UNIVERSITY OF CALIFORNIA

Los Angeles

High-speed Volumetric Functional Imaging with Light Field Microscopy

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Bioengineering

by

Zhaoqiang Wang

2024

ABSTRACT OF THE DISSERTATION

High-speed Volumetric Functional Imaging with Light Field Microscopy

by

Zhaoqiang Wang

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2024

Professor Tzung Hsiai, Co-Chair

Professor Liang Gao, Co-Chair

The continuous advancement in microscopy has been unveiling the hidden world of tissues, cells, and molecules. In the quest for deeper spatiotemporal insights into biological processes, light field microscopy (LFM) has emerged as a powerful and intriguing tool. Unlike traditional imaging systems that capture focused images, LFM records multiplexed signals with single snapshot that encodes information within a three-dimensional (3D) volume. By leveraging computational

reconstruction algorithms, this approach enables the observation of transient volumetric dynamics with remarkable efficiency and speed.

This thesis presents a series of efforts to apply LFM in functional imaging, enabling researchers to monitor real-time changes in live organisms, including ion fluxes, electrical signaling, and cells interactions. The exceptional temporal resolution makes LFM a unique tool to visualize and analyze rapid processes that are difficult to capture with conventional 3D microscopy. We demonstrated calcium imaging of motor neurons in freely moving C. elegans and tracked flowing blood cells *in-vivo* within a beating zebrafish heart. The excessive and unpredictable motion observed in these processes requires capturing hundreds of 3D volumes per second, a demanding but necessary task that provides insights into the underlying mechanisms of neural and cardiac functions. We further moved forward to voltage imaging, a frontier in neuroscience, which directly measures the neural action potential as well as sub-thresholding activities. Our LFM provides kilohertz volumetric imaging on leech ganglion and mouse hippocampus. It measures 7.3 gigavoxels per second in a 3D field of view of $550 \times 550 \times 300\ \mu m^3$, which makes it capable of recording the accurate timing and waveform of neural spikes across entire volume.

These demonstrations are achieved through several innovative redesigns of LFM, detailed in Chapter 3 to 5. The first approach, VCD-LFM, addresses the inherent trade-off between spatial resolution and depth information in light field imaging by introducing a learning-based reconstruction algorithm. By incorporating data priors and constraints, this method aims to mitigate the issues of low spatial resolution and artifacts in conventional LFM without compromising imaging speed. The second approach, Squeezed Light Field Microscopy (SLIM), leverages data redundancy in light fields and revises the optical hardware to achieve kilohertz volume rate. Designed to meet the high-speed demands of voltage imaging, SLIM offers a

powerful and robust imaging tool for sparse volumetric processes. Lastly, the third approach, Light Field Tomography (LIFT), adapts LFM for one-dimensional (1D) measurements through optical Radon transformation. This method enables the use of low-dimensional detectors, such as line sensor, to capture high-dimensional light fields, resulting in enhanced sensitivity, reduced cost and even greater temporal resolution.

The dissertation of Zhaoqiang Wang is approved.

Peng Hu

Jau-Nian Chen

Jun Chen

Tzung Hsiai, Committee Co-Chair

Liang Gao, Committee Co-Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

Acknowledgements

I am incredibly fortunate to have had the opportunity to work with Prof. Liang Gao and Prof. Tzung Hsiai. Their insights, guidance and tremendous support have been instrumental to the completion of this work. Liang shaped me as an optics researcher through countless, generous, humble and open-minded discussions. He created a lab environment that felt like a playground for students, where I found the freedom to explore my ideas and passions with great joy. Tzung motivated me to delve deeper into the purpose of the tools I build. His expertise in cardiology and his keen interest in biological studies have driven my work to become useful for truly biological discovery. He has generously shared his career experiences and perspectives, which have rooted deeply in my mind, shaping me into a better researcher with a profound sense of gratitude, respect, and responsibility.

I would like to thank Prof. Peng Hu, Prof. Jau-Nian Chen, and Prof. Jun Chen for serving on my Ph.D. committee. Their diverse research background and kind support have been invaluable assets to this work, providing guidance and insights that have refined it in many ways.

Additionally, I want to extend my profound gratitude to Prof. Peng Fei, who has been my advisor, my mentor and my friend since my time as an undergraduate in his lab. When he first showed me a Lytro light field camera from his collection, neither of us could have imagined the journey that lay ahead. During those pivotal turning points in my Ph.D., his overseas phone calls were always a harbor for my wandering and hesitant mind. His optimism, perseverance, patience,

and subtlety make him a role model I deeply admire, and the reason I began—and, if fortunate, will continue—my academic career.

This work wouldn't be possible without my collaborators. Lanxin Zhu and Ruixuan Zhao were incredible peers with whom I've had the pleasure of working. They helped me push forward two important projects, VCD-LFM and SLIM, through bushes and mud until the end. I am grateful for their years of tolerance, as well as their faith in me. Prof. Daniel A. Wagenaar endured a year of grueling traffic between Caltech and UCLA to perform voltage imaging with us. Wenjun Kang and Prof. Rongguang Liang generously donated their time and expertise in lens fabrication, iterating through many versions of our lenslet array. Yuan Dong has been a kind helper in the UCLA fish core over the years, teaching me everything about fish handling and often greeting me with delicious oranges from her home garden. I am thankful for the support of the Amazon Science Hub, which generously funded my project through a year-long fellowship. Haofu Liao, as my fellowship mentor, broadened my perspective on deep learning methods through our monthly meetings. I also want to extend my thanks to the collaborators who are willing to work with our microscopes on their animal models, including Prof. Peyman Golshani, Prof. Gerard Wong, Prof. Shangbang Gao, and many others.

Finally, I would like to express my heartfelt appreciation to my friends, my partner and my parents. For the loosened guitar string, a fleeting and fading rock band formed in our spiritual holy war against pandemics; for the storm-laden highways of Texas, chasing the shadow of total eclipse with the beloved; for the silver strands of hair, your timeless and boundless love and encouragement; this dissertation is dedicated to you, to the good memory, to the journey beyond.

Curriculum Vitae

Education

2018 – 2024	Ph.D. Candidate in Bioengineering University of California, Los Angeles
2014 – 2018	B.E. in Optoelectronic Information Science and Engineering Huazhong University of Science and Technology, China

Selected Publications

1. **Wang, Z.**, Zhao, R., Wagenaar, D.A., Kang, W., Lee, C., Schmidt, W., Pammar, A., Zhu, E., Wong, G.C., Liang, R., Hsiai, T. and Gao, L., Kilohertz volumetric imaging of in-vivo dynamics using squeezed light field microscopy. ***bioRxiv*** (2024)
2. **Wang, Z.**, Hsiai, T., Gao, L., Augmented light field tomography through parallel spectral encoding. ***Optica*** (2023)
3. **Wang, Z.**, Zhu, L., Zhang, H., Li, G., Yi, C., Li, Y., Yang, Y., Ding, Y., Zhen, M., Gao, S., Hsiai, T.K. and Fei, P., Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. ***Nature Methods*** (2021)
4. **Wang Z,** Ding Y, Satta S, Roustaei M, Fei P, et al. A hybrid of light-field and light-sheet imaging to study myocardial function and intracardiac blood flow during zebrafish development. ***PLOS Computational Biology*** (2021)

Chapter 1
Introduction

The dynamic nature is a hallmark of life. In its continuous transformation, movement and evolution, the living systems create biological order from the fundamental elements and adapt themselves in the changing environment. This inherent dynamism presents an ill-posed problem to infer true physiological relevance solely from structural imaging of fixed specimens. The unbiased understanding of life must rely on the continuous observation of the spatiotemporal processes displayed by the organism in their natural state.

This thesis focuses on three-dimensional (3D) processes occurring at millisecond scale. This temporal scope encompasses several vital functions that sustain life and coordinate response to external stimuli, such as the cardiovascular processes and neural activities. For instance, zebrafish embryos, a widely used model organism, have a heart rate of up to 120 beats per minute. Within just a few hundred milliseconds, their heart completes a complex electrochemical process involving the synchronized propagation of electrical signals and muscle contraction. The pumped blood interacts with the deforming heart chambers, imposing varying mechanical forces based on the local geometry. To elucidate such cardiac function, visualization requires a sampling rate in the

hundreds of Hz across the entire 3D heart, without interfering with the normal heartbeat. However, high-speed volumetric *in vivo* imaging has yet to become a readily available tool in the biological research community.

Conventional 3D fluorescence microscopes, such as confocal and light-sheet microscopes, capture entire volumes by scanning a focused laser spot or a thin plane, respectively. The speed of these serial approaches is largely constrained by the bandwidth of mechanical scanners and camera sensors, leading to an inherent trade-off between temporal resolution, 3D field of view (FOV), and spatial resolution. Additionally, high-speed imaging inevitably reduces the signal-to-noise ratio (SNR) of measurements. This occurs because the already limited photon budget of fluorescence is further diminished due to shorter exposure times, increasing the demands on the sensitivity and noise performance of the image sensors. Due to these challenges, very few microscope systems have achieved fluorescence detection across 3D volume at a temporal resolution of milliseconds.

To address the challenges, this dissertation discusses the application of a novel detection strategy, namely light field imaging, in the high-speed function microscopy. As a single-shot computational imager, LFM encodes 3D information onto 2D multiplexed measurements within one camera exposure, and digitally reconstructs the 3D volume in post-acquisition. This scanless approach fundamentally alleviates the bandwidth limitations in conventional methods. And the multiplexed signals also simultaneously increase the SNR of single measurement, which makes LFM a promising tool for high-speed fluorescence detection.

In this thesis, I will first introduce the basic concepts of light field microscopy (LFM) in Chapter 2, covering its optical design, reconstruction methods, and the intrinsic trade-offs involved. To address the limitations of LFM, Chapter 3 proposes a novel reconstruction algorithm that significantly enhances spatial resolution without compromising imaging speed. Chapter 4 presents

an innovative hardware design that reduces data load in LFM, enabling kilohertz imaging capabilities. Chapter 5 extends LFM to a line sensor configuration, further increasing imaging speed. We successfully demonstrate volumetric functional imaging at rates ranging from one hundred to three thousand Hz in various animal models, including *C. elegans*, zebrafish larvae, bacteria, and mice. The potential and promise of LFM for high-speed microscopy are thoroughly explored.

Chapter 2
Light field microscopy

Light field, or plenoptic function, refers to a complete description of light in the given space. It attempts to characterize all the visual information, embodied by a seven-dimensional (7D) function that records the intensity of ray of every spatial location (x, y, z), angle (θ, ϕ), wavelength λ and time t[1]. The modern light field photography and microscopy usually deal with the subset of original plenoptic function, where only spatial and angular information are considered. In contrast to conventional 2D imaging, the additional angular information allows us to back trace the light and reconstruct the irradiance image at different depths in post-processing. This 3D reconstruction ability, also known as synthetic refocusing, removes the necessity of depth scanning in traditional 3D imaging, thus dramatically increasing the speed of volumetric detections in various applications such as particle velocimetry, cell imaging, cardiovascular imaging and neural imaging.

2.1 Fundamentals of light field imaging

The monochromatic ray space at a certain time inside the camera can be modeled by a two-plane light field, $\mathcal{L}(u, v, s, t)$, assuming that light doesn't attenuate when traveling in free space. We put

the (u, v) plane at the exit pupil and (s, t) plane at the sensor (**Figure 1a**). With this definition, the irradiance image value that would appear on the sensor will be given by[2]:

$$E(s,t) = \frac{1}{D^2} \iint \mathcal{L}(u,v,s,t)\, A(u,v) \cos^4 \theta \, dudv, \tag{1}$$

where D is the distance between two planes, A is the aperture function that defines one within the pupil and zero outside it, and θ is the angle of incidence that ray (u, v, s, t) makes with the sensor plane.

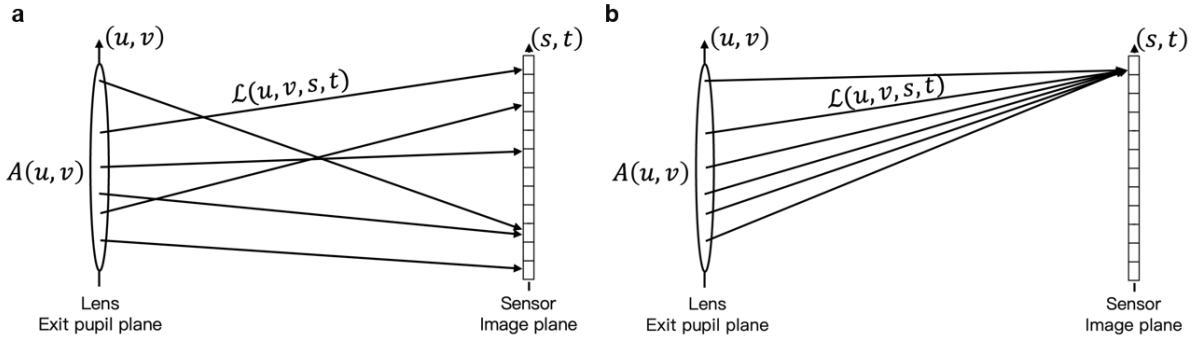

Figure 1. The two-plane parameterization of light field.

With a paraxial approximation and ignoring the constant D, we can further simplify the imaging equation to

$$E(s,t) \approx \iint \mathcal{L}(u,v,s,t)\, A(u,v) dudv \tag{2}$$

It becomes intuitive that the irradiance at a certain location (s, t) is calculated by integrating over all rays from different angles that are indexed by (u, v) (**Figure 1b**). Instead of physical sensor plane, we can also consider a virtual image plane (s', t') that has been shifted away by a ratio α (**Figure 2a**). A new light field function $\mathcal{L}'(u, v, s', t')$ can be defined so that the irradiance

image is computed by equation (2). **Figure 2a** further illustrates the relationship between $\mathcal{L}'$ and $\mathcal{L}$ as

$$\mathcal{L}'(u,v,s',t') = \mathcal{L}\left(u,v,u+\frac{s'-u}{\alpha},v+\frac{t'-v}{\alpha}\right), \tag{3}$$

since they are expressing the same ray.

Therefore, we can write the virtual image on plane (s',t') with original light field function $\mathcal{L}$:

$$E(s',t') \approx \iint \mathcal{L}'(u,v,s',t')\,A(u,v)dudv = \iint \mathcal{L}\left(u,v,u+\frac{s'-u}{\alpha},v+\frac{t'-v}{\alpha}\right)A(u,v)dudv \tag{4}$$

Note that this equation indicates that we are able to calculate irradiance on a shifted virtual image plane in post-processing, given the light field function. The digital refocusing capability allows us to reconstruct focused image on different depths of the scene without taking extra measurements (**Figure 2b**).

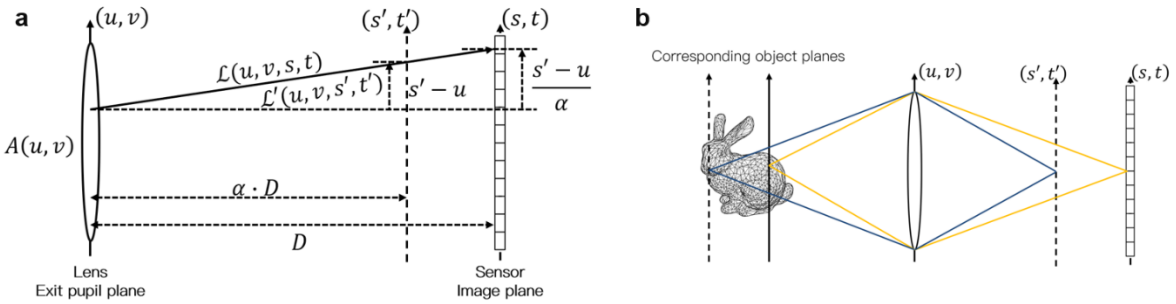

Figure 2. Digital refocusing on virtual image plane with light field.

By fixing coordinate (u,v), the light field function $\mathcal{L}(u_0,v_0,s,t)$ becomes essentially a sub-aperture image captured by placing a 'pinhole' at location (u_0,v_0). It provides a view of the scene with infinite depth of field and at a certain perspective angle. Further examining equation (4)

reveals that digital refocusing is conceptually a summation of laterally shifted sub-aperture images. This shift-and-add approach has been used in early implementations of light field imaging[2–4].

2.2 Light field microscope (LFM)

The post-processing refocus makes light field imaging an appealing tool for high-speed 3D microscopy. However, this claim is under the assumption that light field can be captured more efficiently and faster than an image stack using scanning-based counterparts. Thus, this article only focuses on implementations providing snapshot light field acquisition with single objective and camera. There are mainly two types: conventional/unfocused light field microscope[3,5,6] and Fourier/focused light field microscope[7–9]. They both utilize microlens array to redistribute the light rays of different angles to different pixels, therefore multiplexing a 2D sensor for 4D information.

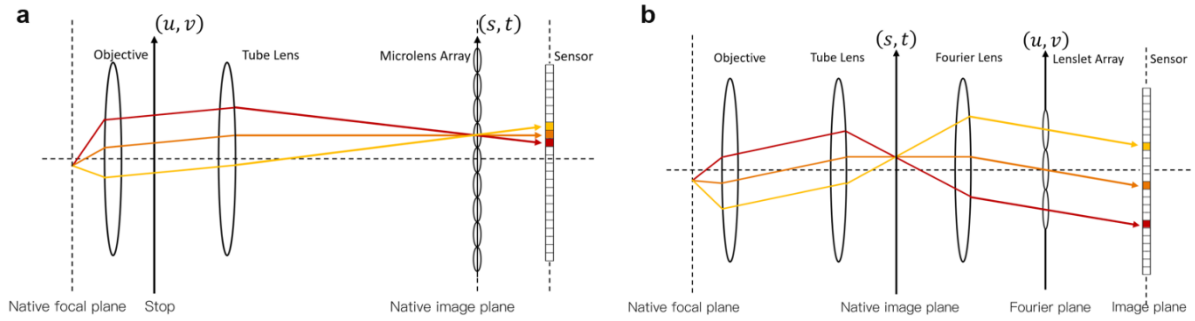

Figure 3. Two types of LFM. (a) Conventional/unfocused LFM; (b) Fourier/focused LFM.

The conventional LFM inserts a microlens array (MLA) at the native image plane. The sensor is then placed at the back focal plane of MLA (**Figure 3a**). The array has a small pitch (usually around one hundred microns) that determines the sampling rate on the spatial components of the light field. In other words, each lenslet provides a (s,t) coordinate in the two-plane

parameterization. By matching the numerical aperture (NA) of the lenslet and objective, the pixels behind the lenslet make one-to-one mapping to different angular components within the light cone.

The Fourier LFM places the MLA at the Fourier (or pupil) plane of the objective. In the cases where the exit pupil (XP) is not accessible, a separate lens, namely Fourier lens, is appended behind the tube lens to relay the XP from the microscope. The sensor, located at the back focal plane of the lenslet, captures an array of sub-aperture images. These images resemble what would have been captured with a normal widefield microscope, but with lower NA (thus larger depth of field) and varying perspective. Compared to conventional LFM, the (u, v) and (s, t) planes are swapped and the location of lenslet now determines the (u, v) coordinate in the previously discussed two-plane parameterization. The lenslets are designed to have a larger aperture (around millimeters) as they decide the field of view (FOV) of the Fourier LFM.

Both modalities attempt to map a 4D light field function onto a 2D sensor. The limited number of pixels introduces a trade-off between the resolution of spatial and angular information, therefore making LFM generally a low spatial resolution technique. Conventional LFM often suffers from severe frequency aliasing due to large MLA pitch, which is addressed by Fourier LFM with lenslets sampling angular rather than spatial components[5,9]. Therefore, the latter can provide less artifacts and smoother images. But Fourier LFM has to deal with the trade-off between the number of sub-aperture images (i.e. angular resolution) and FOV. An array of large lenslets also expects larger objective NA and sensor form factor. Despite their differences in design trade-offs, conventional and Fourier LFM share similar reconstruction algorithms ranging from native shift-and-add[3,8] to iterative deconvolution[5,9]. We can also find shared applications such as neuron imaging in zebrafish brain[6,10] and live cell imaging[11–14]. They both extend to confocal configurations[15,16]. And

conventional LFM demonstrates capability in adaptive optics[17] while Fourier LFM supports DOF extension by using MLA of multiple focal lengths[10].

