Dual-stage-processed 3D network super-resolution for volumetric fluorescence microscopy far beyond throughput limit

Hao Zhang[1+], Yuxuan Zhao[1+], Chunyu Fang[1+], Guo Li[1], Yi Li[1,4], Meng Zhang[2], Yusha Li[2], Peng Wan[2], Tingting Yu[2,3], Yuhui Zhang[2,3], Dan Zhu[2,3], Shangbang Gao[4], Peng Fei[1*]

[1] *School of Optical and Electronic Information-Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, 430074, China.*

[2] *Britton Chance center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China.*

[3] *MoE Key Laboratory for Biomedical Photonics, Huazhong University of Science and Technology, Wuhan, 430074, China*

[4] *College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China.*

[+] *These authors contribute equally to this work*

[*] *Correspondence: feipeng@hust.edu.cn*

Abstract

Though three-dimensional (3D) fluorescence microscopy has been an essential tool for modern life science research, the fundamentally-limited optical throughput prevents its further move towards faster, clearer and higher-throughput applications. We hereby report a dual-stage, mutual-feedback 3D super-resolution network approach that allows gradient deep learning of the microscopy degradation from high-resolution targets to low-resolution images, which might be decimated in conventional 1-stage network approaches. Using single blurred-and-pixelated 3D image as input, our trained network is able to infer a 3D output with notably higher resolution and improved contrast. Therefore, it can push the throughput limit of current 3D fluorescence microscopy from three striking aspects as: notably reducing the acquisition time for accurate mapping of large organs, breaking the diffraction limit for imaging subcellular events with faster, lower-toxicity measurement, and significantly improving temporal resolution for capturing instantaneous biological processes. We demonstrate the success of this approach by imaging vessels/neurons in mouse brain at single-cell resolution and throughput of 6 minutes per whole-brain. We also image microtubules in single *U2OS* cell beyond the diffraction limit based on singe 3D acquisition at 2 Hz volume rate, and map neuronal activities of freely-moving *C. elegans* across 3400 × 800 × 60 μm field-of-view at single-cell resolution and 30 Hz volume rate.

Introduction

A recurring trend in biology is the attempt to extract ever more spatial information from three-dimensional biological specimens in an ever short time. Many millisecond- dynamic cellular processes, such as functional activities occurring in live animals, requires high-speed capture of transient spatial patterns at high resolution. Massive cellular details distributed across very large three-dimensional tissues also need to be obtained within an affordable time. These increasing quests are posing substantial challenge to current 3D light microscopy[1-3]. For either conventional epifluorescence microscopy, such as confocal[4] and two-photon excitation microscopes[5], or the newly-emerged light-sheet fluorescence microscopy (LSFM)[6-12], the optical information these imaging techniques could provide in a certain time, also known as

optical throughput, remains unsatisfactory. For example, through a single acquisition, these microscopes are still incapable of either offering subcellular lateral resolution across large field-of-view (FOV) of meso-scale tissues, or breaking the diffraction limit for resolving organelles in single cell. On the other hand, these methods need to record many planes to comprise a 3D image with high-axial resolution, for which, however, the transient temporal profiles are decimated due to extended acquisition time. To improve the temporal resolution for capturing highly dynamic biological processes, light field microscopy (LFM) particularly permits the retrieval of transient 3D signal distribution through post-processing of 2D light field recorded by single camera snapshot, but inevitably suffers from compromised spatial resolution and presence of reconstruction artefacts. Along another direction, various spatial resolution enhancement techniques, Fourier ptychographic microscopy[13,14], structured illumination microscopy (SIM)[15,16], single molecular localization microscopy[17-19], have been developed in past decade and greatly revolutionized the life science research. At the expense of increased acquisition time, these methods can computationally reconstruct a super-resolved 3D image based on a number of low-resolution measurements that contain correlations in space, frequency, or spectrum domain. Hence, despite either capturing 3D dynamics faster by one snapshot or achieving higher resolution by computing multiple image stacks, the optical throughput of these 3D microscopy methods remains fundamentally limited and has so far prevented more widespread applications.

Unlike abovementioned multi-frame super-resolution techniques, single-image-based super-resolution (SISR) methods can improve the resolution without necessarily increasing the imaging time, thereby improving the throughput. However, these methods suffer from either poor enhancement effect[20-22], or restriction of the signal characteristics[23,24]. Super-resolution based on neural network has recently provided a paradigm shift in light microscopy[25-29], with directly deducing a higher-resolution image based on single input. Despite its advances in stronger SISR capability and end-to-end high reconstruction speed, the lack of super-resolution in three dimensions yet prevents its application to capture cellular events in 3D tissues at desired high throughput. Furthermore, the quest on ever better super-resolution effect also intrinsically drive current network-based super-resolution methods towards newer structure and logic for applications now requiring increasingly higher spatiotemporal performance.

Here we propose a novel imaging strategy based on a dual-stage-processed neural network, termed DSP-Net, which can efficiently recover a 3D fluorescence image stack with improved resolution and higher signal-to-noise ratio. Unlike the conventional one-stage network approaches, our DSP-Net contains two mutual-feedback sub-nets that manage the de-blurring and de-pixelization of 3D image separately, thus fundamentally reversing the microscopy imaging process through which the microscope transform a high-resolution (HR) target into a blurred-and-pixelated image. Using such hierarchical training architecture, the network allows progressive deep learning of the recovery from low-resolution blurred-and-pixelated images, to middle-resolution aliased images, and finally to high-resolution images of the target. By this means, a properly-trained DSP-Net can restore more abundant signal details which might be decimated in conventional one-stage network approaches, leading to stronger super-resolution capability. Furthermore, the end-to-end non-iterative inference by the network enables ultra-high processing throughput suited for large-scale 3D reconstruction of or 4D video processing. To demonstrate the ability of DSP-Net combined with various imaging methods, such as

Confocal, LSFM, N-SIM, to push the throughput limits of 3D fluorescence microscopy, we image neurons activities of freely-moving *C. elegans* across a 3400 × 800 × 60 μm large FOV at single-neuron resolution, extracting the spatiotemporal patterns of neuronal calcium signaling and tracking correlated worm activity at volume rate of 30 Hz, notably faster than conventional Bessel LSFM. We also demonstrate 3D imaging of organelles in single *U2OS* cell at 2 Hz volume rate as well as 100-nm spatial resolution that is unmet by conventional N-SIM, and 3D anatomical mapping of neurons/vessels in whole mouse brain at single-cell resolution as well as ultra-high throughput of 6 minutes per brain, which is over 2-orders-of-magnitude higher as compared to conventional organ 3D imaging methods.

Results

DSP-Net for 3D super resolution.

An end-to-end super-resolution neural network is able to fit complex mapping functions only when sufficient parameters are provided. Practically, owing to the highly nonlinear relationship in the ill-posed SISR problem, it is often difficult to directly map the single low-resolution (LR) input to its high-resolution (HR) target perfectly as expected, especially when they have significant quality gap, or in another word, a notable quality improvement is desired. We mitigate this network limitation by logically solving the SISR problem by two steps as: extracting rich details from the badly decimated LR input and enhancing the true signals while suppressing noises and the background. These two steps also fundamentally correspond to the recovery of digital signal aliasing by camera pixelization, and the restoration of optical blurring by the objective with limited numerical aperture (NA). Accordingly, we create a resolver and an interpolator modules in our DSP-Net, with each containing independent training target corresponding to the solution of two sub-issues. Similar to standard deep learning strategy, our DSP-Net comprises a training and an inference phases (Fig. 1a). At the training phase, the neural network establishes its super-resolution ability from scratch by learning from a bunch of examples (i.e., the training dataset). HR 3D images, regarded as ground truth, were acquired using scanned Bessel light microscopes with high-magnification/ high-NA objectives, or in the case for breaking diffraction limit, by Bessel light-sheet microscope combined with 3D SRRF techniques[30]. We then applied an image degradation model(Supplementary Note 1), including optical blurring, down-sampling and noise addition, to the ground truths to generate synthetic LR images, which was corresponded to our LR measurement under low-magnification/low-NA setups (Fig.1a, step1). Meanwhile, a directly down-scaled version of the ground truths, termed mid-resolution (MR) data, was also generated as aliased results to bridge the two-stage training (Fig.1a, step2). With these synthetic training pairs prepared, the LR images were first fed to the super-resolver for clearing the noises and the backgrounds while retaining the potential high frequency structures. (Fig.1a, step3). The mean square error (MSE) between the MR and the output of the super-resolver, termed MR', was defined as the resolving loss, to evaluate the output quality of the super-resolver (Fig.1a, step 4). The MR' outputs were then sent the interpolator to be up-sampled (4 times in each dimension) into the super-resolution results (SR), with their high-frequency features being extracted and further enhanced (Fig.1a, step 5). We further defined the MSE between the HR ground truths and the SR outputs as the interpolating loss (Fig.1a, step 6). The outputs of interpolator, termed SR, are the final results for computing the interpolating error with the HR targets. At the first stage, the parameters of the resolver were

iteratively optimized via minimizing the resolving error calculated by the mean-square-error between MR' and MR (step 4). At the second stage, both the interpolator and the resolver responded to the interpolating loss so that its minimization will result in the optimization of both sub-nets (step 4, 6). By iteratively minimizing the two loss functions using a gradient descent approach, the parameters of both sub-nets could be optimized gradually, empowering the whole DSP-Net with ability to recover SR outputs from LR inputs with quality very close to HR ground truths. Once the training of DSP-Net converged to an optimal status, blurred, noised and pixelated 3D images captured by low-magnification/low NA (or diffraction-limited) setups could be input to the trained DSP-Net, to be recovered into a higher-resolution, higher-contrast, de-pixelated outputs directly without iterative computation required (Fig. 1a, step7). By substantially increasing the spatial resolution without repetitive imaging/stitching, DSP-Net efficiently improves the imaging throughput, known as the spatial information provided per unit acquisition time, of 3D fluorescence microscopy, which is originally limited by the system optics, regardless of its employed magnification factor or numerical aperture (NA).

