Deep learning light field microscopy for rapid four-dimensional imaging of behaving animals

Zhaoqiang Wang[+], Hao Zhang[+], Yicong Yang[+], Yi Li, Shangbang Gao[*], Peng Fei[*]

Huazhong University of Science and Technology, Wuhan, 430074, China

[+] Equal contributing authors

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

Abstract

We propose a novel approach to implement efficient reconstruction for light field microscopy. This approach is based on convolutional neural network (CNN). By taking a two-dimensional light field raw image as the input, our model is able to output its corresponding three-dimensional volume at high resolution. Compared to the traditional light field reconstruction method, our approach dramatically reduces the computation time cost, and significantly improves the image quality (resolution). As furthermore combing this deep learning light field microscopy with selective light-sheet illumination, we can achieve high-contrast, high-resolution (~1.6 μm) three-dimensional imaging of acting *C.elegans* at a speed of fifty volumes per second. We furthermore apply this technique to the interrogation of beating zebrafish heart, *intoto* visualizing the cardiovascular hemodynamics inside zebrafish embryo by rapid volumetric recording of beating myocardium and blood flow. Our method is demonstrated to be promising in a wide range of biomedical applications such as neuroscience and development, in which high-resolution, high-speed volumetric imaging is highly desired.

Introduction

Light field microscopy (LFM) has recently emerged as a rapid volumetric imaging technique for observing live biological speicimans[1]. Compared to conventional imaging schemes, it captures both the lateral position (x, y) and angle (θ_x, θ_y) of the light reaching the sensor by inserting a microlens array in the native image plane. This enables the camera sensor to record information from a four-dimensional (4-D) light field instead of two-dimensional (2-D) focal plane of the sample in a single snapshot. With obtaining the raw light field information, a series of synthetic focal planes or different perspective views of the sample can be retrieved through post-processing[2-6]. Therefore, light field microscopy eliminates the need of stepwise z-scan which is commonly used for three-dimensional (3-D) microscopy and allows the volumetric imaging of multicellular samples at very high speed. Light field microscopy has delivered promising results for monitoring the transient neuronal activities in various animals, such as *C.elegans*, zebrafish embryo[7,8] and rodent brain[9]. As a reference, it has been proved efficient in brain-wide functional imaging of a freely swimming zebrafish with a volume rate up to 77 Hz[10].

Although light field imaging has been successful for 3-D imaging of behaving organisms, a tradeoff exists between the high temporal resolution by one exposure and high spatial resolution by which finer structures can be discerned. The limited sensor pixels which are originally allocated to the sampling of 2-D lateral information are now spread over the 4-D light field, resulting in a significant decimation of the lateral resolution. Several attempts have been made to address this problem either through optimizing the way light field being recorded, or developing new algorithms to reconstruct more spatial information from the light field. In terms of recording, a phase mask was incorporated to achieve better LFM resolution profile[11] or a customized dual microlens arrays were placed at the rear pupil plane to record light field in forms of sub-images, which is capable of collecting information from a larger depth[10]. Besides these approaches requiring precise design of customized optics, alternatively, post-processing algorithms, including LF deconvolution[7,12] and enhancement through fusion of ground-truth images[13], has been

reported to computationally improve the LF reconstruction quality. These algorithms remain limited to the recovery of sparse signals, and also pose high demand on computation resource for iteratively approaching a high-resolution output. In a more sophisticated method, visual volume reconstruction could be skipped while the neuron signals are statistically demixed directly from the raw light field[8,9]. But this method relies on the signal's fluctuation in time series so that it is insensitive to inactive neurons and incapable of handling moving samples.