2.3 Reconstruction algorithm

The digital refocusing, or 3D reconstruction, aims to synthesize a focal stack of the sample by processing the light field. In contrast to photography, the diffraction effects become considerable in LFM, so a full wave optics treatment is often necessary for light field processing. The most common approach is deconvolution, where we either simulate or experimentally measure the point spread functions (PSFs) to map the relationship between sample space signal and LFM raw measurement. Then we iteratively solve the inverse problem (i.e. estimate original sample from final measurement) using numerical methods such as Richardson-Lucy deconvolution.

The widefield fluorescent microscope is often modeled as an incoherent linear optical system. The diffraction pattern generated by an ideal point source, namely PSF, is the impulse response function. And the intensity image becomes a convolution of the sample feature and PSF. However, the assumption for such model, that the PSF is shift-invariant, becomes invalid in LFM. The light field PSF changes with respect to the 3D position of the point emitter. From a PSF engineering perspective, such variation encodes the 3D position in the detected 2D pattern and is crucial for the 3D reconstruction. In this context, we adopt a more general linear superposition integral to describe the formation of image $f(x)$

$$f(\boldsymbol{x}) = \int |h(\boldsymbol{x}, \boldsymbol{p})|^2 \, g(\boldsymbol{p}) d\boldsymbol{p} \tag{1}$$

where $\boldsymbol{p} = (p_1, p_2, p_3)$ denotes the 3D position of the point source in object space whose intensity is defined as $g(\boldsymbol{p})$, and $\boldsymbol{x} = (x_1, x_2)$ represents the coordinate on sensor plane. The optical transfer

function $h(\boldsymbol{x}, \boldsymbol{p})$ is derived from the wave optic model we are going to discuss soon, where a squared modulus is used to convert to PSF owing to the incoherence nature of fluorescent imaging.

To derive the theoretical PSF, both conventional[5] and Fourier LFM[9] start with the wavefront at the native image plane. For an objective with circular aperture, the complex amplitude $U_i(\boldsymbol{x}, \boldsymbol{p})$ at position $\boldsymbol{x}$ of the image plane produced by a point source at location $\boldsymbol{p}$ in the object space can be described by scalar Debye theory[18]:

$$U_i(\boldsymbol{x}, \boldsymbol{p}) = \frac{M}{f_{obj}^2 \lambda^2} \exp\left(-\frac{iu}{4\sin^2\left(\frac{\alpha}{2}\right)}\right) \int_0^{\alpha} P(\theta) \exp\left(\frac{iusin^2\left(\frac{\theta}{2}\right)}{2\sin^2\left(\frac{\alpha}{2}\right)}\right) J_0\left(\frac{\sin(\theta)}{\sin(\alpha)} v\right) \sin(\theta)\, d\theta \quad (2)$$

where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind. The variables v and u represent normalized radial and axial optical coordinates which are defined as

$$v = k\sqrt{\left(\frac{x_1}{M} - p_1\right)^2 + \left(\frac{x_2}{M} - p_2\right)^2} \sin(\alpha)$$

$$u = 4kp_3 \sin^2\left(\frac{\alpha}{2}\right)$$

The M and f_{obj} are the magnification and the focal length of the objective; the wavenumber $k = 2\pi n/\lambda$ is calculated using the emission wavelength λ and the refractive index n of the immersion medium; the half-angle of the NA is $\alpha = \sin^{-1}(NA/n)$. The $P(\theta)$ is the apodization function of the microscope and $P(\theta) = \sqrt{\cos(\theta)}$ for abbe-sine corrected objectives. The equation (2) only holds for low to moderate NA objective. The high NA objective requires a vectorial diffraction theory instead[14].

For conventional LFM, the MLA will directly modulate the wavefront at native image plane. Considering MLA as a phase mask $\Phi(\boldsymbol{x})$, it can be expressed as a convolution of a single lens $\phi(\boldsymbol{x})$ and an array of unit impulse functions. In the case where a grid array is used, we can use a 2D comb function, such as

$$\Phi(\boldsymbol{x}) = \phi(\boldsymbol{x}) \otimes comb(\boldsymbol{x}/d_{MLA}) = \left(A(\boldsymbol{x}) \exp\left(\frac{-ik}{2f_{MLA}} \|\boldsymbol{x}\|_2^2\right)\right) \otimes comb(\boldsymbol{x}/d_{MLA})$$

given the aperture function $A(\boldsymbol{x})$ of single on-axis lenslet, focal length of the MLA f_{MLA} and the pitch d_{MLA}. The symbol $\otimes$ represents the convolution operator.

Next, a Fresnel diffraction integral[19,20] is used to propagate the modulated wavefront to the sensor plane

$$h(\boldsymbol{x}, \boldsymbol{p}) = \mathcal{F}^{-1}\left(\mathcal{F}\left(U_i(\boldsymbol{x}, \boldsymbol{p})\Phi(\boldsymbol{x})\right) \exp\left(-i\pi\lambda f_{MLA}\left(f_x^2 + f_y^2\right)\right)\right) \tag{3}$$

where f_x and f_y are spatial frequencies along x and y directions in the sensor plane.

For Fourier LFM, the MLA is placed at the pupil plane of the microscope. Therefore, an optical Fourier transformation is applied to $U_i(\boldsymbol{x}, \boldsymbol{p})$ before being modulated by $\Phi(x)$. The final wavefront is therefore calculated by

$$h(\boldsymbol{x}, \boldsymbol{p}) = \mathcal{F}^{-1}\left(\mathcal{F}\left(\mathcal{F}\left(U_i(\boldsymbol{x}, \boldsymbol{p})\right)\Phi(\boldsymbol{x})\right) \exp\left(-i\pi\lambda f_{MLA}\left(f_x^2 + f_y^2\right)\right)\right) \tag{4}$$

While the PSFs of an LFM can be simulated by equation (3) or (4), they can also be experimentally measured by mechanically translating a sub-diffraction fluorescent bead under a real setup[10,15,21]. However, due to the shift-variance nature of LFM PSFs, experimental measurement requires massive scanning steps over the entire 3D FOV. To make it a more practical approach, we usually adopt certain level of assumptions. For example, we assume the Fourier LFM

to have a laterally shift-invariant PSF since the sub-aperture images are essentially widefield image except for lower NA and off-axis location. The sub-diffraction bead is only scanned axially to profile the LFM PSFs[10].

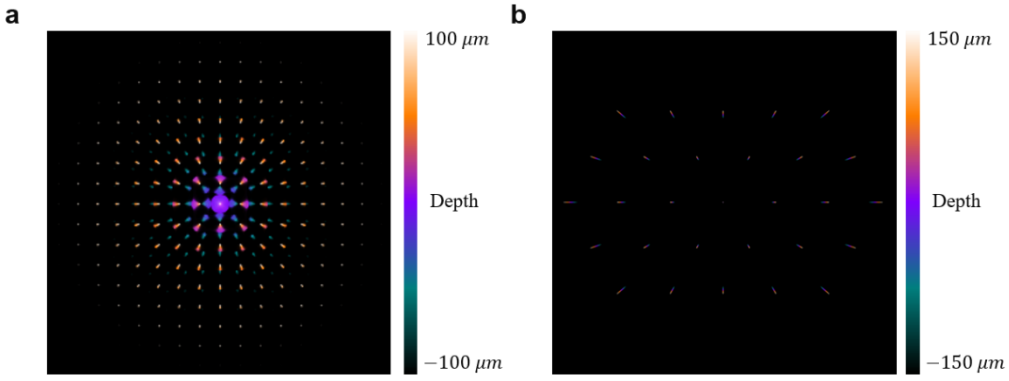

Figure 4. Example PSFs of conventional LFM (a) and Fourier LFM (b). PSFs from different depths are color coded.

With PSF known, we could use equation (1) to model the formation of a light field raw image. To write equation (1) in discrete form, we have the following linear model:

$$\boldsymbol{f} = H\boldsymbol{g} \tag{5}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are vectorized pixels on camera sensor and voxels in the volume being reconstructed. The H is a measuring matrix whose coefficients are dependent on PSFs. In specific, each column of H is the corresponding PSF of a voxel in $\boldsymbol{g}$ and has same size as $\boldsymbol{f}$.

Our reconstruction, which attempts to inversely solve $\boldsymbol{g}$ given H and $\boldsymbol{f}$, is highly ill-posed because of the dimension mismatch (i.e. from 2D to 3D). One typical approach is to incorporate prior knowledge, such as signal sparsity, and solve the following constrained optimization

$$\widehat{\boldsymbol{g}} = \underset{\boldsymbol{g}\geq 0}{\operatorname{argmin}} \|H\boldsymbol{g} - \boldsymbol{f}\|_2^2 + \tau R(\boldsymbol{g}) \tag{6}$$

The function $R(\cdot)$ provides the prior term. For example, for Total Variation (TV) regularization, $R(\boldsymbol{g}) = \|\nabla \boldsymbol{g}\|_1$ and it forces sparsity in 3D gradients of the reconstructed volume. τ is a hyperparameter to tune the weight of regularization term. Problem (6) can be effectively tackled by algorithms like fast iterative shrinkage-thresholding algorithm (FISTA), alternating direction method of multiplier algorithm (ADMM), as demonstrated in relevant works[22–25].

Another important reconstruction method is based on Richardson-Lucy deconvolution. It assumes the photon shot noise dominance and can be formulated as a maximal likelihood estimation (MLE). The Poisson likelihood of the measured light field raw image $\boldsymbol{f}$ given a particular volume $\boldsymbol{g}$ is

$$P(\boldsymbol{f}|\boldsymbol{g}) = \prod_i \frac{(H\boldsymbol{g})_i^{\boldsymbol{f}_i} \exp(-(H\boldsymbol{g})_i)}{\boldsymbol{f}_i!} \tag{7}$$

where i is the index of the pixel in image $\boldsymbol{f}$. Maximizing the log-likelihood of (7) over $\boldsymbol{g}$ yields the following update strategy:

$$\boldsymbol{g}^{k+1} = \boldsymbol{g}^k \times (H^T(\boldsymbol{f} \,/\, H\boldsymbol{g}^k)) \,/\, (H^T\mathbf{1}) \tag{8}$$

The operator $\times$ and $/$ denote element-wise product and division; $\mathbf{1}$ is an all-one vector; H^T is the transpose of measuring matrix H. H and H^T are often referred as forward and backward projector. Through iterate projection between light field and volume domain, the 3D image stack will be estimated. The initialization can start with an all-one volume $\boldsymbol{g}^0$ or an initial backward projection $H^T\mathbf{f}$.

This algorithm is widely used in LFM and known as light field deconvolution (LFD)[5,6,9,10,14,15,26,27]. It proves to be a robust and effective method to reconstruct 3D image stack from light field raw measurement but is also prone to artifacts and heterogeneous resolution. In

chapter 3, we will propose a new reconstruction algorithm, namely VCD-Net, to address the limitations of LFD so that the potentials of LFM can be better exploited for high-speed microscopy. In chapter 4, we will present a novel hardware design of ever faster LFM system, i.e. SLIM. The LFD is modified and improved, and a specific implementation is presented for readers who are interested in more details in LFD. And in chapter 5, we adopt aforementioned optimization algorithm, FISTA, to exemplify other variants of light field reconstruction algorithms.

Chapter 3
Light field microscopy with deep learning reconstruction

3.1 Motivations for a data-driven approach in LFM reconstruction

A LFM captures 3D volume simultaneously with a single snapshot. In contrast to scanning-based methods, such as confocal microscopes and light sheet microscopes, it avoids the mechanical constraints of scanners and provides high-speed volumetric detection up to the camera's frame rate. This distinct capability has enabled the study of transient biological processes in living animals. Representative demonstrations include motor neuron calcium imaging in freely moving *C.elegans* at 50 Hz[6], and whole brain calcium imaging in freely swimming zebrafish larvae at 77 Hz[10].

Despite its advantage in imaging speed, the widespread application of LFM has been impeded by its low and nonuniform spatial resolution and the presence of reconstruction artifacts. The small aperture lenslet in MLA introduces band limiting filtering that blurs the image. Especially in conventional LFM design, the low spatial sampling rate, determined by the MLA pitch divided by microscope magnification, further introduces aliasing in the high frequency content. Moreover, the

spatial sampling pattern varies as different axial locations[5]. Therefore, the major reconstruction algorithm, light field deconvolution (LFD), struggles to restore spatial resolution given such poorly sampled signals. The iterative scheme of deconvolution also imposes heavy computation burden and LFM can easily generate data to overwhelm even offline processing pipelines.

Deep learning (DL) offers a robust framework to address these challenges. Instead of relying on analytical models of image formation and sample priors, a neural network is trained to map raw measurements to final images. During training, the network is fed low-resolution images as inputs and high-resolution images as targets, enabling it to learn the statistical properties of sample structures and to infer higher resolution images from lower resolution ones[28]. In the context of MRI, a deep neural network can simultaneously learn the image formation process and sample priors, effectively reducing noise and reconstruction artifacts[29].

In this chapter, we are going to explore the possibility of a DL approach for LFM reconstruction. The data-driven reconstruction has the potential to include implicit sample priors in its approximation of LFM's image formation process. We demonstrated this algorithm on the motor neuron imaging of moving *C.elegans* and blood flow of a beating zebrafish heart. The efficiency and effectiveness of DL reconstruction will advance LFM into a more powerful tool in instantaneous volumetric imaging of biological processes.

3.2 The design of VCD-LFM

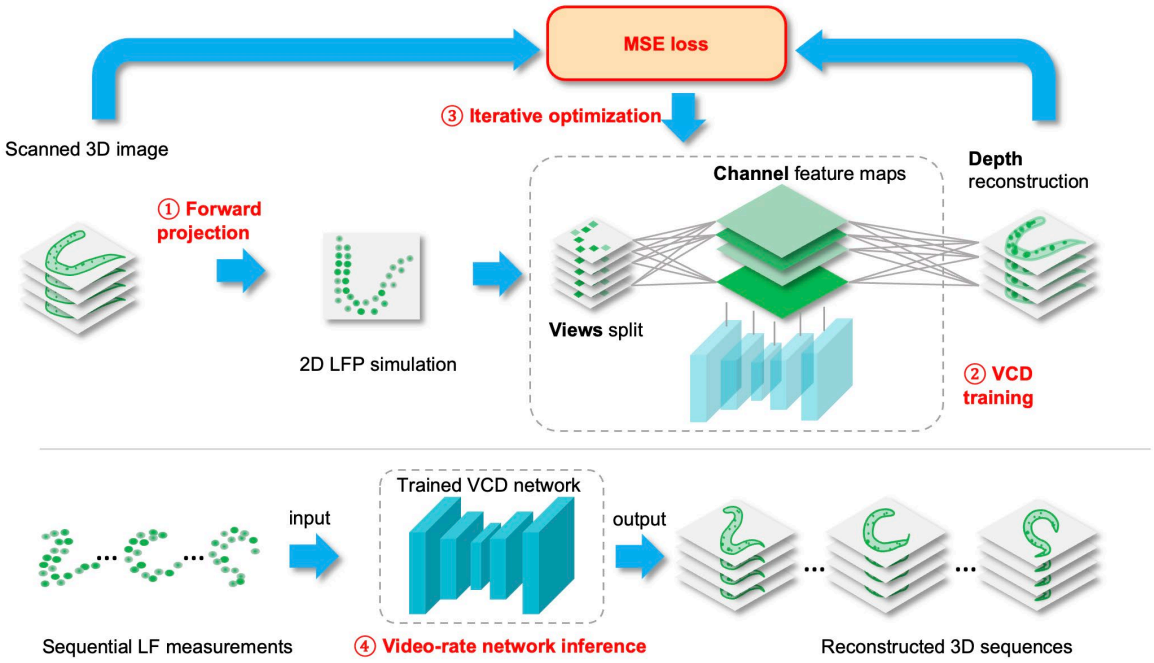

Figure 5. The VCD-Net reconstruction pipeline, containing: (1) forward light-field projection (LFP) from the HR image stacks; (2) VCD transformation of synthetic light-field inputs into intermediate 3D image stacks; (3) network training via iteratively minimizing the difference between VCD inferences and confocal ground truths and (4) inference of 3D images from the recorded light-field images by a trained VCD-Net.

The VCD-LFM[30] attempts to utilize a convolutional neural network (CNN) to directly infer a conventional 3D image stack from the light field raw measurement (**Figure 5**). With massive dataset of paired light field and ground truth sample images, the network is trained to recover the high-frequency features from sample while refraining from structured artifacts that are typical of existing light field reconstruction algorithms.

3.2.1 Dataset

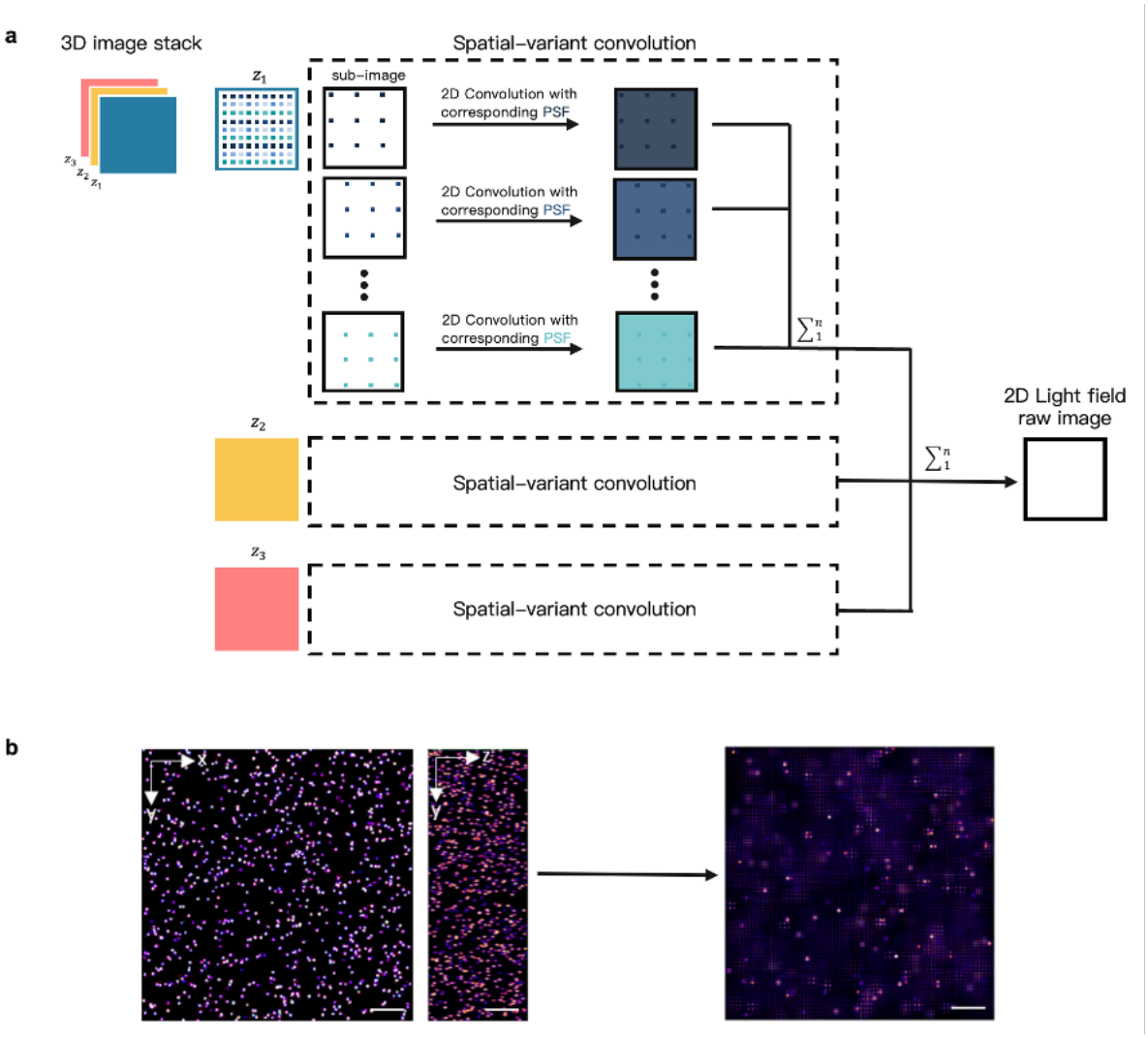

Figure 6. **Light field projection (LFP).** (a) The pipeline to synthesize a 2D light field measurement from 3D image stack. (b) Example fluorescent beads images of 3D ground truth and corresponding synthetic light field raw image.

