Deep learning-enhanced light-field imaging with continuous validation

Nils Wagner*[1,10,11], Fynn Beuttenmueller*[1,12], Nils Norlin[1-3], Jakob Gierten[4,5], Joachim Wittbrodt[4], Martin Weigert[6], Lars Hufnagel[1], Robert Prevedel*[1,7-9] and Anna Kreshuk*[1]

* These authors contributed equally to this work.

[1] Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.
[2] Department of Experimental Medical Science, Lund University, Sweden.
[3] Lund Bioimaging Center, Lund University, Sweden.
[4] Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany.
[5] Department of Pediatric Cardiology, University Hospital Heidelberg, Heidelberg, Germany
[6] Institute of Bioengineering, School of Life Sciences, EPFL, Lausanne, Switzerland
[7] Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.
[8] Epigenetics and Neurobiology Unit, European Molecular Biology Laboratory, Monterotondo, Italy.
[9] Molecular Medicine Partnership Unit (MMPU), European Molecular Biology Laboratory, Heidelberg, Germany
[10] Present address: Computational Molecular Medicine, Technical University of Munich, Munich, Germany
[11] Present address: Munich school for data science (MUDS), Munich, Germany
[12] Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences

Correspondence should be addressed to R.P. (prevedel@embl.de) and A.K. (kreshuk@embl.de).

Light field microscopy (LFM) has emerged as a powerful tool for fast volumetric image acquisition in biology, but its effective throughput and widespread use has been hampered by a computationally demanding and artefact-prone image reconstruction process. Here, we present a novel framework consisting of a hybrid light-field light-sheet microscope and deep learning-based volume reconstruction, where single light-sheet acquisitions continuously serve as training data and validation for the convolutional neural network reconstructing the LFM volume. Our network delivers high-quality reconstructions at video-rate throughput and we demonstrate the capabilities of our approach by imaging medaka heart dynamics and zebrafish neural activity.

Capturing neuronal activity distributed over whole brains, elucidating long-range molecular signaling networks or analyzing structure and function of beating hearts in small animals necessitate imaging methods that are capable of resolving these highly dynamic processes on milli-second time-, and hundreds of micrometer length scales. To address these challenges, several imaging approaches[1] have recently been proposed or optimized, ranging from highly optimized point and line scanning, to selective plane illumination[2] or by reducing the dimensionality of the image acquisition[3]. While the former two are limited by the sequential nature of the image capture, the latter often require sparsely labelled samples or a severely compromised imaging field-of-view (FOV).

A particularly attractive candidate for high-speed three-dimensional (3D) imaging in biology is light-field microscopy (LFM), due to its ability to instantaneously capture 3D spatial information in a single camera frame, thus permitting volumetric imaging limited by the frame-rate of the camera only[4–6]. The exceptional ability to image the 3D distribution of fluorescent emitters over large, hundreds of micrometer-scale FOV with millisecond temporal resolution has opened new avenues in developmental and neuro-biology, such as the recording of whole-brain neuronal activity in several model organisms[6–8] or the visualization of ultra-fast cardiovascular dynamics[9,10]. Technologically, LFM has seen a steady raise in performance over the past years, including approaches to advance its rather low and non-uniform spatial resolution[9] and signal-to-noise[10,11], and to optically[9,12] or computationally[13,14] reduce the presence of image reconstruction artefacts. Yet the widespread use of this optically appealing technique in the life sciences has been hampered by a computationally demanding, iterative image reconstruction process that ideally demands large-scale computational infrastructure as well as data management, and thus practically restricts the effective experimental throughput, especially with respect to long-term recordings.

With the explosive development of deep learning and convolutional neural networks (CNNs), multiple algorithms have recently been proposed with the aim to replace iterative deconvolution procedures such as Richardson-Lucy's by a CNN[15]. In the natural image domain, CNNs are now the primary method for removing motion blur and other artefacts traditionally solved by iterative deconvolution[16]. Similarly, in microscopy several deep learning-based methods have recently been introduced for deblurring, denoising or super-resolution applications[17]. Although these methods demonstrate excellent image reconstruction performance and empirically have been shown to generalize to data similar to the one used in training, no theoretical guarantees on generalization can be given. It is therefore of utmost importance to extensively validate and, if needed, to retrain the CNN for each experimental setting[18]. This requirement presents a problem for many bio-imaging applications, and in particular for dynamic imaging with LFM, as raw light field images are difficult to interpret. For many dynamic biological processes, it is not possible to arrest the activity and acquire a static training volume by confocal or other imaging modalities[19,20]. In that case, training has to be restricted to simulations or iterative algorithm

reconstructions, without showing the CNN any independently acquired volumetric data which truly corresponds to the light field images it aims to reconstruct.

To overcome this limitation, here we present a novel framework for fast and high-fidelity reconstructions of experimental light-field microscopy images, termed HyLFM. Our approach is based on reconstruction by a CNN enhanced by simultaneous acquisition of high-resolution image data for training and validation. Our neural network - which we term HyLFM-Net - is designed for light-field data processing and 3D image reconstruction (see **SI Fig. 1** and **SI Tab. 1** for a detailed architecture description). To avoid potential bias to the previously seen training data and to enable direct validation of the reconstructions from uninterpretable LFM images, we have included an additional, continuous validation mechanism into our LFM imaging setup, thereby achieving and ensuring high-fidelity, trustable reconstructions. Experimentally, this is realized by adding a simultaneous, selective-plane illumination microscopy (SPIM) modality into the LFM setup which continuously produces high-resolution ground truth images of single planes for validation, training or refinement of the CNN. The training can thus be performed both on static sample volumes and dynamically from a single plane that sweeps through the volume during 3D image acquisition. Besides direct training from non-static samples, the latter approach allows for fine-tuning of the network on-the-fly if an inconsistency is found during continuous validation. We demonstrate the capabilities of our HyLFM system by imaging the dynamics of a hatchling (8 dpf) medaka (*Oryzias latipes*) beating heart across a 350×300×150 μm^3 FOV at a volumetric speed of 40–100Hz, as well as calcium-evoked neural activity in 5dpf zebrafish (*Danio rerio*) larvae over 350×280×120 μm^3 FOV at 10Hz. We train HyLFM-Net dynamically on a part of the acquired timelapse and achieve superior image reconstruction quality (and spatial resolution) compared to traditional, iterative LFM deconvolution-based techniques[5,6] and at a video-rate (18Hz) inference throughput (**SI Tab. 2**).

The design of our HyLFM imaging system is conceptually shown in **Fig. 1a** and based on an upright SPIM configuration with dual-illumination (**Online Methods and SI Fig. 2**). This allows to simultaneously or sequentially illuminate the entire sample volume for light-field and/or a single plane for SPIM recording. On the detection side, the objective (Olympus, 20x 0.5NA) collects the excited fluorescence which is split either via a 70/30 beamsplitter or based on wavelength into separate optical paths for SPIM and LFM imaging, respectively. A fast, galvanometric mirror in the illumination path, together with an electrical-tunable lens (ETL) in the SPIM detection path enables to arbitrarily reposition the SPIM excitation and detection planes in the sample volumes at high speed (15ms), in respect to the LFM imaging volume (**Fig. 1b**). An automated image processing pipeline[9] ensures that both LFM and SPIM volumes are co-registered in a common reference volume/coordinate system with high precision, which is an important prerequisite for CNN training and validation. Our ability to simultaneously acquire both 2D and 3D training data is paramount to ensure high-fidelity and reliable CNN light-field reconstructions of arbitrary samples, including data

never seen in previous training. Furthermore, this includes dynamic samples for which the process of interest cannot be arrested to acquire a static training volume at high resolution. This is an important advancement of our system compared to previous artificial intelligence enhanced LFM reconstructions[19–21].

