2021 IEEE International Conference on Computational Photography (ICCP) | 978-1-6654-1952-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCP51581.2021.9466256

Real-Time Light Field 3D Microscopy via Sparsity-Driven Learned Deconvolution

Josue Page Vizcaino, Zeguan Wang, Panagiotis Symvoulidis, Paolo Favaro, *Member, IEEE*,
Burcu Guner-Ataman, Edward S. Boyden, and Tobias Lasser, *Member, IEEE*

Abstract—Light Field Microscopy (LFM) is a scan-less 3D imaging technique capable of capturing fast biological processes, such as neural activity in zebrafish. However, current methods to recover a 3D volume from the raw data require long reconstruction times hampering the usability of the microscope in a closed-loop system. Moreover, because the main focus of zebrafish brain imaging is to isolate and study neural activity, the ideal volumetric reconstruction should be sparse to reveal the dominant signals. Unfortunately, current sparse decomposition methods are computationally intensive and thus introduce substantial delays. This motivates us to introduce a 3D reconstruction method that recovers the spatio-temporally sparse components of an image sequence in real-time. In this work we propose a combination of a neural network (SLNet) that recovers the sparse components of a light field image sequence and a neural network (XLFMNet) for 3D reconstruction. In particular, XLFMNet is able to achieve high data fidelity and to preserve important signals, such as neural potentials, even on previously unobserved samples. We demonstrate successful sparse 3D volumetric reconstructions of the neural activity of live zebrafish, with an imaging span covering $800 \times 800 \times 250 \mu m^3$ at an imaging rate of $24-88$Hz, which provides a 1500 fold speed increase against prior work and enables real-time reconstructions without sacrificing imaging resolution.

Index Terms—Computational Photography, Microscopy, Light Field Imaging, Deconvolution, Sparse Representations, Neural Networks

1 Introduction

Light field microscopy (LFM) is a single shot microscopy technique suitable for rapid 3D imaging applications [1]. The remarkable speed of LFM makes it a powerful tool in neuroscience for high-speed neural activity imaging, especially in small transparent animal models. This capability has been first demonstrated in 2014 by imaging the whole brain of C. Elegans and larval zebrafish [2] and further enhanced in recent years [3]–[11]. It has also been applied to mice [12]–[14] and drosophila [15] neural activity imaging. Such pan-neuronal, high temporal resolution data from LFM has fueled new understandings of the principles underlying key cognitive processes, such as decision making [16].

However, two drawbacks have limited the applicability of LFM. First, compared to scanning 3D imaging methods (e.g., confocal and light sheet microscopy), the inferior spatial resolution of classical LFM methods is sometimes insufficient to resolve, for instance, individual neurons in larval zebrafish brains. Second, current methods for LFM volumetric reconstruction can require days of data processing.

To improve the spatial resolution of LFM, extended field-of-view light field microscopy (XLFM) [17], an optimized light field architecture, was developed recently, also known as Fourier LFM (FLFM) [18]–[20]. XLFM simultaneously records multi-view projection images of a sample through a micro-lens array at its Fourier plane, and has achieved near-cellular resolutions throughout an entire larval zebrafish brain [17]. More recently, sparse decomposition algorithms were applied to XLFM reconstruction to take advantage of the spatio-temporal sparsity of neural activity. This modified computational method, termed sparse decomposition light field microscopy (SDLFM), has further improved the resolution and signal-to-noise ratio in immobilized samples [21].

Several attempts have been made to accelerate the reconstruction process of LFM. For example, the offline data processing time can be greatly reduced by directly estimating the light field “footprints” and activity of individual cells and structures without reconstructing volumes (see SID [22] and Compressive LF (CLF) [23]). Where our method reconstructs the full 3D volume, both SID and CLF reconstruct the neural activity at discrete positions. SID/CLF uses a conventional LFM with a space-variant PSF, where the deconvolution is prone to strong artifacts, while our approach uses a Fourier LFM with a space-invariant PSF. Both SID and CLF require the re-computation of the neural signatures for every new sample, which is not needed for our method, as XLFMNet can be trained on a single fish and used on different specimens. Also, in SID/CLF, the 3D neural positions are computed using a numerically generated PSFs, which is prone to errors due to aberrations and component misalignment, while in our work we used a measured PSF, which alleviates this issue.

An advantage of these methods is that they are com-

- *J. Page and Tobias Lasser are with the Computational Imaging and Inverse Problems Group, Technical University of Munich. Z. Wang, P. Symvoulidis and E. S. Boyden are with the Synthetic Neurobiology Group, Massachusetts Institute of Technology. P. Favaro is with the Computer Vision Group, University of Bern. B. Guner-Ataman is with the McGovern Institute for Brain Research, Massachusetts Institute of Technology.*
E-mail: see https://pvjosue.github.io/
Source-code: https://github.com/pvjosue/SLNet_XLFMNet

978-1-6654-1952-9/21/$31.00 ©2021 IEEE

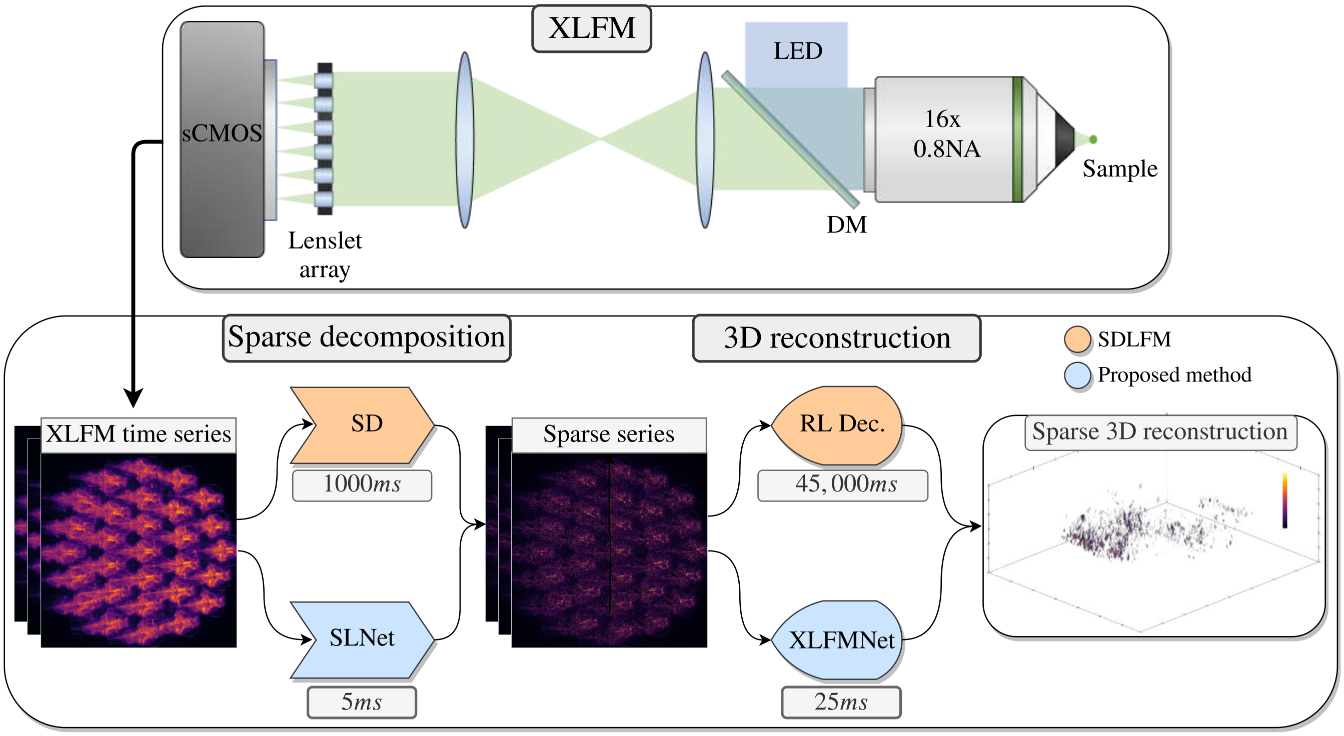


Fig. 1: **Top:** Diagram of the extended field-of-view light field microscope (XLFM) used in this work. The microlens array was conjugated to the back focal plane of the objective lens through a 4-f lens pair. Excitation light from a 470 nm LED was projected on the sample through the objective lens using a dichroic mirror (DM). An sCMOS camera recorded all the sub-images formed behind the microlens array. **Bottom:** A comparison between the state of the art SDLFM reconstruction (in orange) and the proposed method (in blue). Both methods first compute a sparse representation of a XLFM time series stack, and later perform 3D reconstruction of the sparse images.