We first demonstrated the DSP-Net based light-sheet imaging of fluorescence-labeled neurons in mouse brain. The raw 3D image stack, acquired by a Bessel light-sheet microscope using 3.2×/0.2 detection combined with ~3-μm thickness plane illumination, encompassed 2 billion voxels across a 4 × 4 × 2 mm large FOV (Fig. 1b, left). The magnified view (Fig. 1b, LR) shows obviously inadequate resolution due to relatively low NA and large voxel size, as compared to the HR reference acquired by 12.6×/0.5 detection combined with ~1.3-μm plane illumination (Fig. 1b, HR). DSP-Net recovered a 3D image with improved resolution very close to that of the HR reference, and also relatively high reconstruction fidelity (Fig.1b, DSP-Net). Furthermore, we remodeled a few existing super-resolution networks into three dimensions[27,31,33], and then compared their 3D super-resolution performance with our DSP-Net. As shown in Fig.1b, Supplementary Fig. 6 and 7, the DSP-Net successfully yielded 3D images with higher resolution as well as less hallucinations, as compared to those state-of-the-art one-stage networks.

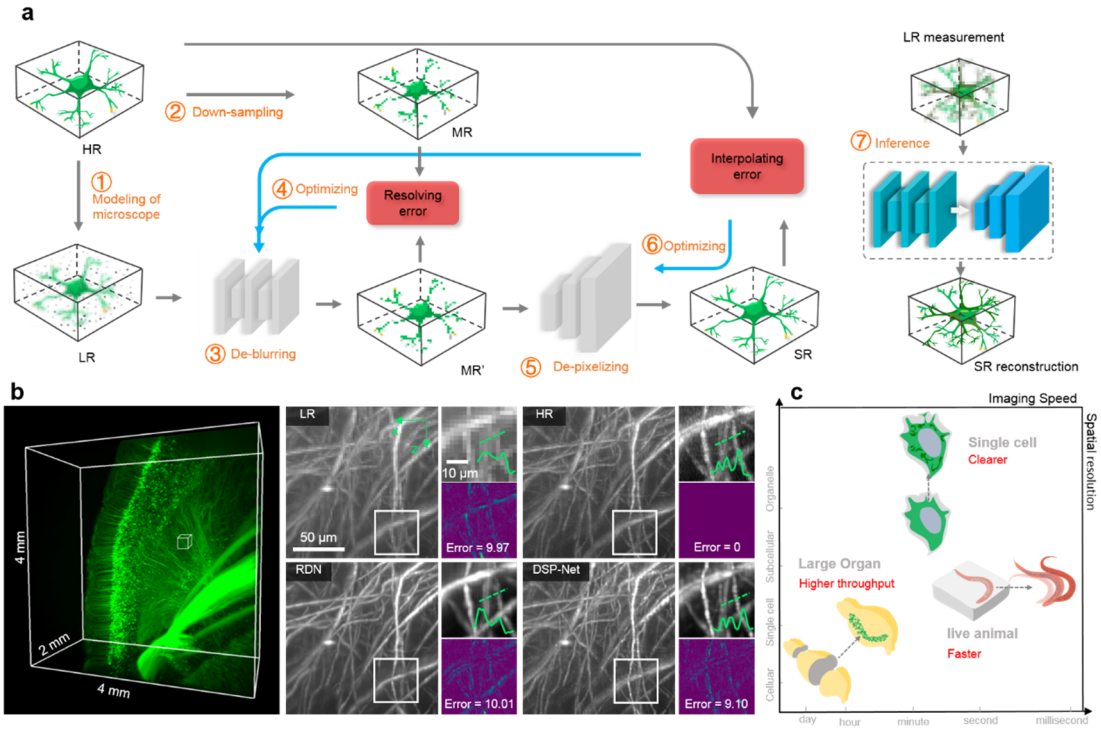

Fig. 1 | DSP-Net and its performance. **a**, The pipeline of DSP-Net super-resolution. First, HR image volumes are acquired using 3D microscopy with high-magnification, high NA optics (or with 3D SRRF computation). A number of noised, blurred, and down-sampled LR volumes are generated from these HR volumes by using a microscope degradation mode (step 1). The corresponding MR volumes are simply 4-times down-sampled version of HRs (step 2), which are used as the targets in the first stage of DSP-Net, termed resolver, to restore the blurred information and reduce the noises for LR inputs (step3). The outputs of resolver, termed MR', are the intermediate results for computing the resolving error with the MR targets. The parameters of the resolver are iteratively optimized via minimizing the resolving error (step 4). The MR' volumes are further input to the second stage of the DSP-Net, termed interpolator, to recover the decimated details for MR'(step 5). The outputs of interpolator, termed SR, are the final results for computing the interpolating error with the HR targets. At the second stage, both the interpolator (SR-HR) and the resolver (MR-MR') are responsible for the interpolating loss, thus minimizing it would result in the optimization of both sub-nets (step 4, 6). After the DSP-Net being trained, LR images experimentally obtained by 3D microscopy with low-NA, low-magnification optics are input to the trained DSP-Net, to instantly reconstruct deblurred, denoised, depixelated image volumes with notable quality improvement. **b**, DSP-Net enabled LSFM results of the fluorescence-labelled neurons in mouse brain. The raw LSFM image stack (2 × 2 × 4 μm voxel) was acquired using a scanned Bessel light-sheet microscope with 3.2× magnification plus 2.7-μm thick laser-sheet illumination. The right panel shows the *xz* maximum intensity projection (1-mm projection depth) of the LR input, the HR ground truth, the SR inference by one-stage RDN network, and the SR inference by our DSP-Net. The line profiles of the resolved structures by each mode are plotted to indicate the achieved resolution. The pixel-wise intensity difference between each mode and the HR ground truth is calculated to evaluate the reconstruction error. **c**, The spatiotemporal map showing the DSP-Net can push the limit of conventional 3D fluorescence microscopy via enabling faster imaging speed for capturing dynamic process, higher spatial resolution for observing subcellular details beyond diffraction limit, and

higher throughput for rapidly mapping whole tissues/organs.

3D imaging of whole mouse brain at single-neuron resolution and minutes time-scale. We imaged GFP-tagged neurons in whole mouse brain (Thy1-GFP-M) at high throughput. The cleared whole brain has a large size of ~10 × 8 × 5 mm with diverse neuronal structures distributed in many brain sub-regions. Even with the use of our self-built Bessel light-sheet microscope that has relatively high acquisition speed (Supplementary Table 2), single-cell-resolution volumetric mapping of whole brain remains highly challenging, owing to the limited system throughput, and thereby requires long-lasting tile imaging, e.g., around 100 tiles in tens of hours, under a high-magnification setup, e.g., 12.6× detection combined with 1.3-μm thin plane illumination (Methods).

In our DSP-Net Bessel light-sheet implementation (Fig. 2a), we first chose high-resolution, high-SNR Bessel images of a few neurons (12.6× setup, 0.4 ms exposure, 5 fps) in isocortex and cerebellum regions to construct training dataset for DSP-Net, and then used the trained network to restore low-resolution, low-SNR images of various types of neurons across the whole brain (3.2× setup, 0.1 ms exposure, 20 fps). In Fig. 2b, we showed the DSP-Net recovered 3D neurons from cortex (b2, middle column), striatum (b3, middle column) and cerebellum (b4, middle column) regions of reconstructed whole brain (b1), and compared the results with raw 3.2× inputs (left columns) as well as HR reference images by the same 12.6× setup used for the acquisition of training data (right columns). We notice that the recovered images by DSP-Net show obviously sharper details of neuron fibers, as compared to the ambiguous LR inputs of 3.2× Bessel sheet, providing near-isotropic 3D resolution as high as that of HR reference of 12.6× Bessel sheet. The linecuts through individual neuronal fibers (Fig. 2b) by each method further confirmed narrower axial and lateral FWHMs for the 3.2× DSP-Net mode, as compared to raw 3.2× Bessel sheet (plots in Fig. 2b). Therefore, DSP-Net enabled Bessel light-sheet microscopy herein shows its strong capability for rapid whole-brain imaging, typically in a few minutes under 3.2× acquisition, at single-neuron resolution, which is close to those by 12.6× HR setup. Its high throughput shown by rapidly achieving high-resolution across large FOV could be over 2-orders-of-magnitude higher than that from conventional Bessel light-sheet (Fig. 2c, Methods). Furthermore, it is shown that the extra robustness of neural network also allows the recovery of various types of neurons in whole brain by merely including a small amount of HR neuron data for training. As a result of high-throughput high-resolution mapping of whole brain, DSP-Net could visualize obviously more nerve fibers in a pyramidal neuron, as compared to the raw 3.2× image, enabling image-based segmentation and tracing of neurons to be performed at system level and in three dimensions. (Fig. 2d).