Deep learning-based image reconstruction methods are commonly trained from "original"-"reconstruction" image pairs, although semi-supervised and self-supervised approaches are now also gaining popularity[22,23]. Here, we follow a fully supervised approach and train HyLFM-Net directly on pairs of SPIM-LFM images. The LFM image which serves as input to the network is transformed into a tensor where the individual pixels of each lenslet are rearranged as channels (**Fig. 1c**). This re-arrangement allows convolution operations to act on angular views, while the projection (1×1 convolution) layers learn to combine information from different angles. The multi-channel 2D images are passed through 2D residual blocks [ResNet] and a transposed convolution. The output goes through a final 2D convolution layer and is then transformed to 3D by reinterpreting network filters as the axial spatial dimension. The 3D images are then further processed by 3D residual blocks and upsampled by transposed convolutions to finally yield the reconstructed 3D volume. For training on single planes, the registration transform between the two detection modalities is encoded into the last network layer to enable direct comparison with the acquired 2D light sheet image (**Online Methods, SI Fig. 1**).

To evaluate and verify the performance of our HyLFM system, we imaged sub-diffraction sized, fluorescent beads suspended in agarose and quantified the improvement in both spatial resolution and overall image quality by comparing it with the standard, iterative light-field deconvolution (LFD – **Fig. 1d–i**). We found that HyLFM-Net correctly inferred the 3D imaging volume from the raw light-field data, yielding high and uniform lateral (1.8±0.2µm) as well as axial (7.1±1.3µm) resolution across the imaging volume (n=4966 beads, **Fig. 1d**), significantly better than what could be obtained by LFD (**Fig. 1i**). Furthermore, HyLFM-Net reconstructions do not suffer from artefacts near the native focal plane that are common in LFD[5] (see arrows in **Fig. 1i** and **SI Fig. 3**). However, it has to be noted that the "PSF" of the signal reconstructed with a CNN depends strongly on the resolution and the shape of the signal in the training data. Therefore, training on small structures such as sub-diffraction sized beads will lead to unnaturally precise signal localizations and, when such a network is applied to a dataset with larger structures (or beads), can lead to erroneous reconstructions. Conversely, training only on large structures can lead to a network which merges small neighboring objects together (**SI Fig. 4**). While empirically the bias to training data can be alleviated by ensuring more diverse training datasets, at the current state of machine learning theory no formal guarantees on network generalization performance can be made. This observation has motivated our hybrid microscope setup, where the network can always be validated on concomitantly acquired high-resolution data and, if necessary, retrained directly on the imaged sample instead of relying on a sufficiently broad and general composition of static

training data, which in practice is typically lacking or time- and/or resource-intensive to produce.

Next, we applied the HyLFM system to the challenging task of imaging a beating medaka fish heart *in-vivo* to show its capability to correctly capture highly dynamic cellular movements in 3D (**Fig. 2a–h** and **SI Vid. 1**). When imaging the dual color expressing cardiomyocytes (*myl7::H2B-eGFP; myl7::H2A-mCherry;* i.e. nuclear-eGFP for LFM-, -mCherry for SPIM detection-path) at a 40-100Hz volume rate, we could visualize the heart at single-cell resolution and free of reconstruction artefacts for both pharmacologically arrested (static) hearts (**Fig. 2a–d,** V-FOV ~350×300×180µm$_3$), and beating (dynamic) hearts (**Fig. 2e–k,** V-FOV ~350×300×150µm$_3$). Here, our HyLFM-Net yielded high image quality metrics (MS-SSIM = 0.926, PSNR = 30.84), compared to SPIM (**Fig. 2h**), and allowed 3D volume inference at 18.2 Hz, which represents at least a 1000-fold reconstruction speed improvement over common LFD[5,6] (**SI Tab. 2**). Note that in order to obtain the above-mentioned image quality metrics, HyLFM-Net was trained on a separate fish heart and applied in inference mode. Being able to perform 3D volume inference at video-rate speed significantly boosts overall experimental imaging throughput and further enables real-time 'quality' control of uninterpretable light field images during image acquisition. Furthermore, if the pre-trained network is not performing sufficiently well, it can then be refined or fine-tuned based on the corresponding single plane SPIM images, or even re-trained from scratch (**Fig. 2f,j**). Importantly, note that the network trained on dynamically acquired SPIM single planes (HyLFM-Net-dyn in **Fig. 2h**) is performing equally well or better than the network trained on full static volumes (HyLFM-Net stat in **Fig. 2h**). This underscores the feasibility of our hybrid 2D/3D imaging approach.

The unique assets of LFM make it a promising method for neural activity imaging in small model organisms. To demonstrate the potential of HyLFM to also deliver quantitatively accurate reconstructions, we imaged 5dpf transgenic larval zebrafish brains expressing the nuclear-confined calcium-indicator GCaMP6s *Tg(elavl3:H2b-GCaMP6s)* (**Fig. 2l–r**). When distributing the excited fluorescence into the LFM and SPIM detection arms we could record whole-volume light field and high-resolution SPIM imaging data at 10 Hz each, over a 350×280×120µm$_3$ volume. Again, the concurrent availability of ground truth data enabled our HyLFM system to faithfully learn and infer not only structural, but also intensity-based information, as demonstrated by the high degree of correlation of Ca^{2+}-signal traces between HyLFM and ground truth data obtained by SPIM or conventional LFD (**Fig. 2o–r**). The ability to rapidly acquire neural activity dynamics from low signal light-field data at hundreds of Hz volume rate should make HyLFM an attractive method for visualizing electrical activity via recently developed voltage-indicators[24], in which kHz-rate volumetric image data needs to be efficiently captured and reconstructed.

In summary, we have demonstrated a new framework for deep-learning based microscopy with continuous ground-truth generation for enhanced reconstruction reliability. Our approach enables light-field imaging with improved spatial resolution and minimal reconstruction artefacts, and compared to previous work based on multiview deconvolution[9], achieves this performance within the relaxed imaging geometry of a standard two-objective SPIM. The ability to reconstruct light-field volumes at sub-second (video-)rate eliminates the main computational hurdle for light field imaging in biology, and we thus expect this to further accelerate the uptake of LFM by the community. The integration of a high-resolution imaging modality into our LFM system further mitigates the omnipresent problem of acquiring appropriate training data, as it can be generated simultaneously and on-the-fly. Furthermore, the on-line availability of single-plane ground truth (SPIM) images distributed over 3D space and time enables continuous CNN output validation and fine-tuning, as early time points of a time-lapse imaging experiment can be used for network training and/or refinement. This new concept to supervised AI-enhanced microscopy also solves the problem of transferability, as the network over time learns on the actual experimental data, and therefore does not require pre-acquisition of training images from particular specimen types with separate microscopes. This is a key advantage of our approach which goes beyond previous work in the field[19–21]. While the HyLFM setup has been developed specifically for light field imaging, the general principle behind it is applicable to other imaging approaches which rely on iterative or trained computational methods for image reconstruction or restoration. Finally, we note that access to time- and resource-efficient light-field reconstructions further facilitates data-intensive, long-term 4D imaging experiments at high throughput as it allows to store volumetric image data in compressed form, i.e. as raw 2D light field images[19]. Given that light-field detection only requires the moderately complex and inexpensive addition of a suitable microlens array into the imaging path and is in principle compatible with any custom or commercial SPIM realization, bears further potential for widespread use of this method in the life sciences.

ACKNOWLEDGMENTS

We would like to thank the EMBL Heidelberg mechanical and electronic workshop for help as well as the IT Services Department for HPC cluster support and Christian Tischer from CBA for his help with volume registration. We further thank M. Majewsky, E. Leist and A. Saraceno for fish husbandry. We thank Krasimir Slanchev and Herwig Baier (MPI Martinsried) for providing calcium reporter zebrafish lines. J.G. was supported by a Research Center for Molecular Medicine (HRCMM) Career Development Fellowship (CDF), the MD/PhD program of the Medical Faculty Heidelberg, the Deutsche Herzstiftung e.V. (S/02/17), and by an Add-On Fellowship for Interdisciplinary Science of the Joachim Herz Stiftung and is grateful to M. Gorenflo for supervision and guidance. N.N acknowledges support from Åke Wiberg foundation and Sten K Johnson foundation. This work was supported by the European Molecular Biology Laboratory (F.B, N.W., N.N., L.H, A.K., R.P.).