Method	Our method	SID [23]	CLF [22]
Microscope type	XLFM	LFM	LFM
Pre-processing time	5h once	7$h/sample$	1$h/sample$
Recon. freq. (Hz)	24-88	30	33
Lateral FOV (μm)	800	900	200
Axial FOV (μm)	250	380	300
Lateral res. (μm)	3.2	20	1-8
Axial res. (μm)	7.7	20	0.5-1.5

TABLE 1: Comparison between the proposed method and state of the art sparse neural activity recovery methods.

putationally more efficient and less memory intensive than reconstructing a full 3D volume. Also, they provide the possibility of imaging in deep scattering medium. For details about the field of view (FOV), the speed and the specifications for each method refer to Table 1.

Recently, deep learning networks were applied to conventional LFM reconstruction and have sped up the process a hundred fold [24]–[26]. However, the generalization capabilities of a neural network for 3D reconstruction of unseen samples is still problematic. The work of Wagner *et al.* [8] presents a system capable of quickly retraining the network when a new sample is presented, but, with the extra degree of complexity of a joint LFM and light sheet microscope. Although this previous work performs at real-time, the need of constant retraining complicates matters.

In this work, we propose SLNet and XLFMNet, two neural networks that can efficiently perform sparse decomposition and volume reconstruction on XLFM raw recordings at high speed (see Fig. 1). We also evaluate the generalization capabilities of the networks to unseen samples through techniques such as reducing the number of parameters and augmenting the training data with estimated noise statistics and spatial transformations.

SLNet is trained with an unsupervised approach, by minimizing a loss function that aims to approximate the input images with a low-rank representation. Making use of this reconstruction a sparse representation can be recovered, as shown in the experiments. This problem is well suited for the chosen network and allows our method to generalize to new samples. XLFMNet learns the convolutional nature of the XLFM image formation model, and it is able to generalize to unseen samples, with sufficient quality to be certain that the network is not modifying the acquired information in unfeasible ways, i.e., through hallucination, which is highly undesired in biomedical data. We evaluate a wide range of neural network settings, and choose the most robust one when reconstructing seen and unseen samples.

The SLNet trained network can perform a temporally and spatially sparse decomposition using three images of the sample at different time-points in an interval of less than 5 milliseconds. The trained XLFMNet reconstructs a 3D volume at $24-88$Hz. This real-time reconstruction capability would not only expand the applicability of LFM, but also unlock novel paradigms of experiments that allow closed-loop feedback control of the experimental parameters, such as instrumental adjustment (e.g., autofocus and tracking), animal stimulus delivery (e.g., visual, auditory, or olfactory), and neuronal activity manipulation (e.g., optogenetics), based on the real-time information of the reconstructed volumes.

We first present the unsupervised training of the SLNet in section 2.2, followed by a description of the XLFMNet parameter ablation and data augmentation in sections 2.3 and 2.4. In sections 2.5 and 2.6 we describe the XLFM microscope and the sample preparation for the experiments.

In the experimental results in section 3.1, we describe our findings on the SLNet training and the effect of the design parameters. Later, in section 3.2 we discuss our results on the XLFMNet ablation and in section 3.3 we analyze the neural activity of seen and unseen zebrafish. In section 3.4, an evaluation of the XLFM generalization capabilities is presented by measuring the full width at half maximum (FWHM) of micro-spheres that are not present in the training set. Finally, section 3.5 presents an estimation of the achieved resolution with the different methods using Fourier domain analysis [27].

2 Methods

2.1 Networks training strategy

The sparse decomposition and 3D deconvolution are performed with two networks trained independently. The SLNet (see section 2.2) is trained in an unsupervised manner using only raw LF images and a crafted loss function. In a separate step, the XLFMNet is trained in a supervised manner (see section 2.3) with sparse images (generated by the SLNet) and their corresponding 3D deconvolutions [21]. The deconvolved volumes are used as ground truth, as their quality is sufficient for single neuron identification in the zebrafish and the main goal of the proposed method is to make this a real-time process.

2.2 Unsupervised Sparse Decomposition (SD)

The robust principal component analysis or sparse decomposition [28] is a method that decomposes a matrix M into its low rank (L) and sparse (S) components, such that $M = L + S$. Let $M_{k,m,r} \in \mathbb{R}_{\geq 0}^{k \times m \times r}$ be a set of images captured at k different times points, with lateral sizes m and r. SD can be used to decompose temporal stacks if we arrange $M_{k,m,r}$ to be $M_{k,mr} \in \mathbb{R}_{\geq 0}^{k \times mr}$ and minimize the optimization problem

$$\begin{aligned} \min_{L,S} \quad & |L|_* + \lambda |S|_1 \\ \text{s.t.} \quad & L + S = M_{k,mr}. \end{aligned} \tag{1}$$

Here $|L|_*$ is the nuclear norm of the low rank component, λ a parameter controlling the degree of sparseness and $|S|_1$ the L_1 norm of the sparse component. This type of constrained optimization is usually solved by means of Augmented Lagrangian methods, such as the Augmented Lagrangian multiplier [29], [30].

However, as previously shown by [31], the constraint can be implicitly fulfilled if we first compute L and with it compute $S = (M - L)_{\geq 0}$, the non-negative result of subtracting the input image and the low rank representation. Let $\mathcal{N}_{\Theta}^{SL}(M_{k,m,r}) \approx L_{k,mr}$ be a neural network with parameters Θ that generates a low rank representation of M. We refer to this network as SLNet. Then $S = (M - \mathcal{N}_{\Theta}^{SL}(M))_{\geq 0}$.

The final loss function becomes

$$\min_{\Theta} \quad |M - \Gamma_\mu \left(\mathcal{N}_{\Theta}^{SL}(M_{k,m,n}) \right)|_1, \tag{2}$$

where $\Gamma_\mu(\cdot)$ is a singular value shrinking operator that enforces a low rank in its output by setting the eigenvalues $\Sigma_{<\mu} = 0$ and by shrinking the remaining ones. The full operator reads

$$\begin{gathered} \Gamma_\mu(X) = U \left[\text{sign}(\Sigma) \cdot \max(|\Sigma| - \mu, 0) \right] V^* \\ \text{where: } X = U \Sigma V^*. \end{gathered} \tag{3}$$

The work of Herrera *et al.* [31] employed four fully connected layers to perform the decomposition, however due to the dimensionality of our images ($k = 3$, $m, n = 2160$) fully connected layers were not possible due to the memory requirement. Hence, our network is implemented using two convolutional layers followed by a single ReLu activation function, as shown in Fig. 2. The threshold μ dictates the degree of rank shrinkage together with the amount of sparseness in S. We explore the effect of varying this parameter in section 3.1.

The weight initialization here is crucial, as larger weights produce an L with entries larger than the entries in M, resulting in zero entries in S and no further training. To ensure this does not happen, we initialize the weights Θ of the SLNet, first using the Kaiming method [32], followed by scaling and a positivity constraint as in $\Theta = 0.1|\Theta|$.

For our decomposition we chose the frames M_{t-100}, M_{t-50} and M_t, where t is the time coordinate. We chose these time shifts based on a grid search analysis, which we make available in the supplementary material. Surprisingly, using 3 frames causes a smaller error, and is computationally cheaper than using 100 frames. Furthermore, employing 3 frames matched the maximum capability of our computing resources, as the memory consumption increases greatly when adding more frames. In practice, to use the SLNet one would start recording frames, and once the images at M_{t-100} and M_{t-50} are available, then the network starts to work in real time, by recovering the sparse components of the 3 images. This could be interpreted as a 100 frame buffer or warm up.

