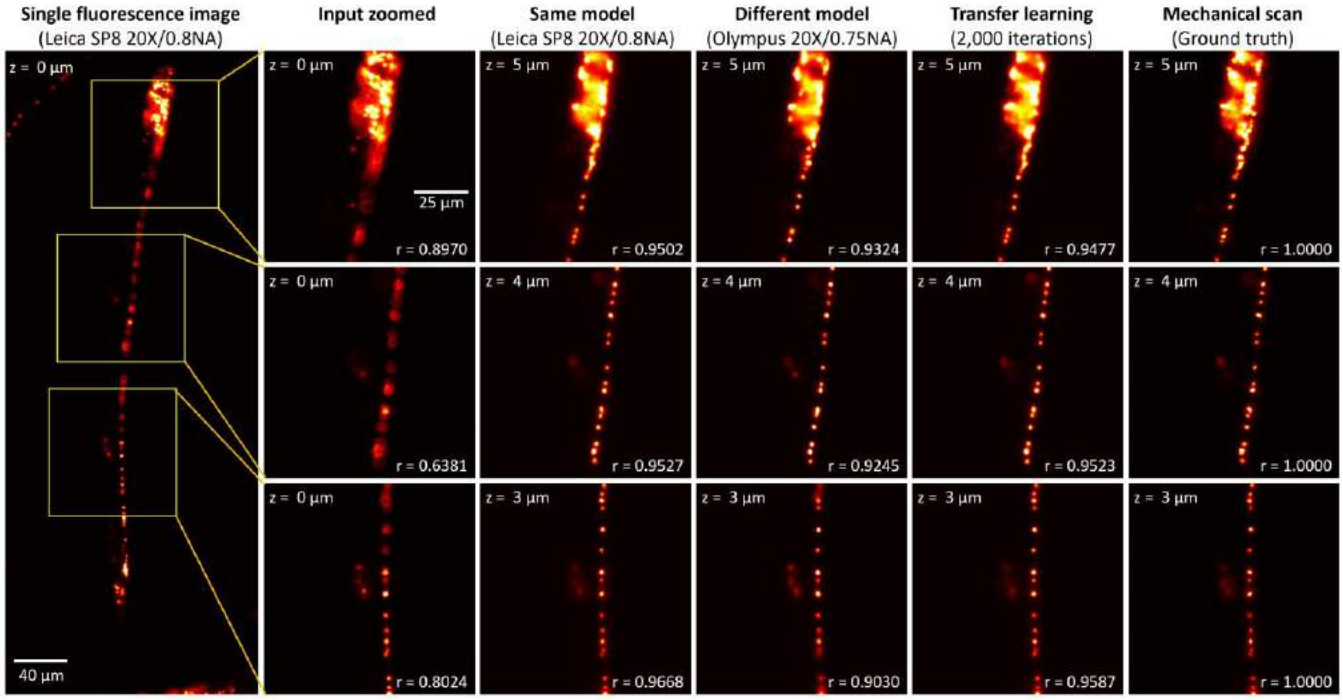
(a) Virtual refocusing of images containing two microbeads under different exposure times from defocused distances of -5, 3 and 4.5 µm, using two *Deep-Z* models trained with images captured at 10 ms and 100 ms exposure times, respectively. (b) Median FWHM values of 91 microbeads imaged inside a sample FOV after the virtual refocusing of an input image across a defocus range of -10 µm to 10 µm by the *Deep-Z* (100 ms) network model. The test images have different exposure times spanning 3 ms to 300 ms. (c) Same as (b), but plotted for the *Deep-Z* (10 ms) model.



Supplementary Figure 8

Deep-Z based virtual refocusing of a different sample type and transfer learning results.

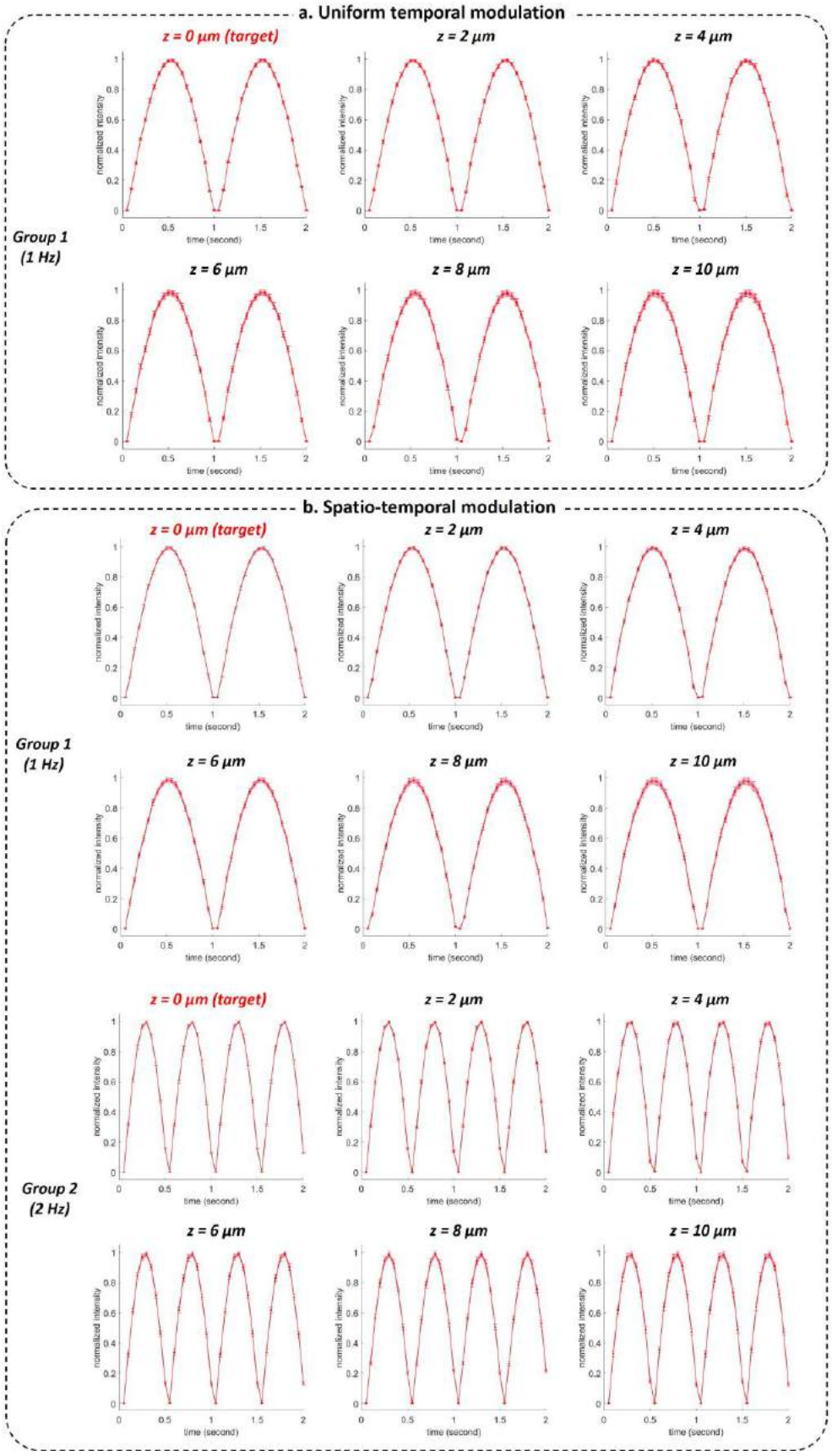
(a) The input image records the neuron activities of a *C. elegans* that is labeled with GFP; the image is captured using a 20×/0.8NA objective under the FITC channel. The input image was virtually refocused using both the optimal worm strain model (denoted as: same model, functional GFP) as well as a different model (denoted as: different model, structural tagRFP); we also report here the results of a transfer learning model which used the different model as its initialization and functional GFP image dataset to refine it after ~ 500 iterations (~30 min of training). (b) A different *C. elegans* sample is shown. The input image records the neuron nuclei labeled with tagRFP imaged using a 20×/0.75NA objective under the Texas Red channel. The input image was virtually refocused using both the exact worm strain model (same model, structural, tagRFP) as well as a different model (different model, 300 nm red beads); we also report here the results of a transfer learning model which used the different model as its initialization and structural ragR FP image dataset to refine it after ~ 4,000 iterations (~6 hours training). Image correlation coefficient (r) is shown at the lower right corner of each image, in reference to the ground truth mechanical scan performed at the corresponding microscope system (Leica and Olympus, respectively). The transfer learning was performed using 20% of the training data and 50% of the validation data, randomly selected from the original data set.



Supplementary Figure 9

Virtual refocusing of a different microscope system and transfer learning results.