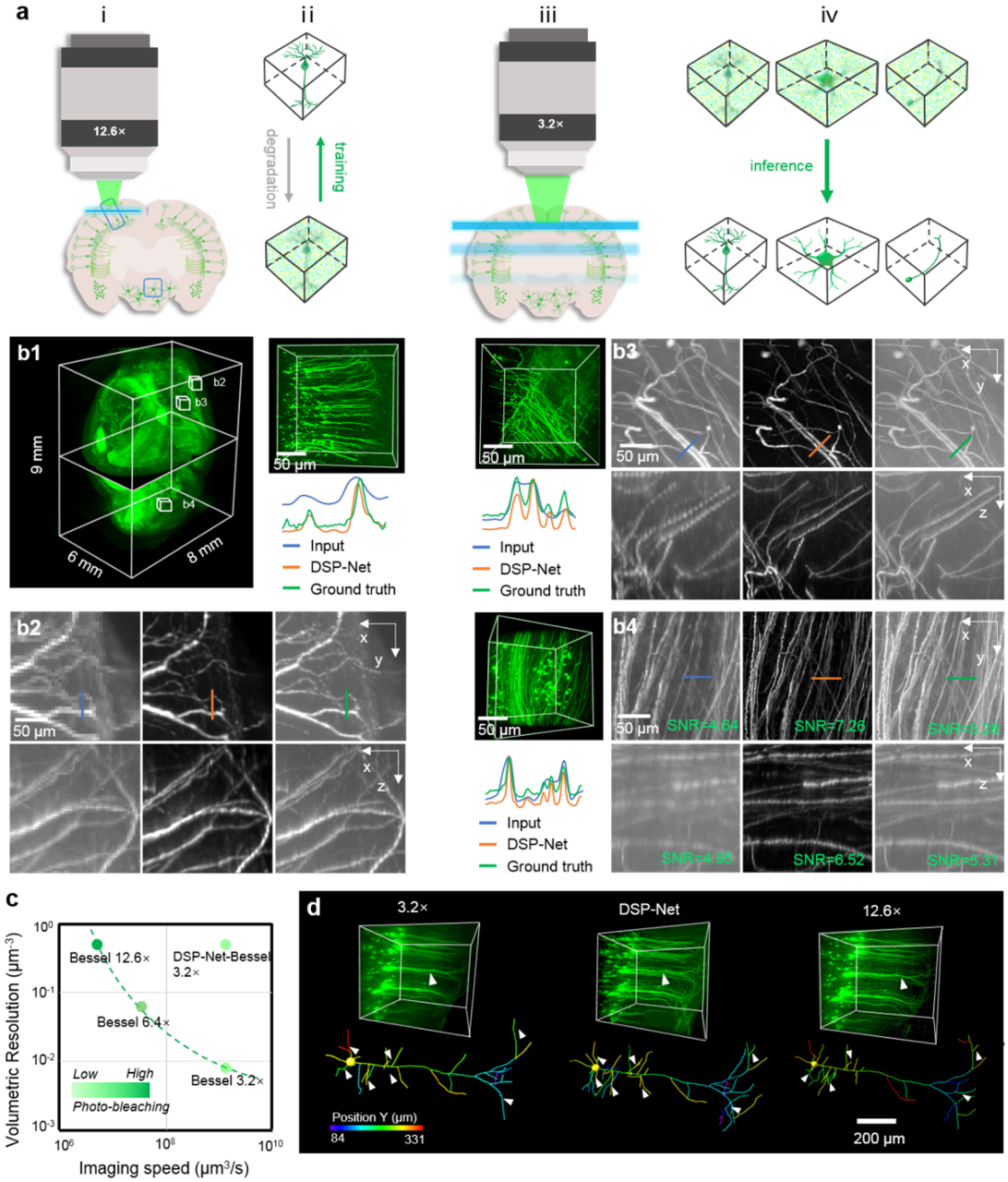

Fig. 2 | DSP-Net for 3D single-cell-resolution imaging of whole mouse brain at minutes time-scale throughput. a, Work flow of whole-brain data acquisition and DSP-Net processing, which includes (i) HR imaging of neurons in a few regions of brain by 12.6× Bessel sheet (0.5 × 0.5 × 1 μm voxel, 0.1 ms exposure, 5 fps). (ii) Network training based on these HR label data. (iii) rapid imaging of neurons in whole brain by 3.2× Bessel sheet (2 × 2 × 4 μm voxel, 0.4 ms exposure, 20 fps), with total 6 tiles acquired in ~6 minutes. (iv) DSP-Net inference to recover a digital whole brain with improved resolution (0.5 × 0.5 × 1 μm voxel) and SNR. **b**, Comparison of three ~200 μm region-of-interests (ROIs) in cortex (**b2**), striatum (**b3**), and cerebellum (**b4**), selected from the 3.2× DSP-Net reconstructed whole brain (**b1**, ~10 × 8 × 5 mm). The magnified views of transverse (*xy*) and coronal (*xz*) planes are shown to compare the detailed neuronal

fibers resolved by raw 3.2× Bessel sheet (left), 3.2× DSP-Net (middle), and 12.6× HR Bessel sheet (right). The line profiles of the same neuron fibers indicate the axial and lateral FWHMs by 3.2× DSP-Net (orange), which are narrower than 3.2× LR inputs (blue), and close to 12.6× HR ground truths (green). Besides the resolution improvement, the low SNR caused by 20 fps acquisition rate at 3.2× (4.64 and 4.95 in b4, left) are also notably increased to a level (7.26 and 6.52 in b4, middle) higher than the SNR of 12.6× HR image with 5 fps acquisition rate (5.28 and 5.31 in b4, right). **c**, The throughput plot comparing the acquisition speed, and volumetric resolution of 3.2× (LR), 6.4× (MR), 12.6× (HR) and DSP-Net enabled 3.2× (SR) Bessel sheet modes. The dash line connecting the data points of LR, MR and HR Bessel sheets represents the throughput limit in conventional light-sheet microscopes. **d**, Improvement for single-neuron tracing brought by DSP-Net. An identical pyramidal neuron in cortex region resolved by 3.2×, 3.2× DSP-Net, and 12.6× was segmented and traced using Imaris.

Rapid 3D super-resolution imaging of single cell with low photo-toxicity.

A key concern in super-resolution fluorescence microscopy is the photon consumption for reconstructing SR images. We demonstrate the ability of DSP-Net by imaging Alexa Fluor 488-labelled microtubules in single *U2OS* cell beyond the diffraction limit, with improved acquisition rate as well as reduced photon budget. We first obtained SR 3D image of microtubules based on super-resolution radial fluctuations (SRRF)[30] computation of thirty diffraction-limited image stacks consecutively acquired by our custom high-magnification Bessel light-sheet microscope (Methods). Due to the noticeable photo-bleaching by repetitive fluorescence emission, the last 30th image stack of the time series dataset obviously shows lowest SNR as well as diffraction-limited resolution, equivalent to the image stack acquired by a lower-intensity laser illumination. Then we used the reconstructed 3D-SRRF image (HR), its corresponding down-sampled version (MR), and diffraction–limited measurement (LR), to construct training dataset for DSP-Net, and then applied the trained network to the SR reconstruction of microtubules in another *U2OS* cell (Methods, Fig. 3a). We demonstrate that for low-resolution, low-SNR input data of microtubules obtained by diffraction-limited optics and low-dose excitation (few photon emission), the DSP-Net enables notable improvement in both spatial resolution and SNR (Fig. 3c). The network was also proven to be highly robust, capable of recovering high-fidelity super-resolution image stack from either 1st or 30th stack of the time-series measurements (Fig. 3c, Supplementary Fig.7). The line profiles of the resolved microtubules by low-SNR, high-SNR inputs, their corresponding DSP-Net outputs, and SRRF result, were plotted in Fig. 3d, to compare the achieved resolution by each method. The quality of DSP-Net recovery is similar with SRRF results (Fig. 3d). We then repetitively imaged a cell 200 times with using low-intensity plane illumination (~0.3 mW at back pupil plane), and compared its photo-bleaching rate with that by regular-intensity plane illumination used for SRRF imaging (~2 mW at back pupil plane). An obviously lower bleaching rate, which is substantially beneficial to live cell imaging, has been shown by low-intensity illumination (Fig. 3e). Meanwhile, our DSP-Net still provided high-quality SR image based on single LR, low-SNR measurement (t=1, 200) (Fig. 3f), with merely consuming ~0.5% photons of SRRF imaging (Fig. 3g). Herein, the combination of our DSP-Net with rapid Bessel-sheet microscopy permits sustained 3D imaging of cells not only beyond the diffraction limit, and also at a high volume rate of ~2 volume s^{-1} (camera rate 150 fps, Fig. 3h).