AUTHOR CONTRIBUTIONS

A.K., L.H. and R.P. conceived the project. N.W. and N.N. built the imaging system and performed experiments with the help of J.G.. J.G. generated transgenic animals under guidance of J.W. F.B., A.K. and M.W. conceived CNN architecture. F.B. and N.W. implemented the CNN and other image processing parts of the computational pipeline and evaluated its performance. A.K. and R.P. led the project and wrote the paper with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests. L.H. is scientific co-founder and employee of Luxendo GmbH (part of Bruker), which makes light sheet-based microscopes commercially available.

REFERENCES

1. Winter, P. W. & Shroff, H. Faster fluorescence microscopy: Advances in high speed biological imaging. *Curr. Opin. Chem. Biol.* **20**, 46–53 (2014).

2. 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 (80-.).* **305**, 1007 LP – 1009 (2004).

3. Lu, R. *et al.* Video-rate volumetric functional imaging of the brain at synaptic resolution. *Nat. Neurosci.* **20**, 620–628 (2017).

4. Levoy, M., Ng, R., Adams, A., Footer, M. & Horowitz, M. Light field microscopy. *ACM Trans. Graph.* **25**, 924 (2006).

5. Broxton, M. *et al.* Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express* **21**, 25418 (2013).

6. Prevedel, R. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* **11**, 727–30 (2014).

7. Aimon, S. *et al.* Fast near-whole–brain imaging in adult Drosophila during responses to stimuli and behavior. *PLOS Biol.* **17**, e2006732 (2019).

8. Nöbauer, T. *et al.* Video rate volumetric Ca2+ imaging across cortex using seeded iterative demixing (SID) microscopy. *Nat. Methods* **14**, 811–818 (2017).

9. Wagner, N. *et al.* Instantaneous isotropic volumetric imaging of fast biological processes. *Nat. Methods* **16**, 497–500 (2019).

10. Truong, T. V *et al.* High-contrast, synchronous volumetric imaging with selective volume illumination microscopy. *Commun. Biol.* **3**, 74 (2020).

11. Taylor, M. A., Nöbauer, T., Pernia-Andrade, A., Schlumm, F. & Vaziri, A. Brain-wide 3D light-field imaging of neuronal activity with speckle-enhanced

resolution. *Optica* **5**, 345 (2018).

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

13. Stefanoiu, A., Page, J., Symvoulidis, P., Westmeyer, G. G. & Lasser, T. Artifact-free deconvolution in light field microscopy. *Opt. Express* **27**, 31644 (2019).

14. Guo, C., Liu, W., Hua, X., Li, H. & Jia, S. Fourier light-field microscopy. *Opt. Express* **27**, 25573 (2019).

15. Guo, M. *et al.* Rapid image deconvolution and multiview fusion for optical microscopy. *Nat. Biotechnol.* 1–10 (2020). doi:10.1038/s41587-020-0560-x

16. Schuler, C. J., Hirsch, M., Harmeling, S. & Scholkopf, B. Learning to Deblur. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 1439–1451 (2016).

17. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215–1225 (2019).

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

19. Page, J., Saltarin, F., Belyaev, Y., Lyck, R. & Favaro, P. Learning to Reconstruct Confocal Microscopy Stacks from Single Light Field Images. *arxiv.org/abs/2003.11004* (2020).

20. Wang, Z. *et al.* Network-based instantaneous recording and video-rate reconstruction of 4D biological dynamics. *bioRxiv* 432807 (2020). doi:10.1101/432807

21. Li, X. *et al.* DeepLFM: Deep Learning-based 3D Reconstruction for Light Field Microscopy. in *Biophotonics Congress: Optics in the Life Sciences Congress 2019 (BODA,BRAIN,NTM,OMA,OMP)* NM3C.2 (Optical Society of America, 2019). doi:10.1364/NTM.2019.NM3C.2

22. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - Learning Denoising From Single Noisy Images. in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2124–2132 (IEEE, 2019). doi:10.1109/CVPR.2019.00223

23. Kobayashi, H., Solak, A. C., Batson, J. & Royer, L. A. Image Deconvolution via Noise-Tolerant Self-Supervised Inversion. *arXiv:2006.06156* (2020).

24. Abdelfattah, A. S. *et al.* Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science (80-.).* **365**, 699–704 (2019).