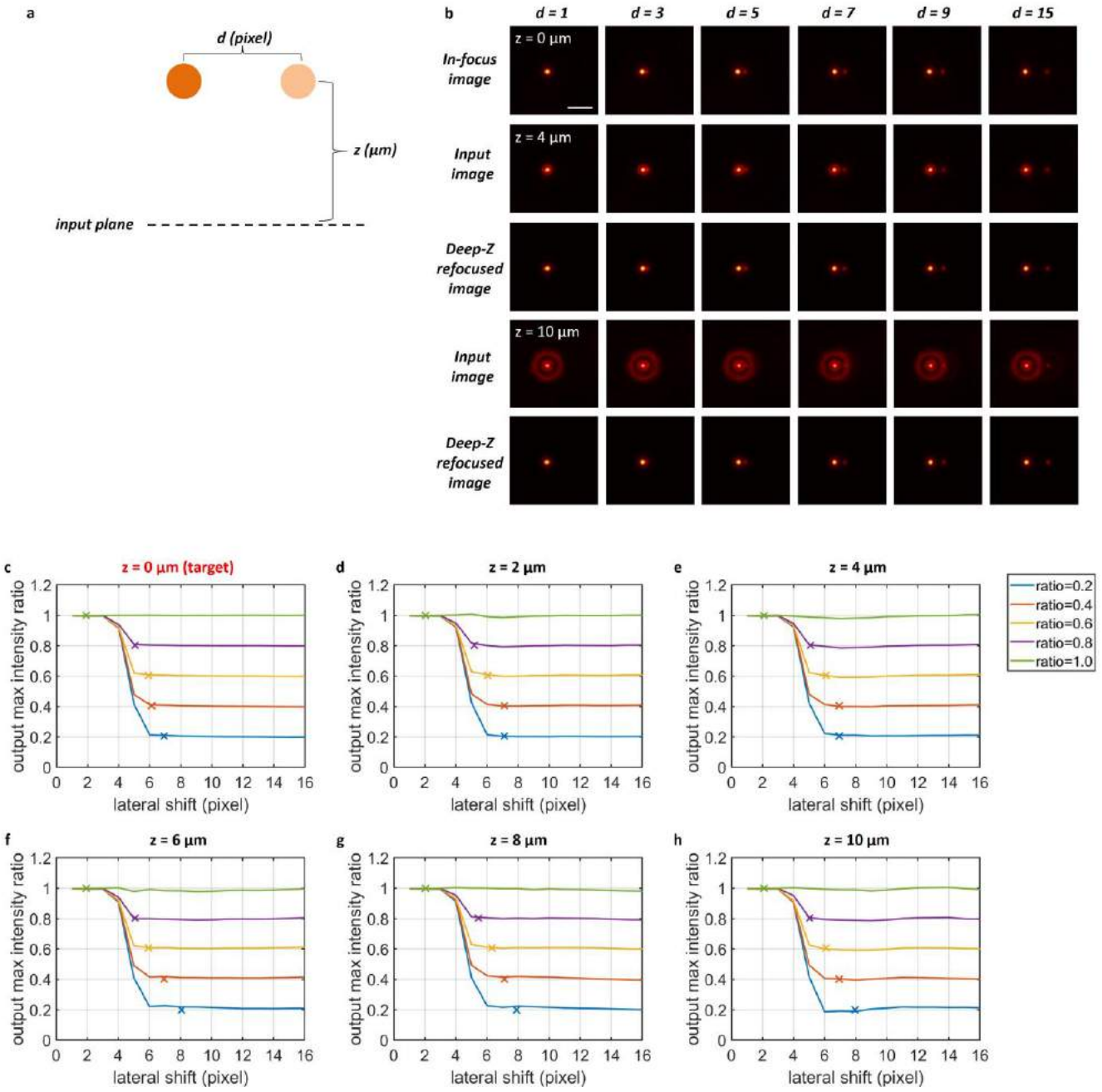
The input image records the *C. elegans* neuronal nuclei labeled with tag GFP, imaged using a Leica SP8 microscope with a 20×/0.8NA objective. The input image was virtually focused using both the exact model (Leica SP8 20×/0.8NA) as well as a different model (denoted as: different model, Olympus 20×/0.75NA); we also report here the results of a transfer learning model using the different model as its initialization and Leica SP8 image dataset to refine it after ~ 2,000 iterations (~40 min training). Image correlation coefficient (r) is shown at the lower right corner of each image, in reference to the ground truth mechanical scan performed at the corresponding microscope system. The transfer learning was performed using 20% of the training data and 50% of the validation data, randomly selected from the original data set.


a. Uniform temporal modulation
z = 0 μm (target)
z = 2 μm
z = 4 μm
Group 1
(1 Hz)
z = 6 μm
z = 8 μm
z = 10 μm
b. Spatio-temporal modulation
z = 0 μm (target)
z = 2 μm
z = 4 μm
Group 1
(1 Hz)
z = 6 μm
z = 8 μm
z = 10 μm
z = 0 μm (target)
z = 2 μm
z = 4 μm
Group 2
(2 Hz)
z = 6 μm
z = 8 μm
z = 10 μm
normalized intensity
time (second)

Supplementary Figure 10

Time-modulated signal reconstruction using *Deep-Z*.

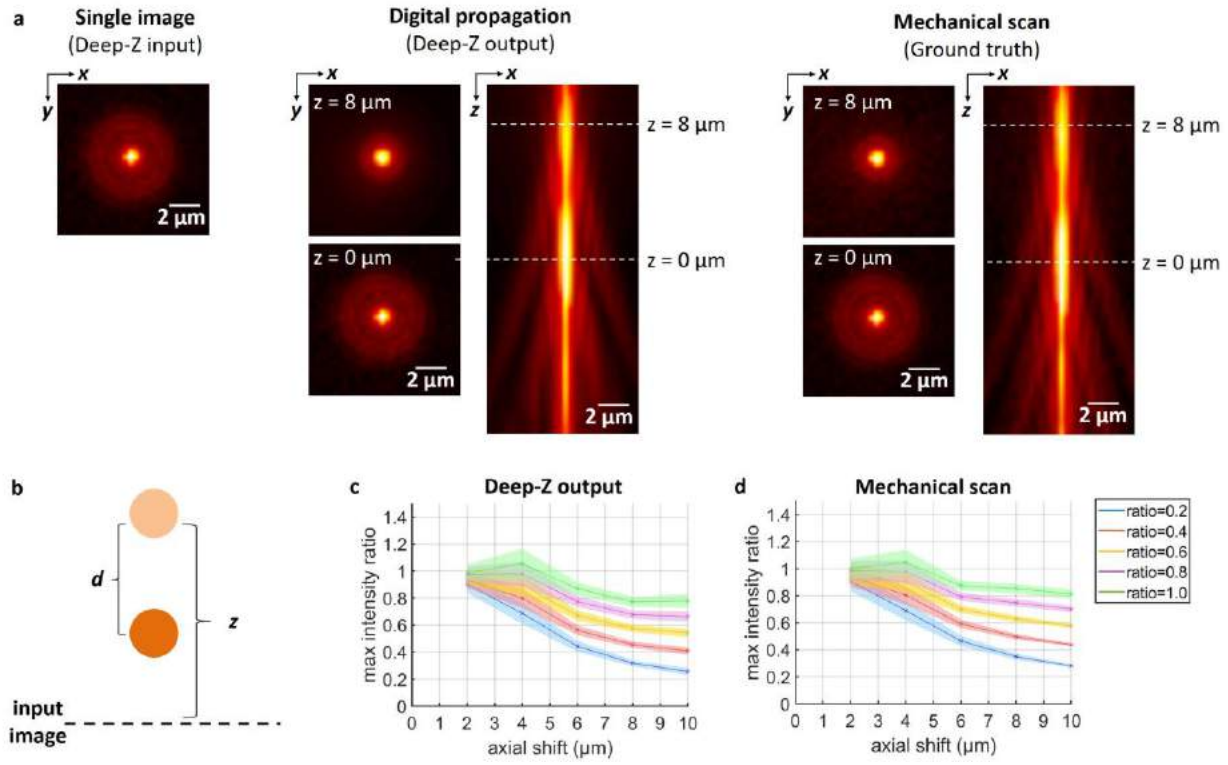
(a) A time-modulated illumination source was used to excite the fluorescence signal of microbeads (300 nm diameter). Time-lapse sequence of the sample was captured under this modulated illumination at the in-focus plane (z = 0 µm) as well as at various defocused planes (z = 2-10 µm) and refocused using *Deep-Z* to digitally reach z = 0 µm. Intensity variations of 297 individual beads inside the FOV (after refocusing) were tracked for each sequence. (b) Based on the video captured in (a), we took every other frame to form an image sequence with twice the frame-rate and modulation frequency, and added it back onto the original sequence with a lateral shift. These defocused and super-imposed images were virtually refocused using *Deep-Z* to digitally reach z=0 µm, in-focus plane. Group 1 contained 297 individual beads inside the FOV with 1 Hz modulation. Group 2 contained the signals of the other (new) beads that are super-imposed on the same FOV with 2 Hz modulation frequency. Each intensity curve was normalized, and the mean (plotted as curve center) and the standard deviation (plotted as error bar) of the 297 curves were plotted for each time-lapse sequence. Virtually-refocused *Deep-Z* output tracks the sinusoidal illumination, very closely following the in-focus reference time-modulation reported in target (z = 0 µm). Please also see the Supplementary Video 5, related to this figure.



Supplementary Figure 11

Deep-Z based virtual refocusing of a laterally shifted weaker fluorescent object next to a stronger object.

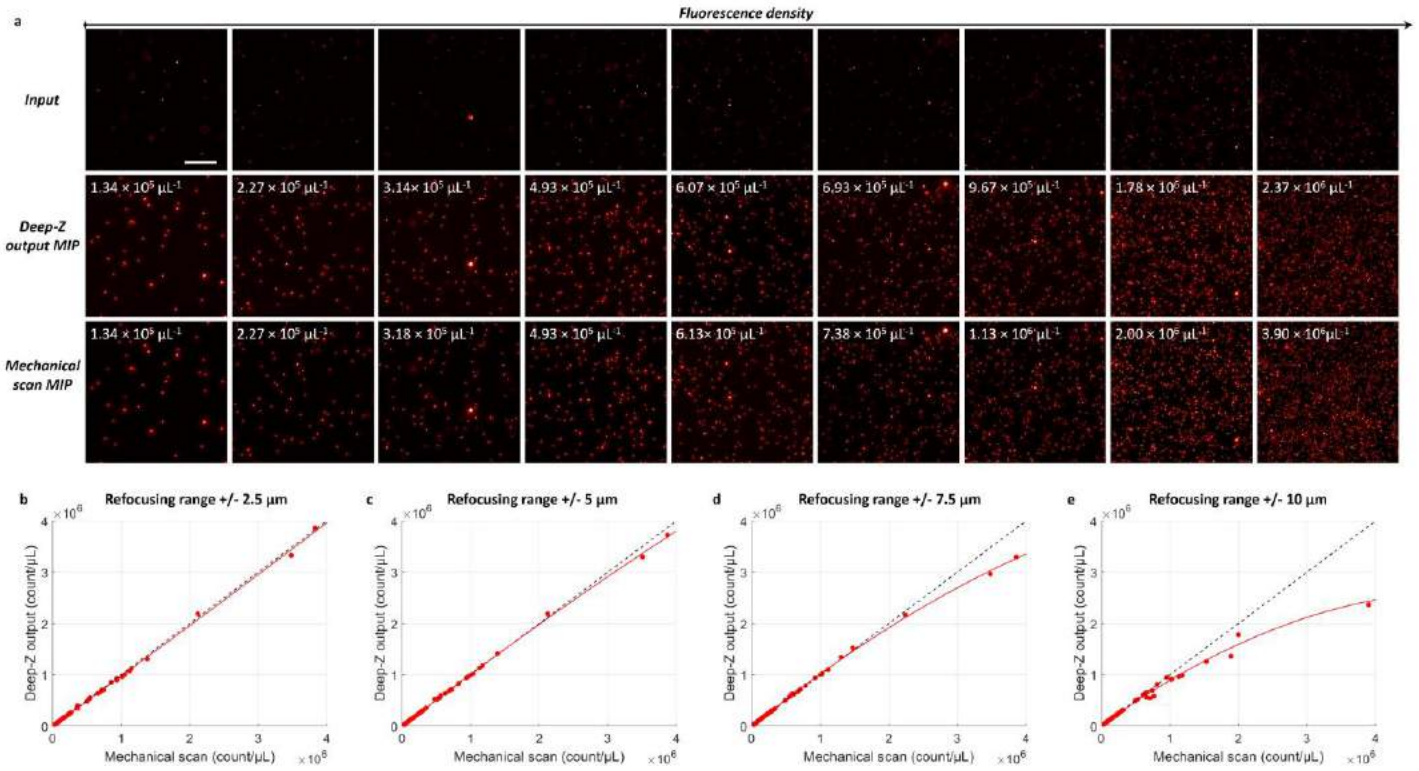
(a) A defocused experimental image (left bead) at plane z was shifted laterally by d pixels to the right and digitally weakened by a pre-determined ratio (right bead), which was then added back to the original image, used as the input image to *Deep-Z*. Scale bar: 5 μm. (b) An example of the generated bead pair with an intensity ratio of 0.2; we show in-focus plane, defocused planes of 4 and 10 μm, and the corresponding virtually-refocused images by *Deep-Z*. (c-h) Average intensity ratio of the shifted and weakened bead signal with respect to the original bead signal for 144 bead pairs inside a FOV, calculated at the virtually refocused plane using different axial defocus distances (z). The crosses "x" in each figure mark the corresponding lateral shift distance, below which the two beads cannot be distinguished from each other, color-coded to represent bead signal intensity ratio (spanning 0.2-1.0).