Here we propose a novel Artificial Neuron Network (ANN) strategy to rapidly recover high-resolution volume from a raw LF image. Compared to existing LF techniques, our ANN-based method achieves better reconstruction quality without compromising LF's advantage in high temporal resolution. We minimize the loss of spatial resolution by incorporating prior structural information of samples so that the network can learn to resolve complex and high-frequency signals from light field. We demonstrate the network's capability by three-dimensionally reconstructing several types of live specimens which are not included the training dataset. Once the LF-ANN network being well trained, our method can quick translate one LF frame into a 3-D volume in a few seconds, showing significant advantage in time lapsed video processing over those iterative optimization algorithms. To obtain optimal results for reconstructing thick specimens, we further use a light-sheet illumination light field (LSILF) geometry to obtain high-contrast LF raw image with less out-of-focus backgrounds. As demonstrated by the 4-D (3-D space plus time) visualization of acting *C. elegan*s, with using a thick light-sheet to selectively illuminate a 60 μm depth for LF recording, our ANN-LSILF method achieves high-fidelity reconstruction of the worm activities at a high resolution of ~ 1.4 μm and high speed of 50 volumes per second, with an ultra-low computation expense compared to the traditional deconvolution method.

Results

Deep learning light filed deconvolution

Aiming to reconstruct a 3D view from a 2D light field projection (LFP), we designed an artificial neural network enabled deconvolution (NED) method. There are two stages involved: the training and the inference. For training, a group of 3D images obtained by light sheet or confocal technique is used as the targets. Then through the forward projection, the simulated 2D LFPs of the 3D targets are generated and used as the input of the neural network. We carefully modeled the point spread function (PSF) of the imaging system, making sure that the LFP simulations were perceptually resemble to the experimental measurements. The network outputs coarse reconstructions of the inputs. The pixel-wise mean-square-error (MSE) between the outputs and the targets are defined as the loss of the training, which is a function of the network parameters. By minimizing the loss function using a gradient descent approach, the parameters get optimized iteratively, and the network is considered to be capable of performing light field deconvolution tasks (the inference stage) on experimental measurements of LFP of new samples. Within a much shorter period of time than the traditional deconvolution method costs, the NED reconstructs the 3D views of the raw inputs, with a higher resolution in both lateral and axial direction.

In order to prove the effectiveness of our ANN-based method, we validated our network on C.*elegans*. The main procedure can be divided into 2 steps. First, we trained the network by inputting high-resolution 3D images of worms, which we acquired from the confocal microscope with 40X objective. (**Figure1. a**) Various parts of worm body, including head and tail where neurons are densely distributed, were recorded to constitute a complete database of worm's structure. These high-resolution 3D images were further transformed into simulated 2D light-field raw images through a optic model[12], which we call forward projection. The light-field raw images and their corresponding high-resolution 3D images made training pairs during learning process. And after an iterative optimization of network parameters, we collected the network as a well-trained model. Second, we tested the model

with empirical light field raw images (**Figure1.b**). Our network is supposed to reconstruct volumes and directly output 3-D images in forms of focal stacks.