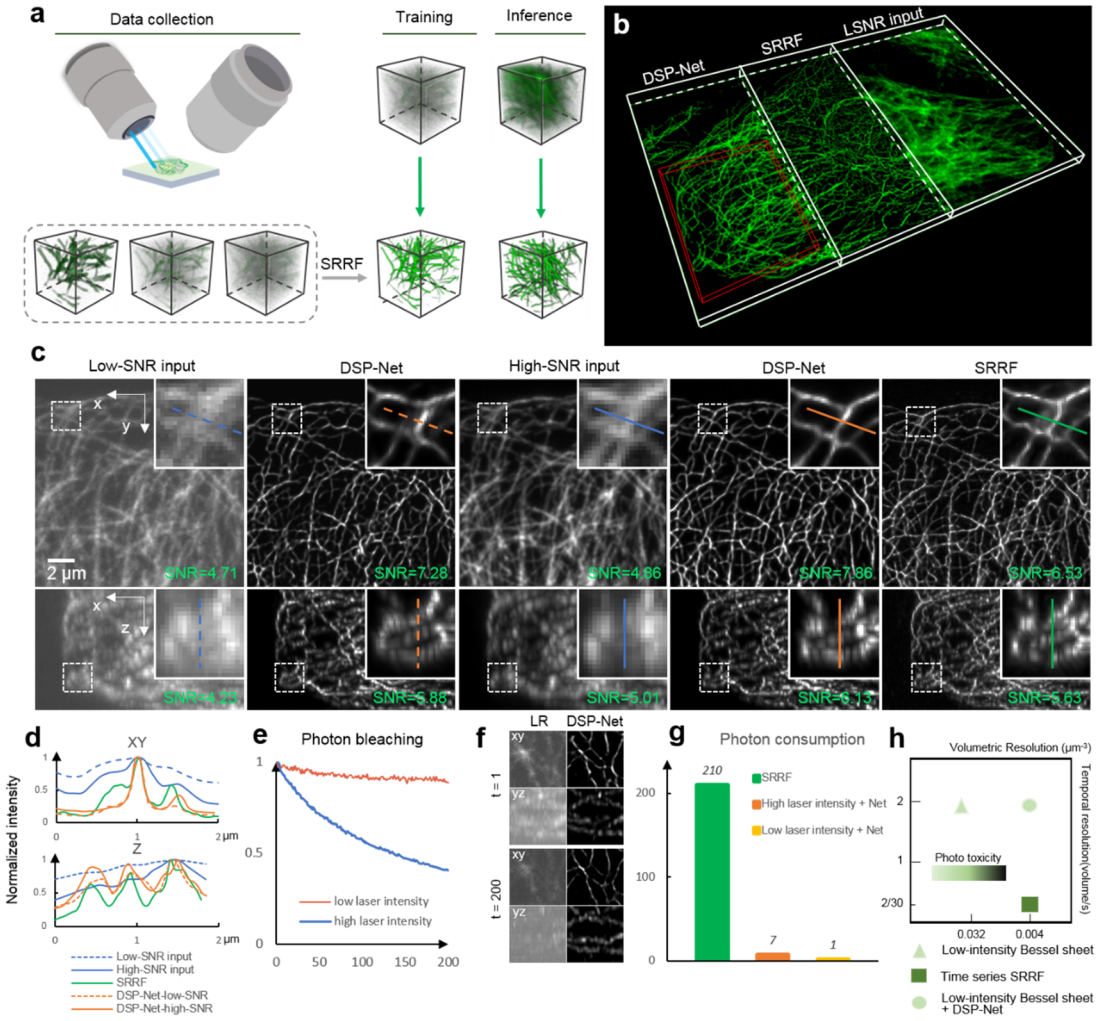

Fig. 3 | DSP-Net for 3D super-resolution imaging of antibody-labeled microtubules (Alexa-488) in single cell with high volume rate and low photon budget/toxicity. a, Work flow of DSP-Net 3D super-resolution imaging including: (i) SRRF reconstruction of microtubules in single cell (Alexa-488) based on 30 diffraction-limited LSFM image stacks consecutively acquired by a high-magnification Bessel light-sheet microscope (60×/1.1 detection objective plus 700 nm thickness Bessel sheet). (ii) DSP network training based on SRRF reconstructed images (HR), 2×-downsampled SRRF images (MR), and the 30th measurement with lowest SNR due to photobleaching (LR). (iii) rapid 3D imaging of microtubules (108 × 108 × 300 nm voxel), at 2 Hz volume rate. The obtained diffraction-limited, low-SNR image stack is super-resolved by trained DSP-Net instantly. **b**, The reconstructed volume rendering of microtubules in a fixed U2OS cell, comparing the overall visual effects by diffraction-limited Bessel sheet, DSP-Net super resolution and SRRF. **c**, Magnified views in *xy* and *yz* planes of the red box region in **b**. Shown from the left to right are: LR (diffraction-limited), low-SNR microtubules (30th measurement); DSP-Net reconstruction of LR, Low-SNR data; LR, high-SNR microtubules (1st measurement); DSP-Net reconstruction of LR, high-SNR data; and SRRF reconstruction. **d**, Line profiles of resolved microtubules by each methods, indicating their achieved lateral and axial resolutions, which are ~110 and 270 nm for DSP-Net, respectively (FWHMs). **e**, Comparison of photo-bleaching rates of microtubules over 200 volumes acquired with different excitation intensity of 0.3 and 2.1 mW. **f**, LR measurement at t = 1 and t = 200 and the corresponding DSP-Net results. **g**, Comparison of the photon consumptions of different

method for reconstructing super-resolution images. **h**, The spatial-temporal map showing the unwilling trade-off between temporal resolution, spatial resolution, and photo-toxicity (triangle and square data points). DSP-Net breaks this spatial-temporal limit (circular data point).

Instantaneous volumetric imaging of neural activities in freely moving *C. elegans*.

We show DSP-Net-enabled 3D microscopy to be capable of capturing dynamic processes in live model organisms by demonstrating the imaging of neuronal activity in freely moving *C.elegans* (Fig. 4). A microfluidic chip was used to permit *C. elegans* (L4-stage) to move inside a large micro-chamber (3300 × 830 × 55 μm, Fig. 4b) freely. A 4×/0.28 objective combined with a fast scanning Gaussian light sheet (Fig. 4a, Supplementary Fig. 13) was used to rapidly image the activities of motor neurons labeled with GCaMP (strain ZM9128 hpIs595[Pacr-2(s)::GCaMP6(f)::wCherry]) at 30-Hz volume-acquisition rate and across 3300 × 830× 55 μm large field-of-view, totally yielding 900 volumes in 30-second observation (Methods). Due to the use of large-FOV objective together with sparse scanning step (4μm) to achieve a high volume rate of recording, the raw image inevitably suffers from ambiguous cellular resolution(Fig. 4c, left), with signals of the adjacent neurons remaining indiscernible. DSP-Net then provided quantitatively-accurate super-resolution to visualize the neuron calcium signaling at single-cell resolution during fast body movement (Fig. 4c, right, Supplementary Fig. 9). The voxel rate of conventional 4× light-sheet is thus expanded by DSP-Net ~ 60 times (Fig. 4d), thereby achieving single-neuron resolution across mesoscale FOV at 30 Hz volume rate. Furthermore, the non-iterative inference of network could sequentially super-resolve 3D images at a high reconstruction rate of ~10 mega voxels per second, suitable for processing sustained biological dynamics which is computationally challenging to conventional methods. Based on the four-dimensional imaging results with sufficiently high spatiotemporal resolution, we identified single specific A- and B- motor neurons that have been associated with motor-program selection, and mapped their calcium activities over time (Fig. 4e, Supplementary Video 1)[38]. Meanwhile, by applying an automatic segmentation of the worm body contours based on the location and morphology of neuron signals[32], we calculated the worm's curvatures related to its locomotion and behavior, thereby allowing classification of the worm motion into forward, backward, or irregular crawling (Fig. 4f, g). The patterns of transient Ca^{2+} signaling were found to be relevant to the switches from forward to backward crawling of the worm, consistent with previous findings[14, 15] (Fig. 4e, f).

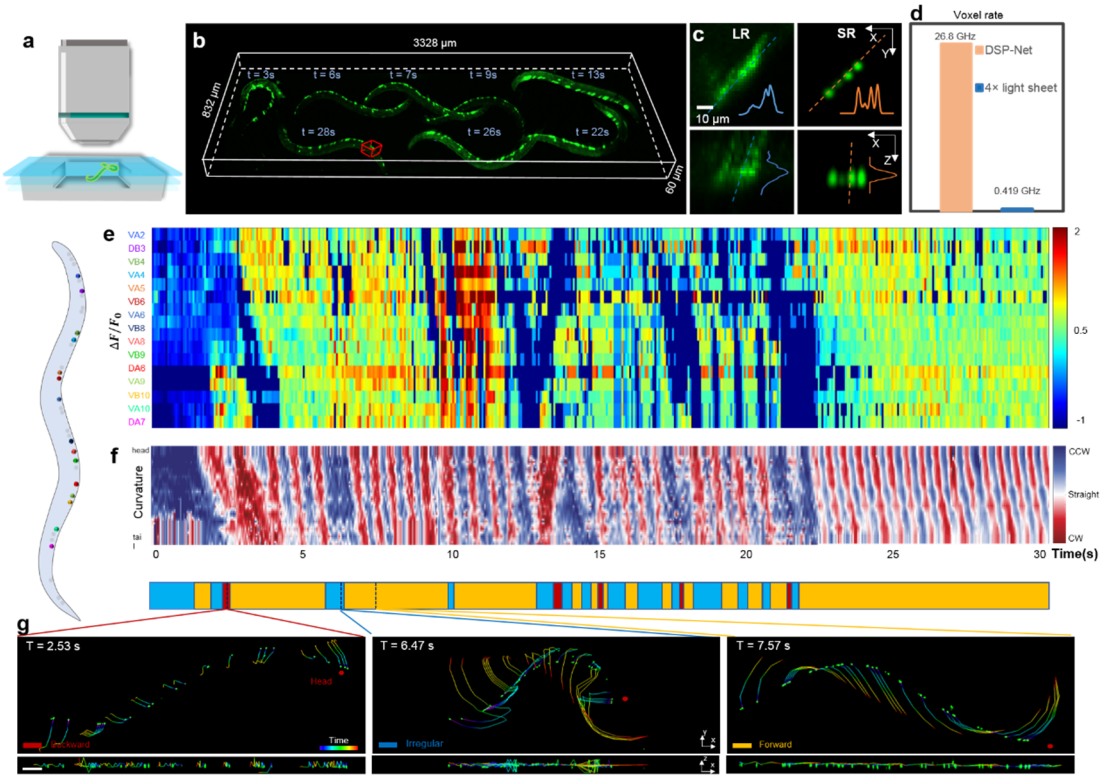

Fig. 4 | Whole-animal Ca^{2+} imaging of freely moving *C. elegans* using DSP-Net. a, Geometry combining fast light-sheet microscopy with microfluidic technique for 3D imaging of motor neuron activities (*hpls595*) of live *C. elegans* freely moving in a micro-chamber (3300 × 830× 55 μm). The acquisition rate is 30 volumes/second. **b**, The 3D trajectory of the *C.elegans* movement during a 30-seconds observation. **c**, Four neurons at one instantaneous volume (red box in **b**) by raw 4×-LSFM and after DSP-Net reconstruction, respectively. It' obvious that DSP-Net restores single-cell resolution and reduces motion artefacts, making possible the quantitative analysis of activities from single neurons. Scale bars, 10μm. **d**, The equivalent imaging throughput of conventional 4× light-sheet and DSP-Net-enabled 4× light-sheet. **e**, Activity of 14 identified motor neurons from the acting worm (left) by tracing the Ca^{2+} transients of individual neuron after being super-resolved by DSP-Net. Each row shows a time-series heat map of an individual neuron with colors indicating percent fluorescence changes ($\Delta F/F_0$). **f**, Kymograms showing the curvature change along the body of acting worm. The body contours of the worm are automatically obtained based on a U-net segmentation. **g**, An ethogram (top panel) based on analysis of curvature change over the entire imaging period classifies the worm behavior (lower panel) into backward (left), irregular (middle) and forward (right) statuses. Three selected volumes with time-coded traces in accordance to the ethogram visualize the forward (left, 300 ms), irregular (middle, 560 ms), and backward (right, 165 ms) crawling tendency of the worm. Scale bars, 50 μm.

Cross-mode 3D super-resolution by DSP-Net.