Supplementary Figure 12

Impact of axial occlusions on *Deep-Z* virtual refocusing performance.

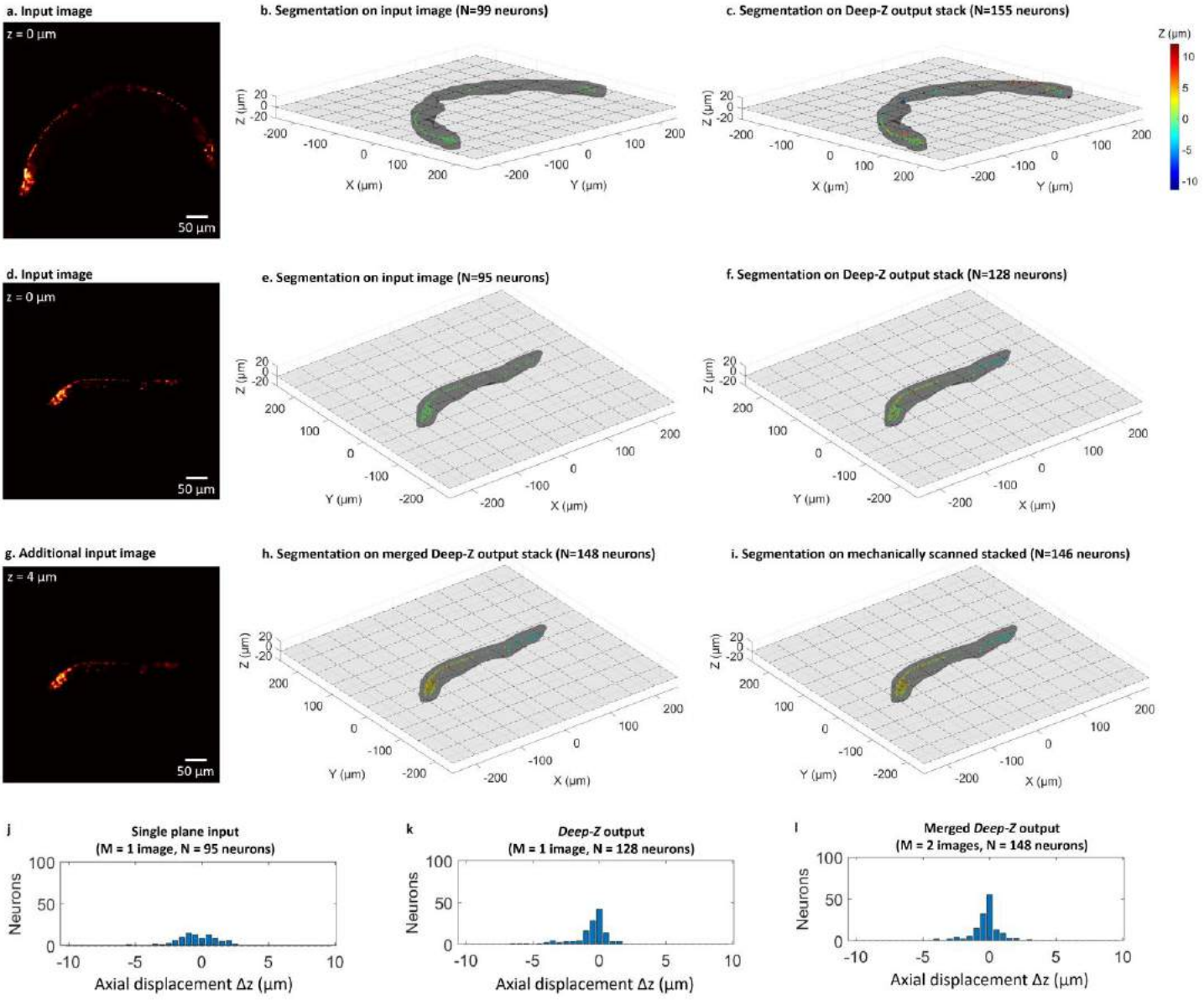
(a) 3D virtual refocusing of two beads that have identical lateral positions but are separated axially by 8 µm; *Deep-Z*, as usual, used a single 2D input image corresponding to the defocused image of the overlapping beads. The virtual refocusing calculated by *Deep-Z* exhibits two maxima representing the two beads along the z-axis, matching the simulated ground truth image stack. (b) Simulation schematic: two defocused images in the same bead image stack with a spacing of *d* was added together, with the higher stack located at a depth of z=8 µm. A single image in the merged image stack was used as the input to *Deep-Z* for virtual refocusing. (c-d) report the average and the standard deviation (represented by transparent colors) of the intensity ratio of the top (i.e., the dimmer) bead signal with respect to the bead intensity in the original stack, calculated for 144 bead pairs inside a FOV, for z = 8 µm with different axial separations (d) and bead intensity ratios (spanning 0.2-1.0).



Supplementary Figure 13

Deep-Z inference results as a function of 3D fluorescent sample density.

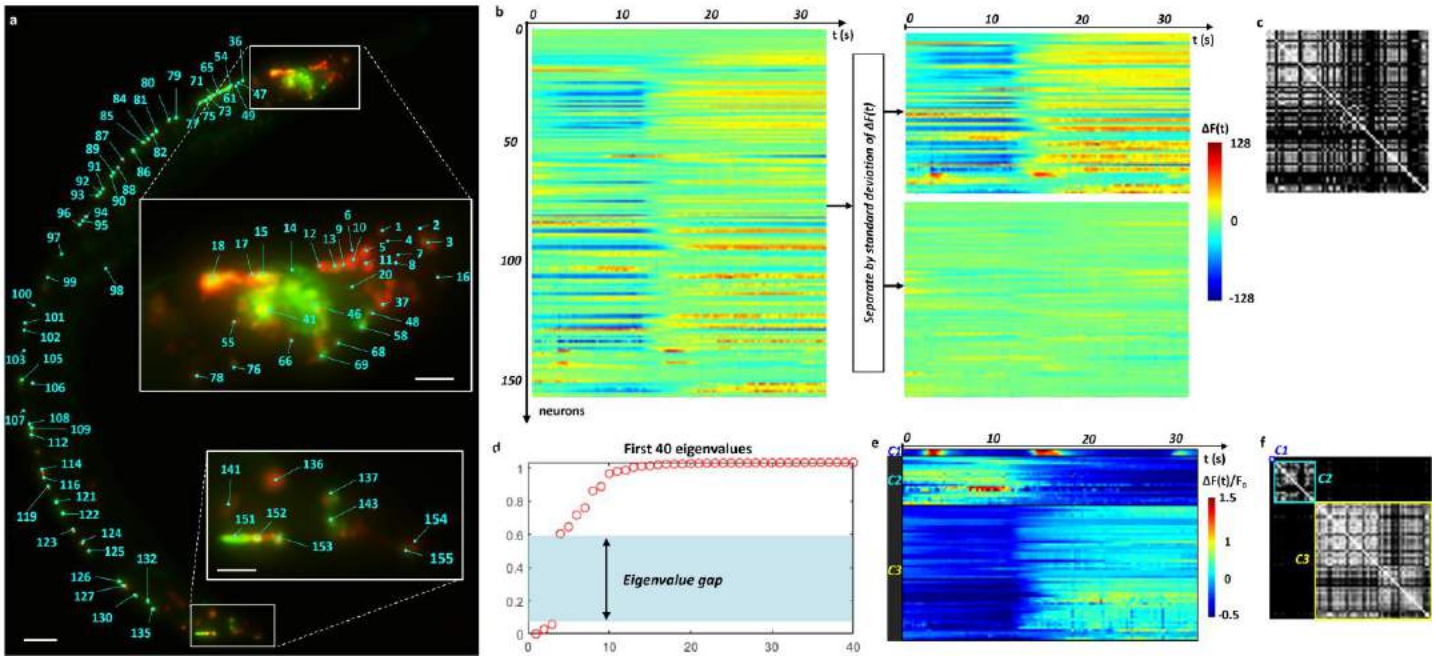
(a) Comparison of *Deep-Z* inference against a mechanically-scanned ground truth image stack over an axial depth of +/- 10 µm with increasing fluorescent bead concentration. The measured bead concentration resulting from the *Deep-Z* output (using a single input image) as well as the mechanically-scanned ground truth (which includes 41 axial images acquired at a scanning step size of 0.5 µm) is shown on the top left corner of each image. MIP: maximal intensity projection along the axial direction. Scale bar: 30 µm. (b -e) Comparison of *Deep-Z* output against the ground truth results as a function of the increasing bead concentration. The red line is a 2nd order polynomial fit to all the data points. The black dotted line represents $y=x$, shown for reference. These particle concentrations were calculated/measured over a FOV of 1536×1536 pixels (500×500 µm^2), i.e. 15-times larger than the specific regions shown in (a).