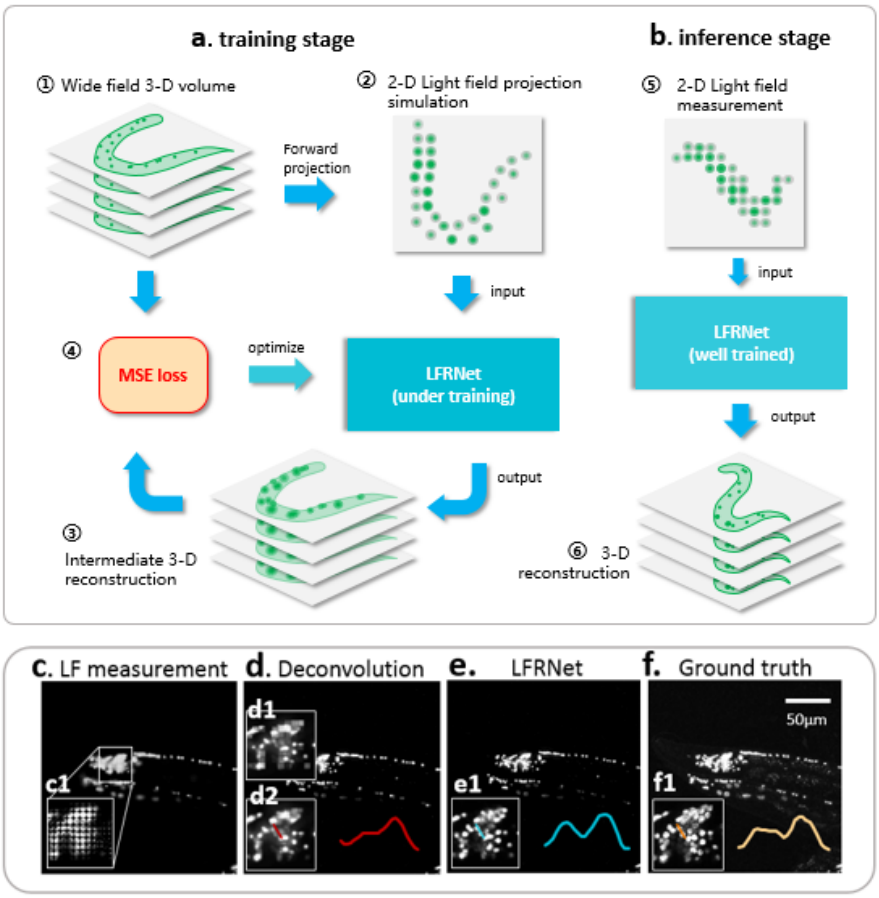

Figure 1. Deep learning approach for light field reconstruction. (a) network training stage. First a high resolution 3-D image of the sample is obtained. Through a forward projection that simulates the light field imaging, a 2-D light field projection of the 3-D sample is generated and used as the input of the network. The network tries to recover the 3-D information from the input and gives a intermediate reconstruction. The original HR 3-D image is defined as the target of the network. The pixel-wise mean-square-error between the target and the network output is calculated as the loss function that measures how close the output is to the target. By iteratively minimizing this loss with a gradient descent technique, the parameters of the network get optimized. (b) Inference stage of the network. While well-trained, the network is ready for the light field reconstructing tasks. A real light field measurement is captured and input to the network. Based on its knowledge from learning examples, the network is capable of immediately generate a 3-D reconstruction that possess both lateral and axial information.(c-f) Characterization of the network by a simulated light field image of C.elegans. (c) Simulated 2-D light field projection, from (f). (d) The maximum z-projection of the output of the traditional deconvolution method, using (c) as the input. (d1) vignette view of the projection (d), where reconstructed layers near the focal plane (z = 0) are included. (d2) layers near the focal plane not included. (e) The max z-projection of the network output. (f) The max z-projection of the ground truth.

We first tested the inference stage on simulated light-field raw image (**Figure 1. c,d,e,f**). Here the input data was also from a forward projection from high-resolution 3D image but excluded from training database. While the high-resolution 3D image served as a ground truth, we also reconstructed using light-field deconvolution[7,12] as a comparison. From a light-field raw image (**Figure 1. c**), both our method and deconvolution recovered worm structures in FOV(~350μm x 350μm x 15μm), which covered the majority of neurons in head region. However,

from the Maximum Intensity Projection (MIP), we observed that our method portrayed in a more detailed way while close neurons are discerned clearly and sharply. And it's free of artifacts near the native focal plane that deconvolution method brought in an intrinsic way.