2.3 3D Deconvolution neural network architecture

In the next stage we employ a 2D U-net [33] for the 3D reconstruction, similar to the one used by Wang *et al.* [26] and Page *et al.* [24], where the depth stacks are stored in the channel dimension. We named this part of our pipeline XLFMNet. Our XLFMNet $\mathcal{N}_{\Theta}^{3D}$ is parametrized by n and w, the number of down/up sample steps and an exponent to control the number of channels in use: The first layer has 2^w channels and any of the n consecutive layers has 2^{w+n} channels. This parametrization allows the exploration of a wide range of different networks, as shown in Fig. 2. We explored networks with $n = \{2, 3, 4, 5\}$ and $w = \{4, 5, 6, 7\}$ in a systematic grid-like fashion.

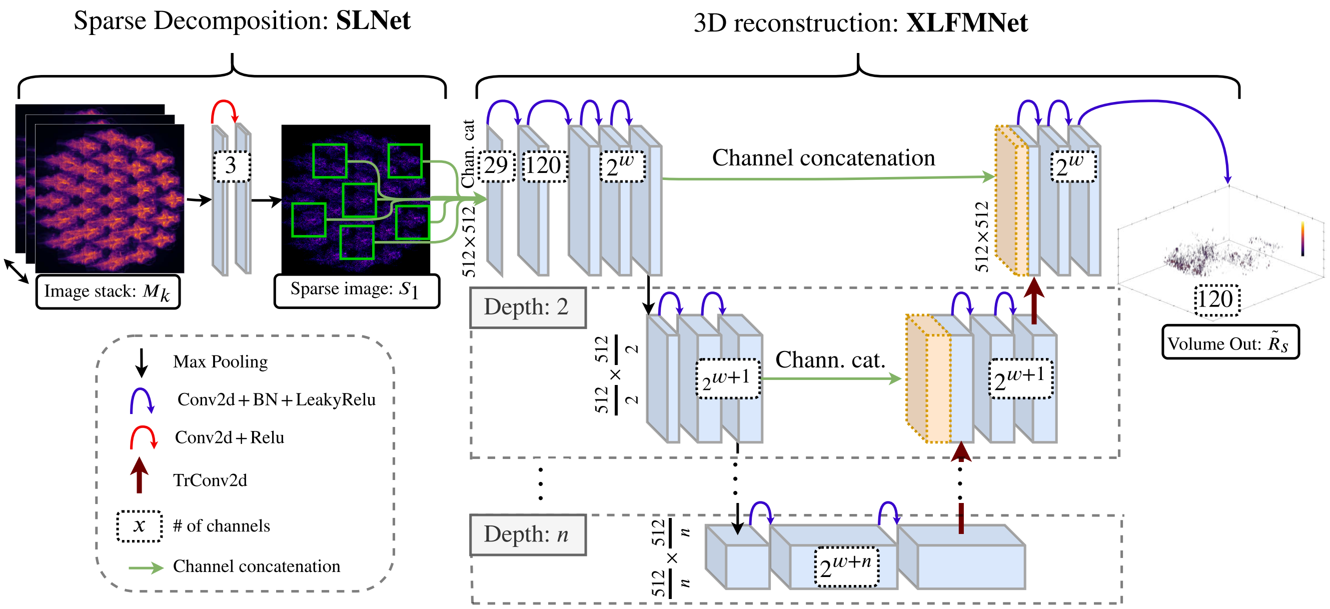


Fig. 2: Proposed network architecture, where the SLNet performs a sparse decomposition of an image time series and XLFMNet the 3D reconstruction of the sparse component. The number of depths can be controlled by the parameter n and the amount of channels per convolutional layer per depth is controlled by the parameter w.

2.3.1 *Network selection criteria towards generalization*

We look for a network that could perform well with unseen images of the same sample (validation set) as well as on unseen samples (testing set). To achieve this we employed several performance criteria:

- PSNR on the test set reconstruction results,
- SSIM in volume space (compared to a conventional deconvolution algorithm),
- SSIM in image space (see below).

The image space metric is computed by forward projecting the reconstructed volume ($\tilde{R}_s$) by means of the image formation model $i_s = \tilde{R}_s \circledast PSF$, and then by comparing the generated image against the sparse image (S) used as input to the XLFMNet. This is possible as the PSF of the XLFM setup was measured prior to the experiments.

2.3.2 *Generalization to non-observed samples*

Generalization in neural networks is not trivial to achieve, as the networks tend to learn the statistics of the observed images only. The degree of overfitting of a network to a dataset depends, among other factors, on the type and size of the dataset and on the number of trainable parameters of the network. To evaluate the generalization capability of our networks we build a testing dataset consisting of several zebrafish and microspheres data acquisitions (see section 2.6). With our architecture parametrization we perform a systematic grid search of the parameters and evaluate the resulting networks' generalization capabilities using a training data set of 100 images. The results for the grid search and the final training can be found in Table 2 and 3, respectively.

2.4 Dataset generation for XLFMNet training

In this section we describe the dataset preparation for the XLFMNet, which contains pairs of XLFM images and sparse volumes (see Fig. 3 for an overview).

An important aspect for achieving generalization is the close resemblance between the data used for training and the data used at validation or test time. However, capturing a dataset of zebrafish with enough variation is a time-consuming task that we try to avoid. In this work, we construct the training dataset by using a single time sequence of a zebrafish, as explained by the following steps (see also Fig. 3):

1) Capture a time series of a fluorescent specimen with the XLFM.
2) Generate a sparse image (S) per frame by means of a pre-trained SLNet.
3) Crop the 29 microlenses images using the coordinates of the sensor response at the central depth from the measured PSF.
4) Apply 3D deconvolution to each time step, using the method from [17].
5) Perform data augmentation to the reconstructed volumes and forward project them to image space.
6) Store the dense images (M) and the sparse volumes (R_s) into the dataset.

The augmentation consisted of random 3D rotations, translation and scaling of the volumes R_m and R_s. However, in order to increase the resemblance of the simulated images to those captured with a real microscope, there are two key aspects to take into consideration:

- **Noise augmentation:** Fluorescence microscopes suffer from noise when acquiring an image, mainly shot noise. Thus, it is important to add the proper shot

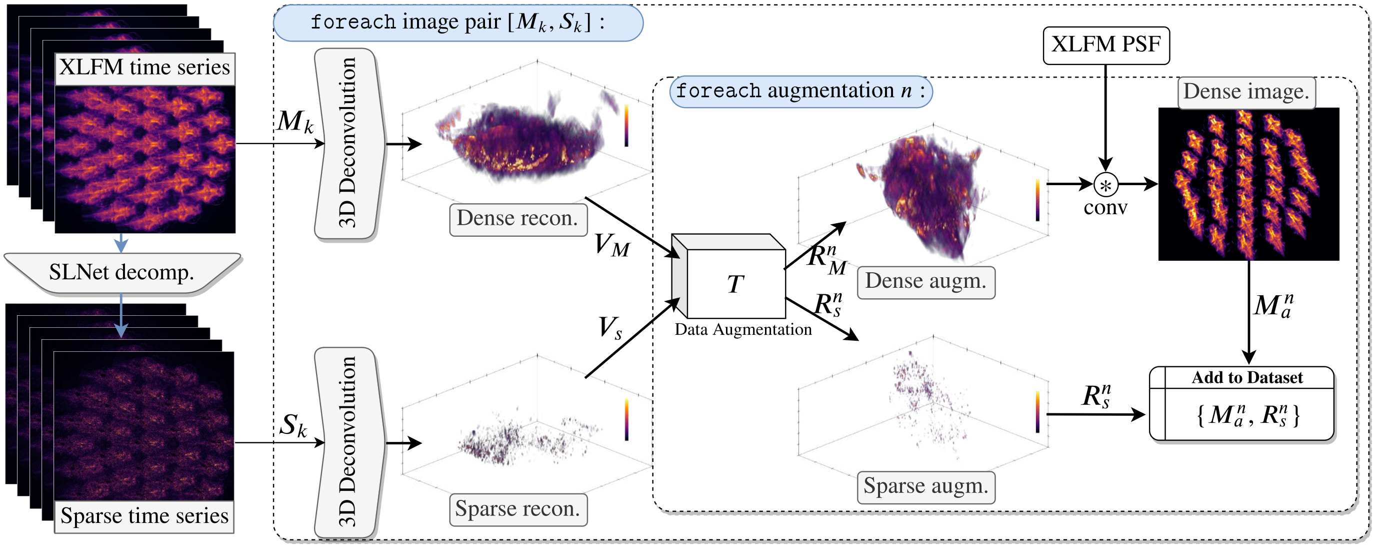


Fig. 3: XLFMNet training data generation pipeline: From left to right, 3D reconstruction is applied to every image k in a time series. The sparse component of the time series is also extracted with the SLNet and reconstructed. In the middle, for each augmentation n, both the dense the sparse reconstructions are fed to T, where the same random transformation is applied to both volumes. The dense augmented volume R_M^n is forward projected to image space. In the last step the dense image M_a^n and the sparse volume R_s^n are stored to be used for training the XLFMNet.