Supplementary Figure 14

C. elegans neuron segmentation comparison.

(a, d) The fluorescence image used as input to *Deep-Z*. (b, e) Segmentation results based on (a, d), respectively. (c, f) Segmentation results based on the virtual image stack (-10 to 10 µm) generated by *Deep-Z* using the input images in (a, d), respectively. (g) An additional fluorescence image, captured at a different axial plane (z = 4 µm). (h) Segmentation results on the merged virtual stack (-10 to 10 µm). The merged image stack was generated by blending the two virtual stacks generated by *Deep-Z* using the input images (d) and (g). (i) Segmentation results based on the mechanically-scanned image stack used as ground truth (acquired at 41 depths with 0.5 µm axial spacing). Each neuron was represented by a small sphere in the segmentation map and the depth information of each neuron was color-coded. (j-l) The detected neuron positions in (e, f, h) were compared with the positions in (i) (see the Supplementary Methods for details), and the axial displacement histograms between the *Deep-Z* results and the mechanically-scanned ground truth results were plotted.



Supplementary Figure 15

C. elegans neuron activity tracking and clustering.

(a) Max intensity projection (MIP) along the axial direction of the median intensity image over time. The red channel (Texas red) labels neuron nuclei and the green channel (FITC) labels neuron calcium activity. A total of 155 neurons were identified in the 3D stack, as labeled here. Scale bar: 25 µm. Scale bar for the zoom-in regions: 10 µm. (b) The intensity of the neuron calcium activity, $\Delta F(t)$, of these 155 neurons is reported over a period of ~ 35 s at ~3.6Hz. Based on a threshold on the standard deviation of each $\Delta F(t)$, we separate neurons to be active (right-top, 70 neurons) and less active (right-bottom, 85 neurons). (c) The similarity matrix of the calcium activity patterns of the top 70 active neurons. (d) The top 40 eigen values of the similarity matrix. An eigen-gap is shown at k=3, which was chosen as the number of clusters according to eigen-gap heuristic (i.e. choose up to the largest eigenvalue before the eigenvalue gap, where the eigenvalues increase significantly). (e) Normalized activity $\Delta F(t)/ F_0$ for the k=3 clusters after the spectral clustering on the 70 active neurons. (f) Similarity matrix after spectral clustering. The spectral clustering rearranged the row and column ordering of the similarity matrix (c) to be block diagonal in (f), which represents three individual clusters of calcium activity patterns.

Supplementary Notes:

Supplementary Note 1: Sample drift-induced defocus compensation using *Deep-Z*

Deep-Z enables the correction for sample drift induced defocus after the image is captured. To demonstrate this, Supplementary Video 4 shows a moving *C. elegans* worm recorded by a wide-field fluorescence microscope with a 20×/0.8NA objective lens (DOF ~ 1 μm). The worm was defocused ~ 2 – 10 μm from the recording plane. Using *Deep-Z*, we can digitally refocus each frame of the input video to different planes up to 10 μm, correcting this sample drift induced defocus (Supplementary Video 4). Such a sample drift is conventionally compensated by actively monitoring the image focus and correcting for it during the measurement, e.g., by using an additional microscope[1]. *Deep-Z*, on the other hand, provides the possibility to compensate sample drift in already-captured 2D fluorescence images.

Supplementary Note 2: Reduced photodamage using *Deep-Z*

Another advantage of *Deep-Z* would be a reduction in photodamage to the sample. Photodamage introduces a challenging tradeoff in applications of fluorescence microscopy in live cell imaging, which sets a practical limitation on the number of images that can be acquired during e.g., a longitudinal experiment. The specific nature of photodamage, in the form of photobleaching and/or phototoxicity, depends on the illumination wavelength, beam profile, exposure time, among many other factors, such as the sample pH and oxygen levels, temperature, fluorophore density and photostability[2–4]. Several strategies for illumination design have been demonstrated to reduce the effects of photodamage, by e.g., adapting the illumination intensity delivered to the specimen as in controlled light exposure microscopy (CLEM)[5] and predictive focus illumination[3], or decoupling the excitation and emission paths, as in selective plane illumination microscopy[6] and among others.

For a widefield fluorescence microscopy experiment, where an axial image stack is acquired, the illumination excites the fluorophores through the entire thickness of the specimen, regardless of the position that is imaged in the objective's focal plane. For example, if one assumes that the sample thickness is relatively small compared to the focal volume of the excitation beam, the entire sample volume is uniformly exited at each axial image acquisition step. This means the total light exposure of a given point within the sample volume is sub-linearly proportional to the number of imaging planes (N_z) that are acquired during a single-pass z-stack. In contrast, *Deep-Z* only requires a single image acquisition step if its axial training range covers the sample depth; in case the sample is thicker or dense, more than one input image might be required for improved *Deep-Z* inference as demonstrated in Supplementary Figure 14h, which, in this case, used two input images to better resolve neuron nuclei in the head region of a *C. elegans*. Therefore, this

reduction, enabled by *Deep-Z*, in the number of axial planes that need to be imaged within a sample volume directly helps to reduce the photodamage to the sample.

To further illustrate this advantage, we performed an additional experiment where we repeatedly imaged in 3D a sample containing fluorescent beads (300 nm diameter, and embedded in ProLong Gold antifade mountant) with N_z=41 axial planes spanning 20 μm depth range (0.5 μm step size) over 180 repeated cycles, which took a total of ~ 30 min. The average fluorescence signal of the nanobeads decayed down to ~80% of its original value at the end of the imaging cycle (see Supplementary Figure 6a). In comparison, to generate a similar virtual image stack, *Deep-Z* only requires to take a single input image, which results in a total imaging time of ~ 15 seconds for 180 repeated cycles, and the average fluorescence signal in the *Deep-Z* generated virtual image stack does not show a visible decay during the same number of imaging cycles (see Supplementary Figure 6b). For imaging of live samples, potentially without a dedicated antifade mountant, the fluorescence signal decay would be more drastic compared to Supplementary Figure 6a due to photodamage and photobleaching, and *Deep-Z* can be used to significantly reduce these negative effects, especially during longitudinal imaging experiments.

Although not demonstrated in this work, the application of *Deep-Z* concept to light sheet microscopy can also be used to reduce the number of imaging planes within the sample, by increasing the axial separation between two successive light sheets using *Deep-Z* 3D inference in between.

In general a reduction in N_z further helps us to reduce photodamage effect if we also take into account hardware-software synchronization times[4] that are required during the axial scan, which introduces additional time overhead if e.g., an arc burner is used as the illumination source; this

illumination overhead can be mostly eliminated when using LEDs for illumination, which have much faster on-off transition times.

In summary, *Deep-Z* has the potential to substantially circumvent the standard photodamage tradeoffs in fluorescence microscopy and enable imaging at higher speeds and/or improved SNR since the illumination intensity can be increased for a given photodamage threshold that is set, offset by the reduced number of axial images that are acquired through the use of *Deep-Z*.

Supplementary Note 3: *Deep-Z* virtual refocusing capability at lower image exposure

To further validate the generalization performance of a pre-trained *Deep-Z* network model under variable exposure conditions (which directly affect the signal-to-noise ratio, SNR), we trained two *Deep-Z* networks using microbead images captured at 10 ms and 100 ms exposure times; we denoted these trained networks as *Deep-Z* (10 ms) and *Deep-Z* (100 ms), respectively, and blindly tested their performance to virtually refocus defocused images captured under different exposure times, varying between 3 ms to 300 ms. Training image data were captured using 300 nm red fluorescent bead samples imaged with a 20×/0.75NA objective lens, same as the micro-bead samples reported in the main text, except that the fluorescence excitation light source was set at 25% power (32.5 mW) and the exposure times were chosen as 10 ms and 100 ms, respectively. Two separate *Deep-Z* networks were trained using the image dataset captured at 10 ms and 100 ms exposure times, where each training image set contained ~ 100,000 image pairs (input and ground truth), and each network was trained for ~ 50 epochs. Testing image data were captured under the same settings except the exposure times varied from 3 ms to 300 ms. The training and testing images were normalized using the same pre-processing algorithm: after image alignment, the input image was similarly first thresholded using a triangular thresholding method (see the Methods section in the main text for details) to separate the sample foreground and background pixels. The mean of the background pixel values was taken as the background fluorescence level and subtracted from the entire image. The images were then normalized such that 1% of the foreground pixels were saturated (above one). This pre-processing step did not further clip or quantize the image. These pre-processed images (in single precision format) were fed into the network directly for training or blind testing.

Examples of these blind testing results are shown in Supplementary Figure 7a, where the input bead images were defocused by -5.0, 3.0, and 4.5 μm. With lower exposure times, the input image quality was compromised by noise and image quantization error due to the lower bit depth. As shown in Supplementary Figure 7a, the *Deep-Z* (100 ms) model can successfully refocus the input images even down to an exposure time of 10 ms. However, the *Deep-Z* (100 ms) model fails to virtually refocus the input images acquired at 3 ms exposure time, giving a blurry output image with background noise. On the other hand, the *Deep-Z* (10 ms) model can successfully refocus input images that were captured at 3 ms exposure times, as illustrated in Supplementary Figure 7. Interestingly, the *Deep-Z* (10 ms) model performs slightly worse for input images that were acquired at higher exposure times; for example the input images acquired at 300 ms exposure time exhibit a slight blur at the output image as demonstrated in the last row of Supplementary Figure 7a.These observations are further confirmed in Supplementary Figure 7b,c by quantifying the median FWHM values of the imaged microbeads, calculated at the *Deep-Z* output images as a function of the refocusing distance. This analysis confirms that *Deep-Z* (100 ms) model cannot successfully refocus the images captured at 3 ms exposure time outside of a narrow defocus window of $\sim [-1\ \mu m, 1\ \mu m]$ (see Supplementary Figure 7b). On the other hand, *Deep-Z* (10 ms) model demonstrates improved refocusing performance for the input images captured at 3ms exposure time (Supplementary Figure 7c). These results indicate that training a *Deep-Z* model with images acquired at exposure times that are relatively close to the expected exposure times of the test images would be important for successful inference. Another important observation is that, compared to the ground truth images, the *Deep-Z* output images also reject the background noise since noise overall does not generalize well during the training phase of the neural network, as also discussed for Supplementary Figure 1.

As also emphasized in the main text, the noise performance of *Deep-Z* can potentially be further enhanced by engineering the microscope's point spread function (PSF) to span an extended depth-of-field, by e.g., inserting a phase mask in the Fourier plane of the microscope, ideally without introducing additional photon losses along the path of the fluorescence signal collection. For example, refer to the Supplementary Video 10 for an experimental demonstration of *Deep-Z* blind inference for objects imaged through a double-helix PSF.

Supplementary Note 4: Robustness of *Deep-Z* to changes in samples and imaging systems

In our results reported so far, the blindly tested samples were inferred with a *Deep-Z* network that has been trained using the same *type* of sample and the same microscope system. Here we evaluate and discuss the performance of *Deep-Z* for different scenarios where a change in the test data distribution is introduced in comparison to the training image set, such as e.g., (1) a different type of sample that is imaged, (2) a different microscope system used for imaging, and (3) a different illumination power or SNR.