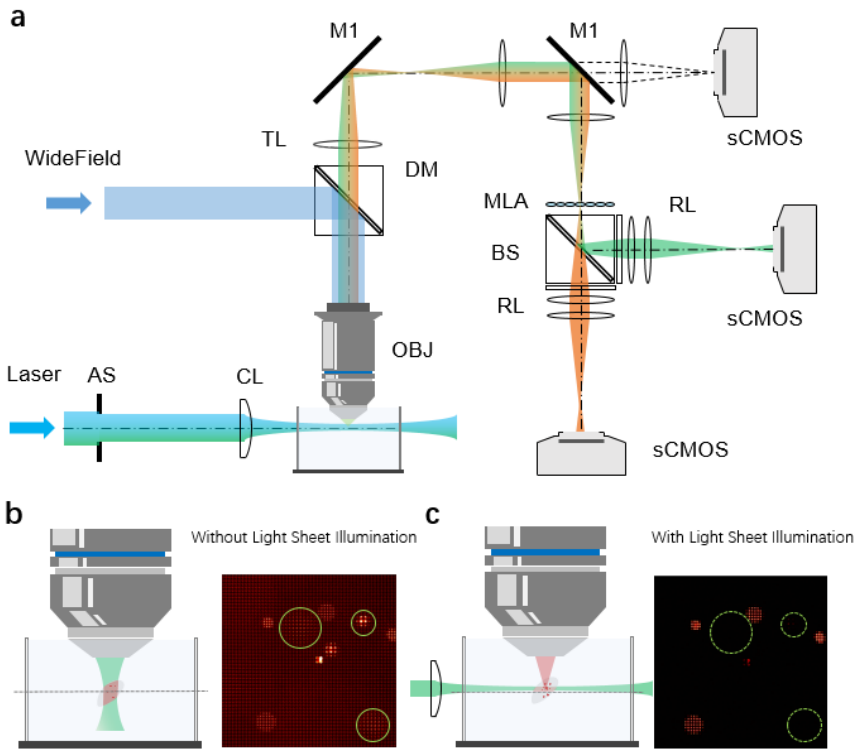

Figure 2. Imaging system schematic. (a) Our light-field imaging system. A thick light sheet generated by a CL (Cylindrical Lens) was implemented to provide selective volumetric illumination. AS (Adjustable Slit) is used to change the thickness of light sheet. OBJ (Objective), DM (Dichromatic Mirror) and TL (Tubelens) are from a commercial microscope (BX51, Olympus). FM (Flip Mirror) directs the light into either a wide-field detection or a light-field detection. The latter one features a MLA (Microlens Array) and RL (Relay Lens). BS (Beam Splitter) is for dual channel imaging. (b) light-field detection of fluorescent beads under wide-field illumination. The whole depth is lit. (c) light-field detection of fluorescent beads under light-sheet illumination. Beads (green circle marker) in positions outsde the selective depth were eliminated.

To validate the reconstruction method on experimental data, we built the light-field imaging system to capture empirical light-field raw images (**Figure 2. a**). A MLA (Microlens Array) was placed at the native image plane to modulate the light so that a 4D light-field could be projected onto a 2D detection sensor. Next we added a light-sheet illumination arm where a Gaussian laser beam was compressed to a thickness of 30-60μm and excite the sample in perpendicular direction to detection arm. This restricts the excited fluorescence in a selected volume. Compared to conventional SPIM[14,15] (Selective Plane Illumination Microscopy) where a thinner light-sheet provides optic sectioning and high z-axis resolution, our design stressed more on the elimination of background noise and unnecessary signals. In Figure2. c, we constrained the illumination region to -30μm-0 and captured the light-field image of fluorescent beads. We can notice beads (marked in green circles) outside this region were removed in comparison with wide-field illumination.

We note that light-field reconstruction is mainly an inverse problem based on a theoretical optic model, or PSF (Point Spread Function) of the system. While a limited computation resource restricts our consideration to a certain subset

of entire PSFs, signals undefined in our model will lead to artifacts and false inference. In this way, a light-sheet illumination optimizes the reconstruction problem by confining the excited signals in a selective volume fully included in computation model. It also increases the contrast of result and brings a clear image, especially in scattering medium like zebrafish larva.