We directly use the DSP-Nets trained by Bessel light-sheet images of brain neurons, or SRRF Bessel light-sheet images of microtubules in cells, to restore the low-resolution images of the neurons by confocal microscope, or diffraction-limited wide-field image of the microtubules by SIM microscope, respectively. As shown in Fig. 5a, 3D images stack of a 100-μm thick mouse brain slice was captured by a confocal microscope (Nikon Ni-E, CFI LWD 16×/0.8 W objective). The artificially degraded version (down sampled 5 times in all three dimensions to

match the voxel size of the LSFM training data) of the confocal stack, considered as the LR inputs (Fig. 5b, LR), was then sent to the abovementioned DSP-Net trained with Bessel light-sheet data, for cross-mode (CM) super-resolution inference. The result, termed as CM-DSP-Net in Fig. 5b, was further compared with the recovered result by another DSP-Net regularly trained with confocal data (DSP-Net), and the 12.6× (down-sampled from 16×) confocal measurement (Fig. 5b, HR). For super-resolution 3D imaging of microtubules in an *U2OS* cell (Fig. 5e), the LR input was a diffraction-limited 3D wide-field image obtained by simply averaging the fifteen SIM exposures at each *z*-plane (Nikon N-SIM, CFI Apo TIRF 100×/1.49 oil, Fig. 5f, input). The abovementioned DSP-Net trained using Bessel light-sheet data was then applied to this wide-field image to recover 3D microtubules beyond diffraction limit (Fig. 5f, CM-DSP-Net). The result was finally compared with super-resolution image obtained by 3D-SIM reconstruction using the 15 patterned exposures (Fig. 5f, 3D-SIM). For both case, these cross-mode DSP-Net could provide extraordinary enhancement to the heterogeneous inputs. By qualitatively analyzing the reconstruction quality, CM-DSP-Net results show spatial resolutions similar with those by regular DSP-Net. At the same time, the high resolution-scaled Pearson coefficient (RSP) as well as low resolution-scaled error (RSE) also indicate the sufficient accuracy of the CM-DSP-Net recovery.

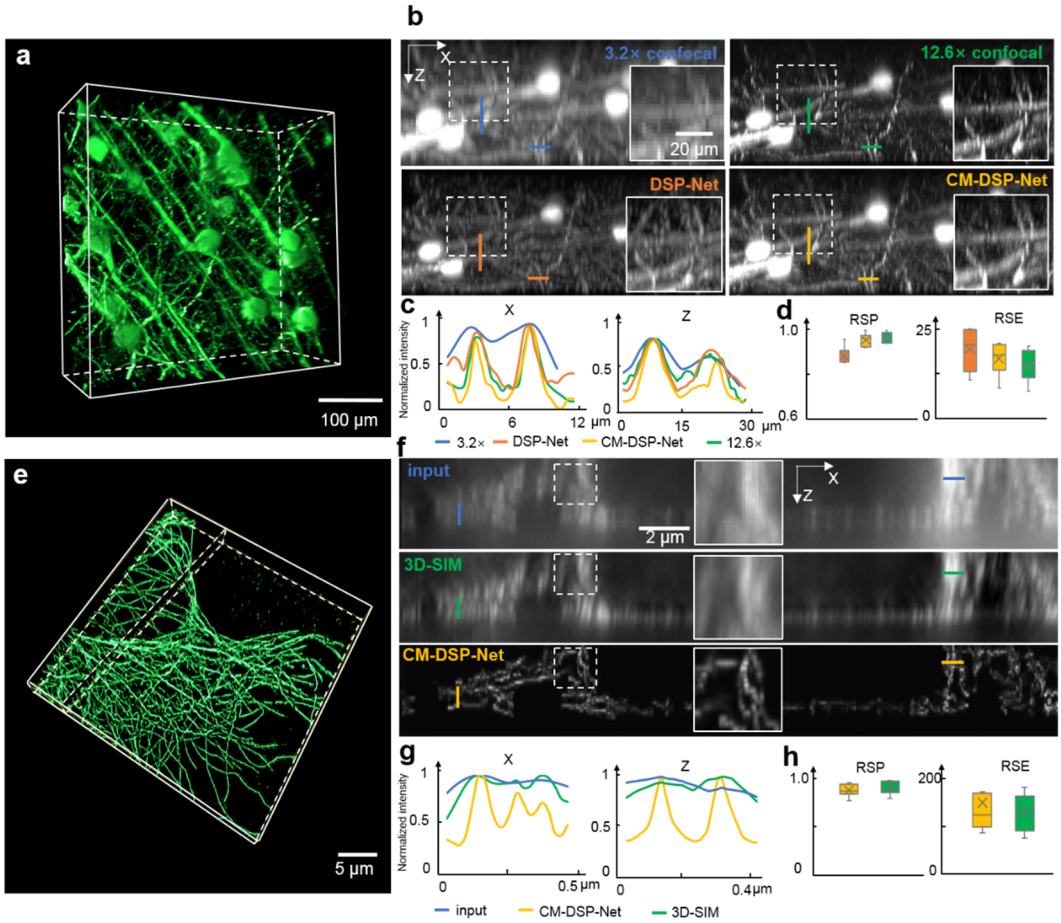

Fig. 5 | Cross-mode 3D super-resolution by DSP-Net. a, 3D confocal image stack of neurons (GFP) in 100-µm thick mouse brain slice recovered by cross-mode DSP-Net (CM-DSP-Net), which is trained with image data obtained by light-sheet microscope. **b**, *xz* projections of the synthetic 3.2× confocal image (blue), 12.6× confocal image (green), super-resolution of 3.2× confocal image by standard DSP-Net (yellow), and CM-DSP-Net (red), respectively. **c**, Normalized intensity plots of the linecuts through neuronal fibers along lateral (*x*) and axial (*z*) directions in **b**. **d**, Comparison of RSP and RSE values of the DSP-Net, CM-DSP-Net and 12.6× HR results calculated with respect to the 3.2× LR images. **e**, CM-DSP-Net reconstruction of the diffraction-limited 3D wide-field images of antibody-labelled microtubules in *U2OS* cell (Alexa-488, 100×/1.49 NA objective). The network was trained with super-resolution images by 3D-SRRF enabled Bessel light-sheet microscopy. **f**, *xz* projections of the raw measurement (blue), 3D-SIM reconstruction (green) and the CM-DSP-Net reconstruction (red), respectively. **g**, Normalized intensity plots of the linecuts through microtubules along lateral (*x*) and axial (*z*) directions in **f**. **h**, Comparison of RSP and RSE values of the CM-DSP-Net and 3D-SIM results calculated with respect to the diffraction-limited microtubule images.

Cross-sample 3D super-resolution by DSP-Net.

In addition to cross-mode capability, we further demonstrate that the DSP-Net also enables cross-sample (CS) super-resolution imaging of brain vessels by a brain neuron-trained network at single cell resolution, and endoplasmic reticulum by a microtubules-trained network beyond the diffraction limit. For both cases, our network shows intrinsically remarkable improvements for new types of signals (Fig. 6a,b,e,f). Furthermore, the quantitative analyses on achieved

resolution and reconstruction errors both verify the reliable cross-sample super-resolution recovery by DSP-Net. It's noted that the robust cross-mode, cross-sample capabilities found in our 3D DSP-Net were also consistent with previous findings from 2D network super-resolution[26].