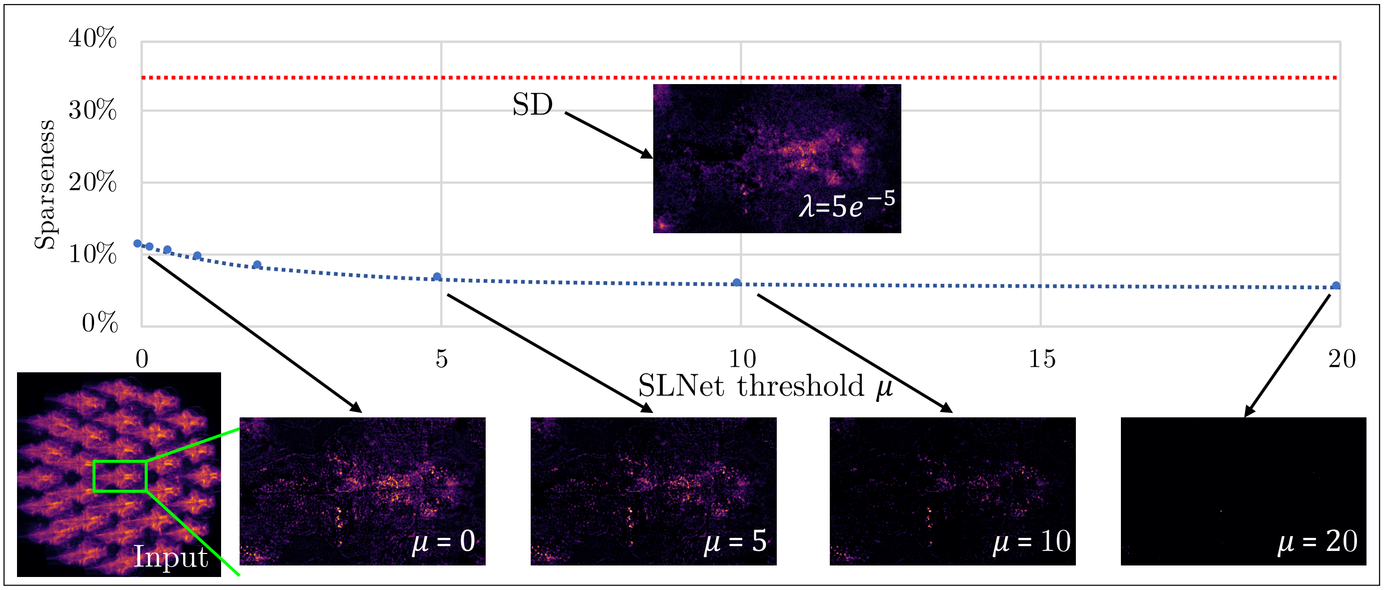


Fig. 4: Comparison between SLNet trained with different μ values and the SD method based on the augmented Lagrangian.

and background noise to the simulated dense images M. Additionally, as the microscope might be used for imaging samples with very low or very high counts, we augmented M by rescaling its pixel intensities to a random signal power (between 30^2 and 70^2 ADU). The noise is then added using the camera specifications.

- **Sample axial distribution:** To avoid training bias towards certain depths we apply a random z translation of the sample, such that in a subset of the samples only a couple of structures are present at depths that are far from the focal plane. We found this to be of crucial importance, due to the nature of the PSF, where the light gathered by the camera pixels decrease for planes far away from the focal plane, often biasing the network towards learning the higher intensity structures near the focal plane. By applying this random shifts, we ensure that there is enough structure in planes far away from the focal plane for the network to learn.

2.5 Extended field-of-view light field microscope (XLFM)

For our imaging hardware (see Fig. 1), we built an XLFM setup as described in [21]. The microscope used a 16× 0.8 NA water dipping objective lens (CFI75 LWD 16 × W,

XLFMNet ablation results						
U-net depth (n)	Channel exponent (w)	# parameters	PSNR volume	SSIM volume (%)	SSIM reproj. (%)	Time (Hz)
2	4	76K	23.92	98.69	61.18	**88.10**
2	5	171K	24.31	98.98	70.29	79.36
2	6	511K	25.36	98.97	71.20	63.37
2	7	1.794M	25.59	**99.00**	**72.64**	45.06
3	4	168K	23.56	98.87	70.31	82.65
3	5	537K	25.34	98.87	67.27	71.45
3	6	1.972M	25.40	98.86	70.08	52.5
3	7	7.631M	**25.99**	98.97	72.43	29.42
4	4	533K	24.49	98.82	67.76	79.38
4	5	1.997M	25.18	98.90	65.54	67.06
4	6	7.809M	25.42	98.92	73.37	46.75
4	7	30.972M	25.91	98.94	72.48	24.00
5	4	1,994K	24.37	98.81	67.73	77.16
5	5	7.835M	25.31	98.96	71.60	62.93
5	6	31.150M	24.92	98.90	70.48	40.04

TABLE 2: Results of XLFMNet ablation study, as described in section 2.3.2. The row in gray is the setting used for our final tests, achieving the best performance in two out of the four performance metrics.

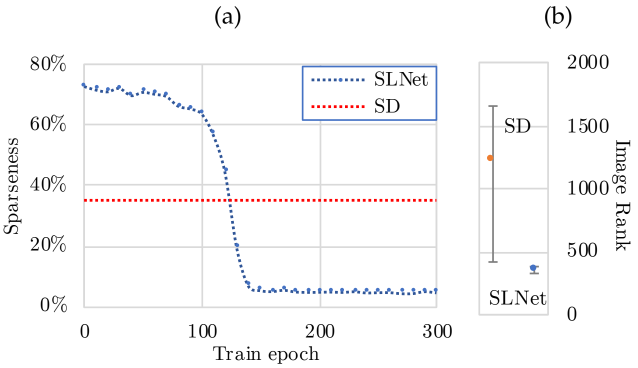


Fig. 5: **(a)** Sparseness progression during the SLNet training with $\mu = 2$ and compared against the SD result using the augmented Lagrangian method. **(b)** Mean rank comparison and min/max results between the SD method and SLNet with $\mu = 2$ when evaluating 12 images in the test set.

Nikon) for excitation and detection. The excitation light generated by a blue LED ($\mu = 470nm$, M470L4, Thorlabs) was collimated and then passed a $480nm$ short-pass filter before being reflected into the back pupil of the objective lens by a dichroic mirror (FF495-Di03-25 × 36, Semrock). A customized microlens array (29 lenses, $f = 35.4mm$ or $36.6mm$) was mounted on a sCMOS camera (Zyla 5.5 sCMOS, Andor), with the camera sensor at the focal plane of the microlenses. The microlens array was conjugated with the back pupil plane of the objective lens through a $4f$ relay lens pair ($f1 = 180mm$, AC508-180-A-ML, Thorlabs; $f2 = 125mm$, PAC074, Newport). A $525/50nm$ band pass filter (FF03-525/50-25, Semrock) was attached on the microlens array for green fluorescent imaging. The system point spread function (PSF) was measured by taking a $600\mu m$ thick image z-stack of a $1\mu m$-diameter green fluorescent bead located at the center of the field of view with an axial step size of $2.5\mu m$.