To create the data for network training, we first obtain high resolution 3D images of stationary samples using synthetic or experimental methods. For example, embryonic zebrafish were sedated to pause the heartbeat to enable data acquisition on cardiac blood cells using commercial laser scanning confocal microscope. Data variation was created by both the large number of fish and different field of view and orientation of the animal.

The spatial-variant point spread functions (PSFs) can be calculated by the scalar wave optics model of LFM (Chapter 2.3). Due to the periodic property of the lenslet array, the PSF also repeats itself by shifting the point source one lenslet pitch laterally, which reduces the necessary computation to a smaller field of view. A linear forward model (Equation (1)) is applied to simulate the formation of light field image from a ground truth 3D stack (**Figure 5**, step 1), i.e. light field projection (LFP). **Figure 6** shows the discrete model of the LFP process.

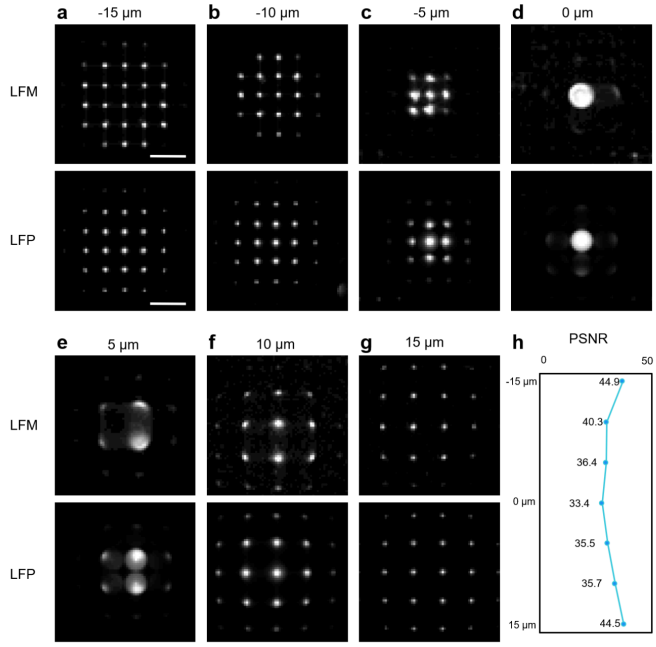

Figure 7. **Comparison between synthetic and experimental light-field raw images.** (**a**)-(**g**) The light-field raw images of identical sub-diffraction beads (z = -15 μm to 15 μm) by experimental LFM (top) and synthetic LFP (light-field projection) (bottom), respectively. (**h**) The PSNR indices of the synthetic and experimental light-field images. The high-level values (greater than 30) indicate the high similarity between the synthetic and experimental data, quantitatively proving the accuracy of the light-field projection model. Scale bar, 5 μm.

The accuracy of this light-field projection is crucial for the VCD-Net to implement reliable data training, and accurate inference afterwards. Therefore, we evaluated the similarity between the synthetic (projection from HR data) and experimental (acquired using LFM) light-field raw images of the same sub-diffraction beads. In **Figure 7**, we show the synthetic and measured PSFs

of the identical beads located at different depths (-15 μm to 15 μm), and quantitatively compared them via calculating the PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity). The synthetic light fields by light-field projection were thus verified to be enough accurate for VCD-Net training.

3.2.2Network

The VCD-Net is a fully convolutional network (FCN) with a 2D U-Net as its main body. Each synthetic light-field image is first rearranged into different views. The features are extracted and incorporated into multiple channels in each convolutional layer. The final output channels are then assigned to a number of planes representing different depths to generate a 3D image stack. By cascaded convolution layers repetitively extracting features, the network generates intermediate 3D reconstructions (**Figure 5**, step 2) for a pixel-wise mean square error (MSE) based loss function to evaluate these outputs in reference to ground-truth. Through iteratively minimizing the loss function (**Figure 5**, step 3), the network gradually optimizes until it could transform the synthetic light fields into 3D images that are similar to the ground-truth images (**Figure 8**). After training on gigavoxels of data, the network is capable of reconstructing sequential experimental measurements of dynamic processes and inferring a sequence of 3D volumes from light field raw images (**Figure 5**, step 4).

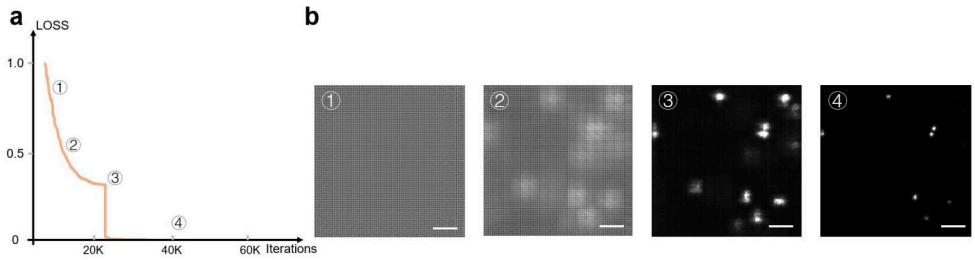

Figure 8. Iterative convergence of the network. (a) The decreasing tendency of the loss function (MSE between the outputs and the targets) during the network training. The parameters of the neural networks were initialized randomly. The training process updated them iteratively so that the network outputs would be more and more similar to the targets, with the decrease of the value of the loss function. (b) The intermediate outputs of the VCD-Net at different training time points shown in a. Scale bar: 10 μm.

After the 2D light field raw image (dimensions $a \times b$, height × width) is reformatted into a stack of different views (dimensions $a/d \times b/d \times d^2$, height × width × views) according to their relative positions to each lenslet center, a series of subpixel convolutional later first interpolates these views to dimension $a \times b \times d^2$ (height × width × views). The following 2D U-Net contains a downsampling path and a symmetric upsampling path with skip connections. Along both paths, each layer has three parameters, n, f and s, denoting the output channels number, the filter size of convolution kernel and the step size of the moving kernel, respectively, as specified in **Figure 9**.

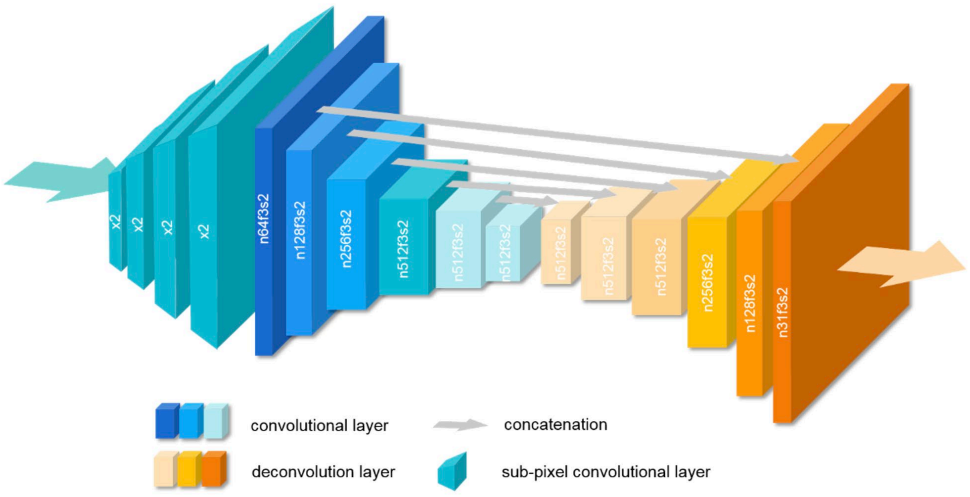

Figure 9. Our approach adopted the well-known U-Net for feature extraction and image reconstruction. Before the U-Net part, we designed a subnet to interpolate the extracted views into its original size (i.e., the size of the light-field measurement). The interpolation net contains 4 subpixel convolutional layers2, with each up-scaling the image by a factor 2. The following U-Net has down-sampling and up-sampling parts for feature encoding and decoding, respectively. The down-sampling part contains down-sampling blocks for feature extraction, where each down-sampling block consists of a convolutional layer, an activation function, a batch normalization and a max pooling. Accordingly, up-sampling blocks in the up-sampling part up-scale the feature maps. Each up-sampling block consists of a convolutional layer, an activation, a batch normalization and an up-sampling operation. Concatenations merge the feature maps of the corresponding down-sampling and up-sampling blocks.

In geometric optic model of light field imaging, a view is formed by the bundle of rays of same angle and observes the sample from a specific perspective (**Figure 10**). The digital refocusing algorithm in the light field photography can be interpreted as a superposition of laterally shifted version of different views. And the multichannel convolution at each layer of the CNN is more than competent for such linear image operations. Therefore, we attempt to mimic the digital refocusing algorithm by manually re-arranging light field raw image into pack of views before the input of VCD-Net.

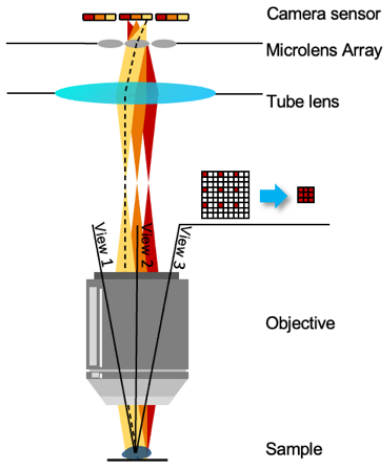

Figure 10. The pixel re-arrangement to convert light field raw measurement to light field views.

The time consumption for the VCD-Net procedure depends on the dataset size and computational resources. As a reference point, the VCD-Net converged after training on 4,580 pairs of blood cell image patches (size 176 × 176 × 51 pixels) with 110 epochs. The time cost was roughly 4 h on a single graphical processing unit. Then the trained network spent around 15 s to reconstruct 450 consecutive volumes (size 341 × 341 × 51 pixels) from acquired light-field videos. This four-dimensional reconstruction throughput was compared to roughly 11.8 h (roughly 42,467 s) by running LFD (eight iterations) on the same workstation. The computation was

performed on a workstation equipped with Intel(R) Core i9-7900X CPU at 3.3 GHz, 128 Gb of RAM and Nvidia GeForce RTX 2080 Ti graphic cards.

3.2.3Optical hardware setups

We have two optical setups to study different types of animals: an epifluorescence setup for *C.elegans* and a selective volume illumination setup for embryonic zebrafish.

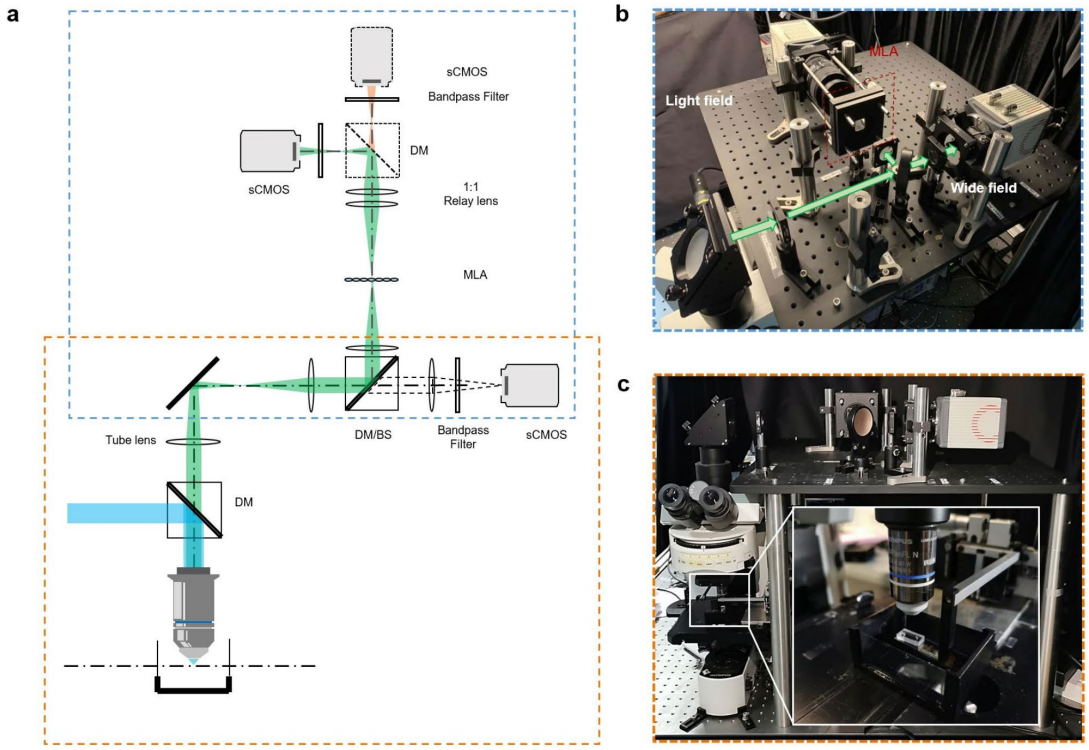

Figure 11. **Epi-illumination light-field and wide-field microscopy setup**. (**a**) Schematic drawing of our epi-illumination light-field and wide-field setup for beads and C. elegans experiments. The fluorescence signals collected by the detection objective (LUMPlanFLN 40×/0.8w NA, Olympus) were either focused onto the camera sensor 1 (Flash 4.0 V2, Hamamatsu) for wide-field imaging or onto the microlens array (APO-Q-P150-F3.5 (633), OKO Optics) for light-field imaging. In the latter case, a 1:1 relay system (AF 60 mm 2.8D, Nikon) was used to focus the camera sensor 2 (Flash 4.0 V2, Hamamatsu) on the back focal plane of MLA. DM (Dichromatic mirror) or BS (Beam splitter) was used for sequential or simultaneous light-field/wide-field imaging, respectively, depending on the experiments. A DM can be optionally added in front of the sCMOS in the light field path for dualcolor light-field imaging. The system was based on an epifluorescence upright microscope (Olympus, BX51), as shown in (**c**), with its customized wide-field and light-field detection paths shown as (**b**). A microfluidic chip (inserted picture) was used to permit the worm acting within the FOV of 40× objective.

The epifluorescence light-field setup (**Figure 11**) was built on an upright microscope (BX51, Olympus). The light-field and wide-field detection paths were appended to the camera port of host microscope, using a flip mirror to switch between two detection modes. A motorized z stage (Z812B, Thorlabs) together with a water chamber were directly mounted onto the microscope stage (xy), to three-dimensionally control the samples inside the chamber. A water immersion objective (LUMPlanFLN ×40/0.8 W, Olympus) was used to collect the epifluorescence signals from samples. For recording the light field, a microlens array (MLA) (APO-Q-P150-F3.5 (633), OKO Optics) was placed at the native image plane to collect the light-field signals. A 1/1 relay system (AF 60 mm 2.8D, Nikon) was used to conjugate the back focal plane of MLA with the camera sensor plane (Flash 4.0 V2, Hamamatsu). The light-field path was optionally extended to dual-channel detection by dividing after MLA and adding an extra camera sensor for the C. elegans experiments.

We also developed an LFM setup based on selective volume illumination (**Figure 12**). Two pairs of beam reducers combined with an adjustable iris were used to generate a scalable rod-like beam (473 or 532 nm), which was finally projected onto the sample through a ×4 illumination objective (Plan Fluor ×4/0.13 W, Nikon) placed perpendicular to the detection path. It confined the fluorescence excitation within the heart region of zebrafish embryo, reducing the excessive emission from out of the volume of interest that could smear the desired signals. This selective volume illumination mode provided light-field image with less background noise and increased contrast[31]. For observing the dynamic process of blood flowing through vessels, we also integrated a standard SPIM channel (473 nm, 4-μm thick laser sheet) to implement the in-situ 3D imaging of static vessels. The illumination paths were aligned, providing double excitation to the sample from its dual sides. The detection path used a water immersion objective (Fluor ×20/0.5 W, Nikon) to collect the fluorescence signals. A dichroic mirror splits the green fluorescent protein (GFP)

(vessels) and DsRed (RBCs) signals for wide-field and light-field detection, respectively, when performing dual-channel imaging. The light-field detection here followed the same design used in the epi-illumination LFM.

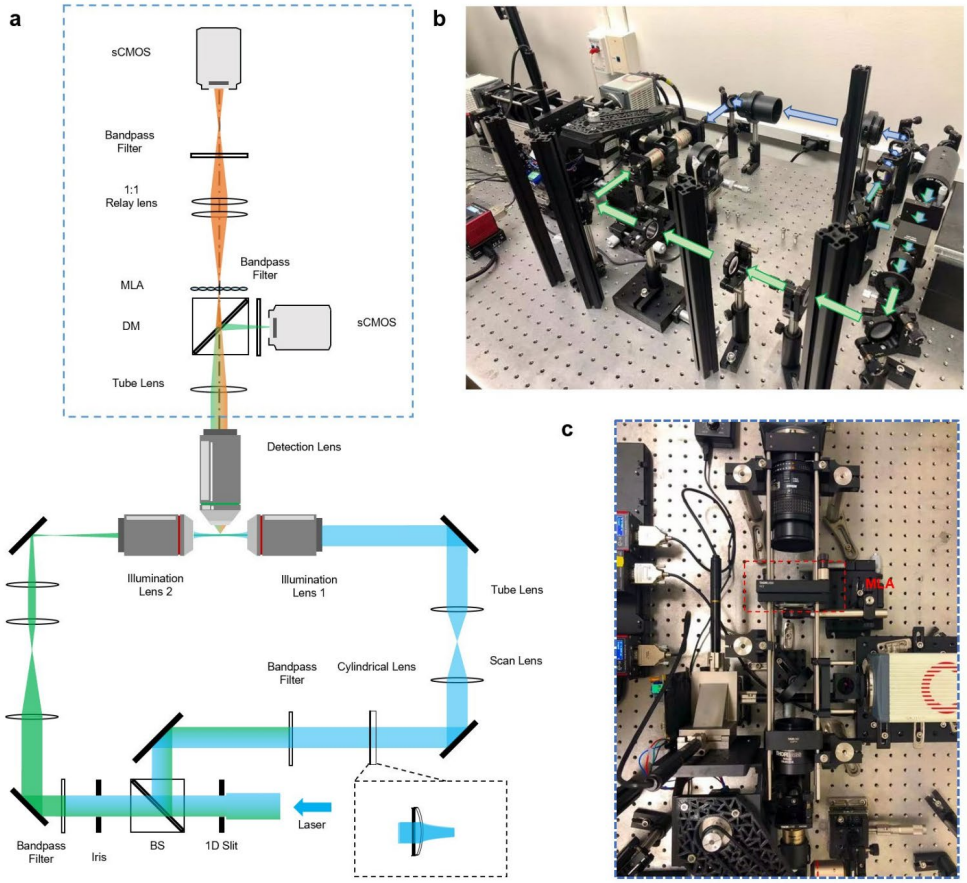

Figure 12. **Selective plane and volume illumination based LFM setup.** (**a**) Schematic drawing of our custom-built selective plane and volume illumination setup for hybrid SPIM and LFM imaging of hemodynamics. A tunable rod-like laser beam (100-200 μm) for volumetric excitation and a thin laser-sheet (~4 μm) for planar excitation could be projected to the sample simultaneously, thereby enabling light-field and light-sheet imaging, respectively. (**b**) Overview of the system. (**c**) Top view of the dual-mode light-field and wide-field detection

3.3 Performance characterization

To demonstrate the capability of the VCD-LFM, we reconstructed subdiffraction beads captured using a ×40/0.8 W objective and quantified the resolution improvement resulting from the network

by comparing the results with those from conventional LFD (**Figure 13**). As verified by 3D wide-field imaging of the same volume, the fluorescence of individual beads was correctly localized throughout the volume.