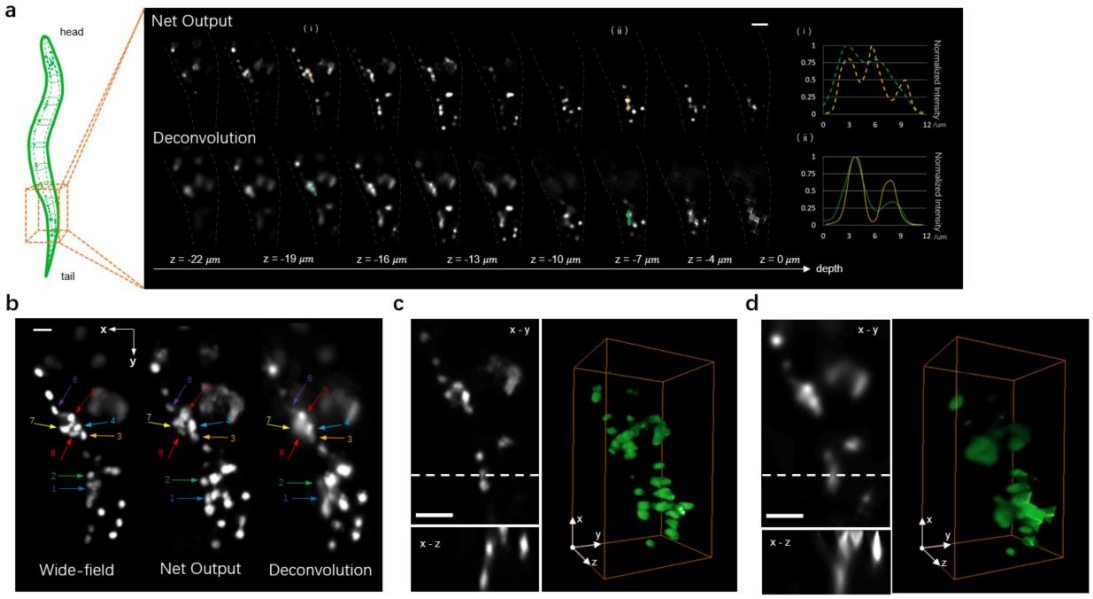

Figure 3 | Reconstruction of C.elegans tail using light-field deconvolution and our method. (a)Synthetic focal planes reconstructed from single exposure (20ms) of light-field microscope. In this capture, the worm was placed off the native focal plane of the objective and a volume of 30μm was selectively illuminated by light sheet. A light-field deconvolved result of 8 iterations was generated from the same light-field raw image, as a comparison of our method. On the right the intensity profile is drawn for regions marked with color lines. The dotted line represents plane at -18μm while solid line represents plane at -7μm. (b) Maximum intensity projection (MIP) of the same region in (a) for a clear illustration of our method's ability to discern dense signals. Arrows and numbers give examples of neurons which actually exist in wide-field (left) and extracted by our method (middle) but blurred by deconvolution (right). Wide-field image was captured using a scanning method for the same ROI but at a different time. A little deformation was caused by movement of worm. Slice of deconvolution at native focal plane is deliberately excluded for better visualization without pixel blocks, which can be noticed in (a). (c) A typical slice of our result and xy cross-section indicated by the dashed lines, as well as a 3D rendering of worm's tail. (d) Comparison of (c) via light-field deconvolution method. Scale bar, 10μm (**a**) (**c**) (**d**), 5μm (**b**).

Using this system, we imaged C.*elegans* of the same strain the network was trained on and demonstrated the result of worm's tail (**Figure 3**). Synthetic focal planes were reconstructed throughout illuminated volume. Our method delivered a high-resolution result while the performance kept on the same level at varying depth. As a comparison, light-field deconvolution showed a deteriorating trend in deeper region and suffered from a grid-like artifacts near native focal plane(z=0μm) (**Figure 3. a**). After capturing the light-field raw image, we also shifted to a standard wide-field detection and recorded the same ROI(region of interest) using a scanning method. This was regarded as a reference to judge the fidelity of our method. Despite minor deformation from worm's movement, we recognized and labeled the corresponding neurons. Our method showed better ability to discern signals in neuron clusters (**Figure 3. b**). It gives excellent portray of the worm's neuron structure both in sectioning plane and 3D rendering (Amira).