2.6 Samples preparation

2.6.1 *Zebrafish preparation for imaging*

Pan-neuronal nuclear localized GCaMP6s Tg(HuC:H2B:GCaMP6s) and pan-neuronal soma localized GCaMP7f Tg(HuC:somaGCaMP7f) [34] zebrafish larvae were imaged at 4–6 days post fertilization. The transgenic larvae were kept at $28°C$ and paralyzed in standard fish water containing $0.25mg/ml$ of pancuronium bromide (Sigma-Aldrich) for 2 min prior to imaging to reduce motion. The paralyzed larvae were then embedded in agar with 0.5% agarose (SeaKem GTG) and 1% low melting point agarose (Sigma-Aldrich) in Petri dishes. Fish water was added to the dishes once the agar solidified. All procedures involving animals at the Massachusetts Institute of Technology (MIT) were conducted in accordance with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the MIT Committee on Animal Care.

2.6.2 *3D fluorescent bead samples preparation*

To better evaluate the performance of our method, we imaged $1\mu m$-diameter green fluorescent beads (ThermoFisher) randomly distributed in 1% agarose (low melting point agarose, Sigma-Aldrich). The stock beads were serially diluted using melted agarose to 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6} of the original concentration. The diluted beads-agar colloid was then transferred to small Petri dishes to gel. The thicknesses of solidified bead samples were approximately $800\mu m$, which were sufficiently large to cover the full axial field of view of the microscope.

3 Experimental Results

3.1 Sparseness threshold and the SLNet

Controlling the degree of sparseness in the reconstructions produced by the SLNet is possible through the term μ from eq. (2). μ dictates how the eigenvalues of the image M get shrunk and thresholded, forming a low rank representation. This behavior was clearly visible in our experiments, where we evaluated different SLNet networks trained on a subset of 20 temporal 2D stacks and tested on 12 unseen ones. We tested networks with a threshold equal

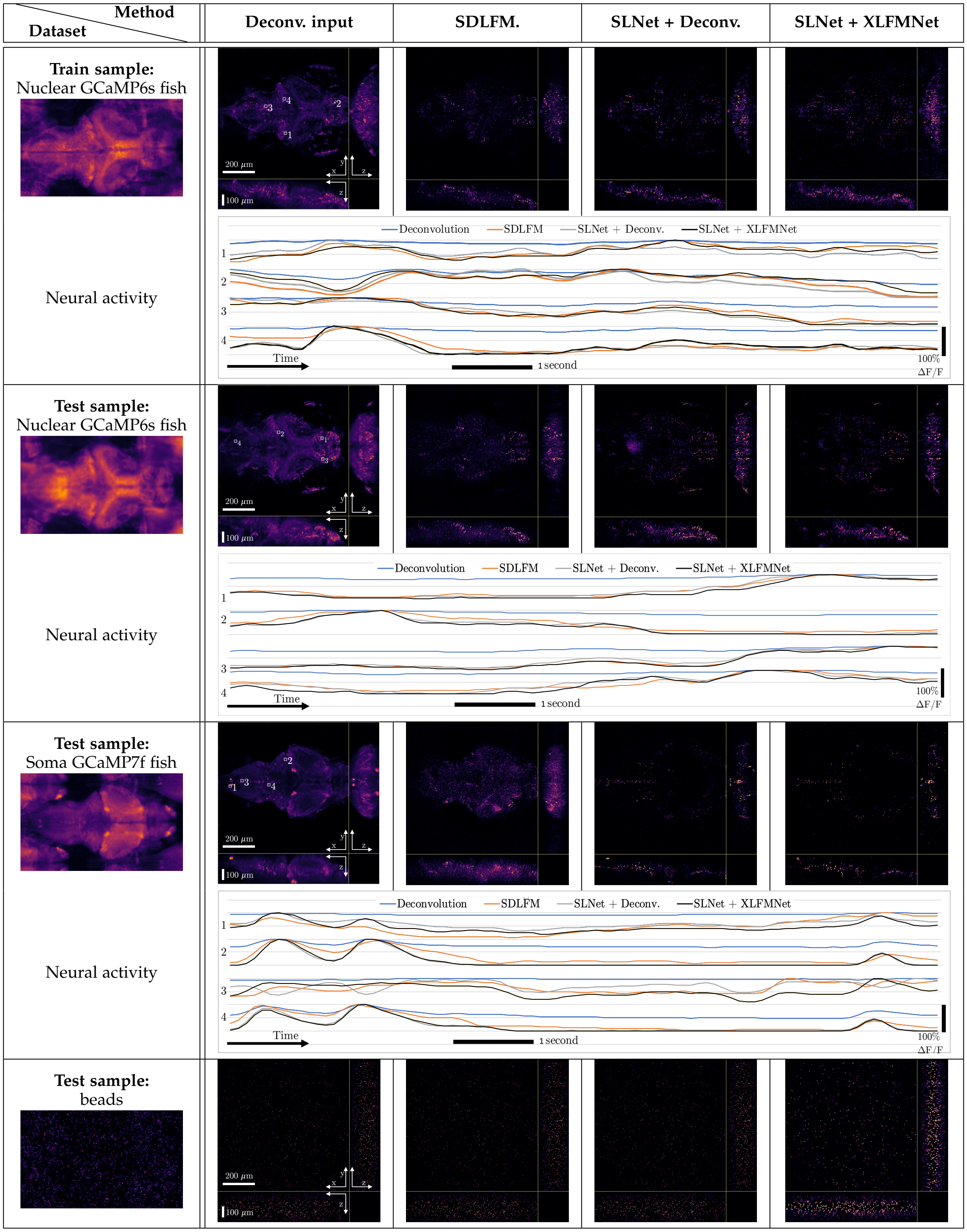


TABLE 3: 3D reconstruction of different samples. 100 frames of the first sample were used for training, as described in section 2.4. The next 100 frames and the first 100 frames of the remaining 3 samples were used for testing. The neural activity sections are taken from the white areas numbered 1 to 4 for each sample.

to $\mu = \{0.0, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0, 20.0\}$. Fig. 4 displays how the different choices of μ influence sparseness, which we define as

$$\text{sparseness} = \frac{\text{\# non-zero elements}}{\text{\# total elements}} \times 100\%. \qquad (4)$$

Our results corroborate the intuition behind the unsupervised training approach (see section 2.2), as the sparseness increases with an increasing μ. This makes this parameter a user-friendly way of controlling the sparseness of the network.

An interesting finding is that even with $\mu = 0.0$, which corresponds to neither shrinkage nor thresholding, the network produces a low rank solution in the spatial domain, which is an excellent starting point for the SLNet to focus mostly on refining the temporal sparseness. Our interpretation is that the blurring nature of convolutions in convolutional neural networks (CNN) is a good fit for this task. In Fig 5 (b) a comparison of rank of the decomposed images with the SD method against the SLNet are presented. The mean, min and max rank is shown for an evaluation of the 12 test images previously described. The rank of the images produced by SLNet are distributed in a small region, which shows the robustness of the proposed method across the sample space.

Fig. 5 shows how the sparseness of a network (with $\mu = 5.0$) decreases across training epochs. We found it useful that the user can decide the level of sparseness for a given application by storing the state of intermediate training steps. However, if the SLNet is trained for too long, eventually the sparseness is too high to be useful, and generates very low-contrast images as a result. We consider that for our images, a sparseness of 5% suffices for the 3D reconstruction to perform optimally.

3.2 Network ablation towards generalization

The network evaluation strategy consists on training on the first 100 frames of a $10Hz$ capture of a single zebrafish, and test in the following 100 frames of the same fish and in two other fish, with different fluorescent labeling and age, as described in section 2.6. As can been seen in Fig. 3 in the supplementary material, our camera's frame rate is higher than the calcium dynamic of both a single event (decay time $> 200ms$) and higher than events of sustained activity (e.g., bursting neurons) that can appear as a long activation. However, higher frame rates up to $40Hz$ can be achieved without modifying the setup.

Training the XLFMNet on a workstation with a Nvidia Quadro RTX 6000 graphic card takes around 5 hours for 500 epochs using the Adam optimizer. The possibility of retraining the network for every new sample is at reach. However, in daily microscopy work, one would avoid retraining the network often. Also, the amount of graphic memory required to train it efficiently is too large for regular computers. In our case, we focus on crafting a network robust enough to work with different samples, without compromising the data fidelity. Hence, reducing the network retraining frequency.