FIGURES

Figure 1

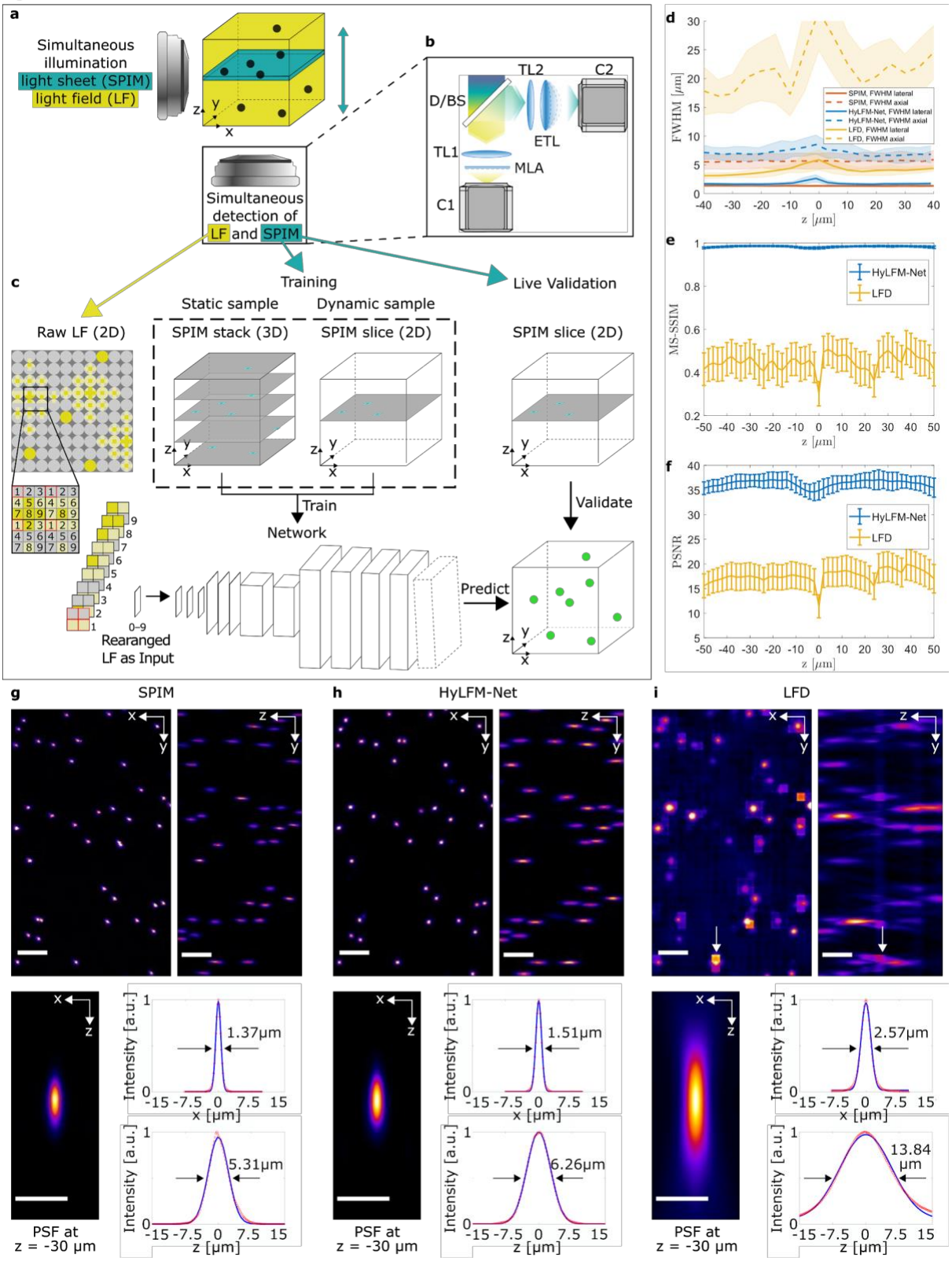


Figure 1: Schematic principle and experimental characterization of HyLFM imaging performance.
(a) Microscope geometry showing simultaneous imaging via SPIM and light field modalities. **(b)** Schematic view of the SPIM and light field detection paths. (D-)BS, (dichroic) beamsplitter; TL, tube lens; MLA, microlens array; ETL, electrotunable lens; C, sCMOS camera. (**c)** HyLFM-Net image reconstruction pipeline. The raw light field image serves as input to a reconstruction CNN, where the lenslet pixels are rearranged channelwise. The last layer encodes the affine transform between the SPIM and LFM spaces. The CNN can be trained either on high-resolution light sheet volumes, for static samples, or on high-resolution light sheet planes, for dynamic samples. The network output can additionally be validated by sweeping light sheet planes. **(d–i)** Evaluation of HyLFM-Net's performance on sub-diffraction sized, fluorescent beads. A 3D SPIM stack **(g)** and the corresponding raw light field image are recorded. **(h)** Same volume is reconstructed by a HyLFM-Net trained on other SPIM / light field pairs. **(i)** Conventional reconstruction based on Lucy-Richardson-type light field deconvolution (LFD). Exemplary artefacts are highlighted by white arrows. **(d)** Lateral and axial resolution as a function of imaging depth for SPIM, HyLFM-Net and LFD, respectively. MS-SSIM **(e)** and PSNR **(e)** image quality metrics across imaging volume comparing HyLFM-Net and LFD with the SPIM ground truth. Scale bars in **(g-i)** are 20µm in whole FOV (top row) and 10µm in PSF close-up (bottom row).

Figure 2

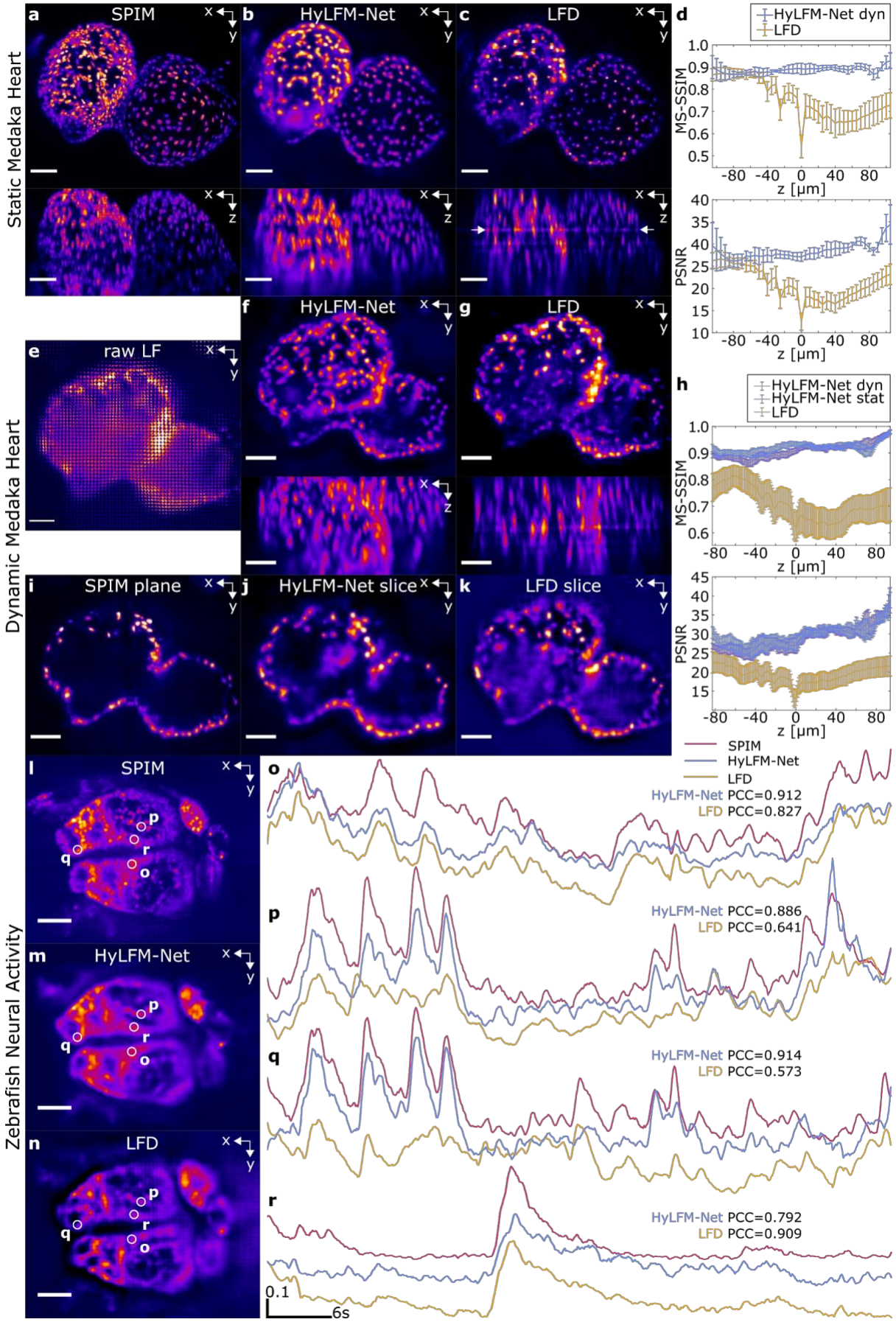


Figure 2: Experimental demonstration of HyLFM on medaka heart and zebrafish neural dynamics.

(a) A static hatchling medaka heart acquired by light sheet (SPIM, maximum intensity projections). The corresponding light field volume reconstructions by **(b)** HyLFM-Net and **(c)** LFD. Note reconstruction artefacts (white arrows) and signal dimming in off-center regions in LFD. **(d)** MS-SSIM and PSNR image quality metrics across imaging volume of HyLFM and LFD compared to SPIM ground truth. **(e)** Example raw light-field (LF) image of a dynamic (beating) medaka heart, acquired at 40Hz. **(f)** Maximum intensity projection of the HyLFM-Net prediction after training on single light sheet planes that continuously swept through the volume during acquisition. Note improved cellular details and absence of artefacts compared to LFD reconstructions in **(g)**. Image quality metrics MS-SSIM and PSNR **(h)** for single plane data in **(i–k)**. **(i)** High resolution light sheet image plane at 16µm depth and same time-point as in **(e).** Single plane validation comparing image quality of HyLFM-Net **(j)** to LFD reconstruction **(k)**. HyLFM-Net stat/dyn refers to networks trained either on full static volumes of the arrested heart as in **(a–c)** (static medaka heart) or on single planes acquired by imaging a beating heart as in **(i–k)** (dynamic medaka heart). Scale bar is 50µm in (**a–n**). See also **SI Videos 1–3**. **(l–n)**: Ca^{2+}-imaging in zebrafish larvae brain using SPIM, HyLFM-Net, and LFD, respectively (mean intensity projection over time). **(o–r)**: Selected Ca^{2+}-traces extracted from regions indicated in **(l–n)**, with Pearson Correlation Coefficients (PCC) comparing LFD and HyLFM-Net to SPIM. Raw traces were pre-filtered (Savitzky-Golay filter order 3, window size 11) and normalized.