Method

Imaging Setup

We built a light field microscope equipped with light sheet illumination around an upright fluorescence microscope (Olympus, BX51). The illumination arm is custom designed and directly mounted on the sample stage. It features [laser], tunable slit and cylindrical lens (Thorlabs). In all imaging experiments, we used a 40X/0.8-NA water objective(Olympus, LUMPlanFLN) to collect the signals, while light field detection arm is appended to camera port of the microscope. A microlens array(OKO Optics) is arranged at the native image plane and further projected onto a Hamamatsu Flash 4.0 sCMOS camera. In C.elegans experiment, a standard GFP filter tube(Olympus) was used whereas zebrafish imaging took advantage of an image slitter(W-View, Hamamatsu) with [filter]. The samples are held by a custom holder, which we scan using a Piezo stage(P12.Z200K, Coremorrow) to obtain adjoining sub-volumes.

Implementation of deep learning light field Deconvlution algorithm

VCD Strategy

In a general convolutional neural network (CNN), the output channels of each convolutional layer represent feature maps extracted by different convolution kernels, which then flows to the next layer as the input to further distill the features. In this way, the network generates a multi-channel output where each channel is a non-linear combination of the original input channels. This mechanism intrinsically agrees with the light field deconvolution: each axial slice of the reconstruction volume arises from a superposition of all angular views of light field projection (LFP), after convolved with the corresponding point spread function (PSF). Thus in our proposal, a View-Channel-Depth (VCD) transformation is naturally used to reconstruct a 3-D volume with depth information from a bunch of angular views of light field projection.

Though the program takes a 2-D LFP as the input, the very first convolutional layer of our net actually deals with a re-formatted 3-D ([height, width, views]) data, in which the 3rd dimension is the extracted and successively arranged angular views from the raw input. The first VCD layer abstracts feature maps from all these views, making the initial transformation from “view” to “channel”. The following convolutional layers keep combining old channels from previous layer and generating new ones, to fully excavate the hidden features. The last layer, finally gives a 3-D output, of which the 3rd dimension is still composites of channels from its predecessor but with a channel number that equals to the depth of the 3-D image to be reconstructed, fulfilling the second transformation from “channel” to “depth”.

Network Structure

Our network comprises an up-scaling part and a feature extracting part (i.e. the VCD transformation module). The former interpolates the angular views to the same lateral size as the LFP. The latter is adapted from the U-net architecture and made up with several convolution layers and deconvolution layers. Each convolution layer and deconvolution layer has three parameters: n, f and s, denoting the output channels number of this layer, the filter size of convolution kernel and the lateral step size when kernel moves, respectively. A convolution layer with s = 2 halves the lateral size of feature maps, while a deconvolution layer with s = 2 will double it. The arrow between two layers

stands for a concatenating operation that put the feature maps of this two layers together, as the input to the follow-up layers.

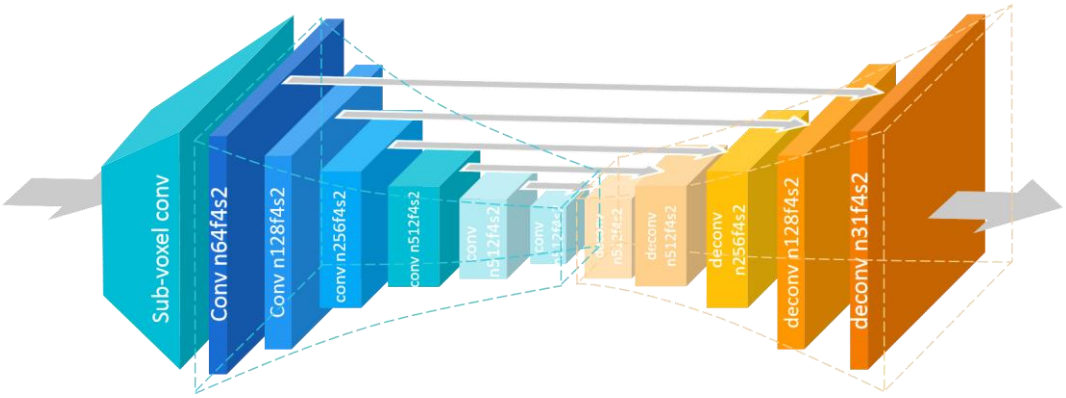

Figure 4. The structure of LFRNet. It comprises a sub-voxel convolution part that interpolates the input into the same size as the raw light field projection, and the VCD transformation module that extracts and merges features into depth information. Each convolutional layer and deconvolution layer is represented with a block, of which the dimension roughly indicates the lateral size and channel numbers of the features maps of the current layer.