The results of the XLFMNet ablation (see Table 2) show that the number of trainable network parameters has a direct relation to the achievable reconstruction quality (e.g., the SSIM of the volume), however, when the parameters are spread across deeper networks, the performance decreases. If we compare, for example, two networks with a similar amount of parameters, but different depths (n) in the U-net, such as the XLFMNet with $n = 2$ and $w = 7$ versus the same network with $n = 5$ and $w = 4$, we find that the first one performs better in all performance metrics.

3.3 3D reconstruction of seen and unseen samples

When evaluating neural networks with biological data it is of crucial importance that the information of the neural activity is preserved, no matter the method used for 3D reconstruction. In other words, the network should not introduce non-physical artifacts that might hamper the analysis of the recovered signal.

In Fig. 3 we evaluated the neural activity across a period of 100 seconds by applying the following methods:

- Deconvolution of the raw image,
- SDLFM (SD + Deconvolution),
- SLNet + Deconvolution,
- SLNet + XLFMNet.

The SDLFM algorithm requires parameter tuning for optimal performance. Based on the original work [21] we used the Frobenious norm to find the best settings for the SD method with the training sample. For more information on this analysis, we refer the reader to the supplementary material. XLFMNet shows consistency with the neural activity detected with the other methods, by showing its reliability even on unseen fish samples.

3.4 Reconstructing fluorescent beads

When applying the different sparse decomposition methods to images of beads, we found that the output is quite similar to the input, as seen in the last row of Fig. 3. However, we can use the beads images to analyze how the XLFMNet infers unseen types of samples. In this experiment, first we applied conventional deconvolution to XLFM images of beads and then computed their FWHM for every detectable bead. The position of the beads was stored to extract the same information from a volume reconstructed with an XLFMNet trained on a single zebrafish.

In Table 4 we compare the FWHM of beads images when 3D reconstructed with deconvolution against XLMNet, together with the detection histogram per depth — in other words, how many beads were detected per depth — which helps us analyze the information trade-off of the proposed method. By looking at the FWHM plots, it is evident that XLFM suffers from resolution loss against deconvolution, mainly in the axial direction. However, as seen in the histogram and the missing beads plot, it keeps a detection rate comparable to that of the deconvolved volume.

3.5 Spatial resolution estimation

To further evaluate the image quality from the SLNet and XLFMNet, we estimated the 3D resolutions of the reconstructed larval zebrafish brain images based on 3D MTF analysis. The analysis results are shown in Fig. 6. We inferred the 3D resolutions from the spatial frequency support regions in the 3D MTFs, as enclosed in white dotted lines in Fig. 6. According to our estimation, XLFM has a lateral resolution of $\sim 4.3\mu m$ and an axial resolution of $\sim 10.0\mu m$, while SDLFM, SLNet+Deconv., and XLFMNet provided similarly enhanced resolutions of $\sim 3.2\mu m$ laterally and $\sim 7.7\mu m$ axially. These resolution values align well with the measured FWHMs of fluorescent beads in Table 4.

4 Discussion

The first contribution of our work is the SLNet, which performs the sparse decomposition of temporal raw XLFM

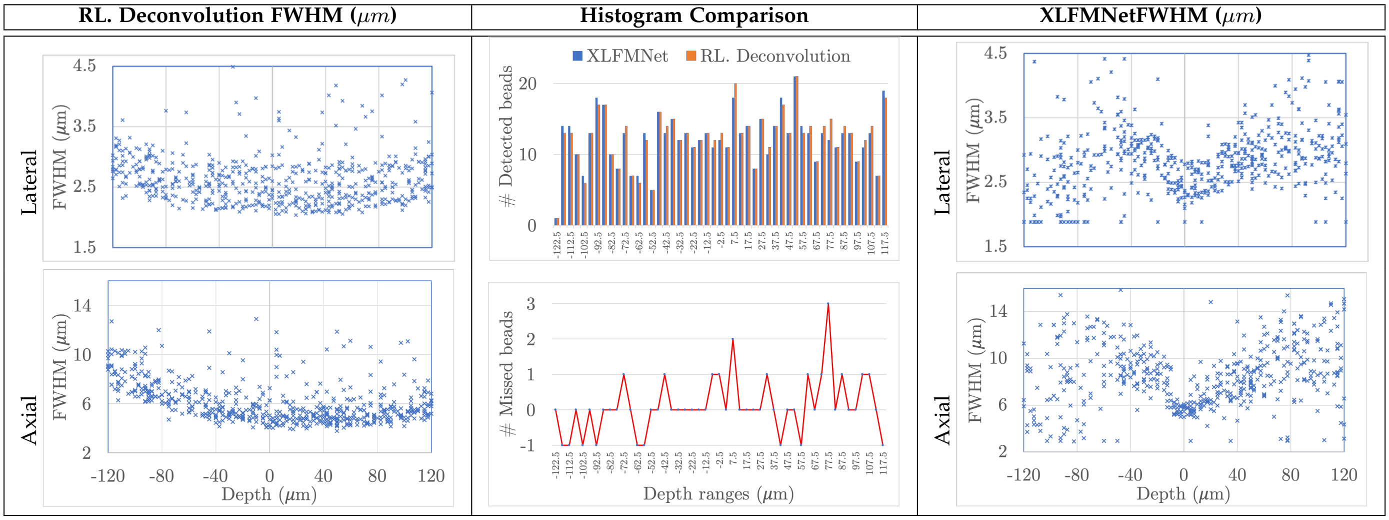


TABLE 4: Beads reconstructed with conventional deconvolution (left) and the proposed XLFMNet (right). The lateral and axial plots show the full width at half maximum for every depth, and a detection rate histogram, i.e., the number of beads found per depth with each method, and missing beads per depth.

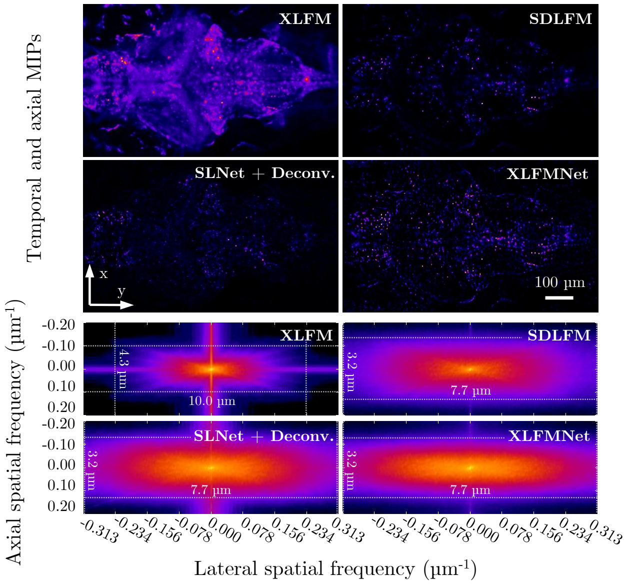


Fig. 6: Spatial resolution estimation for different reconstruction methods. Top panel shows the temporal and axial max intensity projections of reconstructed zebrafish brain volumes. The bottom panel are 3D MTFs (displayed in log scale) that show the spatial frequency support of each method.

stacks. During its training and evaluation we found that CNNs are a great choice for this task, due to their ability to produce blurry images that are already a low rank representation of the input in the spatial domain. Then, the SLNet focuses mostly on finding the sparseness in the temporal domain. Also, we found that the sparseness parameter μ used to shrink the principal components of the images during training is very user friendly as it serves as an intuitive way of controlling the sparseness of the reconstructions. We consider that this approach could be integrated into the microscopy workflow (e.g., as a Micro-Manager or ImageJ plugin). Another option would be to store multiple networks trained with different μ so that the user could decide the level of desired sparseness at test time.

The second contribution is XLFMNet, which reconstructs 3D volumes out of the sparse representations produced by the SLNet. We consider that the ablation study provided useful information regarding the generalization of the network to unseen samples. A network with higher number of parameters performs better than one with a lower number. However, these parameters should not be spread across too many down-convolutional steps. The network that we found to perform the best was the one with two down-convolutional steps ($n = 2$) and $w = 7$ channels, which was able to reconstruct volumes at a rate of $45.05 Hz$.