Regarding the first item, if there is a high level of similarity between the trained sample type and the tested sample type distributions, we expect the performance of the network output to be comparable. As reported in Supplementary Figure 8a, a *Deep-Z* model that was trained to virtually refocus images of tagRFP-labeled *C. elegans* neuron nuclei was blindly tested to virtually refocus the images of GFP-labeled *C. elegans* neuron activity. The output image results of the different model column are quite similar to the output images of the optimal model, trained specifically on GFP-labeled neuron activity images (same model column), as well as the mechanically-scanned ground truth images, with a minor difference in the correlation coefficients of the two sets of output images with respect to the ground truth images of the same samples. Similar conclusions may be drawn for the effectiveness of a *Deep-Z* model blindly tested on images of a different strain of *C. elegans*.

On the other hand, when the training sample type and its optical features are considerably different from the testing samples, noticeable differences in *Deep-Z* performance can be observed. For instance, as shown in Supplementary Figure 8b, a *Deep-Z* model that was trained with 300 nm beads can only partially refocus the images of *C. elegans* neuron nuclei, which are typically 1-5 μm in size, and therefore are not well-represented by the training image dataset

containing only nanobeads. This limitation can be remedied through a transfer learning[7,8] process, where the network trained on one type of sample (e.g., the nanobeads in this example) can be used as an initialization of the network weights and the *Deep-Z* model can be further trained using new images that contain neuron nuclei. Compared to starting from scratch (e.g., randomized initialization), which takes ~ 40,000 iterations (~60 hours) to reach an optimal model, transfer learning can help us achieve an optimal model with only ~4,000 iterations (~6 hours) that successfully refocuses neuron nuclei images, matching the performance of the optimal model (transfer learning column in Supplementary Figure 8). This transfer learning approach can also be applied to image different types of *C. elegans* using earlier models that are refined with new image data in e.g., ~500-1,000 iterations. Another advantage of transfer learning is using less training data; in this case, for example, only 20% of the original training data used for the optimal model was used for transfer learning.

Regarding the second item, i.e., a potential change in the microscope system used for imaging can also adversely affect the inference performance of a previously trained network model. One of the more challenging scenarios for a pre-trained *Deep-Z* network will be when the test images are captured using a different objective lens with a change in the numerical aperture (NA); this directly modifies the 3D PSF profile, making it deviate from the *Deep-Z* learned features, especially along the depth direction. Similar to the changes in the sample type, if the differences in imaging system parameters are small, we expect that a previously trained *Deep-Z* model can be used to virtually refocus images captured by a different microscope to some extent. Supplementary Figure 9 shows an example of this scenario, where a *Deep-Z* model was trained using the images of *C. elegans* neuron nuclei, captured using an *Olympus IX81* microscope with a 20×/0.75NA objective lens, and was blindly tested on images captured using a *Leica SP8*

microscope with 20×/0.8NA objective lens. Stated differently, two different microscopes, manufactured by two different companies, have been used, together with a small NA change between the training and testing phases. As illustrated in Supplementary Figure 9, most of the virtual refocusing results remained successful, in comparison to the optimal model. However, due to these changes in the imaging parameters, a couple of mis-arrangements of the neurons in the virtually refocused images can be seen in the different model output column, which also resulted in a small difference of ~0.02-0.06 between the correlation coefficients of the optimal *Deep-Z* model output and the different model output (both calculated with respect to the corresponding ground truth images acquired using two different microscope systems). As discussed earlier, one can also use transfer learning to further improve these results by taking the initial *Deep-Z* model trained on Olympus IX81 microscope (20×/0.75NA objective) as initialization and further training it for another ~2,000 iterations on a new image dataset captured using the Leica SP8 microscope (20×/0.8NA objective). Similar to the example that we presented earlier, 20% of the original training data used for the optimal model was used for transfer learning in Supplementary Figure 9.

As for the third item, the illumination power, together with the exposure time and the efficiency of the fluorophore, contributes to two major factors: the dynamic range and the SNR of the input images. Since we use a pre-processing step to remove the background fluorescence, also involving a normalization step based on a triangular threshold (see the Methods section of the main text for details), the input images will always be re-normalized to similar signal ranges and therefore illumination power associated dynamic range changes do not pose a major challenge for *Deep-Z*. Furthermore, as detailed earlier, robust virtual refocusing can still be achieved under significantly lower SNR, i.e., with input images acquired at much lower exposure times (see

Supplementary Figure 7). These results and the corresponding analysis reveal that *Deep-Z* is fairly robust to changes observed in the dynamic range and the SNR of the input images. Having emphasized this, training a *Deep-Z* model with images acquired at exposure times that are relatively similar to the expected exposure times of the test images would be recommended for various uses of *Deep-Z*. In fact, the same conclusion applies in general: to achieve the best performance with *Deep-Z* inference results, the neural network should be trained (from scratch or through transfer learning which significantly expedites the training process) using training images obtained with the same microscope system and the same types of samples as expected to be used at the testing phase.

Supplementary Note 5: Time-modulated signal reconstruction using *Deep-Z*

To further test the generalization capability of *Deep-Z*, we conducted an experiment, where the microbead fluorescence is modulated in time, induced by an external time-varying excitation. Training data were captured for 300 nm red fluorescent beads using a 20×/0.75NA objective lens with the Texas Red filter set, same as the microbead samples reported earlier (e.g., Fig.4), except that the fluorescence light source was set at 25% illumination power (32.5 mW) and the exposure time was chosen as 100 ms. Testing data consisted of images of 300 nm red fluorescent beads placed on a single 2D plane (pipetted onto a #1 coverslip) captured using an external light emitting diode (M530L3-C1, Thorlabs) driven by an LED controller (LEDD1B, Thorlabs) modulated by a function generator (SDG2042X, Siglent) that modulated the output current of the LED controller between 0 to 1.2 A following a sinusoidal pattern with a period of 1 s. A Texas Red filter and 100 ms exposure time were used. The same FOV was captured at in-focus plane (z = 0 μm) and five defocus planes (z = 2, 4, 6, 8, 10 μm). At each plane, a two-second video (i.e. two periods of the modulation) was captured at 20 frames per second. Each frame of the defocused planes was then virtually refocused using the trained *Deep-Z* network to digitally reach the focal plane (z = 0 μm). Fluorescence intensity changes of 297 individual beads within the sample FOV captured at z = 0 μm were tracked over the two-second time window. The same 297 beads were also tracked as a function of time using those five virtually refocused time-lapse sequences (using *Deep-Z* output). The intensity curve for each bead was normalized between 0 and 1.

Supplementary Figure 10a reports the time-modulated signal of 297 individual microbeads at the focal plane (z = 0 μm) tracked over a 2 s period at a frame rate of 20 frames per second, plotted with their normalized mean and standard deviation. This curve shows a similar modulation

pattern as the input excitation light, with a slight deviation from a perfect sinusoidal curve due to the nonlinear response of fluorescence. The standard deviation was ~1.0% of the mean signal at each point. Testing the blind inference of *Deep-Z*, the subsequent entries of supplementary Figure 10a reports the same quantities corresponding to the same field-of-view (FOV), but captured at defocused planes (z = 2, 4, 6, 8, 10 μm) and virtually refocused to the focal plane (z = 0 μm) using a *Deep-Z* network trained with images captured *under fixed signal strength*. The mean curves calculated using the virtually-refocused images (z = 2, 4, 6, 8, 10 μm) match very well with the in-focus one (z = 0 μm), whereas the standard deviation increased slightly with increased virtual refocusing distance, which were ~1.0%, 1.1%, 1.7%, 1.9%, and 2.1% of the mean signal for virtual refocusing distances of z = 2, 4, 6, 8, and 10 μm, respectively.

Based on this acquired sequence of images, we further took every other frame to form a new video; by doing so, the down sampled video compressed the original 2 s video to 1 s, forming a group of beads that were modulated at doubled frequency, i.e., 2 Hz. This down-sampled video was repeated, and added back onto the original video, frame-by-frame, with a lateral shift of 8 pixels (2.6 μm). Supplementary Figure 10b shows the *Deep-Z* output on these added images, corresponding to 297 pairs of beads that had the original modulation frequency 1 Hz (first row) and the doubled modulation frequency 2 Hz (second row), masked separately in the same output image sequence. This analysis demonstrates that *Deep-Z* output tracks the sinusoidal illumination well, closely following the in-focus reference time-modulation reported in the first column, same as in Supplementary Figure 10a. We also created Supplementary Video 5 to illustrate an example region of interest containing six pairs of these 1 Hz and 2 Hz emitters, cropped from the input and output FOVs for different defocus planes.

Supplementary Note 6: Impact of the sample density on *Deep-Z* inference

If the fluorescence emitters are too close to each other or if the intensity of one feature is much weaker than the other(s) within a certain FOV, the intensity distribution of the virtually refocused *Deep-Z* images may deviate from the ground truth. To shed more light on this, we first used numerical simulations resulting from experimental data, where we (1) laterally shifted a planar fluorescence image that contained individual 300 nm fluorescent beads, (2) attenuated this shifted image intensity with respect to the original intensity by a ratio (0.2 to 1.0), and (3) added this attenuated and shifted feature back to the original image (see Supplementary Figure 11 for an illustration of this). Based on a spatially-invariant incoherent PSF, this numerical simulation, derived from experimental data, represents an imaging scenario, where there are two individual sets of fluorescent objects that have different signal strengths with respect to each other, also with a varying distance between them. The resulting images, with different defocus distances (see Supplementary Figure 11b) were virtually refocused to the correct focal plane by a *Deep-Z* network that was trained using planar bead samples (see Methods section, Sample Preparation). Supplementary Figure 11b-h demonstrates various examples of bead pairs that were laterally separated by e.g., 1-15 pixels and axially defocused by 0-10 μm, with an intensity ratio that spans 0.2-1.0. To quantify the performance of *Deep-Z* inference for these different input images, we plot in Supplementary Figure 11c-h the average intensity ratio of 144 pairs of dimmer and brighter beads at the virtually refocused plane as a function of the lateral shift (d) and the intensity ratio between the dimmer and the brighter beads, also covering various defocus distances up to 10 μm; in each panel of this figure, we marked the minimal resolvable distance between the two beads by a cross-symbol "x". This figure reveals that larger defocus distances

and smaller ratios require slightly larger lateral shift amount for the bead pairs to be accurately resolved.