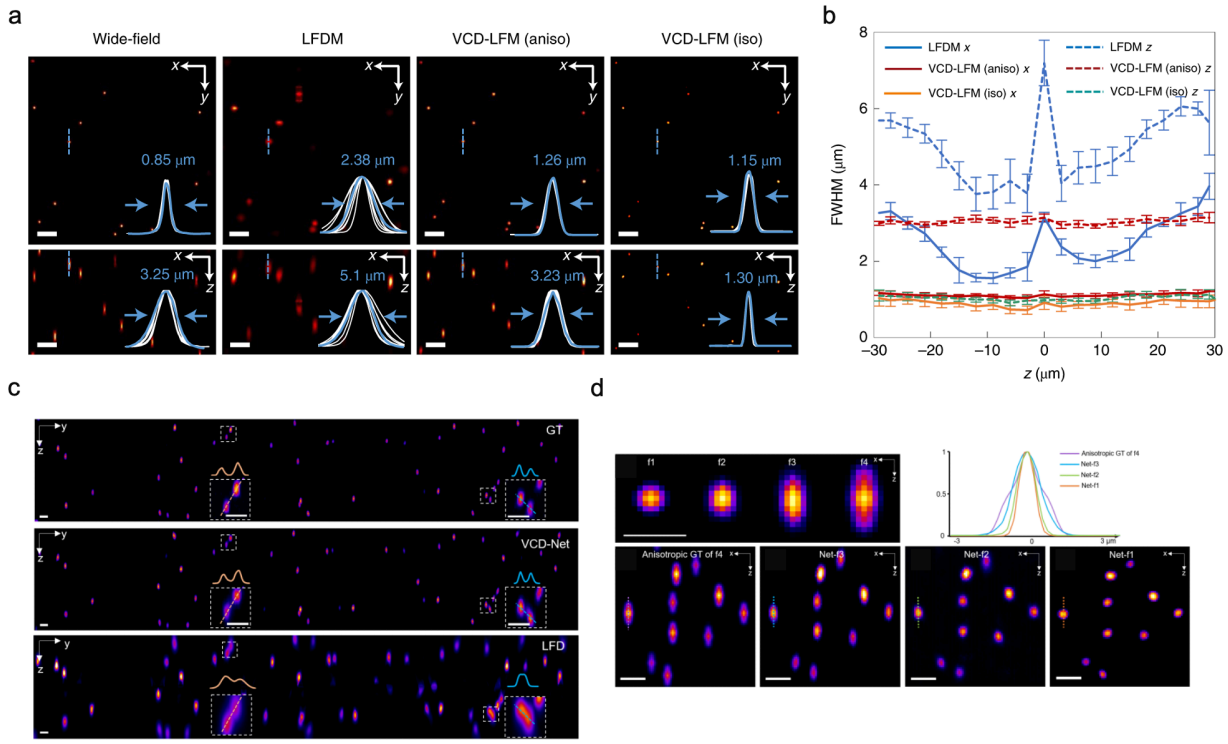

Figure 13. **VCD-LFM and its performance.** (**a**) Maximum intensity projections (MIPs) of the same fluorescent beads and achieved resolution (FWHM) by wide-field microscopy, LFDM and VCD-LFM trained with anisotropic and isotropic HR data, respectively. White lines, intensity profiles of all the resolved beads shown in the MIPs. Blue lines, intensity profiles across a selected bead (indicated by a vertical line) at 20 μm off the focal plane. Scale bars, 10 μm. (**b**) Average axial (dashed lines) and lateral (solid lines) FWHM of the beads across the volumes reconstructed by LFD (n = 2,039 beads), anisotropic (n = 2,527 beads) and isotropic VCD-LFM (n = 2,731 beads), respectively. Center lines represent means and error bars denote standard deviations. (**c**) Cross-sectioning image (y-z plane) of fluorescent beads by 3D wide-field microscopy (GT), VCD-Net and LFD. (**d**) Data-driven reconstruction imposes different constraints on the signals by different fluorescent datasets f1-f4.

The VCD-LFM with a network trained on wide-field 3D image yielded an average resolution of 1.1 μm (x,y) and 3.0 μm (z) (n = 2,527 beads), which was uniform across a 60-μm imaging depth (1.0 μm (x,y) and 2.9 μm (z) at the best plane, 1.3 μm (x,y) and 3.1 μm (z) at the outer edge of the

axial field of view (FOV)) (**Figure 13b**). This demonstrates notable improvements compared to light field deconvolution microscopy (LFDM) on the same image data with an average resolution of 2.6 μm (x,y) and 5.0 μm (z), and a broader distribution (1.6 μm (x,y) and 3.8 μm (z) at the best plane, 4.0 μm (x,y) and 7.0 μm (z) at the outer edge of the axial FOV). We note that the performance of VCD-LFM is dependent on training data; hence, the beads can be reconstructed isotropically (1.0 μm x,y,z) by including higher-resolution data in the training (**Figure 13d**).

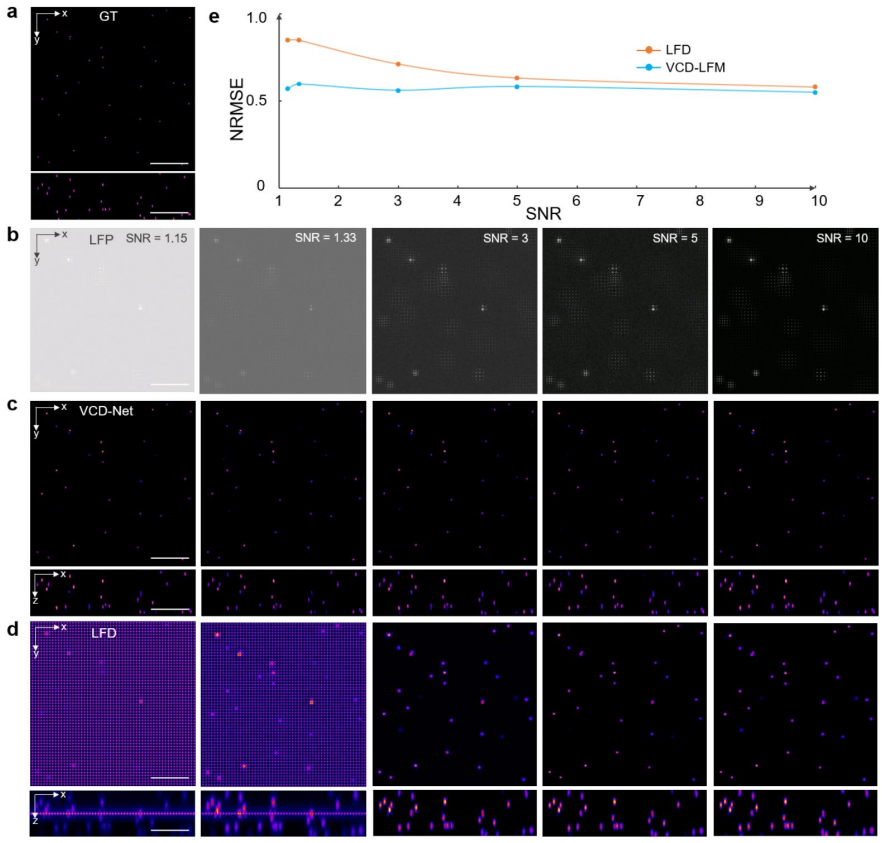

Figure 14. **Comparison of VCD-Net and LFD on recovering low-SNR noisy signals.** (**a**) MIPs in xy and xz planes of high-resolution high-SNR 3D image by light-sheet microscopy. (**b**) Five synthetic light-field images of (a) with different levels of background noise (Gaussian noise at LFP and Poisson noise at final sensor image) added. The calculated SNR is from 1.15 to 10. (**c**)-(**d**) MIPs in xy and xz planes of the 3D reconstructions of these different-SNR light-field images by VCD-Net and LFD, respectively. The VCD-Net was trained with different SNR light-fields with ground truth high-SNR 3D images. (**e**) Reconstruction error, termed normalized root mean square error (NRMSE, lower is better), of VCD-Net and LFD under different SNR conditions. Results in (c) to (e) validate the robustness of VCD-Net for recovering noisy data. Scale bars, 50 μm.

Furthermore, VCD-LFM substantially removed the mosaic-like artifacts near the focal plane that are common in LFDM (**Figure 13**), and it performed well even when the signals were weak with high background noise (**Figure 14**).

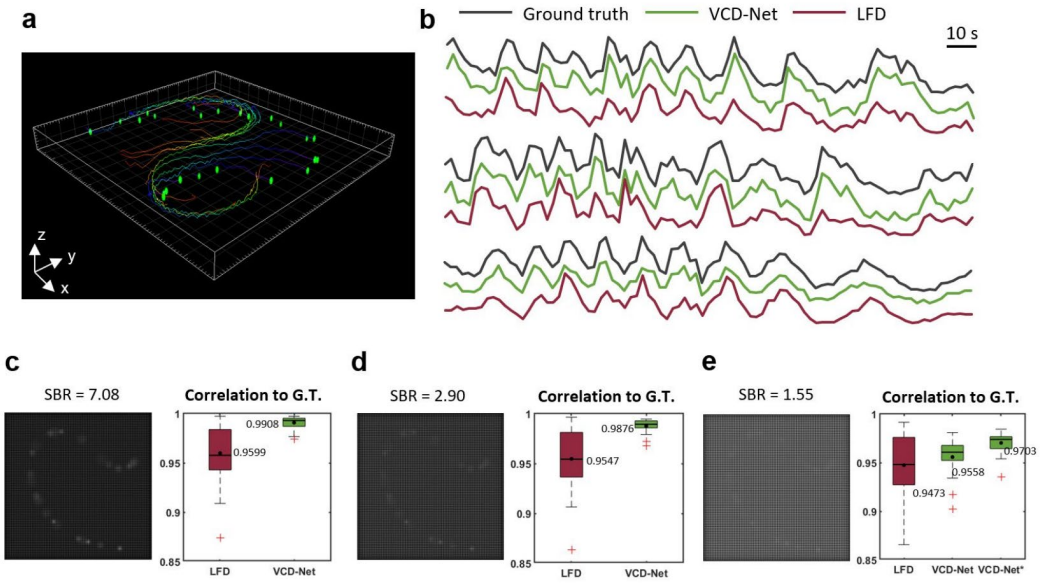

Figure 15. **VCD-Net performance on tracking the signal fluctuation in moving neuron activities.** We generated sequence of volumes containing synthetic point-like neurons with time-varying positions and intensities, to simulate the ground truth status of a GCaMP-labelled freely moving worm. The positions were collected from empirical data (Supplementary Video 2) and calcium signals were acquired following measurements in published work3. The synthetic light-field movie was generated and then input into the trained VCD-Net to reconstruct a 3D movie. (**a**) Trajectories of neurons (neuron candidate in green dot; trajectory color indicates the time: purple, beginning; red, end) within the first 5 seconds. (**b**) Intensity fluctuations of three selected signals, which simulated the calcium signaling of VA10, VA11 and DA7 neurons of the worm, in ground truth (black lines, upper), VCD-Net (green lines, middle) and LFD (red lines, lower) reconstructions, were compared throughout a 180-second period. (**c**),(**d**),(**e**) Comparative analysis correlating the ground-truth signals with VCD-Net and LFD signals reconstructed from light-field images at different signal-to-background ratios (SBR, 7.08 to 1.55), with consideration to samples with strong baseline background and noise. Gaussian and Poisson noise were added while the neural signal magnitude was altered to achieve different SBRs. Boxplots show the median as thick lines, and the 25th and 75th percentiles as box edges. Whisker: maximum and minimum after excluding outliers (red cross). Dot and number, the mean correlation. The VCD-Net was trained on high-SBR data and applied to all the three different SBR levels. VCD-Net* was specifically trained on low-SBR data. The signal traces shown in (b) were based on the results in (e).

To further validate the accuracy of reconstructed signals, we applied VCD-Net to the reconstruction of synthetic firing neurons that are adjacent to each other. The improved image quality achieved by VCD-Net suppressed signal cross-talk from blurring and artifacts, and thus

contributed to accurate recovery of signal fluctuations when recording the activity of densely labeled neurons. We further validated the reconstruction accuracy of VCD-Net on both static and moving neurons with varying signal magnitude and density (**Figure 15, 16**).

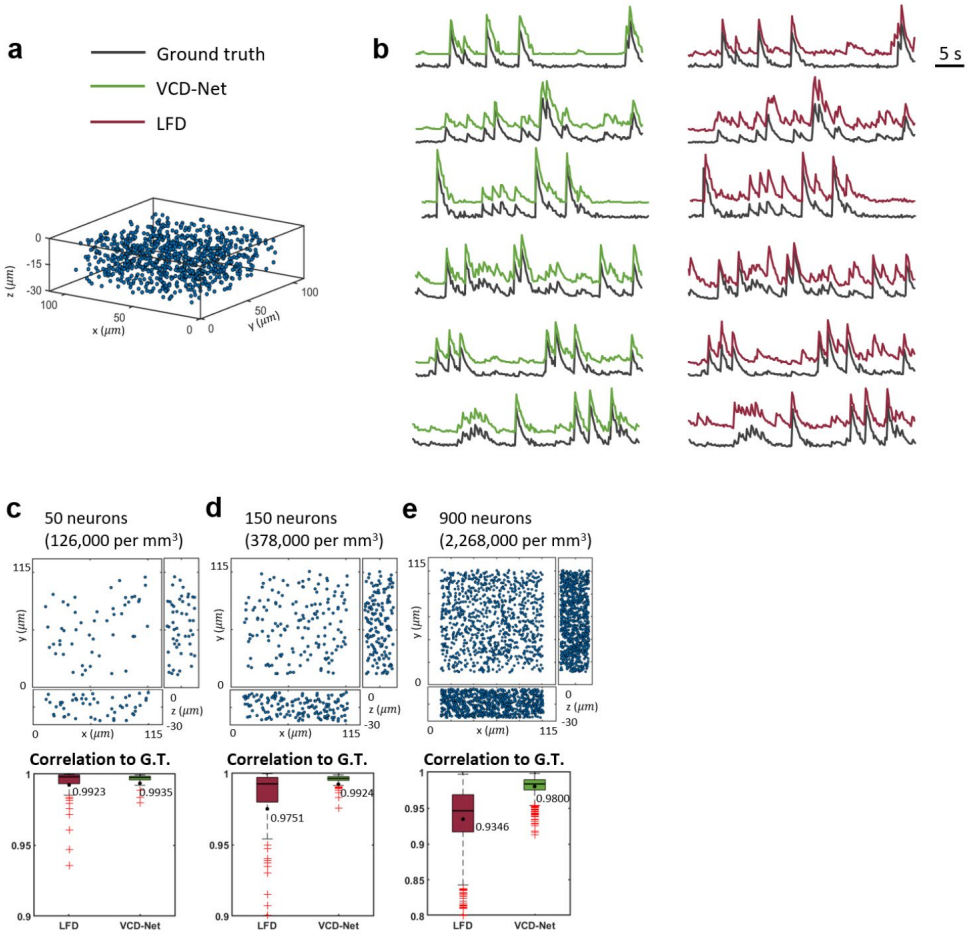

Figure 16. **VCD-Net performance on static neuron activities with different signal densities**. (**a**) Synthetic firing neurons with fixed positions and various densities were generated. Neurons (spheres of 8 $\mu\mu\mu$ diameter) were randomly seeded in the volume of 116 × 116 × 30 $\mu\mu mm3$ with signals simulated as Poissonian spike trains (firing rate 0.1 Hz, 200 time step at sampling rate 5 Hz) convolved with an exponentially decaying kernel (mean decay time constant 1.2s). (**b**) After reconstruction by VCD-Net (green lines, left panel) and LFD (red lines, right panel), neuron signals have been measured to compare with the Ground truth (black lines). 6 example traces are shown. (**c**),(**d**),(**e**) Intensity accuracy of reconstruction at different neuron densities. Boxplots show the median as thick lines, and the 25th and 75th percentiles as box edges. Whisker: maximum and minimum after excluding outliers (red cross). Dot and number, the mean correlation. The signal intensity traces shown in (b) were based on the results in (e).

3.4 Imaging of locomotion-associated neural activity in moving C. elegans

We demonstrated VCD-LFM by capturing dynamic processes in live animals such as the neuronal activity in moving *C. elegans*. A microfluidic chip was used to permit *C. elegans* (L4 stage) to rapidly move inside a micro-chamber ($300 \times 300 \times 50\ \mu m^3$, **Figure 17**). We used the epifluorescence setup with a 40×/0.8 water objective and imaged the activity of fluorescently labeled motor neurons (ZM9128 *hpIs595[Pacr-2(s)::GCaMP6(f)::wCherry]*) at a 100-Hz acquisition rate, yielding 6,000 light fields in both green and red channels in a 1-min observation. VCD-Net reconstructed the neuronal signals at single-cell resolution during fast body movement (**Figure 17**). In contrast, LFD suffered from suboptimal cellular resolution and deteriorated image quality around the focal plane, as shown on the same image data.

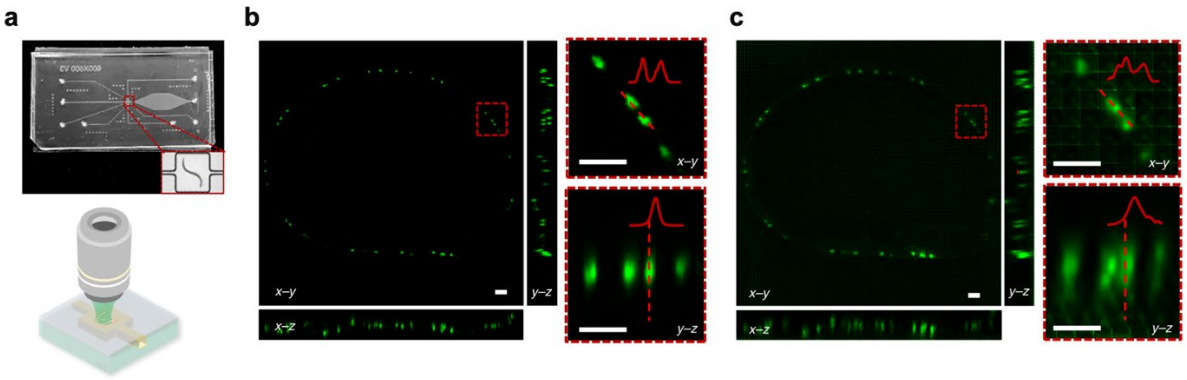

Figure 17. **Whole-animal Ca^{2+} imaging of moving *C. elegans* using VCD-LFM**. (**a**) Configuration combining LFM with a microfluidic technique for imaging motor neuron activity in L4-stage *C. elegans* (strain ZM9128 *hpIs595[Pacr-2(s)::GCaMP6(f)::wCherry]*) moving in a microfluidic chip ($300 \times 300 \times 50\ \mu m^3$, top panel) at a 100-Hz recording rate. (**b**),(**c**) MIPs as well as magnified views of the indicated regions of one instantaneous volume reconstructed by VCD (**b**) and LFD (**c**), respectively. The data shown are representative of $n = 10$ independent *C. elegans*. Scale bars, 10 μm.

We identified A and B motor neurons that have been associated with motor-program selection and mapped their calcium activity over time (**Figure 18**) after ratiometrically correcting the

calcium signals (GCaMP6(f)) using red fluorescent protein (RFP) baselines (wCherry) to remove motion noise (**Figure 18b**). By applying an automatic segmentation of the worm body contours, we calculated the worm's velocity and curvatures related to its locomotion and behavior (**Figure 19**), thereby allowing classification of the worm motion into forward, backward or irregular crawling (**Figure 18b-e**).