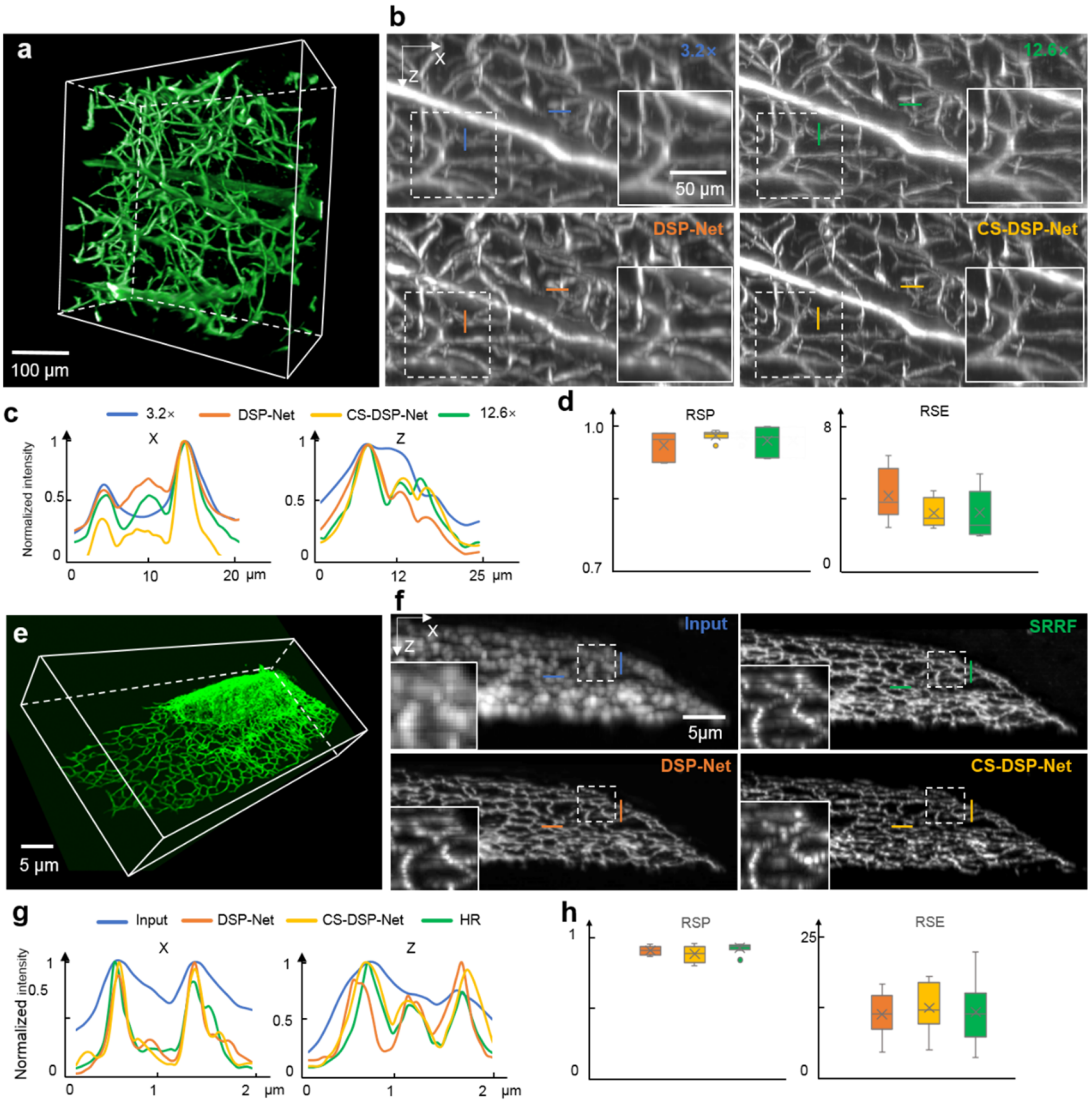

Fig. 6 | Cross-sample 3D super-resolution by DSP-Net. a, 3D Bessel sheet image stack of blood vessels (Alexa-647) in mouse brain recovered by cross-sample DSP-Net (CS-DSP-Net), which is trained with 3D neuron images by the same Bessel light-sheet microscope. **b**, *xz* projections of the 3.2× Bessel sheet image (blue),12.6× Bessel sheet image (green), super-resolution of 3.2× image by standard DSP-Net (yellow), and CS-DSP-Net (red), respectively. **c**, Normalized intensity plots of the linecuts through vessels along lateral (*x*) and axial (*z*) directions in **b**. **d**, Comparison of RSP and RSE values of the DSP-Net, CS-DSP-Net and 12.6× HR results calculated with respect to the 3.2× LR images. **e**, CS-DSP-Net reconstruction of the diffraction-limited 3D wide-field images of antibody-labelled endoplasmic reticulum (the antibody labeled EGFP) in *U2OS* cell (100×/1.49 NA objective). The network was trained by microtubule images. **f**, *xz* projections of the raw diffraction-limited image of endoplasmic reticulum (blue), SRRF reconstruction (green), DSP-Net reconstruction (yellow) and CS-DSP-Net reconstruction (red), respectively. **g**, Normalized intensity plots of the linecuts through endoplasmic reticulum along lateral (*x*) and axial (*z*) directions in **f**. **h**, Comparison of RSP and RSE values of the DSP-Net, CS-DSP-Net and 3D-SRRF results

calculated with respect to the diffraction-limited endoplasmic reticulum images.

Discussion

In summary, the marriage of DSP-Net with cutting-edge 3D microscopy enables high spatiotemporal resolution volumetric imaging of various types of specimens far beyond the optical throughput limit. Highly inclined, LSFM has been a powerful tool that can image samples in three dimensions at high speed and low phototocixity, although its throughput remains limited by the system optics that shows limited NA as well as certain degree of pixelization. The combination of our all-3D DSP-Net in this case greatly push its limit for achieving ultra-high optical throughput which is unmet before. Also unlike current network super-resolution method directly pairing the high-resolution label data with severely degraded low-resolution measurements, DSP-Net defines the SISR problem into two sub-issues of de-blurring and de-pixelization, which fundamentally invert the microscopy imaging process that transforms a high-resolution target into a blurred-and-pixelated image, and solve them successively. This step-by-step strategy permits the network to extract more abundant features to create strong correlations between HR targets, MR aliased images and degraded LR measurements, capable of creating a more accurate mapping of LR measurements to HR targets, which might be decimated in conventional one-stage deep learning networks. The experimental results have sufficiently verified that our DSP-Net can recover the 3D fluorescence image with more details super-resolved as compared to other state-of-art super-resolution networks. Benefitting from the mutual-feedback mechanism of the dual-stage training, DSP-Net also requires less epochs and similar time consumption to reach convergence as compared to conventional super-resolution networks, showing very high training efficiency. The image improvement by current DSP-Net is ~4× in each dimension. This value could be further increased, to achieve larger super-resolution factor, by possibly combing progressive growing methods[33] if the equally accumulated artefacts and hallucinations could be suppressed at the same time. Thus, to implement this training strategy, a larger number of 3D images captured at various magnification is necessary to establish reliable targets for every growth stage of the progressive training.

The increased spatiotemporal information provided by DSP-Net LSFM, together with its non-iterative rapid image reconstruction, are crucial for high-throughput imaging of large-size samples or sustained recording of biological dynamics at high resolution. These advances have been demonstrated by notably increasing the throughput for 3D high-resolution mapping of clarified large organs, breaking the diffraction limit for imaging intracellular organelles with faster, lower photo-toxicity measurement, and significantly improving the temporal resolution for capturing instantaneous biological process. Besides the combination with LSFM, DSP-Net has also shown its ability in cross- mode and cross-sample recovery, indicating the extra flexibility of DSP-Net which allows a network pre-trained with a relatively small amount of desired HR data by a relatively specialized system, e.g. fine neurons by Bessel light-sheet, could be extensively applied to various types of samples showing similar structural characteristics, e.g., line-like vessels, microtubules, that are obtained by other imaging modes, such as commercialized confocal and wide-field microscopes, which might be more readily accessible. Therefore, DSP-Net is believed to be transformative, and could potentially bring new insight

for current computation-based imaging methods to be widely applied for biomedical applications that are unceasingly requiring ever higher spatiotemporal performance.

References

1 Pawley, J. B. Handbook of Biological Confocal Microscopy. *Journal of Biomedical Optics* **25**, 029902 (1995).

2 Denk, W., ., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73-76 (1990).

3 Jan, H., Jim, S., Filippo, D. B., Joachim, W. & Stelzer, E. H. K. Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* **305**, 1007-1009 (2004).

4 Davidovits, P. & Egger, M. D. Scanning Laser Microscope. *Nature* **223**, 831-831 (1969).

5 Helmchen, F. & Denk, W. Deep tissue two-photon microscopy. *Nature Methods* **2**, 932-940, doi:10.1038/nmeth818 (2005).

6 Keller, P. J., Schmidt, A. D., Wittbrodt, J. & Stelzer, E. H. K. Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy. *Science* **322**, 1065-1069 (2008).

7 Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods* **10**, 413 (2013).

8 Verveer, P. J. *et al.* High-resolution three-dimensional imaging of large specimens with light sheet-based microscopy. *Nature Methods* **4**, 311-313 (2007).

9 Huisken, J., Swoger, J., Bene, F. D. & Wittbrodt, J. Optical Sectioning Deep inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **305**, 1007-1009 (2004).

10 Ritter, J. G., Veith, R., Veenendaal, A., Siebrasse, J. P. & Kubitscheck, U. Light Sheet Microscopy for Single Molecule Tracking in Living Tissue. *Plos One* **5**, e11639 (2010).

11 Tomer, R., Khairy, K., Amat, F. & Keller, P. J. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nature Methods* **9**, 755 (2012).

12 Keller, P. J. *et al.* Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nature Methods* **7**, 637 (2010).

13 G, Z., R, H. & C, Y. Wide-field, high-resolution Fourier ptychographic microscopy. *Nature Photonics* **7**, 739-745 (2013).

14 Ou, X., Zheng, G. & Yang, C. Embedded pupil function recovery for Fourier ptychographic microscopy. *Optics Express* **22**, 4960-4972 (2014).

15 Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy* **198**, 82-87 (2010).

16 Gustafsson, M. G. L. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13081-13086 (2005).

17 Hell, S. W. & Wichmann, J., . Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters* **19**, 780-782 (1994).

18 Rust, M. J., Bates, M. & Zhuang, X. Stochastic optical reconstruction microscopy (STORM) provides sub-diffraction-limit image resolution. *Nature Methods* **3**, 793 (2006).

19 Eric, B. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645 (2006).

20 Freeman, W. T., Jones, T. R. & Pasztor, E. C. Example-based super-resolution. *Computer Graphics & Applications IEEE* **22**, 56-65 (2002).

21 Duchon, C. E. Lanczos Filtering in One and Two Dimensions. *J.appl.meteor* **18**, 1016-1022 (1979).

22 Li, X., . & Orchard, M. T. New edge-directed interpolation. *IEEE Trans Image Process* **10**, 1521-1527 (2001).

23 Kwang In, K. & Younghee, K. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **32**, 1127 (2010).

24 Weisheng, D., Lei, Z., Guangming, S. & Xiaolin, W. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing* **20**, 1838-1857 (2011).

25 Rivenson, Y. *et al.* Deep learning microscopy. *Optica* **4**, 1437-1443, doi:10.1364/OPTICA.4.001437 (2017).

26 Wang, H. *et al.* Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103-110, doi:10.1038/s41592-018-0239-0 (2019).

27 Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090-1097, doi:10.1038/s41592-018-0216-7 (2018).

28 Zhang, H. *et al.* High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. *Biomedical Optics Express* **10**, 1044-1063, doi:10.1364/BOE.10.001044 (2019).

29 Ozcan, A. *et al.* Deep learning microscopy. *Optica* **4**, 1437- (2017).

30 Gustafsson, N. *et al.* Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications* **7**, 12471 (2016).

31 Zhang, Y., Tian, Y., Kong, Y., Zhong, B. & Fu, Y. Residual Dense Network for Image Restoration. *arXiv e-prints* (2018). <https://ui.adsabs.harvard.edu/abs/2018arXiv181210477Z>.

32 Wang, Z. *et al.* Network-based instantaneous recording and video-rate reconstruction of 4D biological dynamics. *bioRxiv*, 432807, doi:10.1101/432807 (2019).

33 Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv e-prints* (2017). <https://ui.adsabs.harvard.edu/abs/2017arXiv171010196K>.

34 Haris, M., Shakhnarovich, G. & Ukita, N. Deep Back-Projection Networks For Super-Resolution. *arXiv e-prints* (2018). <https://ui.adsabs.harvard.edu/abs/2018arXiv180302735H>.

35 Shi, W. *et al.* in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1874-1883.

36 Dong, H. *et al.* TensorLayer: A Versatile Library for Efficient Deep Learning Development. *arXiv e-prints* (2017). <https://ui.adsabs.harvard.edu/abs/2017arXiv170708551D>.

37 Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods*, S1046202316303346.

38 Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods* **16**, 67-70, doi:10.1038/s41592-018-0261-2 (2019).

39 Restif, C. *et al.* CeleST: Computer Vision Software for Quantitative Analysis of C. elegans Swim Behavior Reveals Novel Features of Locomotion. *Plos Computational Biology* **10**, e1003702.

40 Culley, S. *et al.* Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature Methods* **15**, 263, doi:10.1038/nmeth.4605

https://www.nature.com/articles/nmeth.4605#supplementary-information (2018).