Next, we examined the impact of occlusions in the axial direction, which can be more challenging to resolve. For this, we created new numerical simulations, also resulting from experimental data, where this time we axially shifted a planar fluorescent bead image stack and added them back to the corresponding original image stack with different intensity ratios (see Supplementary Figure 12b for an illustration of this). To accurately represent the inference task, the deep network was trained via transfer learning with an augmented dataset containing axially-overlapping objects. Supplementary Figure 12a demonstrates our *Deep-Z* results for a pair of beads located at z = 0 and z = 8 μm respectively. The network was able to successfully refocus these two beads separately, inferring two intensity maxima along the z-axis at z = 0 μm and z = 8 μm, very well matching the simulated mechanically-scanned image stack (ground truth). Supplementary Figures 12c, d plot the average and the standard deviation of the intensity ratio of the top (i.e., the dimmer) bead and the lower bead (i.e., the bead in the original stack) for 144 individual bead pairs inside a sample FOV, corresponding to z=8 μm with different axial separations (*d*, see Supplementary Figure 12b), for both the virtually refocused *Deep-Z* image stack and the simulated ground truth image stack, respectively. The results in Supplementary Figure 12c and d are similar, having rather small discrepancies in the exact intensity ratio values. Our results might be further improved by potentially using a 3D convolutional neural network architecture [9].

To further understand the impact of the axial refocusing distance and the density of the fluorescent sample on *Deep-Z* 3D inference, we performed additional imaging experiments corresponding to 3D bead samples with different densities of particles, which was adjusted by

mixing 2.5 μL red fluorescent bead (300 nm) solution at various concentrations with 10 μL ProLong Gold antifade mountant (P10144, ThermoFisher) on a glass slide. After covering the sample with a thin coverslip, the sample naturally resulted in a 3D sample volume, with 300 nm fluorescent beads spanning an axial range of ~20-30 μm. Different samples, corresponding to different bead densities, were axially scanned using a 20×/0.75NA objective lens using the Texas Red channel. To get the optimal performance, a *Deep-Z* network was trained with transfer learning (initialized with the original bead network) using 6 image stacks (2048 × 2048 pixels) captured from one of the samples. Another 54 non-overlapping image stacks (1536 × 1536 pixels) were used for blind testing; within each image stack, 41 axial planes spanning +/- 10 μm with 0.5 μm step size were used as ground truth (mechanically-scanned), and the middle plane (z=0 μm) was used as the input image to *Deep-Z*, which generated the virtually refocused output image stack, spanning the same depth range as the ground truth images. Thresholding was applied to the ground truth and *Deep-Z* output image stacks, where each connected region after thresholding represents a 300 nm bead. Supplementary Figure 13a illustrates the input images and the maximal intensity projection (MIP) of the ground truth image stack as well as the *Deep-Z* output image stack corresponding to some of the non-overlapping sample regions used for blind testing. At lower particle concentrations (below 0.5×10^6 μL^{-1}), the *Deep-Z* output image stack results match very well with the mechanically-scanned ground truth results over our training range of +/- 10 μm axial defocus. With larger particle concentrations, the *Deep-Z* output gradually loses its capability to refocus and retrieve all the individual beads, resulting in under-counting of the fluorescent beads.

In fact, this refocusing capability of *Deep-Z* not only depends on the concentration of the fluorescent objects, but also depends on the refocusing axial distance. To quantify this, we plot in

Supplementary Figure 13b-e the fluorescent particle density measured using the mechanically-scanned ground truth image stack as well as the *Deep-Z* virtually refocused image stack as a function of the axial defocus distance, i.e., ±2.5 μm, ±5 μm, ±7.5 μm and ±10 μm from the input plane (z=0 μm), respectively. For example, for a virtual refocusing range of ±2.5 μm, the *Deep-Z* output image stack (using a single input image at z=0 μm) closely matches the ground truth results even for the highest tested sample density (~ 4×10^6 μL^{-1}); on the other hand, at larger virtual refocusing distances *Deep-Z* suffers from some under-counting of the fluorescent beads (see e.g., Supplementary Figures 13c-e). This is also consistent with our analysis reported earlier (e.g., Supplementary Figures 11-12), where the increased density of the beads in the sample results in axial occlusions and partially affects the virtual refocusing fidelity of *Deep-Z*.

In our analysis reported so far, we mainly focused on fluorescent particles to be able to quantify different trade-off mechanisms in *Deep-Z* inference; in these examples that we reported, the training image data did not include strong variations in the signal intensities of the particles or axial occlusions that existed in the testing data; this was an important disadvantage for *Deep-Z*. However, we believe that a *Deep-Z* model that is trained with the correct type of samples (matching the test sample type and its 3D structure) will have an easier task in its blind inference and virtual refocusing performance since the training images will naturally contain relevant 3D structures, better representing the feature distribution expected in the test samples.

Supplementary Note 7: Additional analysis on *C. elegans* neuron segmentation

As summarized in Supplementary Table SN1, using an additional neuron segmentation algorithm (named TrackMate[10]) that is based on the Laplacian of Gaussian (LoG) method, resulted in 147 neurons for the *Deep-Z* output, 175 neurons in the target image stack (mechanically scanned) and 169 in the *Deep-Z* merged stack (only 2 axial planes used as input images) for the same worm shown in Supplementary Figure 14d-i, revealing a close match between *Deep-Z* results and the results obtained with a mechanically scanned image stack, with a relatively small depth difference (Δz = -0.243 ± 0.706 μm; see Supplementary Table SN1). The neuron segmentation results obtained by TrackMate (Supplementary Table SN1), in comparison to the ones shown in Table 1 in the main text, also show some inconsistency in the neuron segmentation itself (meaning that there might not be a single ground truth method). Furthermore, as shown in Supplementary Table SN1, a change in the center plane of the image stack (i.e., $z = 0$ μm v.s. $z = -2.5$ μm, where both image stacks cover the entire body of the worm) also changes the resulting number of segmented neurons, even for the ground truth, using M = 41 mechanically scanned images; this once again illustrates the challenging nature of the neuron segmentation task itself.

Neuron locations in Supplementary Figure 14 were compared by first matching the pairs of neurons from two different methods (e.g., *Deep-Z* results vs. mechanically-scanned ground truth). Matching two groups of segmented neurons (Ω_1, Ω_2), represented by their spatial coordinates, was considered as a bipartite graph minimal cost matching problem, i.e.:

$$\underset{x_e}{\operatorname{argmin}} \sum_{e} c_e \cdot x_e$$

$$\text{s.t.} \sum_{e \in \delta(u_1)} x_e = 1, \quad \text{for } \forall u_1 \in \Omega_1$$

$$\sum_{e \in \delta(u_2)} x_e \leq 1, \quad \text{for } \forall u_2 \in \Omega_2$$
$$x_e \in \{0, 1\}$$

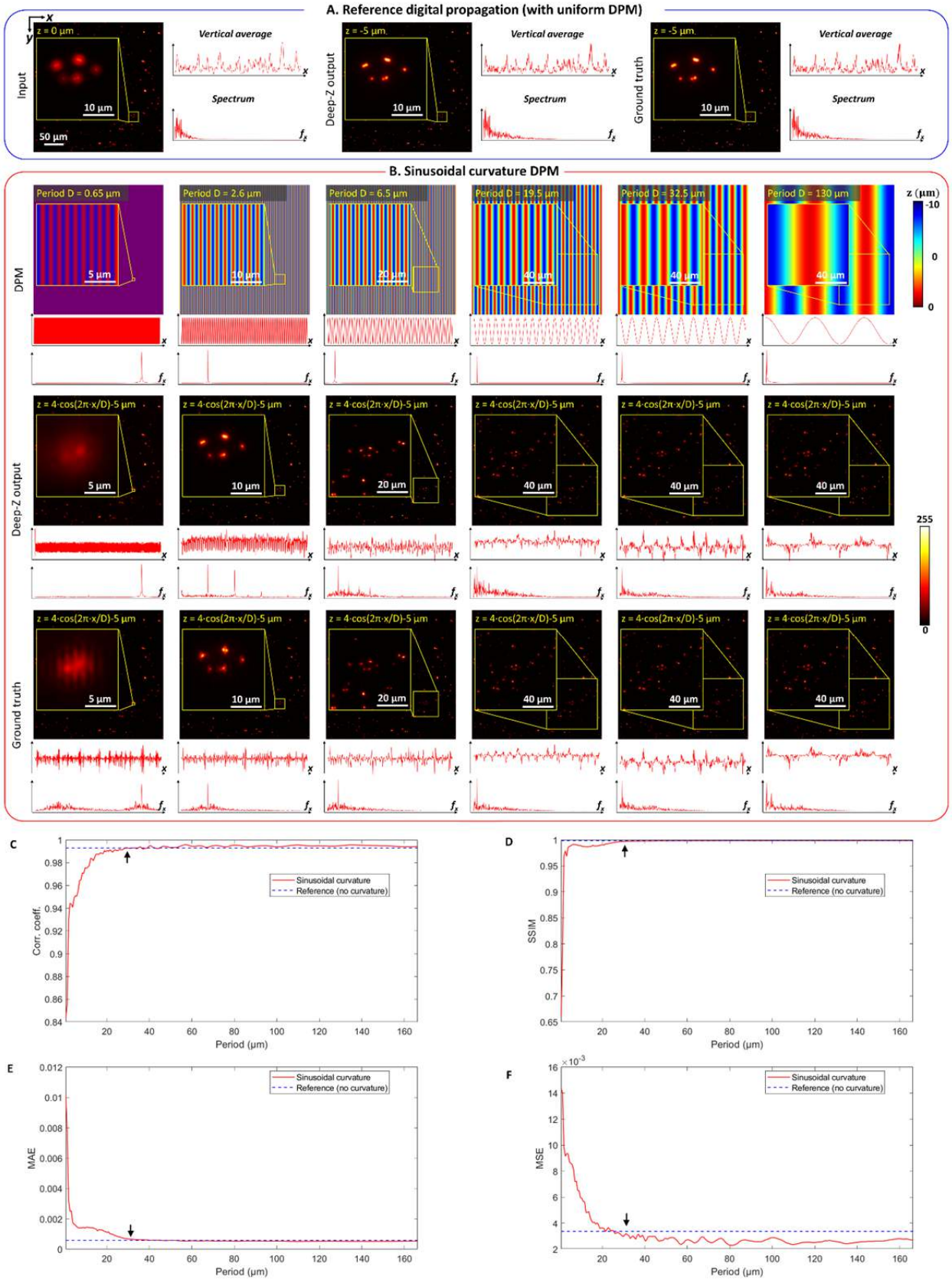
where $x_e = 1$ represents that the edge between the two groups of neurons (Ω_1, Ω_2) were included in the match. The cost on edge $e = (u_1, u_2)$ is defined based on the Manhattan distance between $u_1 \in \Omega_1, u_2 \in \Omega_2$, i.e., $c_e = |x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|$. Because the problem satisfies totally unimodular condition, the above integer constraint $x_e \in \{0,1\}$can be relaxed to linear constraint $x \geq 0$ without changing the optimal solution, and the problem was solved by linear programming using Matlab function linporg. Then the axial distances between each paired neurons (Δz) were calculated and their distributions were plotted and reported.