When evaluating the network with different samples, in Table 3 we found out that when trained on a zebrafish, it works well for other unseen zebrafish, and preserves the neural potentials with a similar pace as the other reconstruction methods. However, when imaging beads that are substantially different than fish, the network is able to reconstruct with high fidelity the central depths, but loses contrast when approaching depths farther away from the focal plane. The intuition behind this is that the beads that are far away spread their energy in a larger sensor area, which dims the individual pixels substantially. Nevertheless, the network could still detect the beads even at far away planes, as seen in the detection rate histogram comparison in Table 4, where the number of detected beads is quite similar as the one from the conventional deconvolution method.

A possible improvement would be to retrain a batch normalization block at the entrance of the network for every new sample type, while keeping the XLFMNet frozen. We leave this for future work.

5 Conclusion

In this manuscript we discuss a novel 3D sparse reconstruction method for XLFM images, which achieve a real-time performance ($45Hz$) on large volumes ($800 \times 800 \times 250 \mu m^3$) and can be used to infer previously unseen samples by preserving data fidelity and image quality.

Having such an algorithm operating in real-time, opens the possibility of a large array of closed loop experiments, where the biological sample stimulation and measurement are closely related (e.g., visual, auditory and olfactory stimuli, as well as optogenetics). We leave these directions as future work.

Acknowledgments

J. Page is supported by the Deutsche Forschungsgemeinschaft (LA 3264/2-1). Z. Wang acknowledges the Alana Down Syndrome Center. E. S. Boyden acknowledges Lisa Yang, NIH R01DA029639, NIH R01MH12297101A1, NIH RF1NS113287, NSF 1848029, NIH 1R01DA045549, and John Doerr. P. Favaro acknowledges the interdisciplinary project funding UniBE ID Grant 2018 of the University of Bern. We thank Owen Randlett at Claude Bernard University Lyon 1 and Florian Engert at Harvard University for sharing zebrafish brain image datasets.

References

[1] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," in *ACM SIGGRAPH 2006 Papers*, 2006, pp. 924–934.

[2] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden *et al.*, "Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy," *Nature methods*, vol. 11, no. 7, pp. 727–730, 2014.

[3] D. Wang, S. Xu, P. Pant, E. Redington, S. Soltanian-Zadeh, S. Farsiu, and Y. Gong, "Hybrid light-sheet and light-field microscope for high resolution and large volume neuroimaging," *Biomedical Optics Express*, vol. 10, no. 12, 2019.

[4] N. C. Pégard, H.-Y. Liu, N. Antipa, M. Gerlock, H. Adesnik, and L. Waller, "Compressive light-field microscopy for 3D neural activity recording," *Optica*, vol. 3, no. 5, 2016.

[5] C. Cruz Perez, A. Lauri, P. Symvoulidis, M. Cappetta, A. Erdmann, and G. G. Westmeyer, "Calcium neuroimaging in behaving zebrafish larvae using a turn-key light field camera," *Journal of Biomedical Optics*, vol. 20, no. 9, 2015.

[6] Z. Zhang, L. Bai, L. Cong, P. Yu, T. Zhang, W. Shi, F. Li, J. Du, and K. Wang, "Capturing volumetric dynamics at high speed in the brain by confocal light field microscopy," *bioRxiv*, 2020.

[7] Z. Wang, Y. Ding, S. Satta, M. Roustaei, P. Fei, and T. K. Hsiai, "A hybrid of light-field and light-sheet imaging to decouple myocardial biomechanics from intracardiac flow dynamics," *bioRxiv*, 2020.

[8] N. Wagner, F. Beuttenmueller, N. Norlin, J. Gierten, J. Wittbrodt, M. Weigert, L. Hufnagel, R. Prevedel, and A. Kreshuk, "Deep learning-enhanced light-field imaging with continuous validation," *bioRxiv*, 2020.

[9] N. Wagner, N. Norlin, J. Gierten, and G. D. Medeiros, "Instantaneous isotropic volumetric imaging of fast biological processes," *Nature Methods*, vol. 16, no. 6, pp. 497–500, 2019.

[10] J. Wu, Z. Lu, H. Qiao, X. Zhang, K. Zhanghao, H. Xie, T. Yan, G. Zhang, X. Li, Z. Jiang, X. Lin, L. Fang, B. Zhou, J. Fan, P. Xi, and Q. Dai, "3D observation of large-scale subcellular dynamics in vivo at the millisecond scale," *bioRxiv*, 2019.

[11] M. Shaw, H. Zhan, M. Elmi, V. Pawar, C. Essmann, and M. A. Srinivasan, "Three-dimensional behavioural phenotyping of freely moving C. Elegans using quantitative light field microscopy," *PLoS ONE*, vol. 13, no. 7, 2018.

[12] L. M. Grosenick, M. Broxton, C. K. Kim, C. Liston, B. Poole, S. Yang, A. S. Andalman, E. Scharff, N. Cohen, O. Yizhar, C. Ramakrishnan, S. Ganguli, P. Suppes, M. Levoy, and K. Deisseroth, "Identification Of Cellular-Activity Dynamics Across Large Tissue Volumes In The Mammalian Brain," *bioRxiv*, 2017.

[13] O. Skocek, T. Nöbauer, L. Weilguny, F. Martínez Traub, C. N. Xia, M. I. Molodtsov, A. Grama, M. Yamagata, D. Aharoni, D. D. Cox, P. Golshani, and A. Vaziri, "High-speed volumetric imaging of neuronal activity in freely moving rodents," *Nature Methods*, vol. 15, no. 6, 2018.

[14] P. Quicke, C. L. Howe, P. Song, H. V. Jadan, C. Song, T. Knöpfel, M. Neil, P. L. Dragotti, S. R. Schultz, and A. J. Foust, "Subcellular resolution three-dimensional light-field imaging with genetically encoded voltage indicators," *Neurophotonics*, vol. 7, no. 3, 2020.

[15] S. Aimon, T. Katsuki, T. Jia, L. Grosenick, M. Broxton, K. Deisseroth, T. J. Sejnowski, and R. J. Greenspan, "Fast near-whole-brain imaging in adult drosophila during responses to stimuli and behavior," *PLoS Biology*, vol. 17, no. 2, 2019.

[16] Q. Lin, J. Manley, M. Helmreich, F. Schlumm, J. M. Li, D. N. Robson, F. Engert, A. Schier, T. Nöbauer, and A. Vaziri, "Cerebellar neurodynamics predict decision timing and outcome on the single-trial level," *Cell*, vol. 180, no. 3, pp. 536–551, 2020.

[17] L. Cong, Z. Wang, Y. Chai, W. Hang, C. Shang, W. Yang, L. Bai, J. Du, K. Wang, and Q. Wen, "Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (*Danio rerio*)," *eLife*, vol. 6, p. e28158, sep 2017.

[18] G. Scrofani, J. Sola-Pikabea, A. Llavador, E. Sanchez-Ortiga, J. C. Barreiro, G. Saavedra, J. Garcia-Sucerquia, and M. Martínez-Corral, "FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples," *Biomedical Optics Express*, vol. 9, no. 1, 2018.

[19] C. Guo, W. Liu, and S. Jia, "Fourier-domain light-field microscopy," *Biophotonics Congress: Optics in the Life Sciences Congress*, 2019.

[20] W. Liu, C. Guo, X. Hua, and S. Jia, "Fourier light-field microscopy: An integral model and experimental verification," *Biophotonics Congress: Optics in the Life Sciences Congress*, vol. 27, no. 18, pp. 25 573–25 594, 2019.

[21] Y.-G. Yoon, Z. Wang, N. Pak, D. Park, P. Dai, J. S. Kang, H.-J. Suk, P. Symvoulidis, B. Guner-Ataman, K. Wang, and E. S. Boyden, "Sparse decomposition light-field microscopy for high speed imaging of neuronal activity," *Optica*, vol. 7, no. 10, p. 1457, 2020.

[22] N. C. Pégard, H.-Y. Liu, N. Antipa, M. Gerlock, H. Adesnik, and L. Waller, "Compressive light-field microscopy for 3d neural activity recording," *Optica*, vol. 3, no. 5, pp. 517–524, 2016.