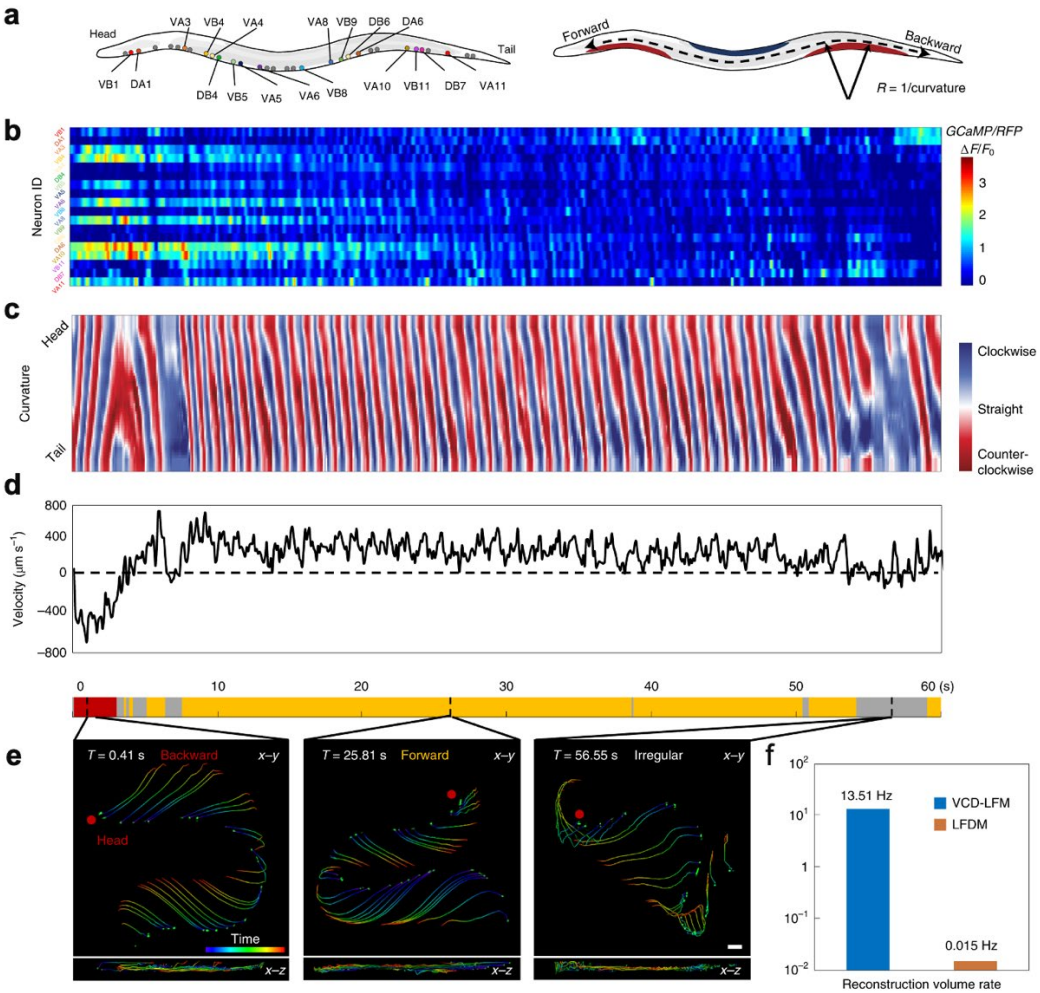

Figure 18. **Motion-correlated neuron activities of moving *C. elegans*.** (**a**) Schematics of the worm with identified motor neurons labeled (left) and body curvature annotated (right). (**b**) Heatmap visualizing the neuronal activity of 18 identified motor neurons during a 1-min observation of the moving worm. Each row shows the Ca^{2+} signal fluctuation of an individual neuron with color indicating the percentage fluorescence changes ($\Delta F/F_0$), where F is ratiometrically corrected by the ratio of GCaMP6(f) fluorescence to wCherry fluorescence. (**c**) Curvature kymograms along the body of the moving worm. (**d**) Velocity plot shows the displacement in the direction of body. An ethogram describing the worm behavior over time (lower panel) is obtained by analyzing the curvature and velocity change. (**e**) Selected volumes with time-coded traces

(duration left and middle, 150 ms; right, 500 ms) in accordance with the ethogram visualizing the backward (left), forward (middle) and irregular (right) crawling tendency of the worm. Scale bars, 20 μm. (**f**) The reconstruction throughput of VCD-LFM and LFDM for processing the same *C. elegans* light-field video.

We found the patterns of transient calcium signaling to be relevant to the switches from forward to backward crawling of the worm, which is consistent with previous findings. Furthermore, the noniterative VCD reconstruction could sequentially recover 3D images at a volume rate of 13.5 Hz, roughly 900 times faster than the iterative LFD method (**Figure 18f**). Our VCD-LFM thus demonstrated advantages for visualizing sustained biological dynamics, which is computationally challenging using conventional deconvolution methods.

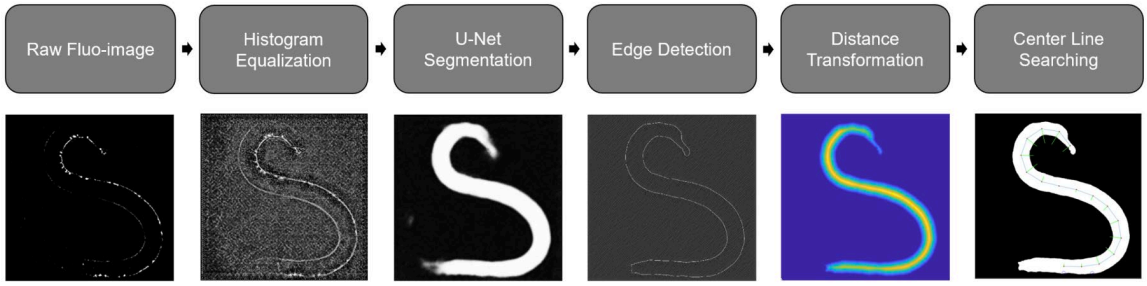

Figure 19. Pipeline for behavior analysis of C. elegans behavior. Steps of image processing to extract center line from raw fluorescence image for behavior analysis. The body curvature and movement velocity were then inferred from the center line.

3.5 Imaging of fast dynamics in the beating zebrafish heart

We captured the cardiac hemodynamics in the beating zebrafish heart. To reduce the background from body tissue, we generated a rod-like laser beam to selectively illuminate the heart region, and recorded the light-field video at a 200-Hz volume rate using a 20×/0.5 water objective. We reconstructed red blood cells (RBCs) (labeled with *Tg(gata1a:dsRed))* and beating cardiomyocytes (nuclei were labeled with *Tg(myl7:nls-gfp)*) in four dimensions using VCD-Net

with resolution, structural similarity and processing throughput notably better than conventional LFD approaches (**Figure 20**). VCD-LFM reconstruction with single-cell resolution permitted quantitative investigation of transient cardiac hemodynamics. We tracked 19 individual RBCs throughout the entire cardiac cycle of 415 ms, during which the blood was pumped in and out of the ventricle at a speed of over 3,000 μm s−1 (**Figure 21**).

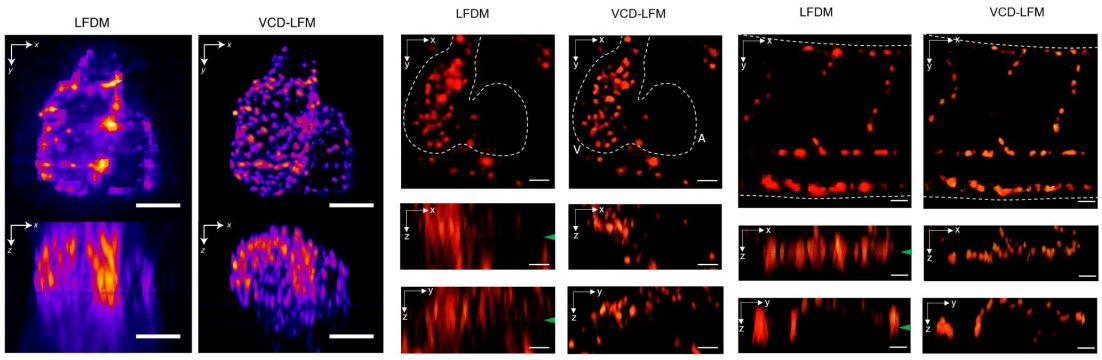

Figure 20. Imaging of beating heart and hemodynamics in embryonic zebrafish using VCD-LFM and previous LFDM

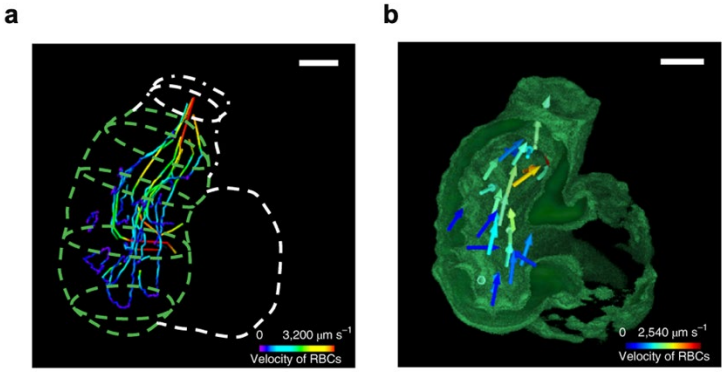

Figure 21. **Blood cell tracking**. (**a**) Tracks of 19 single RBCs throughout the cardiac cycle. A static heart is outlined for reference. (**b**) Velocity map computed from two consecutive volumes of RBCs during systole.

Furthermore, through a combination of VCD-LFM and selective plane illumination microscopy (SPIM), we visualized and analyzed both blood flow and myocardium in the

embryonic zebrafish heart *Tg(gata1a:dsRed; cmlc2:gfp)*[32]. The combination features a sequential imaging pipeline along with a modified retrospective gating method for hybrid light-field and light-sheet imaging (**Figure 22**). Using the LFM for capturing the traveling blood cells, we acquired light fields at up to 200 frames per second (fps). The heart was selectively illuminated by a rod-shaped laser beam to eliminate background noise and to enhance image contrast (**Figure 22a,** upper panel). We used VCD-Net for 3-D reconstruction of the blood cells acquired from the raw

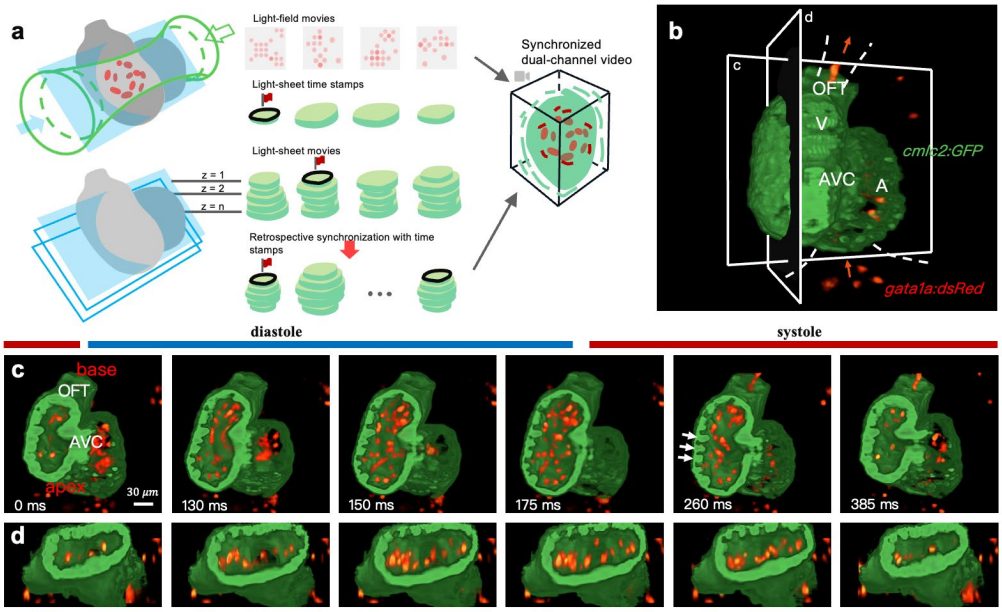

Figure 22. A pipeline for high speed imaging for ventricular contraction and intracardiac flow dynamics. (**a**) The integration of light-sheet and light-field microscopy captures the contracting myocardium (*cmlc2*:*GFP*) and traveling blood cells (*gata1a*:*dsRed*) at 200 volumes per second. The light field-generated blood cell images (upper illustration) are synchronized with the light sheet-generated myocardial cross-section images (lower illustration). (**b**) GFP-labeled cardiomyocyte light chain *(cmlc2*) and dsRed-labeled blood cells (*gata1a*) are simultaneously visualized in a 3-D reconstructed embryonic heart. The red arrows indicate the direction of blood flow. A: atrium; V: ventricle. (**c**),(**d**) A time sequence of a cardiac cycle is illustrated in the coronal (C) and sagittal (D) plane, respectively. During diastole (light-blue bar), the blood cells traverse the atrioventricular canal (AVC) into ventricle. During systole, the blood cells travel through the outflow tract (OFT).

2-D light-field sequences. This algorithm reconstructed an equivalent 3-D imaging speed at 200 vps. Using the LSFM for imaging cardiac contraction, we captured the cross section of beating

myocardium with high spatial resolution and signal-to-noise ratio (SNR). At each scanning step, we acquired the time sequence at the same frame rate as LFM at 200 fps. By adopting a modified retrospective gating method[33–36], we synchronized all the sequences from both LSFM and LFMs at the identical cardiac phases, and we performed 3-D reconstruction of the contracting myocardium in synchrony with the blood flow (**Figure 22a**, lower panel, **Figure 22b-d**). We demonstrated the atrium and ventricle, along with the intracardiac blood cells, during a cardiac cycle in the *Tg(cmlc2:GFP; gata1a:dsRed)* transgenic line from 3 to 5 dpf . Thus, the reconstructed dual-channel zebrafish heart enabled dynamic and simultaneous analyses of the myocardial displacement and blood flow.

Images acquired by our pipeline embraced the time- and position-dependent information for vector analysis of the myocardial displacement (**Figure 23**). We computed the myocardial displacement between two consecutive frames by using the deformable image registration (DIR) to infer the voxel-based vector field (**Figure 23b**). Each vector, divided by the time interval, indicated the direction and magnitude of myocardial velocity (μm/s). A heatmap could reveal the spatial variations of the myocardial velocity (**Figure 23c**). The increase in velocity toward the apex and atrioventricular canal (AVC) suggested a large displacement and myocardial kinetic energy, as previously reported[37]. This image-based analysis provided a robust platform to demonstrate the 3-D myocardial motion in an unsupervised manner, while the previous methods (e.g. strain analysis, fractional shortening) often involved human-annotation[33,38]. Therefore, the result acquired from our pipeline reduced inter-observer variations and increased experimental reproducibility.

To address the intracardiac blood flow, we employed the Particle Tracking Velocimetry (PTV) to track the individual cells during two cardiac cycles. We mapped the trajectories of 81 blood cells

traversing from AVC to the outflow tract (OFT) (**Figure 23d**), and extracted the velocity vectors of the individual cells frame-by-frame (**Figure 23e**). We integrated DIR with PTV techniques to merge the vectors for myocardial velocity (**Figure 23b**) and the traveling blood cells (**Figure 23e**), capturing the synchronized myocardial contration with intracardiac hemodynamics (**Figure 23f**).

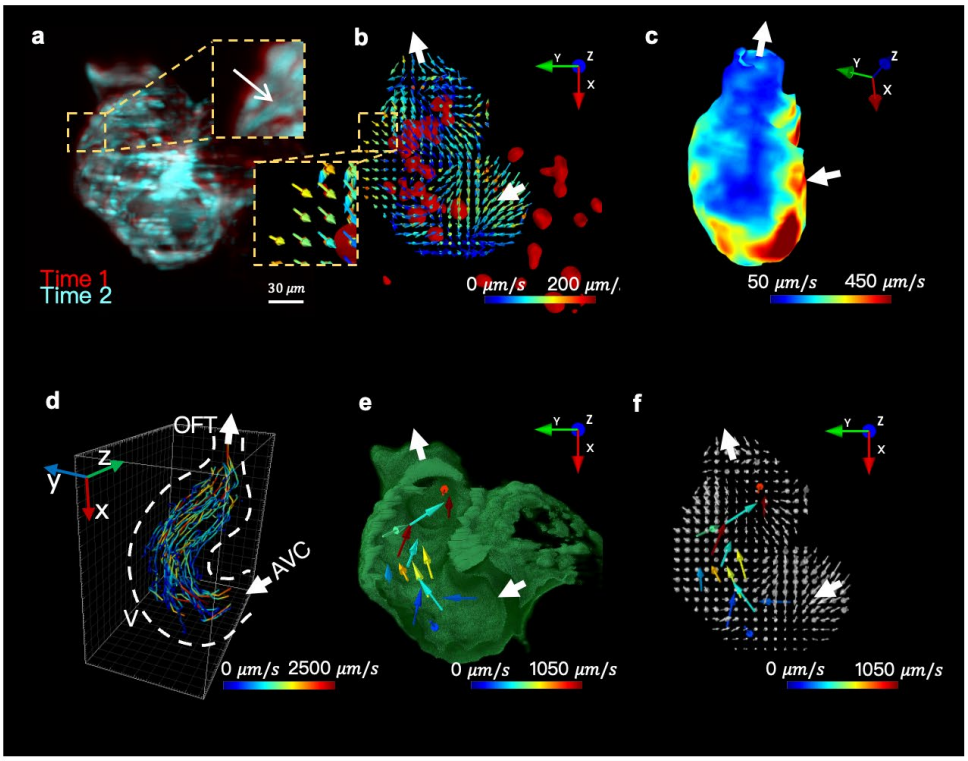

Figure 23. **Post-imaging computation to reconstruct 3-D vector fields for myocardial displacement and blood flow.** (**a**) Images of the myocardium from 2 different time points are overlaid to demonstrate the myocardial displacement. The arrow indicates the direction of displacement. (**b**) Deformable image registration (DIR) is used to infer the myocardial displacement between frames. Each vector indicates the direction and magnitude of the displacement velocity as color-coded by the magnitude ($\mu m/s$) in the corresponding voxel. Intracardiac blood cells (red) and flow direction (white arrow) are superimposed with the vector field. (**c**) A representative heatmap depicts the segmental variations in the magnitude of the myocardial velocity. (**d**) Particle tracking velocity follows the trajectories of the blood cells during the cardiac cycles. V: ventricle; OFT: outflow tract; AVC: atrioventricular canal. (**e**) The vector field represents the direction and speed of the blood cells. Dissected myocardium (green) and flow direction (white arrow) are superimposed with the vector field. (**f**) The velocity vector fields for myocardial displacement and the traveling blood cells are merged.

To quantify the flow pattern, we defined the flow direction along the centerline that was fitted to the distribution of the traveling blood cells (**Figure 24a**). We projected the velocity vectors along the flow centerline to generate a mean velocity (**Figure 24c**, upper panel). The negative values during the end of systole and diastole represented flow reversal across the valves, resulting in a decrease in the net forward flow at 3 days post fertilization (dpf). The regurgitation was reported to be a biomechanical cue to valvulogenesis[39].

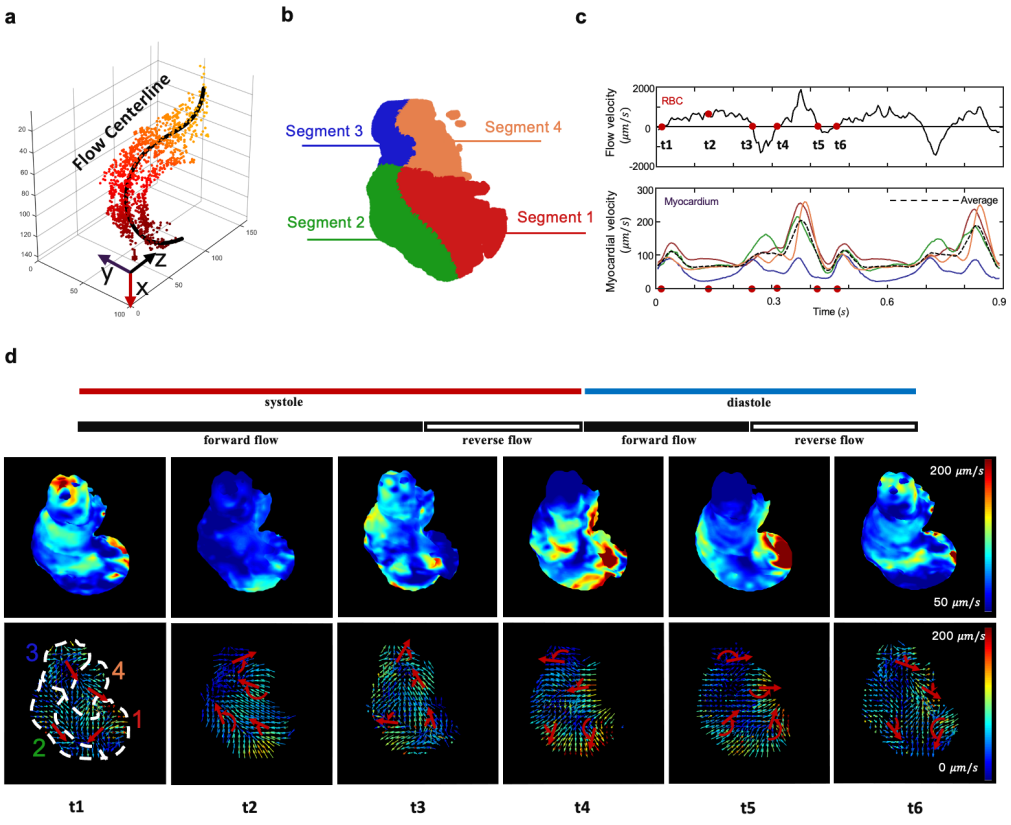

Figure 24. Frame-to-frame analyses of the myocardial contraction and intracardiac blood flow at 3 dpf. (**a**) The trajectory of the blood cells is depicted as the flow centerline, starting from the AV canal (dark colors) to the outflow tract (light colors). (**b**) The myocardium is divided into 4 segments to assess regional variations in the vector field. (**c**) The velocity of individual blood cells is projected onto the flow centerline, and the mean flow velocity is plotted as a function of time. Myocardial velocity in each segment is quantified as the magnitude-weighted mean vector for each segment, as plotted in the color line. The black dotted line demonstrates the average velocity magnitude of the entire ventricle. (**d**) Six representative time points (from t_1 to t_6) of the heatmaps are presented for the myocardium (upper panels, color-coded in vector magnitude) and vector maps (lower panels). Red arrows indicate the rotational direction of the mean vector for each segment.