C. *elegans* experiments.

To map the neurons in C.*elegans*, we chose worms at young adult stage while stain used in this study was: QW1217 hpls491(Prgef-1::GCaMP6); hpls467(Prab-3::NLS-RFP). It was maintained at 22 ℃ on NGM plates with OP50 E.coli as food. The worm was immobilized by levamisole and further embedded in 1% agarose. We generated a 30um-thick light sheet and moved the focal plane of objective to the bottom of light sheet so the illumination region spanning from -30um to 0um. The worm was then placed in this region by a z-axis stage motivated by the actuator [ZST225B, Thorlabs] to make the neurons perfectly excited while the background was greatly eliminated to improve contrast. We used a 40X water objective and a standard RFP filter set. For each worm, we first captured light field image, and then quickly switched to normal widefield detection, scanned the worm body for a reference for light field reconstruction result.

In prior to reconstruction, a network model should be well trained. We collected the static high resolution 3D images through a confocal laser scanning microscope [FLUOVIEW FV3000, Olympus]. The samples are worms of the same strain as previously discussed and also immobilized. We randomly imaged about 24 worms under 40X magnification and transformed them into simulated light field raw images through a wave optic model[12]. Then we trained our network through pairs of light field image and corresponding 3D images until the loss function is stabilizing. Finally we saved the network for all the empirical data reconstruction.
After reconstruction, we map the neurons in C.*elegans* with the reference of WormAtlas.

Discussion

Reference

1 Levoy, M., Ng, R., Adams, A., Footer, M. & Horowitz, M. in *ACM SIGGRAPH 2006 Papers*

924-934 (ACM, Boston, Massachusetts, 2006).

2 Levoy, M., Ren, N., Adams, A., Footer, M. & Horowitz, M. 924-934.

3 Hoffmann & Maximilian. Light field deconvolution microscopy for optical recording of neuronal population activity. (2014).

4 Pégard, N., Liu, H. Y., Antipa, N., Waller, L. & Adesnik, H. in *Imaging Systems and Applications.* JTh4A.3.

5 Pégard, N. C. *et al.* Compressive light-field microscopy for 3D neural activity recording. *Optica* **3**, 517 (2016).

6 Prevedel, R. *et al.* Simultaneous whole-animal 3D-imaging of neuronal activity using light-field microscopy. *Nature Methods* **11**, 727-730 (2014).

7 Prevedel, R. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature Methods* **11**, 727, doi:10.1038/nmeth.2964 https://www.nature.com/articles/nmeth.2964#supplementary-information (2014).

8 Pégard, N. C. *et al.* Compressive light-field microscopy for 3D neural activity recording. *Optica* **3**, 517-524, doi:10.1364/OPTICA.3.000517 (2016).

9 Nöbauer, T. *et al.* Video rate volumetric Ca2+ imaging across cortex using seeded iterative demixing (SID) microscopy. *Nature Methods* **14**, 811, doi:10.1038/nmeth.4341 https://www.nature.com/articles/nmeth.4341#supplementary-information (2017).

10 Cong, L. *et al.* Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio). *eLife* **6**, e28158, doi:10.7554/eLife.28158 (2017).

11 Cohen, N. *et al.* Enhancing the performance of the light field microscope using wavefront coding. *Opt. Express* **22**, 24817-24839 (2014).

12 Broxton, M. *et al.* Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express* **21**, 25418-25439, doi:10.1364/OE.21.025418 (2013).

13 Lu, C. H., Muenzel, S. & Fleischer, J. in *Frontiers in Optics.*

14 Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E. H. K. Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy. *Science* **305**, 1007-1009, doi:10.1126/science.1100035 (2004).

15 Huisken, J. & Stainier, D. Y. R. Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM). *Opt. Lett.* **32**, 2608-2610, doi:10.1364/OL.32.002608 (2007).