Online methods

Hybrid LFM-SPIM imaging setup

The microscope consists of one illumination and one detection objective, orthogonal to each other (see **SI Fig. 2**). The illumination sources are continuous-wave lasers (λ = 488 nm, 20mW, Omicron, and λ = 561 nm, 50 mW, Cobolt). We use a 10x0.3 NA (Nikon CFI Plan Fluor 10XW) water dipping objective for illumination and a 20x0.5 NA (Olympus UMPLFLN 20XW) water dipping objective for detection. For the latter, a tube lens with focal length of 200mm (Nikon MXA20696) yields an effective magnification of 22.5x. Two illumination paths, combined by a dichroic mirror (D1), enable simultaneous dual color light-sheet and light-field illumination. For single-color calcium imaging, two separate 488 nm excitation lasers were used and the dichroic mirror (D1) was replaced by a non-polarising beamsplitter (Thorlabs, 70:30). A digital light sheet was generated using one of the axes of a 2D galvo pair (Cambridge technology) combined with a scan lens (SL) and tube lens (TL1, 200mm). To achieve a selective volume illumination for light-field excitation, the laser beam was first expanded and the central plateau of the (Gaussian) illumination profile was used to illuminate the entire volumetric FOV at once. The lateral extensions of the volumetric FOV could be adapted by changing the size of an aperture (2D slit) that was used to crop out the central region of the laser beam. A lens (TL2, 300mm) focused the light on the objective back aperture. On the detection side, the microscope has two arms separated by either a dichroic mirror (D2) for dual color imaging or a non-polarising beamsplitter (Thorlabs, 2”, 70:30 ratio) for single-color (e.g. calcium) imaging. For the light field detection arm, light first passes through a chromatic filter (BP2), and thereafter through a microlens array (pitch 125 µm and focal length 3.125 mm, RPC photonics MLA s125 f25) mounted in a six-axis kinematic mount (Thorlabs, K6XS) allowing fine adjustment of the array in respect to the optical axis. The microlens array is subsequently imaged onto a 4.2 megapixel (2,048×2,048 pixels) sCMOS camera (Andor Zyla) using a 1:1 relay macro lens objective (Nikon AF-S 105 mm 2.8 G VR IF-ED Micro). The second detection arm is used for recording of the light sheet image modality. The light sheet images were acquired by displacing the illuminated light sheet with respect to the focal plane with a galvanometric mirror while refocusing the detection plane remotely using an electrically tunable lens (ETL, Optotune EL-10-30). Here, the fluorescence first passes through a band pass filter (BP1), and the primary image plane is then relayed through the remote focusing unit using a 100mm lens (RL1) to a lens pair consisting of a -75mm offset lens (OL), and the ETL combined with another 100mm lens (RL2). Finally, two 150mm lenses (RL3, RL4) in a 4f configuration as well as a 1:1 macro lens (Nikon AF-S 105 mm 2.8 G VR IF-ED Micro) relay the image plane onto a second sCMOS camera (Andor Zyla). Precise sample positioning was enabled by a composite xyz linear positioning stage (Newport M-562-XYZ) together with a piezo stage (Nanos LPS-30-30-1-V2_61-S-N and controller MC101) and a small rotation stage (Standa, 7R128). Further information on the setup can be found in **SI Fig. 2**.

HyLFM image acquisition and registration

The training data for our HyLFM-Net network consists of SPIM data (ground truth) and the corresponding 2D light field image (input). Two options can be pursued to acquire high-resolution training data in our HyLFM setup: 1) The SPIM imaging plane remains stationary in order to sample the dynamics (e.g. multiple heart beats) over time and obtain sufficient variability in training data (e.g. heart shapes during the beat cycle). 2) The SPIM modality continuously loops in 3D in order to acquire enough variability at each respective z plane. The second option minimizes potential photobleaching and -toxicity effects and was thus chosen for the majority of the experiments. Due to experimental imperfections, the FOV of the two detection paths might not overlap completely. In order to register the light sheet data to the light field volume, we acquired a light field image of fluorescent beads and a light sheet stack of the same volume by displacing the light sheet with respect to the focal plane with a galvanometric mirror and refocusing on the illuminated plane with the ETL. The light field volume was then reconstructed from the recorded light field image using Richardson-Lucy deconvolution (LFD) as in Ref.[6]. The two corresponding volumes were registered with the Multiview Reconstruction Plugin in Fiji[25], yielding the affine transformation that maps the light field volume to the light sheet stack. In dynamic training this affine transformation is then used within the final layer of the network and only the slice for which a light sheet equivalent has been acquired is sampled from the predicted volume during training. This routine also allows for an easy comparison between SPIM images and volumes reconstructed by the HyLFM-Net and LFD. We used the affine transformation that was computed from the fluorescent bead sample throughout all experiments.

HyLFM-Net for light-field reconstruction

The input to the network is a 3-dimensional tensor. The original 2D light field images, composed of up to 70×85 lenslets, 19×19 pixels each, are rearranged to contain 361 (19^2) channels, with each channel corresponding to an angular view, i.e. same pixels of each lenslet. The input is normalized by its 5.0th and 99.8th percentile without clipping. For training and evaluation, the light sheet target images are normalized by their 5.0th and 99.9th percentile. During training several data augmentations are applied. These include addition of Gaussian or Poisson noise, joined random rescaling of light field input image and light sheet target image, lateral axis flipping, as well as joined random 90-degree rotations (applied before rearranging light field to 361 channels). The full network architecture is shown in **SI Fig. 1** and **SI Tab. 1**. Briefly, the rearranged 3D input tensor is passed through two or three residual blocks interlayered with transposed convolutional layers scaling up by factor 2 in the lateral dimensions. The output of the last residual block undergoes a final 2-dimensional convolution, after which its channel dimension is re-interpreted as an axial dimension and a smaller channel dimension. This 4D tensor is passed through 3D residual blocks and transposed convolutional layers, further upsampling in the two lateral dimensions. For predictions aligned with the SPIM data, the last layer of the network encodes the registration of the reconstructed LF volume to the SPIM volume. For static volumes

the SPIM volume was transformed instead. In dynamic training only the one slice, for which a light sheet equivalent has been acquired, is sampled from the predicted volume. The network is trained either with L2 Loss or with a weighted, smooth L1 Loss (down-weighting non-peak-signal-pixels with a decaying weight). Only for the sparse bead data this choice has a significant impact on network convergence, as with the L2 loss the network converges to only predict background. The Adam optimizer is used with the learning rate set between 1.0e-5 and 3.0e-4. The networks have been trained for 26.5, 121.2, 48, and 89.8 hours for the beads, static heart, dynamic heart and neural activity tasks respectively, which were determined by observing the smooth L1 validation loss (for beads) or the MS-SSIM validation score (heart and brain). All training was done on a single NVIDIA GeForce RTX 2080 Ti GPU. Training data volumes measured 137 261.2×400.2×100 μm^3 stacks for beads, 111 339.0×394.6×245 μm^3 and 34 311.3×483.6×245 μm^3 for the static heart, 13747 339.0×394.6 μm^2 and 5925 277.9×283.5 μm^2 slices for the dynamic heart, as well as 26164 355.7×439.1 μm^2 slices for the neural activity (see also training data acquisition). Full network code using the PyTorch 1.4.0 framework can be found at https://github.com/kreshuklab/lnet.

Reconstruction quality analysis

To quantify the microscope's performances in terms of spatial resolution, we imaged a 3D distribution of 0.1 µm sized fluorescent beads (TetraSpeck, Thermo Fisher Scientific) embedded in agarose. The Fiji plugin 'Multiview-Reconstruction'[25] was used to detect 3D bead positions in the recorded light sheet stacks. The same positions were then used to fit a 3D Gaussian and to compare the full width at half maximum (FWHM) in SPIM, LFD and HyLFM-Net prediction volumes respectively. In order to investigate bias to training data and shape priors we imaged 4 µm sized fluorescent beads (TetraSpeck, Thermo Fisher Scientific), cross-applied trained deep neural networks and computed FWHM for all possible combinations (see **SI Fig. 4**). We computed MS-SSIM and PSNR values per z plane for light field and network predictions, using light sheet planes as the reference, for the fluorescent beads and the medaka heart respectively. The following values were used for MS-SSIM computations: NumScales=5, ScaleWeights=fspecial('gaussian', [1, numScales], 1), Sigma=1.5 (Ref.[26])

Fish husbandry and transgenic lines

All medaka fish are maintained in closed stocks at Heidelberg University. Medaka (*Oryzias latipes*) husbandry (permit number 35–9185.64/BH Wittbrodt) and experiments (permit number 35-9185.81/G-145/15 Wittbrodt) were performed according to local animal welfare standards (Tierschutzgesetz §11, Abs. 1, Nr. 1) and by European Union animal welfare guidelines. The fish facility is under the supervision of the local representative of the animal welfare agency. Medaka was raised and maintained as described previously[27]. For in-vivo imaging, embryos were kept in 165 mg/l 1-phenyl-2-thiourea (PTU) in embryo rearing medium (ERM) from 1 dpf until imaging to inhibit pigmentation. For heart imaging, the following transgenic medaka

lines were crossed: *myl7::H2B-eGFP* (see Ref.[9]) and *myl7::H2A-mCherry*. For the generation of the *myl7::H2A-mCherry* transgenic medaka line the *myl7::eGFP* cassette of the pDestTol2CG plasmid (http://tol2kit.genetics.utah.edu/index.php/PDestTol2CG) was replaced by a *myl7::H2A-mCherry* cassette and the modified plasmid was co-injected with Tol2 transposase mRNA into wild-type stock Cab embryos as described earlier[28]. The calcium imaging experiments were performed using a zebrafish (*Danio rerio*) line with a nuclear-localized calcium sensor (*Tg(elavl3:H2B-GCaMP6s)*).