Acknowledgements

We thank Di Jin at MIT, Shaoqun Zeng and Xiang Bai at Huazhong University of Science &Technology for their helpful advices on the project design. We thank Yanyi Huang at Peking University and Jianbin Wang at Tsinghua University for their discussions and comments on the manuscript. This work was supported by the following grants: National Key R&D program of China (2017YFA0700500 to P.F. and D.Z.), National Natural Science Foundation of China (21874052 to P.F., 31671052 to S.G., 61860206009 for D.Z., 31770924 for Y.Z.), and the Junior Thousand Talents Program of China (P.F. and S.G.). Confocal experiment was done at the Wuhan National Laboratory for Optoelectronics at Huazhong University of Science and Technology and SIM experiment was done at the State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University.

Author contributions

P.F. and H.Z. conceived the idea. P.F., S.G., Y.Z., and D.Z. oversaw the project. C.F. and Y.Z. developed the optical setups. H.Z. developed the programs. Y.L., P.W, Y.L., T.Y., and M.Z. conducted the sample preparation experiments. C.F., Y.Z., and G.L. conducted the microfluidic and optical experiments. H.Z., C.F., Y.Z., G.L., and P.F. processed the images. H.Z., C.F., Y.Z., and P.F. analyzed the data and wrote the paper.

Competing interests

The authors declare no conflicts of interest in this article.

Methods

1. Light-sheet microscopy setups

1.1 High-magnification Bessel-sheet microscope. Our high-magnification Bessel-sheet microscope for single-cell imaging was built based on an upright microscope (BX51, Olympus). Fiber-coupled lasers with four excitation wavelengths (405, 488, 589, 637 nm) were first collimated by a collimator (Thorlabs, F280FC-A) and then sent to an acousto-optic tunable filter (AOTF) to control their intensities. The 1st-order diffraction beam after AOTF was selected and expanded (Thorlabs, CBE05-A) to 10 mm diameter and then sent to an axicon (Thorlabs, AX252-A) to generate annular Bessel beam at the front focal plane of the achromatic lens (Thorlabs, AC254-075-A-ML). The interference pattern was focused onto a galvanometer (Thorlabs, GVS112), which was combined with a scan-lens (Thorlabs, CLS-SL) to scan the Bessel beam into a sheet. Finally, the Bessel sheet was compressed for 20 folds through a 4-f system containing a tube lens (Thorlabs, TTL-200) and an objective lens (Thorlabs, MY20×-803) to obtain the final illuminating Bessel-sheet with FWHM thickness of ~700 nm. In the detection path of upright microscope, a water dipping objective (Olympus, LUMFLN 60XW 60×, 1.1 NA) was used to collect the fluorescence emission from the fluorophore-tagged U2OS cell. To eliminate the emission excited by the side lobes of the Bessel beam, the rolling shutter of the detection sCMOS camera (Hamamatsu, ORCA Flash 4.0 V3) was synchronized with the beam scanning to form an electronic confocal slit so that only the fluorescence excited by the central maximum of the Bessel beam could be detected. Finally, we obtained a high-quality Bessel sheet illumination with ~700 nm uniform thickness covering the entire illuminated range of the U2OS cell. For 3D imaging, the cells attached to the cover glass was mounted onto a customized holder and vertically (z) scanned by a piezoelectric stage (P-737, Physik Instrumente). The synchronized imaging process was controlled by a self-built LABVIEW program. Supplementary Fig. 12 shows the optical layout and photographs of the system.

1.2 Macro-view Bessel-sheet microscope. Our custom macro-view Bessel-sheet microscope for mouse brain imaging was based on a tunable dual-side plane illumination and a zoomable detection. With splitting the collimated beam into 2 paths by a 50/50 beam splitter (CCM1-BS013/M, Thorlabs), two synchronized scanning Bessel sheets were generated at both sides of the sample using the similar way described above. A pair of long working distance objectives (Mitutoyo, 10×/0.28 or 5×/0.14) combined with tube lenses (ITL200, Nikon) were used to generate final dual-side Bessel sheets with illumination range of 1 or 4 mm (~1.3 or 2.7 μm thickness) adapted to the detection FOV of 12.6× or 3.2× (1 or 4 μm lateral resolution) from a zoomable upright microscope (MVX10, Olympus). During image acquisition, the Bessel sheets were synchronized with the rolling shutter of sCMOS camera (Hamamatsu ORCA-Flash4.0 V2) at 20 fps for recording 3.2× LR images, or alternatively at 5 fps for recording 12.6× HR images. A motorized *xyz* stage (Z825B, Thorlabs combined with SST59D3306, Huatian) moved the sample in three dimensions to enable image stitching for the entire brain. A LabVIEW (National Instruments) program was developed to synchronize the parts to automatically implement the beam scanning, image acquisition and tile-stitching in the proper order. Supplementary Fig. 11 shows the optical layout and photographs of the system.

1.3 Gaussian light-sheet microscope. Besides the Bessel-sheets, a SPIM path was also built

604 for high-speed imaging of live *C. elegans* acting in microfluidic chamber. The 3.3 mm
605 collimated lasers were first sent to a pair of cylindrical lenses (Thorlabs, LJ1695RM-A and
606 LJ1567RM-A) to expand the beam height to ~4 mm, which is sufficient to cover the length the
607 chamber (3 mm). Then through another focusing cylindrical lens (Thorlabs, LJ1703RM-A), a
608 Gaussian light-sheet with FWHM thickness of ~10 μm and confocal range of ~900 μm
609 matched to the width of the chamber (0.8 mm) was finally generate to illuminate the *C.elegans*.
610 The sample together with the microfluidic chip were moved by a motorized stage along z
611 direction (Z825B, Thorlabs) for enabling fast 3D imaging.

612 **2. Dual-stage-processed network (DSP-Net)**

613 **2.1 Network architecture.** Our DSP-Net consists of a resolver that enhances the details and
614 reduces noises for the low-quality LR images, and an interpolator that de-pixelates the
615 intermediate output of the resolver by further extracting and permuting features. We adopted
616 the deep back-projection network (DBPN)[34] as the resolver (Supplementary Fig. 1), for its
617 outstanding feature extracting ability brought by alternate up-sampling and down-sampling
618 operations. We modified all the layers to allow the processing of 3D input, and removed the
619 last deconvolution layer to keep the image size of output unchanged. For the interpolator
620 (Supplementary Fig. 2), residual dense network (RDN)[31] was used with structures also modified
621 for 3D computation. To up-sample the 3D images in RDN interpolator, we further introduced a
622 sub-voxel convolution layer (Supplementary Fig. 2), which was derived from previously
623 reported efficient ESPCNN[35]. The DSP-Net implemented in Python using Tensoflow and
624 Tensorlayer[36]. The custom code for training and inference together with the example data can
625 be found at https://github.com/xinDW/DVSR.

626 **2.2 Network training and inference procedure.** The DSP-Net training was based on semi-
627 synthetic dataset. For mouse brain, the HR images of mouse brain were first experimentally
628 obtained by 12.6× Bessel-sheet setup while the synthetic LR images were artificially generated
629 by degrading the HR images using a model which simulates the optical blurring of microscope
630 and pixelization of camera (details see Supplementary Note 1). For single-cell, the LR images
631 were simply the diffraction-limited measurements under 60× Bessel-sheet. The corresponding
632 HR images were generated by SRRF[30] computation based on 30 stacks of diffraction-limited
633 measurements. For live *C.elegans*, the synthetic HR and LR 3D images of microbeads were
634 generated to represent the point-like neuronal signals. All abovementioned experiment used a
635 directly down-sampled version of the HR data as the MR data, which still contains aliased high-
636 resolution details that might be completely lost in the LR data. At the training stage, LR data
637 were imported into the resolver of the DSP-Net, and sequentially convolved with the filters of
638 each convolutional layer. The pixel-wise mean-square-errors (MSE) between the resolver's
639 outputs and the MR data, termed as the resolving loss, works as cost function of the parameters
640 in each filter. A gradient descent method was used to efficiently minimize the resolving loss by
641 optimizing the distribution of these parameters. The converged outputs of the resolver continues
642 to flow into the interpolator, where the image size increases via extending and re-arranging
643 feature maps extracted by the filters in the convolutional layers, and the following sub-voxel
644 convolutional layers (Supplementary Fig. 2). At the 2nd stage, the dual sub-nets are mutual-
645 feedback. The minimization of MSE between the converged outputs of interpolator and the HR
646 ground truths could lead to the optimization of the parameters not only in the 2nd-stage

interpolator, but also those in the 1st-stage resolver. Finally, the well-trained DSP-Net with optimized parameters in both sub-nets was used to deduce super-resolution 3D image stacks from the LR 3D inputs.

At the inference stage, the entire LR image stack were automatically subdivided into multiple blocks for being processed block by block (typically 100 × 100 × 100 voxels) , owing to the memory limit of the GPUs. The super-resolved patches were automatically stitched together into a complete SR image stack automatically. The data implementation details were elucidated in Supplementary Table 1.

3. Mouse brain experiment

3.1 Imaging. Neuron-tagged transgenetic whole mouse brain (Thy1-GFP-M) and vessel-labelled whole mouse brain (Alexa-647) were imaged using our custom macro-view Bessel-sheet microscope. For obtaining adequate HR image stacks for network training, we first imaged several brain sub-regions, such as isocortex, striatum, hippocampus, and cerebellum, using 12.6× detection combined with 1.3 μm-thickness Bessel-sheet illumination. The acquired 3D image of each sub-region encompassed 2048×2048×500 voxels across a 1.1 × 1.1 × 0.5 mm volume size. The total imaging time is ~30 minutes for 20 regions with 5 fps acquisition rate. 10 patches of HR image (2048 × 2048 × 500 for each) were pre-processed to generate the HR-MR-LR pairs for network training.

Then the whole-brain imaging was implemented using 3.2× detection plus 2.7 μm-thickness Bessel-sheet illumination. We sequentially imaged 6 blocks (2 × 3) of the brain with each one containing 2048 × 2048 × 1250 voxels across a 4 × 4 × 5 mm volume size. With such a low-magnification setup, the total imaging time for a whole brain is down to ~6 minutes with 20 fps acquisition rate. The 6 blocks were then stitched into an LR whole brain using Grid/collection stitching plugin of ImageJ. The stitched LR whole brain was taken as input for in-parallel super-resolution inference by the trained DSP-Net. Finally, a super-resolved whole brain output encompassing 3.2 teravoxels across 10 × 8 × 5 mm^3 volume size was obtained.