The ventricle was divided into 4 segments with reference to the flow centerline to compare with the myocardial velocity in the individual segment (**Figure 24b**). We calculated the magnitude-weighted mean vectors for each segment (**Figure 24c**, lower panel, color lines), as compared with the global average magnitude (black dotted line). These myocardial vectors quantified the spatial variations in myocardial motion during a cardiac cycle. Using the heatmap and velocity field, we demonstrated the time-dependent myocardial velocity. We further revealed that the myocardial velocity vectors in each segment rotated clockwise or counterclockwise during a cardiac cycle (**Figure 24d**).

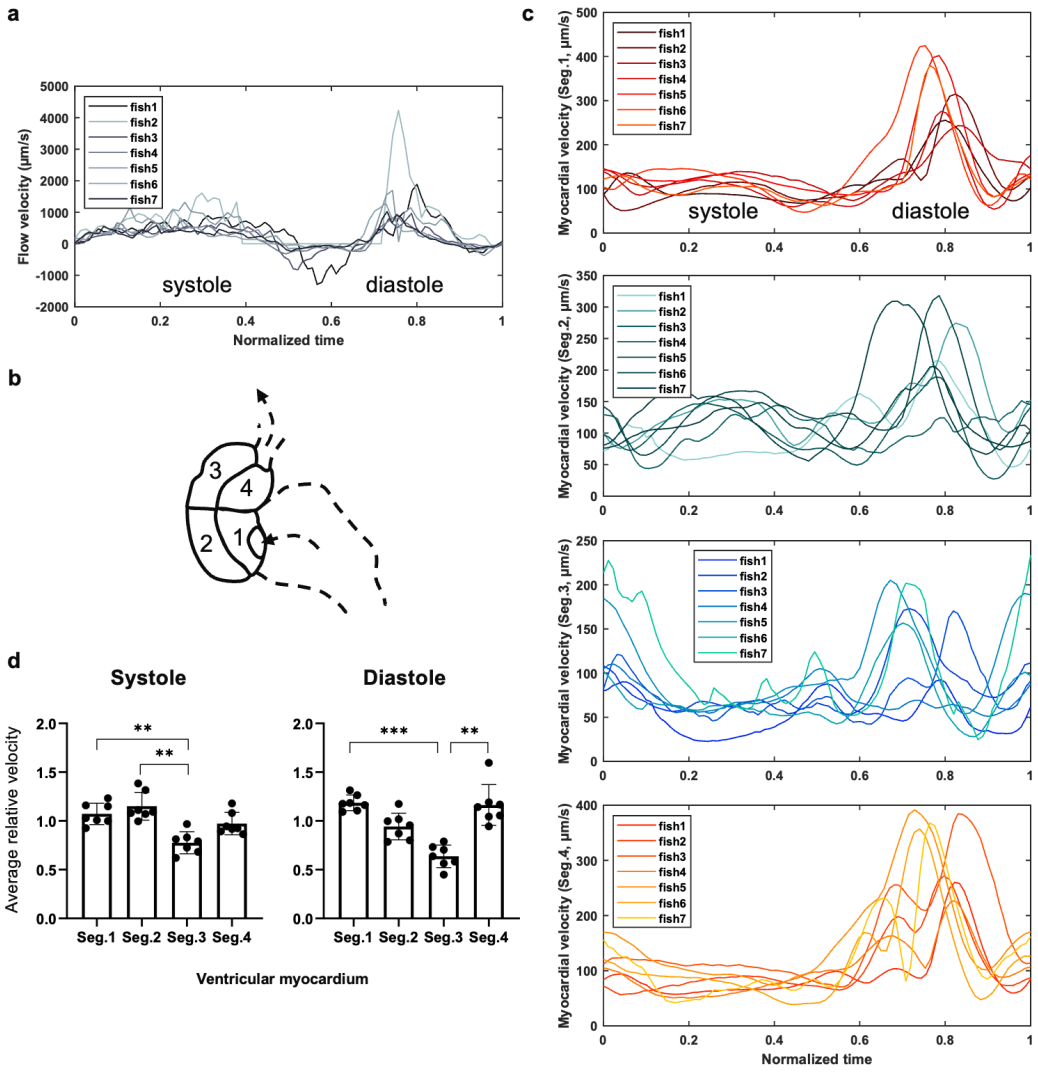

Figure 25. Velocity profiles in myocardial displacement and intracardiac blood flow at 3 dpf. (**a**) Blood flow velocity is depicted during a cardiac cycle. To compare different heart rates, we normalize the time to one cardiac cycle for each zebrafish embryo. (**b**) The ventricle is divided into 4 segments. Arrows indicate the flow direction. The dotted line delineates the atrium. (**c**) The tracing of the myocardial velocity is normalized to one cardiac cycle in each segment. (**d**) The mean of relative displacement velocity is depicted in each segment during systole and diastole. The relative velocity in each segment is compared as the ratio of the segmental velocity to the mean of global velocity during a cardiac cycle. Kruskal-Wallis test and Dunn's multiple comparisons test are performed in GraphPad Prism (** $p \leq 0.01$; *** $p \leq 0.001$, n = 7).

At 3 dpf, we observed similar temporal profiles in the blood flow velocity among several zebrafish, where blood flow was steady and slow during systole, but instantaneous and rapid during diastole (**Figure 25a**) in synchrony with the myocardial velocity (**Figure 25b,c**). To minimize the fish-to-fish variation, we normalized the segmental velocity of the myocardium to the global velocity for the individual zebrafish hearts during systole and diastole. We observed that myocardial contraction (systole) was more prominent in segments 1 and 2 (near the apex) as compared to segments 3 and 4, whereas myocardial relaxation (diastole) was more prominent in segments 1 and 4 (near the atrium) as compared to segments 2 and 3 (**Figure 25d**).

3.6 Comparison between direct reconstruction and post-reconstruction image enhancement

We further compared our 2D-3D VCD-Net reconstruction with a 2D-3D-better3D procedure accomplished by LFD reconstruction combined with state-of-the-art deep-learning image restorations (iso-CARE, 3DCARE)[28]. The computation workflows, including the data training and image application, of VCD-Net, LFD + iso-CARE, and LFD + 3D-CARE were illustrated in **Figure 26**. The sparse pointlike cardiomyocyte nuclei and densely-labeled continuous myocardium reconstructed by these three procedures were visually and quantitatively compared in **Figure 26b**, **c** and **d**, **e**, respectively. Though LFD + iso-CARE didn't necessarily require HR labelling data for training, this strategy showed limited enhancement to LFD, for either myocyte

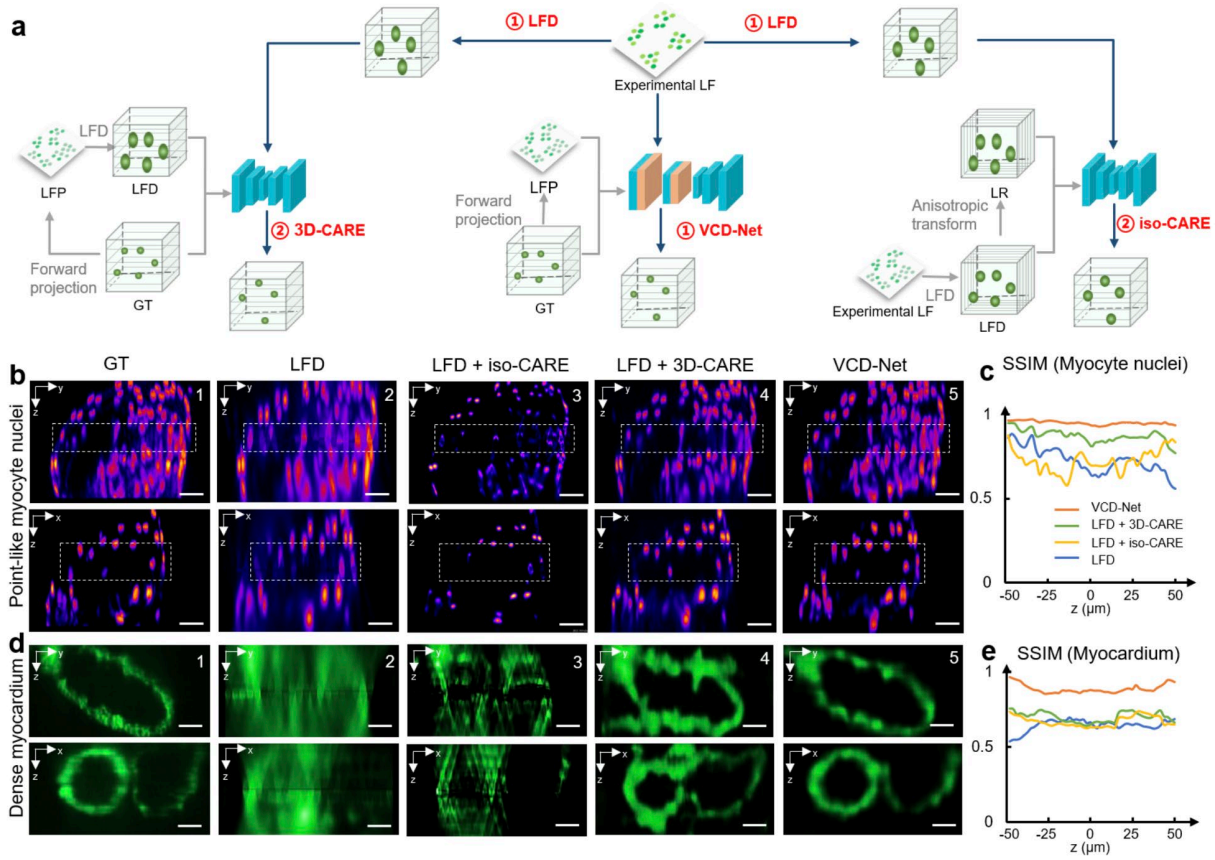

Figure 26. **Comparison between VCD-Net and LFD + deep-learning image restoration.** (**a**) Workflows of LFD + 3D-CARE (left), VCD-Net (middle) and LFD + iso-CARE (right) reconstruction procedures. (**b**) Comparative MIPs in y-z (top) and x-z (bottom) planes of the labeled cardiomyocyte nuclei by confocal microscope (GT, b1), LFD (b2), LFD plus iso-CARE (b3), LFD plus 3DCARE (b4), and our VCD-Net (b5) from left to right, respectively. White boxes show the regions near the native focal plane. (**c**) Structure similarity (SSIM) curve across the depth of the reconstructions, indicating the accuracy of each method when reconstructing non-dense nuclei signals. (**d**) Comparative MIPs in y-z (top) and x-z (bottom) planes of the densely-labeled myocardium by light-sheet microscope (GT, d1), LFD (d2), LFD plus iso-CARE (d3), LFD plus 3D-CARE (d4), and our VCD-Net (d5) from left to right, respectively. (**e**) SSIM curve of each method when reconstructing dense myocardium signals. Scale bar, 25 μm

nuclei or myocardium (**Figure 26b**3, **Figure 26d**3), owing to the still suboptimal lateral quality of LFD result (**Figure 26b**2, **Figure 26d**2). Unsurprisingly, the addition of 3D-CARE, which was trained by HR labelling data and corresponding synthetic LFD images, showed more significant improvement to LFD results (**Figure 26b**4, **Figure 26d**4). However, while LFD + 3D-CARE could reconstruct non-dense myocyte nuclei signals with similar quality (**Figure 26b**4, **Figure 26b**5)

except for a few lost signals near the focal plane (white boxes), its recovery on densely-labeled myocardium was still far worse, as compared to VCD-Net (**Figure 26d**4, **Figure 26d**5). Noticeable restoration hallucinations arose from the excessive artefacts in LFD results, and also caused unacceptably low structure similarity (SSIM) to the ground-truth data (**Figure 26e**).

Methods	Data preparation time	Model training time	Reconstruction speed
LFD	/	/	~90*s* per volume
LFD + iso-CARE	~20*h* (LFD)	~3.3*h*	~90.4*s* per volume
LFD + 3D-CARE	~22.2*h* (LFD + LFP)	~9.4*h*	~90.5*s* per volume
VCD-Net	~2.2*h* (LFP)	~5*h*	~0.03*s* per volume

Figure 27. Table comparing the efficiency of 4 light-field reconstruction strategies, in terms of the model implementation time and reconstruction speed. The data preparation and model training time of each method was calculated based on the same 800 groups of myocyte data. The reconstruction time of each approach was calculated by recovering a 3D output of myocytes (330*330*51) from the same light-field input (330*330). All the tests were performed on the same workstation equipped with RTX 2080 Ti graphic card.

It's also noted that while these combinatorial recovery strategies show more or less enhancement to LFD results, they have issue of low application efficiency. Since they need to first go through the iterative LFD for each frame of a light-field video, the processing speed is even lower than LFD only, over threeorder-lower than VCD reconstruction, and thus impractical for many applications (**Figure 27**). Furthermore, such additional deep-learning restoration to LFD results also requires a lot of LFD data prepared for model training, which is more time consuming than generating LFP data for VCD-Net training (~10× slower, **Figure 27**). Therefore, from the perspectives of reconstruction quality, reconstruction speed and model implementation, which are all important factors for a computational imaging technique, we have shown the notable advantages of our VCD-Net approach, even when compared with the combination of two established restoration methods. Especially under dense labelling conditions, only VCD-Net can

realize accurate light-field reconstruction at high resolution, which remains unachievable by alternative approaches.

3.7 The generalization ability of VCD-Net

Generalization ability has been an important characteristic of a deep-learning-based model. In this section, we demonstrate hybrid data training and multi-sample recovery, and cross-sample or transfer learning applications, to evaluate the generalization ability of VCD-Net.

3.7.1 Performance of VCD-Net trained on hybrid cardiac data

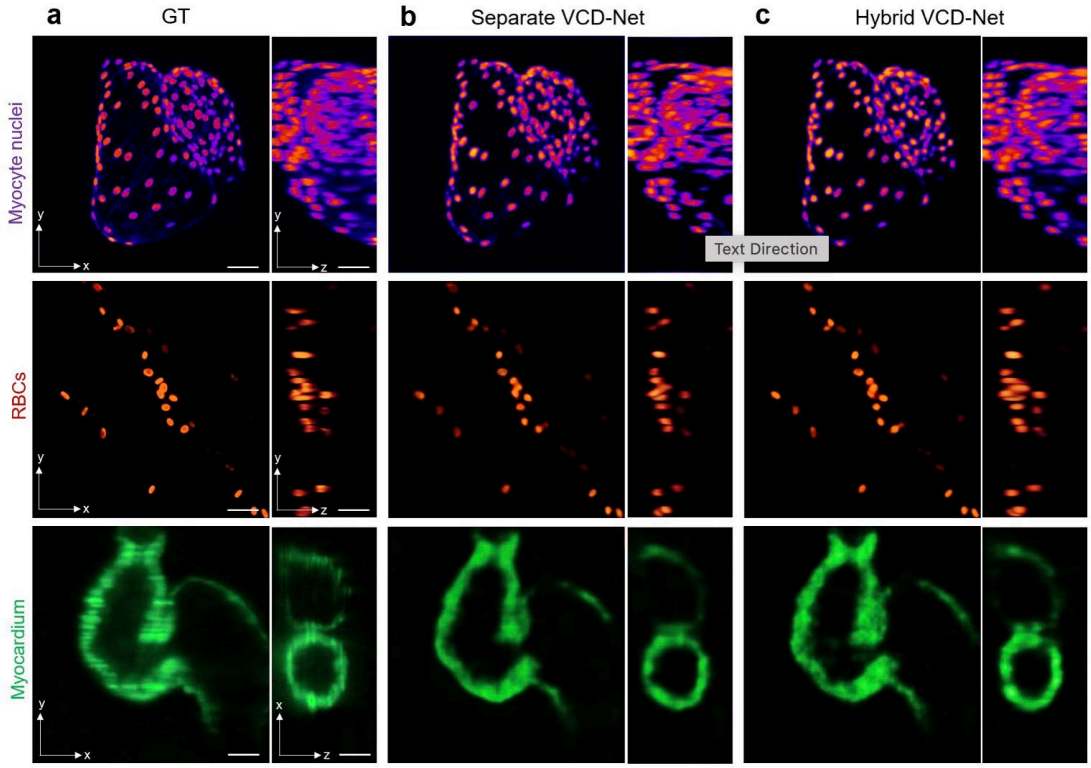

Figure 28. **Performance of VCD-Net trained on hybrid cardiac data.** (**a**) High resolution confocal images of the myocyte nuclei (row 1), RBCs (row 2), and light-sheet microscopy image of myocardium (row 3). (**b**),(**c**) Reconstructions of the same samples by separate VCD-Nets with each trained on a single type of data, and hybrid VCD-Net trained on mixed cardiac datasets, respectively. Scale bar, 30 μm.

Unlike the conventional network implementation based on a single type of sample data, here a hybrid cardiac VCD-Net was first trained on mixed datasets containing GT-LFP image pairs of myocyte nuclei, red blood cells and myocardium (**Figure 28**), and then applied to the light-field reconstruction of all the three types of samples. The results by such a hybrid VCD-Net (**Figure 28c**) were compared with those by three individual VCD-Nets that were trained on myocytes nuclei, RBCs and myocardium separately (**Figure 28b**). The similar reconstruction quality by the hybrid network indicates that the VCD-Net could generalize well when trained on many datasets and applied to many samples.