Medaka imaging

Medaka larvae were imaged 1–3 days after hatching. Hatchlings were anesthetized in 150 mg/l disodium phosphate-buffered (pH 7,3) tricaine and mounted in 1 % low-melting agarose (in ERM) containing 150 mg/l tricaine. For light field volume reconstructions with Richardson-Lucy deconvolution, a light-field PSF was chosen that yielded 49 distinct *axial* planes, spaced 5 µm apart after 8 iterations of deconvolution. To create a static medaka heart for HyLFM-Net training, *myl7::H2B-eGFP*, *myl7::H2A-mCherry* transgenic medaka hatchlings were sedated with 150 mg/l tricaine and the heart was pharmacologically arrested with 40 mM 2,3-butanedione 2-monoxime (BDM). Pre-treated hatchlings were mounted in 1% low-melting agarose (in ERM) containing 150 mg/l tricaine and 40 mM BDM. In the case of re-onset of cardiac contractions, BDM stock solution (100 mM) was titrated into the sample chamber until the heart stopped beating. Then light field images and light sheet stacks were acquired subsequently for the same position of the static heart. In order to get sufficient variability for training data we imaged the heart at multiple positions and at different angles. For this purpose, a linear piezo stage was used, which displaced the static heart diagonally to the detection objective, thereby assuring variations in two coordinates simultaneously, while the sample angle was modified manually with a rotation stage. For imaging the beating heart, we simultaneously acquired pairs of light field and light sheet images at all z-planes in the volume.

Zebrafish calcium imaging

Zebrafish calcium imaging was performed using a nuclear localised calcium reporter Tg(elavl3:H2b-GCaMP6s). The zebrafish embryos were mounted according to previous work[9] in 1 % low-melting agarose and imaged 5 days after fertilization using alternatingly SPIM and light-field illumination with 10Hz for both modalities.
Training and validation data were recorded by alternatingly acquiring light-field and SPIM images at 10Hz each (50ms exposure time). To cover the whole volume the SPIM plane was swept along the axial dimension.

Code availability

The neural network code with routines for training and inference are available at https://github.com/kreshuklab/lnet.

Data availability

The datasets generated and/or analysed during the current study will be made publicly available at the point of publication.

METHODS ONLY REFERENCES

25. Preibisch, S. *et al.* Efficient Bayesian-based multiview deconvolution. *Nat. Methods* **11**, 645–648 (2014).

26. Wang, Z., Simoncelli, E. P. & Bovik, A. C. Multiscale structural similarity for image quality assessment. in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003* 1398–1402 (IEEE). doi:10.1109/ACSSC.2003.1292216

27. Koster, R., Stick, R., Loosli, F. & Wittbrodt, J. Medaka spalt acts as a target gene of hedgehog signaling. *Development* **124**, 3147 LP – 3156 (1997).

28. Rembold, M., Lahiri, K., Foulkes, N. S. & Wittbrodt, J. Transgenesis in fish: efficient selection of transgenic fish by co-injection with a fluorescent reporter construct. *Nat. Protoc.* **1**, 1133–1139 (2006).

Supplementary Information

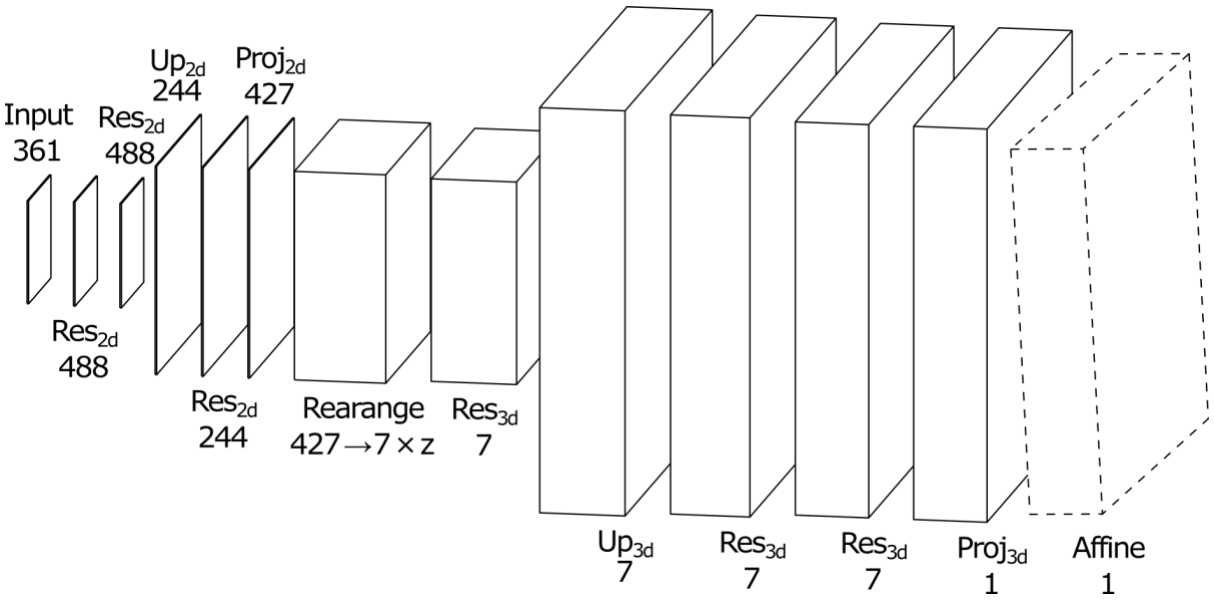
Deep learning-enhanced light-field imaging with continuous validation

Nils Wagner*[1,10,11], Fynn Beuttenmueller*[1,12], Nils Norlin[1-3], Jakob Gierten[4,5], Joachim Wittbrodt[4], Martin Weigert[6], Lars Hufnagel[1], Robert Prevedel*[1,7-9] and Anna Kreshuk*[1]

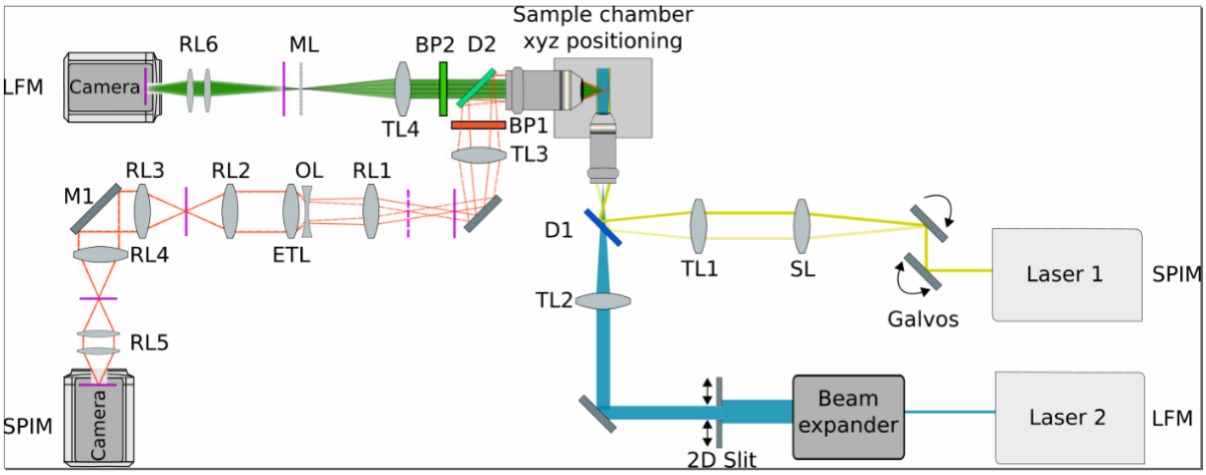
* These authors contributed equally to this work.