[23] T. Nöbauer, O. Skocek, A. J. Pernía-Andrade, L. Weilguny, F. M. Traub, M. I. Molodtsov, and A. Vaziri, "Video rate volumetric ca 2+ imaging across cortex using seeded iterative demixing (sid) microscopy," *Nature methods*, vol. 14, no. 8, p. 811, 2017.

[24] J. Page, F. Saltarin, Y. Belyaev, R. Lyck, and P. Favaro, "Learning to reconstruct confocal microscopy stacks from single light field images," *arXiv*, 2020.

[25] X. Li, H. Qiao, J. Wu, Z. Lu, T. Yan, R. Zhang, X. Zhang, and Q. Dai, "DeepLFM : Deep Learning-based 3D Reconstruction for Light Field Microscopy," *Biophotonics Congress: Optics in the Life Sciences Congress*, 2019.

[26] Z. Wang, H. Zhang, L. Zhu, G. Li, Y. Li, Y. Yang, S. Gao, T. K. Hsiai, and P. Fei, "Network-based instantaneous recording and video-rate reconstruction of 4D biological dynamics," *bioRxiv*, 2019.

[27] R. Mizutani, R. Saiga, S. Takekoshi, C. Inomoto, N. Nakamura, M. Itokawa, M. Arai, K. Oshima, A. Takeuchi, K. Uesugi *et al.*, "A method for estimating spatial resolution of real image in the fourier domain," *Journal of microscopy*, vol. 261, no. 1, pp. 57–66, 2016.

[28] Q. Wang, Q. X. Gao, G. Sun, and C. Ding, "Double robust principal component analysis," *Neurocomputing*, vol. 391, pp. 119–128, 2020.

[29] L. Z., C. M., W. L., and Y. Ma., "The augmented lagrange multiplier method for exact recovery of a corrupted low-rank matrices." *Mathematical Programming, submitted*, 2009.

[30] Y. Xiaming and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *preprint*, 2009.

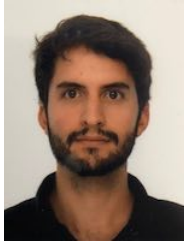
[31] C. Herrera, F. Krach, A. Kratsios, P. Ruyssen, and J. Teichmann, "Denise: Deep learning based robust pca for positive semidefinite matrices," 2020.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Com-*

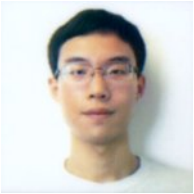
puting and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.

[34] O. A. Shemesh, C. Linghu, K. D. Piatkevich, D. Goodwin, O. T. Celiker, H. J. Gritton, M. F. Romano, R. Gao, C.-C. J. Yu, H.-A. Tseng *et al.*, "Precision calcium imaging of dense neural populations via a cell-body-targeted calcium indicator," *Neuron*, vol. 107, no. 3, pp. 470–486, 2020.



Josue Page Vizcaino was born in San Cristobal, Mexico, in 1990. He received his Bachelor degree from the Centro de Enseñanza Técnica Industrial, Mexico, in 2014, and the M.Sc. degree in biomedical computing from the Technical University of Munich in 2017.

Between 2011 to 2015 founded and worked as interactive programmer in Visualma. From 2014 to 2015 he was a graphics hardware engineer at Intel, Guadalajara, Mexico. He is currently doing his Ph.D with the Computational Imaging and Inverse Problems Group at the Technical University of Munich.



Zeguan Wang was born in Hebei, China, in 1995. He received his BS degree in applied physics from University of Science and Technology of China, Hefei, China, in 2018, and a MS degree in media arts and sciences from MIT, Cambridge, USA, in 2020.

From 2016 to 2017, he was a visiting student at the Institute of Neuroscience, CAS, in Shanghai, China. Since 2018, he has been a research assistant at MIT, focusing on optical technologies for neuroscience. He is currently a PhD student and an Alana Fellow at MIT.

Mr. Wang received Guo Moruo Scholarship and National Scholarship as an undergraduate.

Panagiotis Symvoulidis was born in Athens, Greece, in 1987. He received his Diploma M.Eng degree in electrical engineering and computer science from the National Technical University of Athens, Greece, in 2011, and the Dr. Eng. degree from Technical University of Munich, Munich, Germany in 2018.

During the period of 2011 to 2018, he held appointments as Research Scientist at Helmholtz Zentrum Munich, Technical University of Munich, and Klinikum Rechts der Isar in Munich, Germany. He engineered imaging systems and pipelines for both pre-clinical and clinical applications. From 2018 and until now he is PostDoctoral Associate at Massachusetts Institute of Technology, Cambridge, MA, USA, focusing on brain imaging-related techniques and discoveries.

Dr. Symvoulidis was involved in the activities of IEEE Student Branch of the National Technical University of Athens as an undergrad and served as Treasurer of the local Board in 2009.

Burcu Guner-Ataman was born in Izmit, Turkey in 1976. She received her BS degree in Molecular Biology and Genetics from the Bosphorus University, Istanbul, Turkey in 2001. She received her Ph.D. in Cellular and Molecular Biology from University of Massachusetts, Amherst in 2008, where studied the role of morphogen gradients in the forebrain and spinal cord patterning. She held Postdoctoral Fellow and Instructor appointments in the Cardiovascular Research Center at Massachusetts General Hospital, Boston, MA between 2008-2018. She addressed molecular mechanisms underlying congenital heart defects.

She is currently a Research Scientist at the Synthetic Neurobiology group at Massachusetts Institute of Technology, Cambridge, MA, where she is focused on developing molecular tools to address how the brain functions in health and disease.

Dr. Guner-Ataman received a Postdoctoral Fellowship and a Scientist Development Grant from the American Heart Association.

Paolo Favaro (Member, IEEE) was born in Montebelluna (TV), Italy, in 1973. He received the Laurea degree (B.Sc. + M.Sc.) in computer engineering from Università di Padova, Padova, Italy, in 1999, and the M.Sc. and Ph.D. degree in electrical engineering from Washington University in St. Louis, St. Louis, MO, USA, in 2003 and 2004 respectively. In 2004 he was a postdoctoral researcher in the computer science department of the University of California, Los Angeles, Los Angeles, CA, USA, and subsequently in Cambridge University, Cambridge, UK.

Between the end of 2004 and 2006 he worked in medical imaging at Siemens Corporate Research, Princeton, USA. From 2006 to 2011 he was Lecturer and then Reader at Heriot-Watt University and Honorary Fellow at the University of Edinburgh, Edinburgh, UK. In 2012 he became full professor at Universität Bern, Bern, Switzerland.

Prof. Favaro is associate editor for the IEEE Transactions on Pattern Recognition and Machine Intelligence. His research interests are in computer vision, computational photography, machine learning, signal and image processing, estimation theory, inverse problems and variational techniques.

Edward S. Boyden, III was born in Plano, TX, USA, in 1979. He earned two BS degrees in physics and electrical engineering and computer science from MIT, in 1999, the MEng degree in electrical engineering and computer science from MIT, in 1999, and the PhD degree in neurosciences from Stanford, in 2005.

From 2006 to 2020, he was a visiting scientst, then assistant professor, then associate professor, then full professor, at MIT, working on tools for analyzing and controlling biological systems. From 2020 to the current day, he has been Y Eva Tan Professor in Neurotechnology at MIT and an investigator of the Howard Hughes Medical Institute (Cambridge, MA, USA), working on ways to solve the brain and other complex systems.

Dr. Boyden is an elected member of the National Academy of Sciences, and winner of the Canada Gairdner International Award and the Breakthrough Prize in Life Sciences.

Tobias Lasser (Member, IEEE) was born in Munich, Germany, in 1979. He received the Diplom-Informatiker degree and the Diplom-Mathematiker degree from the Technical University of Munich, Munich, Germany, in 2006 and 2008, respectively, the Dr. rer. nat. degree in informatics from the Technical University of Munich in 2011, and the Dr. rer. nat. habil. degree in informatics from the Technical University of Munich in 2017.

From 2005 to 2006 he was with the Massachusetts General Hospital, Harvard University, Boston, MA, USA. From 2006 to 2020 he was with the Department of Informatics, Technical University of Munich, Munich, Germany. He is currently a lecturer with the Department of Informatics and the Munich School of BioEngineering at the Technical University of Munich.

Dr. Lasser is currently an associate editor for the IEEE Transactions on Computational Imaging.