3.7.2 Cross-sample and transfer learning applications of VCD-Net

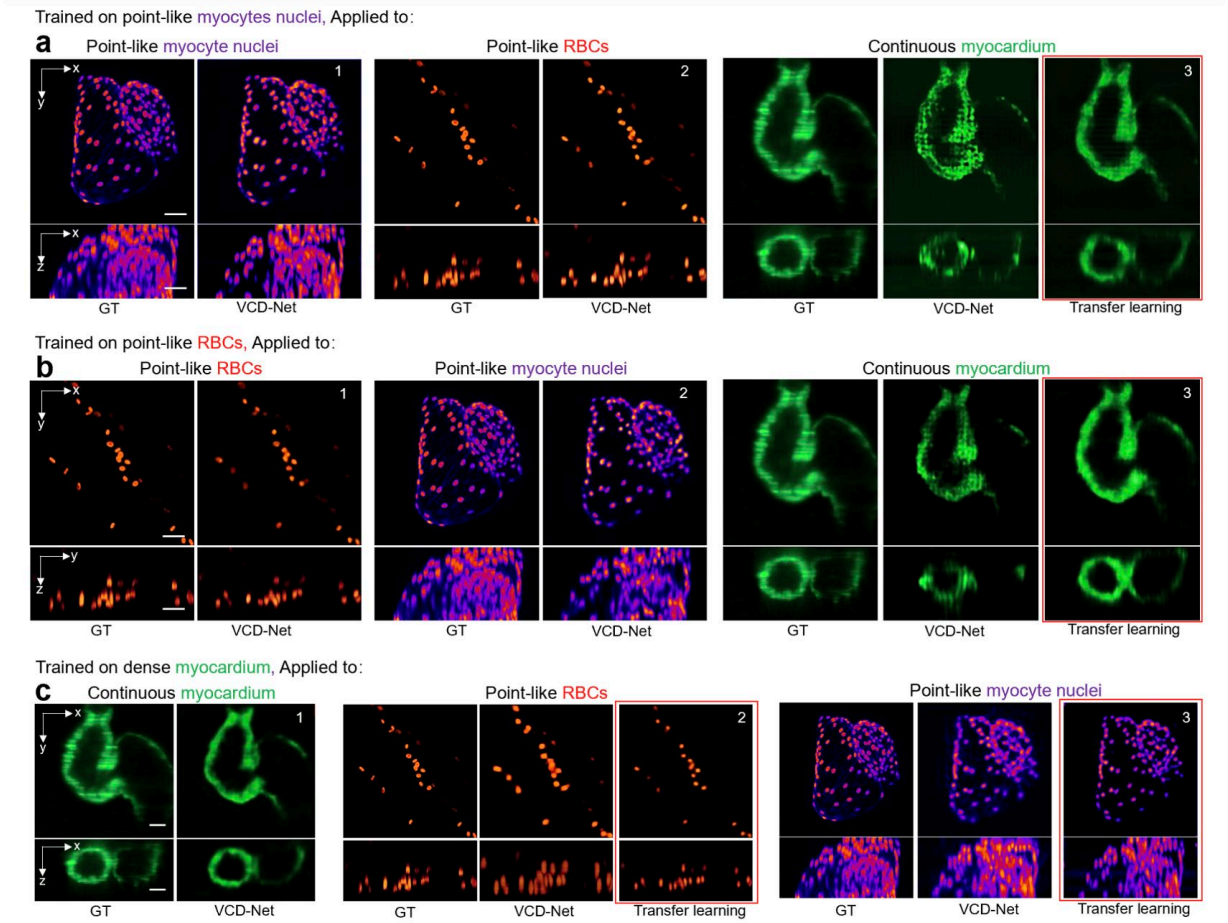

Figure 29. **Cross-sample and transfer learning applications of VCD-Net.** (**a**)-(**c**) Cross applications to myocyte nuclei, RBCs, and myocardium, by VCD-Net models which are solely trained on myocyte nuclei, RBCs, and myocardium, respectively. This limitation from different signal types can be overcome by introducing transfer learning (red boxes). Scale bar, 30 μm

We trained 3 VCD-Nets, using image pairs of the sparse red blood cells, mid-density myocyte nuclei, and highly dense myocardium, respectively, and applied each of networks to the reconstruction of all the three types of samples (**Figure 29**). As compared to the high-quality reconstruction for the same types of samples (a1, b1, c1), the quality of cross-applications was also acceptable (a2, b2) when recovering the similar types of signals (point-like myocyte nuclei network for point-like RBCs in a2, or vice versa b2). At the same time, the reconstruction quality was severely compromised when applying pointsignals-trained networks to continuous myocardium data (a3 middle, b3 middle), or vice versa (c2 middle, c3 middle).

To overcome the barrier between such different signal types, we also introduced transfer learning, which leveraged the knowledge already learned by the previously trained network and thus required much fewer training data and iterations, to further enhance the generalization ability of VCD-Net. In practice, we saved the best checkpoints of the pre-trained point-signal-based and continuous-signal-based VCD models, and then trained them using small amount (~20%) of data from continuous sample and point-like samples, respectively. As shown in the right columns of a3, b3, c2, c3, after transfer learning applied, the previously corrupted reconstructions caused by style-mismatching were mitigated. Sufficiently accurate reconstructions have been provided, as compared to the GT data. Therefore, the VCD-Net could be highly generalizable when trained on one type of data and applied to another, especially when a transfer learning based on small amount of target data involved.

3.8 Discussion

The VCD-LFM achieves real-time recording and video-rate reconstruction of instantaneous 3D processes in whole moving C. elegans. Combined with efficient network-based locomotion analysis, it offers an efficient pipeline for the study of sustained worm neural activity and related locomotion behavior at high throughput. The robust performance of VCD-LFM enables the suppression of signal artifacts, elimination of motion blurs and accurate quantification of calcium signaling. For cardiovascular imaging, recent scanning-based approaches for volumetrically imaging in zebrafish larvae have required complicated optics, an ultra-fast camera and high-intensity excitation[34,40]. In contrast, our method based on a relatively simple system and easily adoptable deep-learning framework offers a compelling solution for investigating the dynamic properties and functions of the cardiovascular system. Therefore, VCD-LFM could be a valuable tool for studying dynamics on fast timescales, potentially benefiting a variety of applications such as behaviorally relevant neuronal activity studies and dysfunctions of the heart and blood transport system in model organisms.

In summary, we introduced a VCD-LFM approach and demonstrated its ability to image transient biological dynamics with improved spatial resolution, minimal reconstruction artifacts and increased reconstruction throughput compared to conventional LFDM approaches. The network-based VCD computational model is robust, versatile and ready for widespread application. While VCD-LFM improves the reconstruction quality from one originally determined by the optical system to one that can be optimized via the training procedure, it requires the preacquisition of a considerable number of training images. We expect this will improve with the continued development of deep-learning technique, which aims for strong generalization ability and weak training supervision, thus allowing the model implementation with much fewer training data

required. Aside from the combination with a basic LFM setup[5,6], we note that VCD-Net is compatible with modified LFM modalities, such as a dual-objective setup[26] or a Fourier LFM setup[41]. Finally, we expect that VCD-LFM could potentially bring new insights for computational imaging techniques by raising the possibility of restoring image beyond the system optical limit rather than just approaching it and showing the capability of increasing the image dimension while minimally compromising the image quality. Taken together, we can further push the spatiotemporal limits for in toto observation of dynamic biological processes.

Chapter 4
Light field microscopy with squeezed camera readout

4.1 Motivations for kilohertz volumetric microscopy

High-speed fluorescence microscopy has been playing an indispensable role in revealing the dynamic interplay and functionality among cells in their native environment. With continuous improvements in fluorescent markers, many transient biological processes, such as the blood flow[42] and neural action potentials[43–45], become trackable and, thus, demand microscopy with an ever higher spatiotemporal resolution. Traditional three-dimensional (3D) imaging tools, such as confocal microscopy, light sheet microscopy, and two-photon microscopy, heavily rely on scanning to acquire a volumetric image. Despite advancement in beam shaping[46,47], remote refocusing mechanisms[48] and detection geometry[40], there persists an inherent trade-off between temporal resolution, the 3D field-of-view (FOV), and spatial resolution in these techniques. This constraint marks a significant challenge to obtain optimal performance across a large 3D field of view (FOV) for robust ultrafast detection exceeding kilohertz (kHz).

Computational imaging mitigates this trade-off by encoding high-dimensional information, such as depth[49], time[50], and spectra[51–53], into two-dimensional (2D) multiplexed camera measurements. Among these techniques, light-field microscopy (LFM) excelled in various biological applications, including observation of neural activity in freely moving animals[6,10,54] and visualization of hemodynamics in the brain[15] and heart[26,30,55]. By simultaneously collecting the spatial and angular information of light rays, LFM enables volumetric reconstruction *post hoc* from snapshot measurements. Without scanning, the sensor bandwidth becomes the primary bottleneck for LFM 3D imaging speed. While modern scientific Complementary Metal-Oxide Semiconductor (sCMOS) sensors typically offer a full framerate lower than 100 Hz, increasing the imaging speed can be achieved by reading out only selected low-format regions of interest (ROI). However, this approach comes at the cost of sacrificing either the spatial and/or angular components associated with the field of view (FOV) and axial resolution.

The integration of ultra-high-speed cameras[40,48,56] and event cameras[57] holds promise for providing higher bandwidths to LFM. However, their current limitations in sensitivity and noise performance present challenges, especially for photon-starved applications like imaging genetically encoded voltage indicators (GEVIs).[48] On the other hand, the compressibility of four-dimensional (4D) (two spatial dimensions plus two angular dimensions) light fields has been leveraged for compressive detection. Coded masks[58–60] and diffusers[61] are employed to modulate and integrate the spatio-angular components originally recorded by distinct pixels. Sparse nonlocal measurements can also be utilized across different angular views to acquire light fields with sensors of arbitrary formats.[22,62,63] Nevertheless, compressive imaging relies on sparsity priors and optimization algorithms for signal recovery from the sub-Nyquist measurement dataset. These

methods are primarily validated on natural photographic scenes under ample lighting conditions. Their robustness and effectiveness in high-speed fluorescence microscopy remain elusive.

To address the unmet need for high-speed light-field imaging, we present herein Squeezed LIght field Microscopy (SLIM)[64], which allows the capture of 3D fluorescent signals at kilohertz volume rates in a highly data-efficient manner. SLIM operates by acquiring an array of rotated 2D sub-aperture images. An anamorphic relay system applies anisotropic scaling, effectively 'squeezing' the image along one spatial axis. This allows the camera sensor to detect the light field using only a low-format letterbox-shaped ROI. Leveraging the row-by-row readout architecture of modern CMOS sensors, SLIM achieves a more than fivefold increase in acquisition rate compared to traditional LFM. Each squeezed sub-aperture image complements the others, facilitating high-fidelity 3D reconstruction from compressed measurement while preserving the nominal FOV and spatial resolution.

We demonstrated SLIM by capturing the flowing red blood cells in an embryonic zebrafish at 1,000 volumes per second (vps), neural voltage spikes in leech ganglions at 800 vps and planktonic *Vibrio cholerae* swimming motility at 200 vps. SLIM enables tracking high-speed cellular motion and millisecond membrane action potentials across a 550 μm FOV within a 300 μm depth range. Furthermore, we showcased that the high framerate of SLIM could be exploited to enhance axial resolution when combined with multi-layer scanning light-sheet microscopy. This allows for imaging densely labeled structures, previously challenging with LFM, such as contracting myocardium in a zebrafish, at 4,800 frames per second (fps), leading to a volume rate of 300 vps.

4.2 Principle and design of SLIM

In a typical SLIM camera (**Figure 30**), the input scene is imaged by a combined system consisting of an array of dove prisms and lenslets (**Figure 30i**, **Figure 31**). Each dove prism within the array is rotated at a distinct angle relative to its optical axis. This arrangement gives rise to an array of perspective images, each rotated at twice the angle of its corresponding dove prism's rotation, all converging at an intermediate image plane situated behind the lenslets. Subsequently, these rotated perspective images are further processed through an anamorphic relay system consisting of two cylindrical doublets with orthogonal optical axes (**Figure 30ii**, **Figure 32**). This relay system imparts anisotropic scaling to the image array, where the images experience de-magnification (0.2X) along one spatial axis while preserving the original magnification along the orthogonal direction. Finally, the rescaled image array is acquired by a 2D camera, where we read out only pixel rows that receive light signals (referred to as active readout ROI in **Figure 30**).

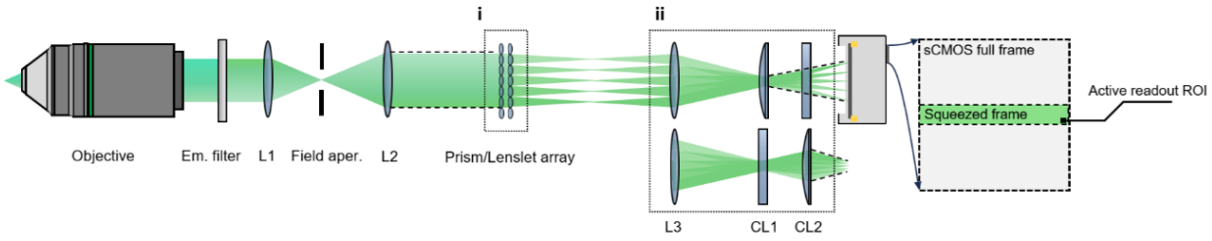

Figure 30. **Schematic of SLIM detection system.** L1-L3, achromatic doublet. CL1-CL2, achromatic cylindrical doublet. SLIM records light field at kilohertz frame rates by using a reduced active readout ROI on the camera.

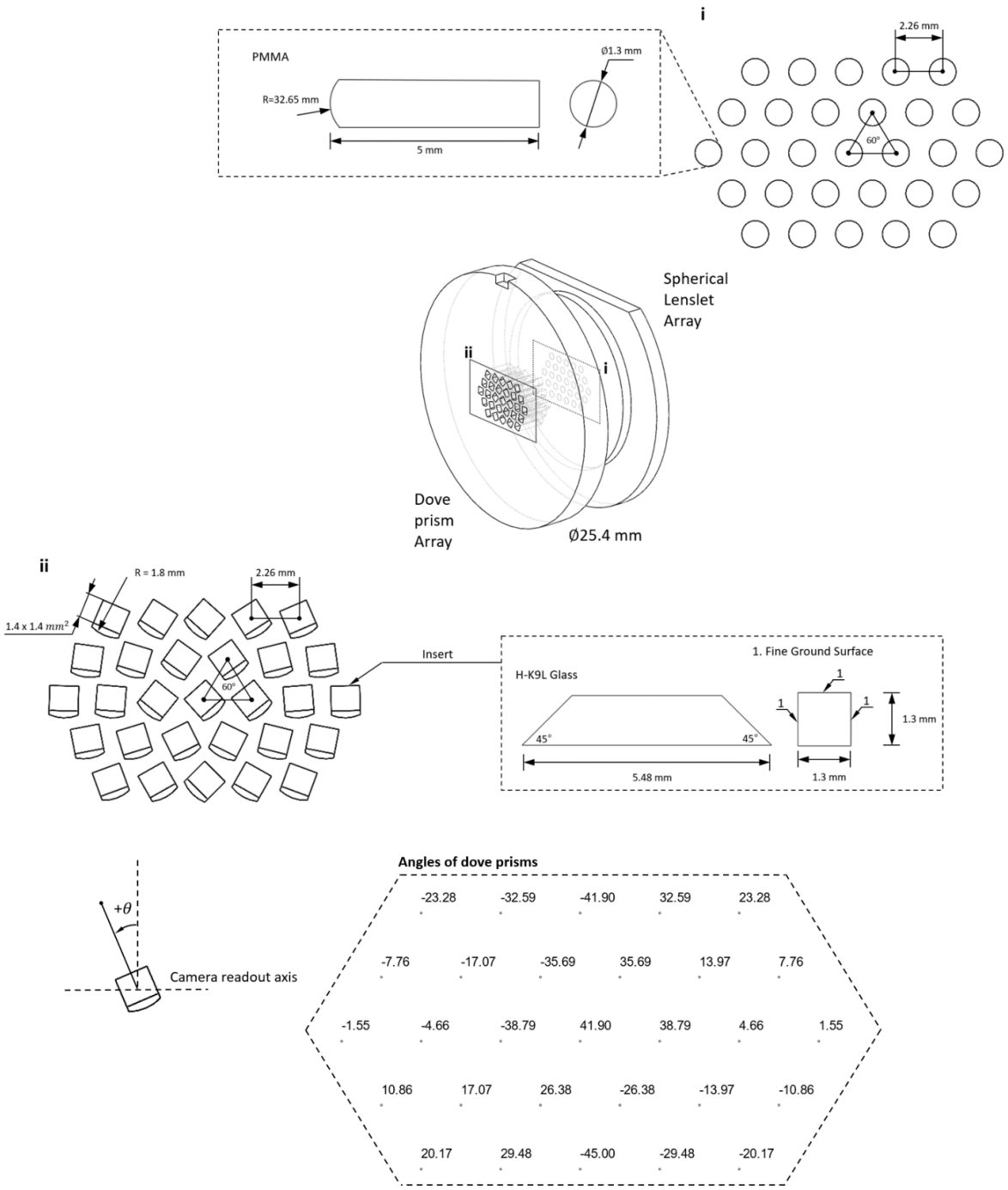

Figure 31. Design of lenslet and dove prism array. (i) The lenslet array is in-house fabricated on a PMMA substrate. (ii) The holder of dove prims, the dove prism, and the angle arrangement of dove prisms.

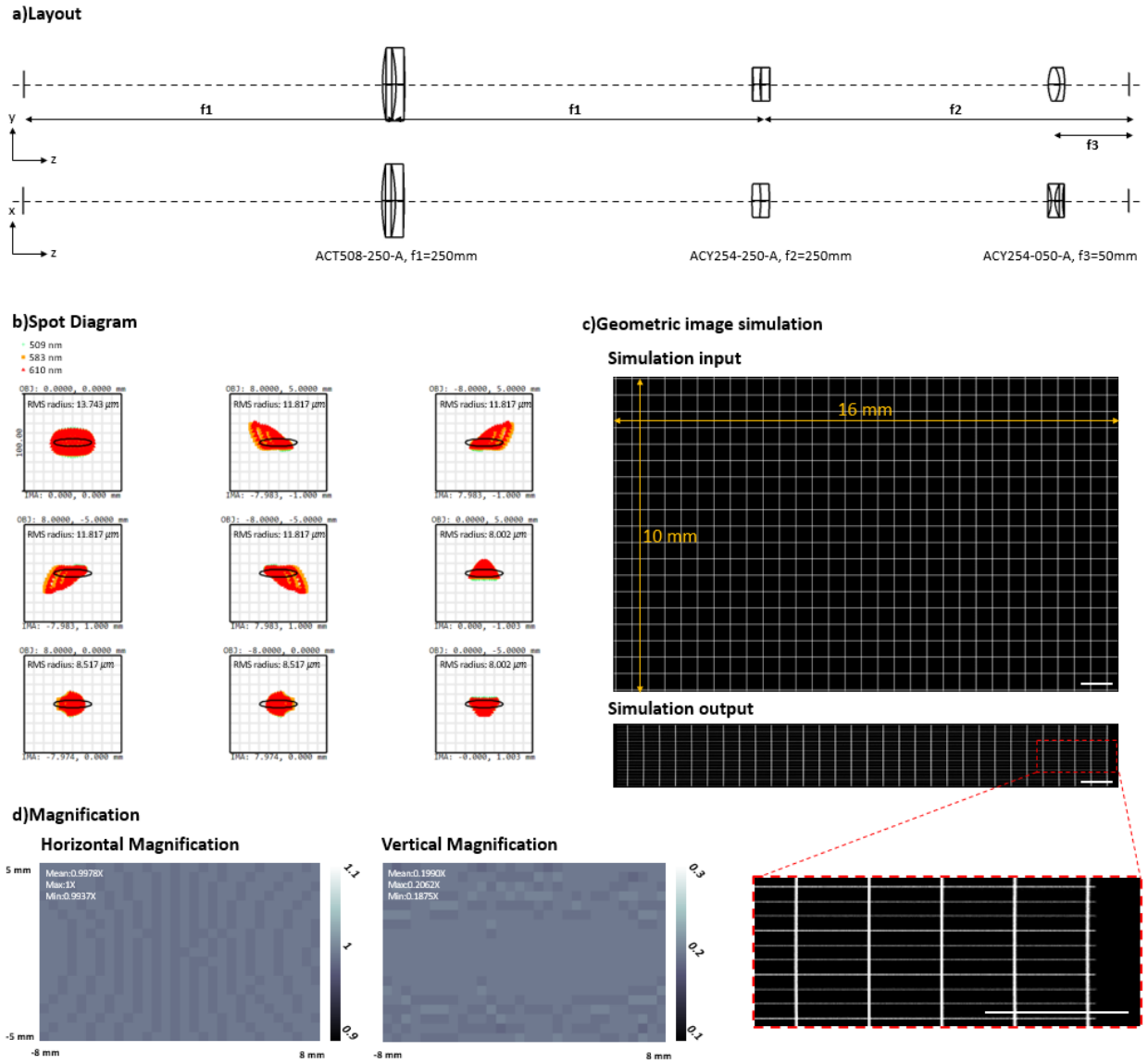

Figure 32. Design of anamorphic relay lens. (**a**) Layout of the relay system, consisting of one spherical achromatic doublet and two orthogonally placed cylindrical achromat doublets, where off-the-shelf components (Thorlabs) were used. (**b**) Spot diagram shows aberration (within two times the pixel size, 6.5 μm) across the field of view of the lenslet array image (16 mm × 10 mm). (**c**) OpticStudio Zemax simulations on an image of a grid. The red dotted box shows the zoom-in picture. Scale bar, 1 mm. (**d**) Horizontal and vertical magnification measured from the grid simulation. The results show a uniform and constant scaling factor across the field of view, which ensures correct image transformation for all sub-aperture images.

One of the key advantages of employing squeezed optical mapping is the improved readout speed. Modern CMOS sensors are equipped with parallel analog-to-digital converters (ADCs) for

each column of pixels, ensuring consistent frame rates regardless of the number of pixel columns being readout. The frame rate is, therefore, solely determined by and inversely proportional to the number of pixel rows being readout. For example, on the Kinetix sCMOS developed by Teledyne, using an ROI of 200×3200 pixels allows SLIM to capture a 19 sub-aperture image array at 1,326 fps and 7,476 fps in 16-bit and 8-bit mode, respectively. In contrast, the full-frame mode achieves frame rates of only 83 fps (16-bit) and 500 fps (8-bit).