1 Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.
2 Department of Experimental Medical Science, Lund University, Sweden.
3 Lund Bioimaging Center, Lund University, Sweden.
4 Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany.
5 Department of Pediatric Cardiology, University Hospital Heidelberg, Heidelberg, Germany
6 Institute of Bioengineering, School of Life Sciences, EPFL, Lausanne, Switzerland
7 Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.
8 Epigenetics and Neurobiology Unit, European Molecular Biology Laboratory, Monterotondo, Italy.
9 Molecular Medicine Partnership Unit (MMPU), European Molecular Biology Laboratory, Heidelberg, Germany
10 Present address: Computational Molecular Medicine, Technical University of Munich, Munich, Germany
11 Present address: Munich school for data science (MUDS), Munich, Germany
12 Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences

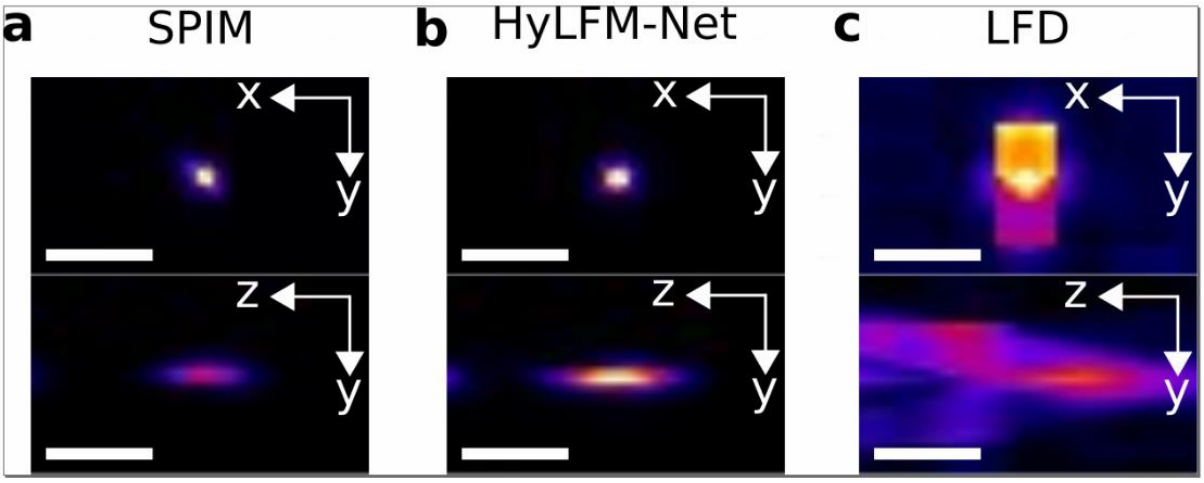
Correspondence should be addressed to R.P. (prevedel@embl.de) and A.K. (kreshuk@embl.de).



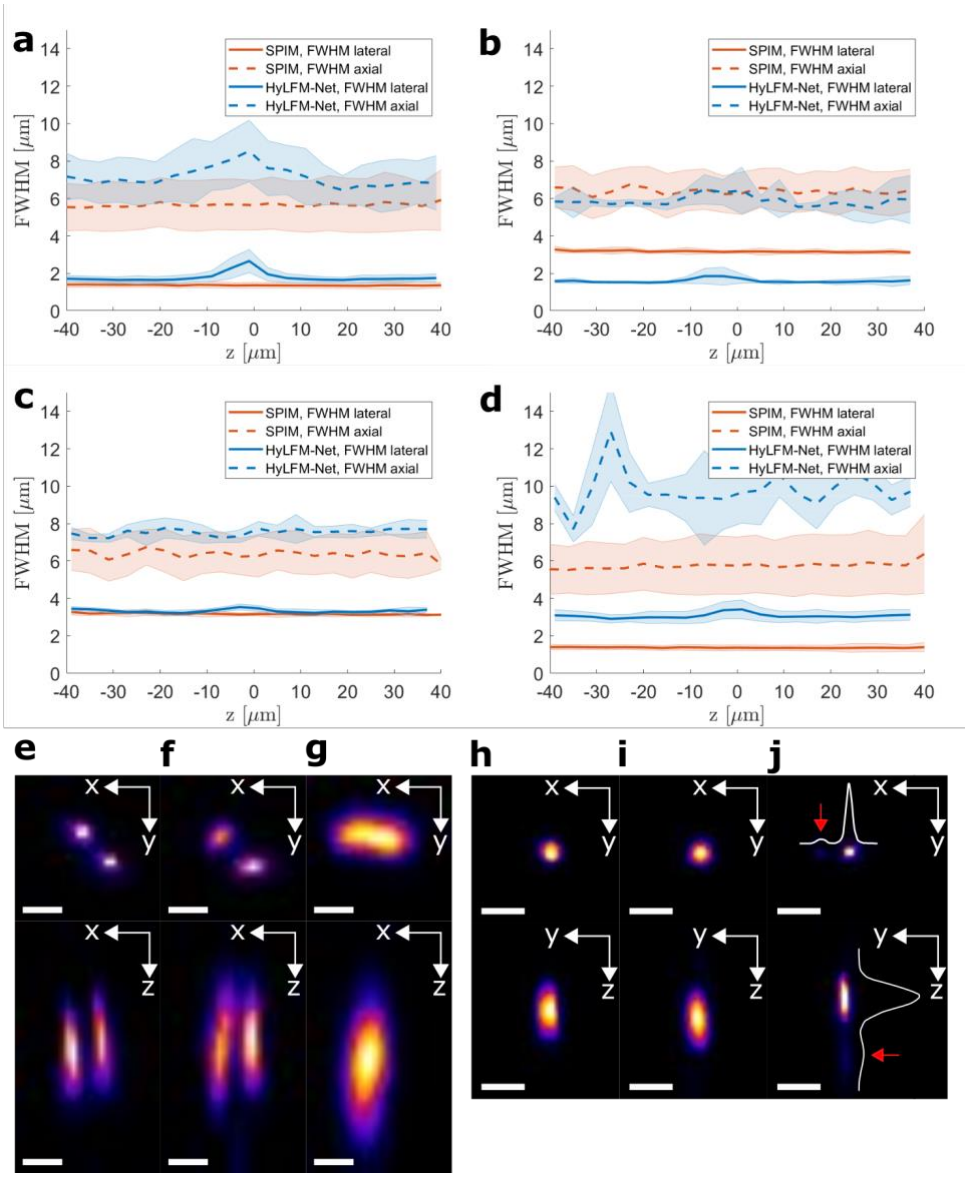
Supplementary Figure 1: Network architecture. $Res_{2/3d}$:residual blocks with 2d or 3d convolutions with kernel size (3×)3×3. Residual blocks contain an additional projection layer (1×1 or 1×1×1 convolution) if the number of input channels is different from the number of output channels. $Up_{2/3d}$: transposed convolution layers with kernel size (3×)2×2 and stride (1×)2×2. $Proj_{2d/3d}$: projection layers (1×1 or 1×1×1 convolutions). The numbers always correspond to the number of channels. With 19×19 pixel lenslets (n_{num}=19) the rearranged light field input image has 19^2=361 channels. The affine transformation layer at the end is only part of the network when training on dynamic, single plane targets; otherwise, in inference mode it might be used in post-processing to yield a SPIM aligned prediction, or the inverse affine transformation is applied to the SPIM target for static samples to avoid unnecessary computations.