3.2 speed and throughput calculation. The speed and throughput for imaging whole mouse brain is calculated by:

$$Imaging\ speed = \frac{Imaging\ volume}{Acquisition\ time}\ (mm^3/second) \quad (1)$$

and

$$Imaging\ throughput = \frac{Imaging\ volume}{(Acquisition\ time)*(volumetric\ resolution)}\ (voxels/second) \quad (2)$$

We compared the achieved optical throughput for imaging a whole mouse brain by 3.2×, 6.4×, 12.6×, and DSP-Net-enabled 3.2× Bessel-sheets. To maintain an SNR high enough for distinguishing fine neuronal details, such as apical dendrites, the acquisition rate was 20 fps for 3.2× imaging, 10 fps for 6.4×, and 5 fps for 12.6× imaging. Under 3.2× setup, 6 tiles (2 × 3) with each one containing 2048 × 2048 × 1250 voxels across a 4 × 4 × 5 mm volume size were stitched together to form a whole brain dataset which finally encompassed 3891 × 5734 × 1250 voxels across 10 × 8 × 5 mm^3 volume size. The total imaging time was around 6 minutes. For

6.4× and 12.6×, we stitched 20 (4 × 5) and 96 (8 × 12) tiles, respectively, with each one containing 7577 × 9420 × 2500 and 14950 × 22323 × 5000 voxels, respectively. The total imaging time was around 83 minutes for 6.4× acquisition and 1600 minutes for 12.6× acquisition. After the 4× DSP-Net super-resolution applied to 3.2× Bessel sheet, the achieved volumetric resolution was around 64 times higher than that from regular 3.2× Bessel-sheet, finally yielding ~5 gigavoxels/s throughput which is over 200-times higher than that by 12.6× Bessel-sheet.

3.3 Neuron tracing. We chose a 0.5 × 0.4 × 0.2 mm^3 volume in isocortex region to demonstrate the resolution enhancement from our DSP-Net can substantially benefit the neuron identification/tracing with improved accuracy. The data was converted and opened in Imaris software (Bitplane, USA). The Autopath mode of the Filament module was used to trace the same single neuron in LR, SR and HR images semi-automatically. We first assigned one point on the neuron to initiate the tracing. Then Imaris automatically calculated the pathway in accordance with the image data, reconstructed the 3D morphology and linked it with the previous part. This automatic procedure would repeat several times with manual adjustment involved until the neuron was finally correctly traced. Because the performance of the tracing highly relies on the resolution and SNR of the source image, the DSP-Net result allows obviously more accurate neuron tracing with finer fibers identified, as shown in Fig. 2d.

4. *U2OS* cell experiment

4.1 Imaging. The antibody-labeled microtubules and antibody-labeled endoplasmic reticulum in fixed *U2OS* cells were imaged by our custom high-magnification Bessel-sheet microscope. A thin Bessel-sheet provided high-axial-resolution optical sectioning (~700 nm) of the cell, while a 60×/1.1 water dipping objective collected the fluorescence signals with a lateral resolution of ~270 nm (pixel size 108 nm). During image acquisition, we repetitively scanned the cell using a piezo actuator (step size 300 nm) with excitation intensity ~2 mW at the pupil plane of illuminating objective, obtaining 30 consecutive image stacks (2048×512×70 voxels for each stack) to reconstruct a super-resolution 3D image with a 3D SRRF computation procedure[30]. Then we used the SRRF results, 2× down-sampled SRRF results, and the diffraction-limited images of the 30th acquisition as the HR, MR and LR data, respectively, for network training. 686 groups of such HR-MR-LR patch pairs with size 96×96×32 for each were included. For network validation, the same 60× Bessel sheet was applied to another *U2OS* cell with imposing regular-intensity excitation (~2 mW), and low-intensity excitation (~0.3 mW), to obtain LR, high-SNR and LR, low-SNR cell images, respectively.

4.2 quantifications of spatiotemporal resolution and photon consumption. The fixed cell were imaged with step size of 0.3 μm. Typically, when imaging adherent cells in a light-sheet fluorescence microscopy format, as their elongated morphologies require □200 image planes per image volume. Limited by the response time of piezo, volume acquisition rates was up to 2 Hz, which results in over 10s for obtaining a super-resolution image stack, compromising the temporal resolution to over 10 seconds for 1 volume. At the same time, the repetitive image also put heavier photon burden to the sample. Using our DSP-Net, over 2-fold higher spatial resolution was achieved in all three dimensions as compared to the diffraction-limited Bessel imaging, with merely imaging one volume of the cell. From another point of view, it could

surpass the diffraction limit with realizing 30-times higher temporal resolution, as compared to time series SRRF.

To ensure the reconstruction quality of SRRF, relatively high-intensity excitation are always required to collect the 30 image stacks. In contrast, DSP-Net permits 7-fold lower-intensity excitation for the cell, owing to the strong capability of DSP-Net for recovering low-SNR raw image. Combined with the super resolution based on single measurement for DSP-Net rather than 30 measurements for SRRF, the total photon consumption for low laser intensity is over 2-orders-of-magnitude lower than the time series SRRF.

5. *C.elegans* experiment

5.1 Imaging. *C. elegans* were cultured on standard Nematode Growth Medium (NGM) plates seeded with OP50 and maintained at 22 °C incubators. The strain ZM9128 hpIs595[Pacr-2(s)::GCaMP6(f)::wCherry] of *C. elegans*, which expressed the calcium indicators to A-and B-class motor neurons, was used to detect the Ca^{2+} signaling in moving animals. For imaging neural activities, late L4 stage worms were loaded into a microfluidic chip with size of 3300 × 830 × 55 μm, allowing the worm to freely move within the FOV of a 4× objective (Olympus, XLFluor 4×/340, 0.28 NA, WD = 29.5 mm). During image acquisition, a 10 μm-thick Gaussian light-sheet illuminated the worm in the chamber from the side of the chip. The chip was mounted to a z-piezo stage, for being scanned at high speed. The sCMOS camera captured the plane images at a high speed of 400 fps (2048× 512 pixel) and with a step size of 4.5 μm. The system as such permitted large-scale volumetric imaging (2048 × 512 × 13 voxel) at 30 volume per second (VPS) acquisition rate across 3300 × 830 × 55 μm FOV. Due to the low phototoxicity provided by light-sheet illumination, we could continuously image the moving *C.elegans* for 30 seconds, recording the various worm behaviors such as forward, irregular and backward crawling, in the obtained 900 consecutive image stacks. The sequential 3D image stacks were then sent to the trained DSP-Net for be processed at high processing rate of 10-mega voxels per second by using a single RTX 2080 Ti graphic card.

5.2 Quantitative Analysis of Ca^{2+} dynamics and behavior of acting *C. elegans*.

Following the previously reported method[37] we applied Ca^{2+} signal tracking to the 4D DSP-Net output. After the individual neurons were segmented in each image stack and automatically tracked throughout the entire 4D dataset, a manual correction was further applied to improve the accuracy of tracing. To extract the fluctuation of Ca^{2+} signals ΔF/F0, we calculated ΔF/F0 = (F(t)-F0)/F0 with F(t) being the fluorescence intensity of certain neuron at one time point, and F0 being the mean value of F(t) over the entire period. To analyze the worm behaviors in response to the neural activities, we first extracted the dynamic worm profiles using a U-Net based image segmentation[38]. Then the center lines of the worm were further extracted to calculate the time-varying curvatures of the worm[39], which were finally used to classify the worm behaviors as forward, backward and irregular crawling.

6. Cross sample experiments

A vessel-labelled whole mouse brain (Alexa-647) was imaged with the macro-view Bessel light-sheet microscope using the same 3.2× illumination and detection configurations mentioned above. Then the obtained image stack of vessels was recovered by the DSP-Nets

trained with homogeneous vessel data and heterogeneous neuron data (Thy1-GFP-M).

Following the same logic, we also imaged the antibody-labeled endoplasmic reticulum in fixed *U2OS* cells (the antibody labeled EGFP) using the same conditions for imaging the microtubules. The diffraction-limited image stack of endoplasmic reticulum was super-resolved by DSP-Nets trained with endoplasmic reticulum data and microtubule data (Thy1-GFP-M).

7. Cross mode experiments

Two coronal slices (unclarified, 300-μm thickness) of a transgenetic mouse brain (Thy1-GFP-M) were imaged by a confocal microscope (Nikon Ni-E) using a 16×/0.8 water-dipping objective (CFI LWD 16×/0.8W). The voxel resolution was 0.31 × 0.31 × 1 μm in each acquired 3D image stack (512 × 512 × 300 voxels). We then re-sampled the confocal stacks with matching their voxel resolution to those of 3.2× (2 × 2 × 4 μm) and 12.6× (0.5 × 0.5 × 1 μm) Bessel-sheet image stacks. The re-sampled 12.6× confocal image stacks were used as HR images for DSP-Net training and LR images were degraded from 12.6× as mentioned before. The MR images were down-sampled directly from 12.6×. At validation stage, the re-sampled 3.2× confocal image stack was recovered by DSP-Nets trained with confocal and Bessel-sheet data.

We imaged antibody-labeled microtubules in fixed *U2OS* cells with a Nikon N-SIM microscope using an apochromatic oil immersion objective (Nikon, CFI Apo TIRF 100×/1.49 oil). The raw wide-field images were acquired under conditions of five pattern phases spaced by 2p/5, three pattern orientations spaced 60 degree apart, and a step size of 200 nm. Then the 3D SIM images were reconstructed based on the acquired wide-field images using a commercial software (NIS-Elements), and the diffraction limited wide-field images were obtained simply by an average of the fifteen measurements at each plane.

8. Image quality evaluation

The signal-to-noise ratio (SNR) is computed as

$$SNR = \frac{\mu_{sig}}{\sigma_{bg}}$$

where μ_{sig} being the average intensity of the signal and σ_{bg} being the standard deviation of the background.

The resolution-scaled error (RSE) and resolution-scaled Pearson (RSP)[40] between the SR (or HR) images and the LR inputs were used to quantify the reliability of reconstruction. We obtained the averaged values based on the analysis of 30 randomly-selected regions using the Nanoj Squirrel ImageJ plugin.