	TrackMate segmentation (centered at Z = 0 μm)			*TrackMate segmentation (centered at Z = -2.5 μm)*		
	N (neurons)	*Δz (mean ± std, μm)*	*|Δz| (mean ± std, μm)*	*N (neurons)*	*Δz (mean ± std, μm)*	*|Δz| (mean ± std, μm)*
Input image (M = 1 image)	7	0.078 ± 0.357	0.273 ± 0.219	6	0.120 ± 0.181	0.189 ± 0.083
Deep-Z output stack (M = 1 image)	147	-0.243 ± 0.706	0.445 ± 0.600	146	-0.577 ± 0.922	0.689 ± 0.841
Merged stack (M = 2 images)	169	-0.102 ± 0.569	0.386 ± 0.430	179	0.581 ± 1.352	0.863 ± 1.191
Mechanical scan stack (M = 41 images)	175	*(Ground truth)*	*(Ground truth)*	177	*(Ground truth)*	*(Ground truth)*

Supplementary Table SN1. Neuron segmentation results for a *C. elegans* worm using TrackMate [10]**.** Two image stacks each spanning +/- 10 μm were used here, one centered at z = 0 μm and the other at z = -2.5 μm, respectively. Both of these image stacks cover the entire body of the worm, and are axially shifted with respect to each other by 5 frames. TrackMate was used to segment neurons with identical parameters on both sets of 3D image stacks. The segmented neuron locations were also compared against the ground truth (i.e., the corresponding mechanically scanned image stack, M = 41), reporting the axial error, Δz, as well as the absolute axial error, |Δz|, in the form of mean ± standard deviation (std), respectively.

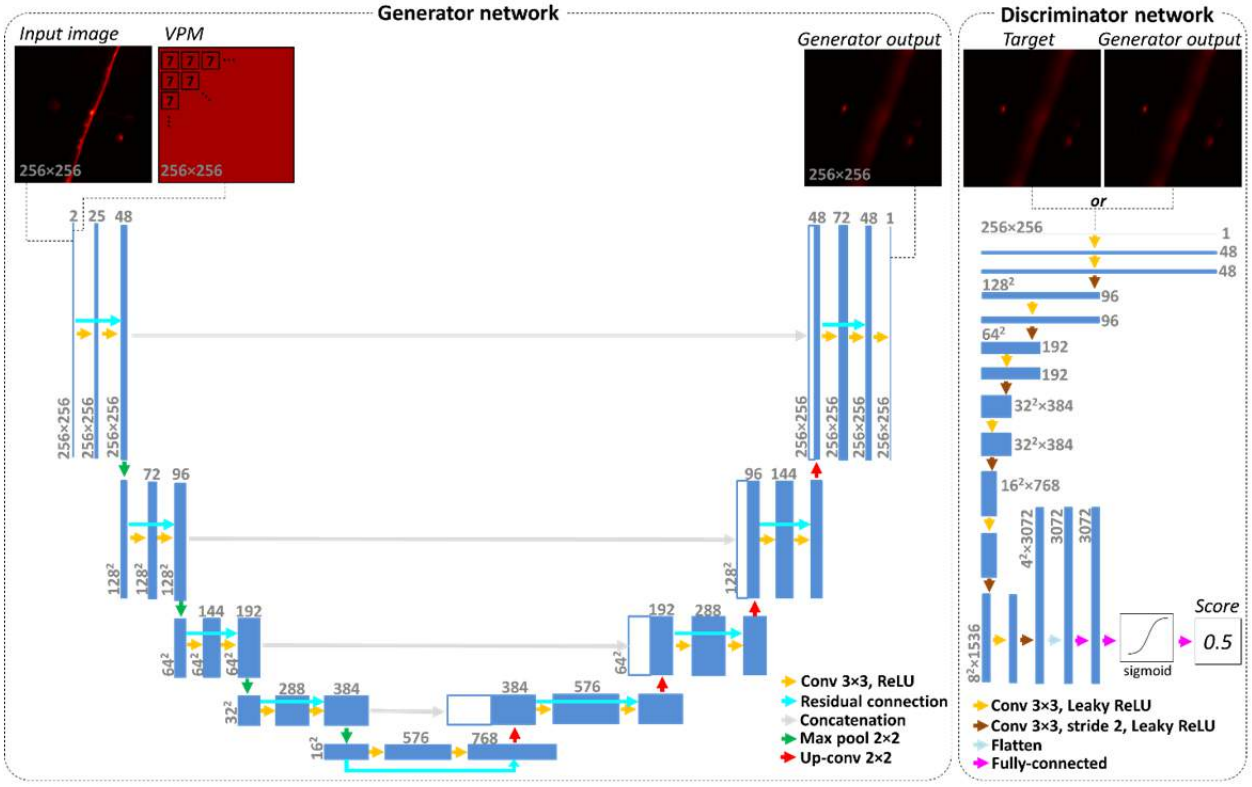
Supplementary Note 8: *Deep-Z* based aberration correction using spatially non-uniform DPMs

To evaluate the limitations of *Deep-Z* based aberration correction using spatially non-uniform DPMs, we quantified the 3D surface curvature that a DPM can have without generating artifacts. For this, we used a series of DPMs that consisted of 3D sinusoidal patterns with lateral periods of $D = 1, 2, \ldots, 256$ pixels along the x-direction (with a pixel size of 0.325 μm) and an axial oscillation range of 8 μm, i.e., a sinusoidal depth span of -1 μm to -9 μm with respect to the input plane. Each one of these 3D sinusoidal DPMs was appended on an input fluorescence image that was fed into the *Deep-Z* network. The network output at each sinusoidal 3D surface defined by the corresponding DPM was then compared against the images that were interpolated in 3D using an axially-scanned z-stack with a scanning step size of 0.5 μm, which formed the ground truth images that we used for comparison. As summarized in Supplementary Figure SN1, the *Deep-Z* network can reliably refocus the input fluorescence image onto 3D surfaces defined by sinusoidal DPMs when the period of the modulation is > 100 pixels (i.e., > 32 μm in object space). For faster oscillating DPMs, with periods smaller than 32 μm, the network output images at the corresponding 3D surfaces exhibit background modulation at these high-frequencies and their harmonics as illustrated in the spectrum analysis reported in Supplementary Figure SN1. These higher harmonic artifacts and the background modulation disappear for lower frequency DPMs, which define sinusoidal 3D surfaces at the output with a lateral period of > 32 μm and an axial range of 8 μm.


A. Reference digital propagation (with uniform DPM)
Input
Deep-Z output
Ground truth
Vertical average
Spectrum
B. Sinusoidal curvature DPM
DPM
Deep-Z output
Ground truth
Period D = 0.65 μm
Period D = 2.6 μm
Period D = 6.5 μm
Period D = 19.5 μm
Period D = 32.5 μm
Period D = 130 μm
z (μm)
z = 4·cos(2π·x/D)-5 μm
C
Corr. coeff.
D
SSIM
E
MAE
F
MSE
Sinusoidal curvature
Reference (no curvature)
Period (μm)

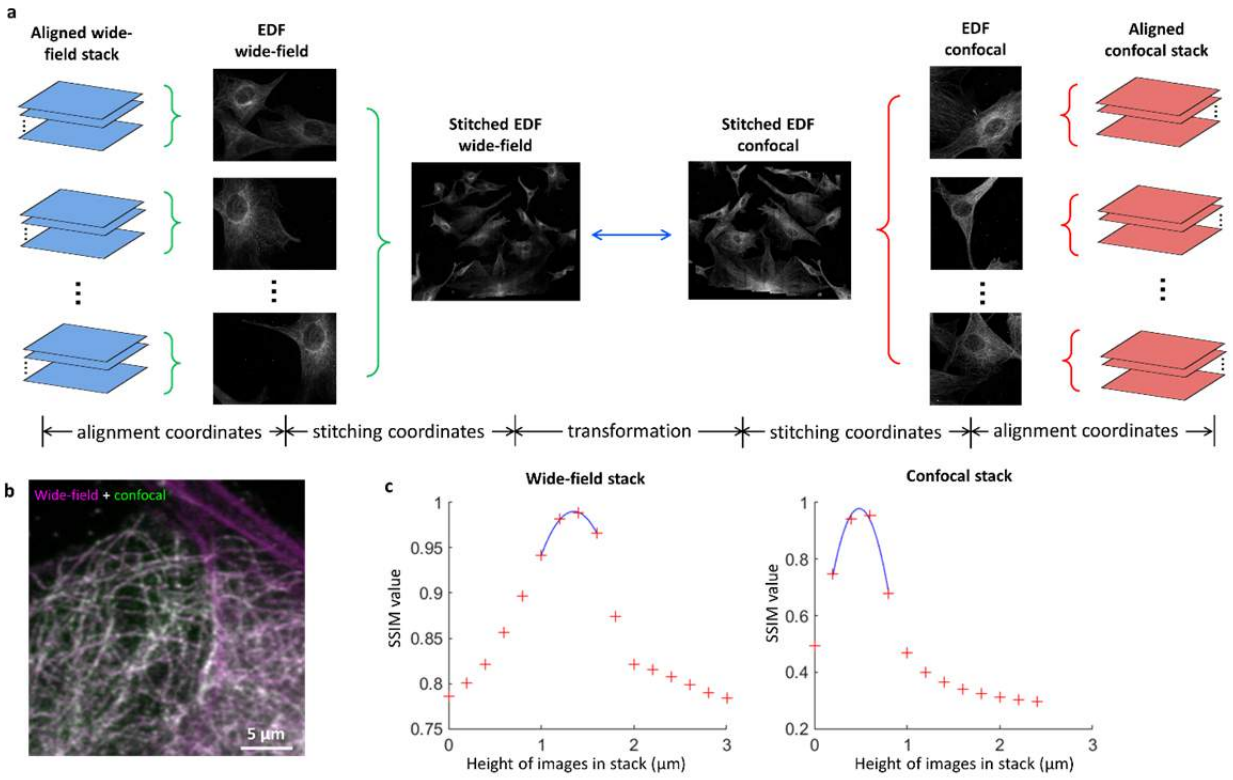
Supplementary Figure SN1. DPM surface curvature analysis. (a) A fluorescent sample consisting of 300 nm fluorescent beads was digitally refocused to a plane 5 μm above the sample by appending a DPM with uniform entries. The ground truth is captured using mechanical scanning at the same plane. Vertical average (i.e., the pixel average along the y-axis of the image) and its spatial frequency spectrum (i.e., the Fourier transform of the vertical average with the zero-frequency removed) are shown next to the corresponding images. (b) Digital refocusing of the same input fluorescence image by appending a DPM that defines a sinusoidal 3D surface with varying periods, from 0.65 μm to 130 μm along the x-axis, with an axial oscillation range of 8 μm, i.e., a sinusoidal depth span of -1 μm to -9 μm with respect to the input plane. The ground truth images were bicubic-interpolated in 3D from a z-scanned stack with 0.5 μm axial spacing. Vertical average of each DPM and the corresponding spatial frequency spectrum are shown below each DPM. Vertical average of the difference images (i.e., the resulting *Deep-Z* image minus the reference *Deep-Z* image in (a) as well as the ground truth image minus the reference ground truth image in (a)) and the corresponding spectra are shown below each image. (c-f) Correlation coefficient (Corr. Coeff.), structural similarity index (SSIM), mean absolute error (MAE) and mean square error (MSE) were used to compare *Deep-Z* output images against the ground truth images (calculated for a single imaging FOV of 1024×1024 pixels) at the same 3D sinusoidal surfaces defined by the corresponding DPMs, with varying periods from 0.65 μm to 170 μm along the x-axis. Reliable *Deep-Z* focusing onto sinusoidal 3D surfaces can be achieved for lateral modulation periods greater than ~32 μm (corresponding to ~ 100 pixels), as marked by the black arrows in (c-f). the same conclusion is also confirmed by our results and spatial frequency analysis reported in (b). (c-f) was calculated from a single image FOV.