Supplementary Figure 2: LFM-SPIM optical setup. Schematic 2D drawing of the LFM-SPIM setup showing the main opto-mechanical components. The sample is illuminated through a single illumination objective with two excitation beam paths (ocra, light sheet illumination and blue, light field selective volume illumination) combined by a dichroic mirror (D1). The fluorescence is detected by an orthogonally oriented detection objective and optically separated onto two detection arms with a dichroic mirror (D2). Bandpass filters (BP1 and BP2) are placed in front of a tube lens (TL3,TL4) for the respective detection path. For the light field detection path (green), the tube lens (TL4) focuses on the microlens array (ML) and the image plane (shown in magenta) displaced by one microlens focal length is relayed by a 1-1 relay lens system (RL6) to an image plane coinciding with the camera sensor (shown in magenta). For the light sheet detection path, a combination of several relay lenses (RL1 to RL4), a 1:1 macro lens (RL5) together with a lens pair consisting of an offset lens (OL) and an electrically tunable lens (ETL) is used to image two axially displaced objective focal planes (shown in magenta, dotted and solid) to a common image plane at the sensor. The refocusing is achieved by applying different currents on the ETL. The mirror M1 is placed at a Fourier plane, such that the FOV of the light sheet path can be laterally aligned to fit the light field detection FOV. For single color imaging, the dichroic mirrors D1 and D2 are replaced by beamsplitters. See Methods for details.



Supplementary Figure 3: HyLFM-Net provides artifact free deconvolution of LFM data. **(a)** Ground truth single light sheet image. **(b)** Subdiffraction beads reconstructed by LFM-Net and **(c)** iterative light field deconvolution (LFD). The volume reconstructed by LFD shows clear signs of artifacts which are removed in HyLFM-Net reconstructed volumes. Scale bar is 10 µm.



Supplementary Figure 4: Cross-application of trained deep neural networks can reveal bias to training data. We created two kinds of samples, one with small (0.1μm) and one with medium-sized (4μm) beads suspended in agarose. In **(a)**, HyLFM-Net was trained on small beads and applied to small beads. FWHM of the beads in the reconstructed volume shows good agreement with SPIM measurements. The same effect is observed in **(c)**, where the network is trained on large beads and applied to large beads. However, when a mismatch between the training and test data is present, the network predictions are biased towards the training data. In **(b)**, HyLFM-Net was trained on small beads and used to reconstruct a volume with large beads, resulting in erroneously small bead reconstructions. Similarly, in **(d)**, HyLFM-Net trained on large beads and used to reconstruct a volume with small beads produces erroneously large objects. **(e)** SPIM image of 0.1μm beads, **(f)** reconstructions of HyLFM-Net from **(a)**, trained on small beads, **(g)** reconstructions from HyLFM-Net from **(d)**, trained on large beads. **(h)** SPIM image of 4μm beads, **(i)** reconstructions of HyLFM-Net from **(c)**, trained on large beads, **(j)** reconstructions of HyLFM-Net from **(b)**, trained on small beads. Line profile is shown to highlight a reconstruction error (red arrows), where the network reconstructs very small beads (as found in the training data) and produces an additional erroneous peak where none is present in the ground truth SPIM volume. Scale bar 2μm in **(e–g)**, and 10μm in **(h–j)**.

Medaka heart	network architecture		image dimensions	
layer name	c	z [px]	y [px]	x [px]
input	361	1	65	75
Res_2d	488	1	65	75
Res_2d	488	1	65	75
Up_2d	244	1	130	150
Res_2d	244	1	130	150
Proj_2d	427	1	130	150
Rearrange	7	61	130	150
Res_3d	7	57	126	146
Up_3d	7	57	252	292
Res_3d	7	53	248	288
Res_3d	7	49	244	284
Proj_3d	1	49	244	284

zebrafish brain	network architecture		image dimensions	
layer name	c	z [px]	y [px]	x [px]
input	361	1	70	83
Res_2d	488	1	70	83
Res_2d	488	1	70	83
Up_2d	244	1	140	166
Res_2d	244	1	140	166
Proj_2d	427	1	140	166
Rearrange	7	61	140	166
Res_3d	7	57	136	162
Up_3d	7	57	272	324
Res_3d	7	53	268	320
Res_3d	7	49	264	316
Proj_3d	1	49	264	316

beads	network architecture		image dimensions	
layer name	c	z [px]	y [px]	x [px]
input	361	1	49	74
Res_2d	976	1	49	74
Res_2d	976	1	49	74
Up_2d	488	1	98	148
Res_2d	488	1	98	148
Up_2d	244	1	196	296
Res_2d	244	1	196	296
Proj_2d	441	1	196	296
Rearrange	7	63	196	296
Res_3d	7	59	192	292
Up_3d	7	59	384	584
Res_3d	7	55	380	580
Res_3d	7	51	376	576
Proj_3d	1	51	376	576

Supplementary Table 1: Network architecture and image dimensions.
Network layers and their corresponding image/tensor dimensions for Medaka heart, zebrafish brain, and beads samples. For Medaka heart samples also slightly smaller image dimensions were used (e.g. **SI Video 2&3**). Layer names as indicated in **SI Fig. 1**. For training on single plane images, an additional layer enacting the affine transformation was appended after the final projection.

Name	LFM image size [px]	Prediction size [px]	Prediction size [µm]	Sample rate [Hz]	Input rate [Mpx/s]	Output rate [Mvx/s]
HyLFM-Net		without i/o to/from hard drive (avg. of 3 runs of 1000 samples)				
heart	1235×1425	244×284×49	339.0×394.6×245.0	18,2	32,1	61,9
brain	1292×1577	256×316×49	355.7×439.1×140.0	15,3	31,1	60,5
beads	931×1406	376×576×51	261.2×400.2×100.0	5,4	7,1	59,6
HyLFM-Net		with i/o to/from hard drive (avg. of 3 runs of 100 samples)				
heart	1235x1425	244×284×49	339.0×394.6×245.0	8,2	14,4	31,3
brain	1292x1577	256×316×49	355.7×439.1×140.0	9,4	19,1	41,5
beads	931x1406	376×576×51	261.2×400.2×100.0	4,3	5,7	47,7
LFD		without i/o to/from hard drive (avg. of 3 runs of 1 sample with 8 iterations)				
heart	1235×1425	1235×1425×49	377.9×416.8×245.0	9,5E-4	1,0E-3	5,1E-2
brain	1292×1577	1292×1577×49	377.9×461.3×245.0	9,0E-4	1,1E-3	5,3E-2
beads	931×1406	931×1406×49	272.3×411.3×245.0	9,3E-4	1,0E-3	5,3E-2

Supplementary Table 2: Inference speed on single GPU. Speed of HyLFM-Net inference and LFD deconvolution on a Nvidia GeForce RTX 2080 Ti. For LFD we used the following settings: magnification 22.5x, n_{num}=19. For HyLFM-Net without i/o and LFD a mini batch size of 1 was chosen. For HyLFM-Net with i/o mini batch sizes of 8, 5, and 1 for heart, brain, and beads were chosen, respectively.

Supplementary Video 1: Volumetric HyLFM reconstruction of a beating Medaka heart at 40Hz.
Volumetric reconstruction of the medaka heart at 40Hz image acquisition speed shown in **Fig. 2e–k**. The cyan plane corresponds to the sweeping SPIM image plane. The panels from left to right show the overlay of the light sheet plane with the respective plane from the prediction/reconstruction volume, a projection of the prediction/reconstruction volume rotated by 45 degrees around the y-axis, and a maximum projection of prediction/reconstruction volume along z-axis, along y-axis, and along x-axis, respectively. Scale bar 30 µm.

Supplementary Video 2: Single plane HyLFM reconstruction of a beating Medaka heart at 56Hz.
Single plane comparison of SPIM ground truth to the corresponding plane of the prediction volume of HyLFM-Net and reconstruction volume of LFD at indicated axial positions of the Medaka heart at 56Hz image acquisition speed. See also **Fig. 2i–k**. Scale bar 30 µm.

Supplementary Video 3: Single plane HyLFM reconstruction of a beating Medaka heart at 100Hz.
Single plane comparison of SPIM ground truth to the corresponding plane of the prediction volume of HyLFM-Net at indicated axial positions of the Medaka heart at 100Hz image acquisition speed. Scale bar 30 µm.