Supplementary Note 9: Generator and discriminator network structures used in *Deep-Z*



Supplementary Figure SN2. Generator and discriminator network structures used in *Deep-Z*. ReLU: rectified linear unit. Conv: convolutional layer.

Supplementary Note 10: Cross-modality z-stack registration



Supplementary Figure SN3. *Deep-Z+* training phase: the registration of a wide-field fluorescence z-stack against a confocal z-stack. (a) Registration in the lateral direction. Both the wide-field and the confocal z-stacks were first self-aligned and extended depth of field (EDF) images were calculated for each stack. The EDF images were stitched spatially and the stitched EDF images from wide-field were aligned to those of confocal microscopy images. The spatial transformations, from stitching to the EDF alignment, were used as consecutive transformations to associate the wide-field stack to the confocal stack. (b) Non-empty wide-field ROIs of 256×256 pixels and the corresponding confocal ROIs were cropped from the EDF image, which were further aligned. The image here shows an example overlay of the registered image pair, with wide-field image in magenta and the corresponding confocal image in green. (c) To align the wide-field and confocal stacks in the axial direction, focus curves in the wide-field stack and

the confocal stack were calculated and compared based on the corresponding SSIM values (see Methods for details).

References:

1. Nguyen, J. P. *et al.* Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* **113**, E1074–E1081 (2016).
2. Hoebe, R. A., Voort, H. T. M. V. D., Stap, J., Noorden, C. J. F. V. & Manders, E. M. M. Quantitative determination of the reduction of phototoxicity and photobleaching by controlled light exposure microscopy. *J. Microsc.* **231**, 9–20 (2008).
3. Schilling, Z. *et al.* Predictive-focus illumination for reducing photodamage in live-cell microscopy. *J. Microsc.* **246**, 160–167 (2012).
4. Magidson, V. & Khodjakov, A. Chapter 23 - Circumventing Photodamage in Live-Cell Microscopy. in *Methods in Cell Biology* (eds. Sluder, G. & Wolf, D. E.) **114**, 545–560 (Academic Press, 2013).
5. Hoebe, R. A. *et al.* Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nat. Biotechnol.* **25**, 249–253 (2007).
6. Huisken, J., Swoger, J., Bene, F. D., Wittbrodt, J. & Stelzer, E. H. K. Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **305**, 1007–1009 (2004).
7. Hoo-Chang, S. *et al.* Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **35**, 1285–1298 (2016).
8. Weiss, K., Khoshgoftaar, T. M. & Wang, D. A survey of transfer learning. *J. Big Data* **3**, 9 (2016).
9. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *ArXiv160606650 Cs* (2016).

10. Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods San Diego Calif* **115**, 80–90 (2017).

Supplementary Video Captions:

Supplementary Video 1. *Deep-Z* inference comparison against the images captured using a mechanical axial scan of a *C. elegans* sample.

Left: a single fluorescence image of the *C. elegans* was captured at the reference plane (z = 0 μm), used as the *Deep-Z* input. Middle: the input image was appended with different DPMs and passed through the trained *Deep-Z* network to digitally refocus the input image to a series of planes, from -10 μm to 10 μm (with a step size of 0.5 μm) with respect to the reference plane. Right: mechanical axial scan of the same *C. elegans* at the same series of planes, used for comparison.

Supplementary Video 2. 3D visualization of a *C. elegans* using *Deep-Z* inference.

Left: a single fluorescence image of a *C. elegans* was captured and used as *Deep-Z* input. *Middle*: the input fluorescence image was digitally refocused using *Deep-Z* to a series of planes, from -10 μm to 10 μm (with an axial step size of 0.5 μm) to generate a 3D stack. This 3D stack was rotated around the vertical axis of the input image, spanning 360° with a step size of 2°. Maximum intensity projection of the 3D volume at each rotated angle is shown in the video, which was generated using the ImageJ plugin "Volume Viewer". *Right*: the same 3D stack was deconvolved using the Lucy-Richardson deconvolution regularized by total variation in ImageJ plugin "DeconvolutionLab2". The deconvolved 3D stack was rotated and displayed in the same way as the middle video.

Supplementary Video 3. *Deep-Z* 3D inference from a 2D video containing four moving *C. elegans*.

A fluorescence video containing four *C. elegans* was recorded at a single plane (z = 0 μm) at 10 frames per second for 10 seconds. Each frame was digitally refocused using *Deep-Z* to a series of planes at z = -6, -4, -2, 2, and 4 μm away from the input plane, generating virtual videos at these different depths in 3D. The input video and the *Deep-Z* generated videos were played simultaneously at one frame per second, i.e., 10-fold slowed down.

Supplementary Video 4. *Deep-Z* 3D inference from a 2D video containing a defocused moving *C. elegans*.

A fluorescence video was captured at a single focal plane (z = 0 μm) at 3 frames per second for a duration of 18 seconds. Each frame was digitally refocused using *Deep-Z* to a series of planes at z = 2, 4, 6, 8 and 10 μm away from the input focal plane, generating virtual videos at these different depths in 3D. In the input video, the worm was mostly defocused due to sample drift and motion. Using *Deep-Z*, neurons are rapidly refocused at these virtual planes in 3D.

Supplementary Video 5. *Deep-Z* based refocusing of spatio-temporally modulated bead images.

Videos contain two groups of 300 nm bead emitters (sinusoidally-modulated at 1 Hz and 2 Hz, respectively). Deep-Z was used to digitally refocus the defocused videos to virtually reach z = 0 μm plane. An example region of interest containing six pairs of such emitters was cropped and shown in this video. Also see Supplementary Figure 10.

Supplementary Video 6. Tracking of neuron calcium activity events in 3D from a single 2D fluorescence video.

A fluorescence video containing a fixed *C. elegans* was recorded at a single focal plane (z = 0 μm) at ~3.6 frames per second for ~35 seconds. The video contained two color channels: the red channel represents the Texas Red fluorescence targeting the RFP-tagged neuron nuclei and the green channel represents the FITC fluorescence targeting the GFP-tagged neuron activities. Each frame was digitally refocused using *Deep-Z* to a series of planes from -10 to 10 μm with 0.5 μm step size, for each one of the fluorescence channels. MIP was applied along the axial direction to generate an extended depth of field image for each frame. 2× zoom-ins of the head and tail regions were shown for better visualization. The input video and the *Deep-Z* generated MIP video were played simultaneously at ~3.6 frames per second.

Supplementary Video 7. Locations of the detected neurons in 3D.

Individual neuron locations were isolated from the *Deep-Z* inferred 3D stack of the neuron calcium activity video. The isolated neuron locations were plotted in 3D. A rough shape of the worm was also plotted, which was generated by thresholding the auto-fluorescence of the worm in the *Deep-Z* generated 3D stack. The 3D plot was rotated around the z-axis at one degree steps for 360 degrees, and viewed at 15 degrees tilt. Each neuron was color-coded according to its depth (z) location, as indicated by the color-bar on the right.

Supplementary Video 8. *Deep-Z* based 3D structural imaging of *C. elegans* at 100 Hz.

A *C. elegans* worm was imaged using a 20×/0.8NA objective lens under the Texas Red channel to capture its tag-RFP signal, labeling the neuron nuclei. A fluorescence video was captured at a single focal plane (z = 0 μm) at 100 full frames per second for a duration of 10 seconds using the stream mode of the camera. Each frame was digitally refocused using *Deep-Z* to a series of planes at z = -2, 2, 4, 6, 8 and 10 μm away from the input focal plane, generating virtual videos at these different depths in 3D, as well as a corresponding MIP video (over an axial depth range of +/- 10 μm).

Supplementary Video 9. *Deep-Z* based 3D functional imaging of *C. elegans* at 100 Hz.

A *C. elegans* worm was imaged using a 20×/0.8NA objective lens under the FITC channel to capture its GFP signal that labels its calcium activity. A fluorescence video was captured at a single focal plane (z = 0 μm) at 100 full frames per second for a duration of 10 seconds using the stream mode of the camera. Each frame was digitally refocused using *Deep-Z* to a series of planes at z = -2, 2, 4, 6, 8 and 10 μm away from the input focal plane, generating virtual videos at these different depths in 3D, as well as a corresponding MIP video (over an axial depth range of +/- 10 μm).

Supplementary Video 10. *Deep-Z* based refocusing of microscopic objects imaged with a 3D engineered point-spread function (PSF).

Left: a single fluorescence image of 300 nm red fluorescence beads with a 3D engineered double-helix PSF was captured at the reference plane (z = 0 μm), which was used as the *Deep-Z* input. Middle: the input image on the left was appended with different DPMs and passed through a trained *Deep-Z* network to digitally refocus the input image to a series of planes, from -13 μm to 10 μm (with a step size of 0.2 μm) with respect to the reference plane. Right: mechanical axial scan of the same sample at the same corresponding planes, used for comparison.