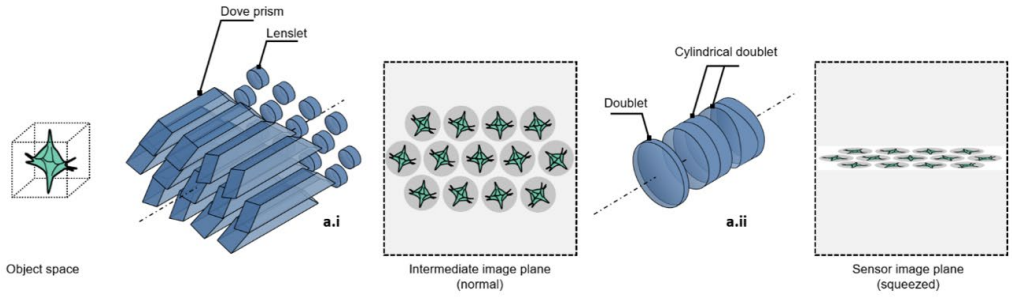

Figure 33. The optical transformation in SLIM comprises image rotation and squeezing, performed by a dove prism/lenslet array (**a.i**) and a customized anamorphic relay (**a.ii**).

The forward model of SLIM is illustrated in **Figure 33**. Similar to Fourier LFM (FLFM)[9,10], SLIM can be conceptualized as a tomographic system, where each sub-aperture image is essentially a parallel projection of a 3D volume along a line of sight at the sub-aperture's view angle[3,5]. However, unlike FLFM, where these sub-apertures images are directly captured by a 2D camera, SLIM applies in-plane rotation and vertical scaling operations to these images before recording.

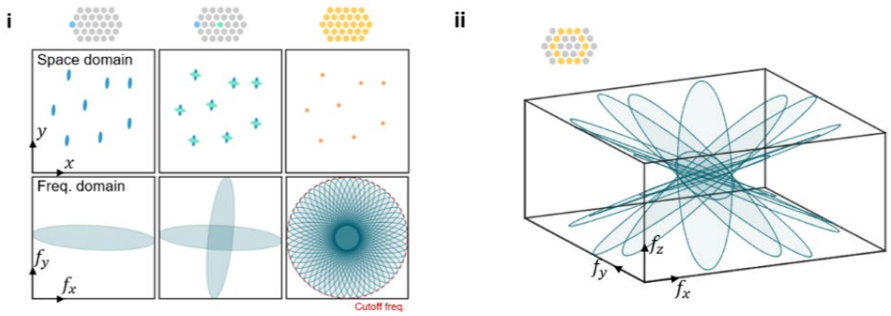

Figure 34. The frequency analysis of SLIM. (**i**) Each sub-aperture, after reversing the squeezing and rotation, gives an image with anisotropic spatial resolution but complementary to others. By merging different sub-aperture images, SLIM estimates the original features. (**ii**) Transfer functions of SLIM by analyzing sub-aperture images as geometrical projections. Each gives an elliptical slice in the 3D frequency space, depending on its rotation angle and sub-aperture location.

In the 3D spatial frequency space, the Fourier spectrum of a SLIM sub-aperture image manifests as a 2D elliptical slice (Fourier Slice Theorem). The short axis of this ellipse corresponds to the low-resolution sampling along the squeezing direction. By using an array of sub-aperture images rotated at complementary angles, SLIM fills in the missing high-frequency information. This process results in a synthesized power spectrum with a bandwidth that approximates that of the original unsqueezed FLFM (**Figure 34**). In addition, the rotation angles of sub-aperture images are carefully crafted to maximize the horizontal projections of their 3D point spread functions (PSFs) (**Figure 35**). In other words, when imaging a 3D object, the sub-aperture images of SLIM

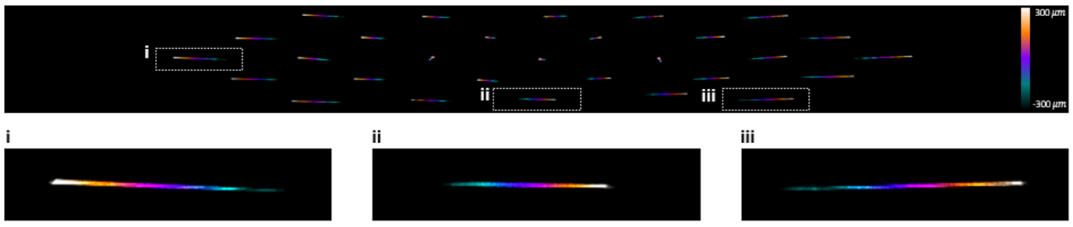

Figure 35. x-y MIP of the SLIM PSFs with depth color-coded.

exhibit lateral disparity shifts due to their view angle difference. We optimize the rotation angle of each sub-aperture image to align its disparity shift with the unsqueezed spatial axis (*i.e.*, camera pixel row direction), thereby maximizing the samplings of disparity and consequently enhancing the axial resolution. This forward model can be further extended to wave optics by using a sum of 2D convolutions between the sample sliced at each depth and the corresponding sub-aperture PSF[9,10]. Through an iterative deconvolution algorithm, SLIM reconstructs the 3D fluorescence distribution by fusing all sub-aperture images.

Figure 36a,b show the 3D reconstruction of fluorescent beads of sub-diffraction size imaged by SLIM. At a magnification of 3.6X, the imaging volume spans a 3D space of Ø550 μm × 300 μm with a spatial resolution of 3.6 μm laterally and 6.0 μm axially (**Figure 36c-f**). Despite a different sampling strategy, SLIM inherits the first-order parameters (*e.g.*, resolution, FOV, and depth range) of FLFM. More importantly, its highly efficient data acquisition manner alleviates the strain on the camera's bandwidth, thereby enabling kilohertz frame rates.

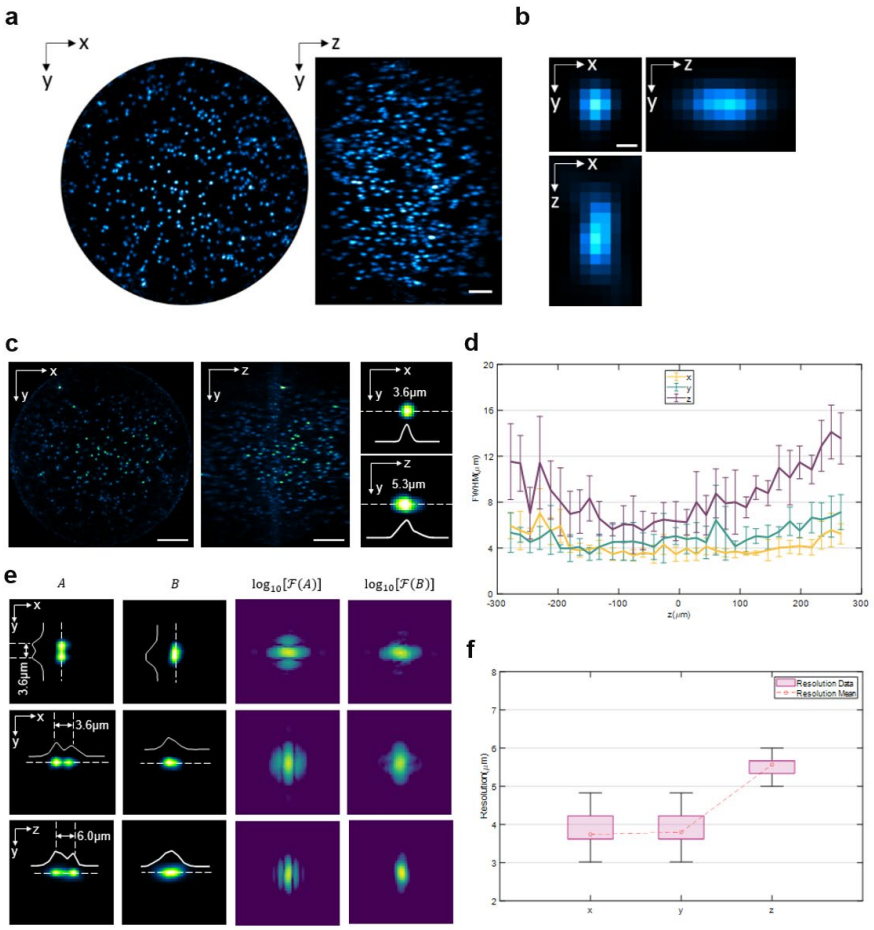

Figure 36. SLIM performance demonstrated on fluorescent beads. (**a**) 3D MIPs of fluorescent beads. Scale bar, 50 μm. (**b**) Cross-sections of single bead showing representative PSFs. Scale bar, 3 μm. (**c**) Fluorescent beads captured by statistic analysis. (**d**) Average axial (z) and lateral (x,y) FWHM of the beads across the volumes reconstructed by SLIM. Center lines represent means and error bars denote standard deviations. (**e**) Analysis of the reconstructed images of two virtually separated beads obtained by SLIM with cross-section profiles along the dashed lines. We imaged the same 1μm bead at two positions, adjusting the interval gradually using a piezo translation stage. By combining the images obtained from these two positions, we created two virtually separated beads with an arbitrary distance. The first and second columns (A and B) are the reconstructed images of resolved beads (A) and unresolved beads (B). The third and fourth columns are the Fourier analysis of the first and second columns using function: f(x)=log(|F(x)|), where F(x) represents the Fourier transform. (**f**) FWHM resolution of two virtually separated beads obtained by SLIM along three directions.

4.2.1 Optical setups

We implemented selective volume illumination to suppress fluorescence outside the imaging volume (**Figure 37**). Depending on the specific applications, we have the option to choose between two variants: 1. Scanning light sheet generated by a high-power laser source, capable of delivering

up to 500 mW (**Figure 37i**); 2. Static LED featuring ultra-low-noise performance (**Figure 37ii**). The light sheet scanning is controlled by a galvo-mirror, and we synchronize the camera's exposure with a complete scan for synthetic volumetric illumination. On the other hand, we shape the LED's beam using a slit and project its conjugate image to the sample. The two methods share the same orthogonal illumination geometry.

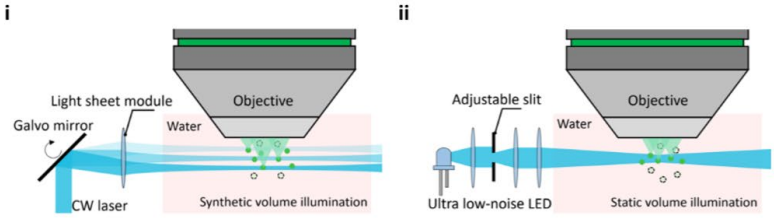

Figure 37. Selective volume illumination with a scanning light sheet (i) and LED (ii).

The detection setup features a 20X water-dipping objective (N20X-PFH, Olympus XLUMPLFLN 20X, 1.0 NA). A 4F relay system (AC508-180-A, AC508-200-A, Thorlabs) forms a conjugate plane of the objective's back pupil, accommodating a customized dove prism and a spherical lenslet array. The dove prism (aperture length: 1.3 mm, material: H-K9L, fabricated by Changchun Sunday Optics) is positioned anteriorly to the plano-convex lenslet (aperture diameter: 1.3 mm, focal length: 36 mm, material: PMMA, fabricated in-house). Each pair generates a rotated sub-aperture image with a magnification of 3.6X and NA of 0.065. In total, 29 pairs are utilized and securely housed in 3D-printed mechanical holders (refer to **Figure 31** for detailed designs of the prism, lenslet, and holder). Our anamorphic relay system comprises a spherical achromat doublet (ACT508-250-A, Thorlabs) and two orthogonally oriented cylindrical achromat doublets (ACY254-250-A, ACY254-50-A, Thorlabs). The back focal planes of two cylindrical lenses are

co-located, producing an image with an anisotropic scaling factor (**Figure 32**). A sCMOS camera (Kinetix, Teledyne) captures the final image, with a 320×3200 pixels ROI covering all 29 sub-aperture images, or a 200×3200 pixels ROI for 19 sub-aperture images. The maximal readout speeds for two ROIs are 830 fps and 1,326 fps in 16-bit dynamic range mode and 4,790 fps and 7,476 fps in 8-bit speed mode. The 29 sub-aperture configuration collects around 50% more light than the 19 sub-aperture one.

The illumination sources include blue and green continuous lasers (MBL-FN-473-500mW and MGL-III-532-300mW, CNI Laser) and an ultra-low-noise blue LED (UHP-T-470SR Prizmatix). For scanning light sheet setup, we use a knife-edge mirror (MRAK25-G01, Thorlabs) to combine two beams with adjustable spacing and a galvo-mirror (GVS011, Thorlabs) to scan them together. Planar illumination is formed perpendicular to the detection axis by a cylindrical lens and a dry objective (RMS4X-PF, Olympus 4X, 0.13 NA). We use a sawtooth function to drive the galvo-mirror. In synthetic volume illumination configuration, we block one light sheet beam. The camera is then triggered at the beginning of the sawtooth waveform and exposed for the entire scan. In scanning plane illumination configuration, we use two light sheet beams and trigger the camera multiple times during a scan. The static LED setup shares the same illumination objective and

perpendicular geometry. We built a Koehler illumination system and used an adjustable slit as a field aperture. The conjugate image of the slit is relayed to the sample, and the slit controls the depth range of the beam. The LED provides ultra-stable illumination power and thus suppresses excitation source noise during our voltage imaging experiments. See **Figure 38** for system schematics.

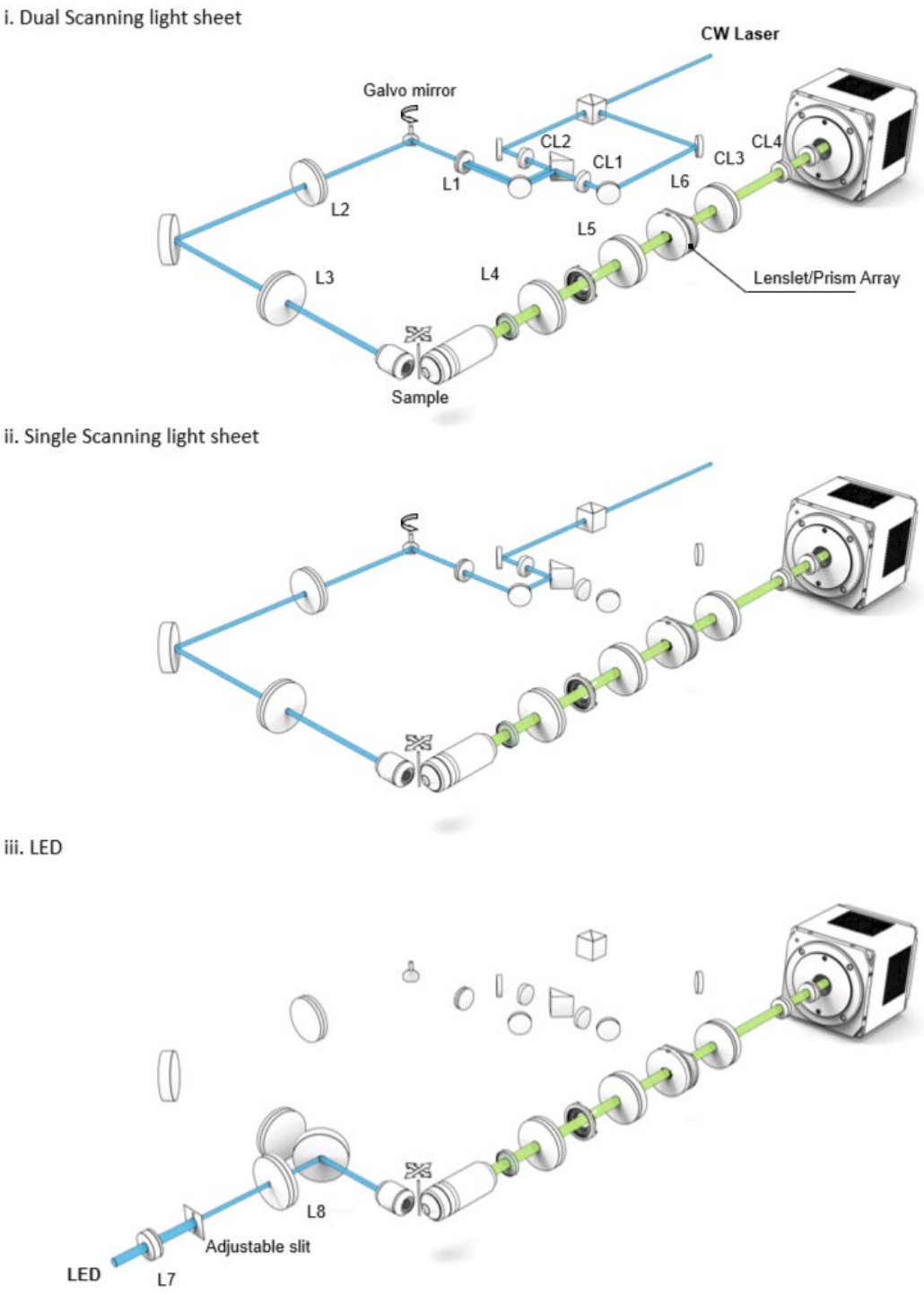

Figure 38. Schematics of the microscope with different illumination systems. L1-L8, lens; CL1-CL2, cylindrical lens.

4.2.2Reconstruction algorithm

The light field of the fluorescent sample is acquired by dividing the objective's back pupil with a lenslet array and recording a group of sub-aperture images. Depending on their sub-aperture locations, they display disparity, that is the distinct displacement shown by the same signal. After calibrating the displacement at every axial position, the formation of each sub-aperture image can be modeled as a sum of laterally shifted depth slices. We replace the shifting operator with a convolution with PSF to account for both the diffraction and displacement. A dove prism is a truncated right-angle prism and used to rotate the incident beam. The rotation of the prism around its longitudinal axis causes the beam to rotate at twice the rate of the prism's rotation. By placing a dove prism array in the infinity space between the objective and lenslet array, we apply varying in-plane rotations to sub-aperture images. Finally, we adopt an anamorphic relay system to introduce anisotropic scaling to the image array: we de-magnify (squeeze) the image in the direction perpendicular to the camera read-out axis while maintaining the original scale in the other direction. This one-axis scaling and the aforementioned in-plane rotation are both directly applied to the 3D fluorescent image in our model.

Given a 3D fluorescence distribution $O(x, y, z)$ and system PSF, the formation of camera measurement $I(x, y, v)$ can be modeled as:

$$I(x, y, v) = \sum_{z} \big(S \cdot R(v) \cdot O(x, y, z)\big) \otimes PSF(x, y, z, v) \tag{9}$$

where v is the index of sub-aperture, $\otimes$ represents the 2D convolution, R(v) applies sub-aperture-dependent rotation from the dove prism, and S introduces image scaling from the anamorphic relay system.

The volume reconstruction algorithm was derived from Richardson-Lucy deconvolution[6,9,10,65]. Based on the forward model (8), the 3D fluorescence distribution $O(x, y, z)$ is iteratively solved from the camera measurements $I(x, y, v)$ and empirical point spread functions PSF. The rotation angles and image scaling factor are pre-calibrated as known priors in the reconstruction. The measurement patch $I(x, y, v)$ is cropped from the raw sensor image according to the center location of each sub-aperture image. We experimentally measure the point spread functions by

Algorithm 1 SLIM reconstruction using Richardson Lucy deconvolution

$s,\ \theta(N_v),\ PSF(N_x, sN_y, N_z, N_v),\ im(N_x, sN_y, N_v)$
▷ Input squeezing factor s, rotation angle θ , normalized PSF, and measured light field views im
$vol^0(N_x, N_y, N_z) = vol_{init}$ ▷ Initialize volume
for iterations **do**
    $\overline{im} = \mathrm{FP}(vol^k)$ ▷ Forward project the current volume estimation
    $vol^{k+1} = vol^k \cdot \mathrm{BP}(im/\overline{im})$ ▷ Backward project the error and update volume
end for

function FP(vol) ▷ Forward project the volume of dimensions N_x, N_y, N_z
    Initialize $im = zeros(N_x, s \cdot N_y, N_v)$
    for each v in N_v **do**
        Apply in-plane rotation ($\theta(v)$) to entire vol
        Resize image at scale s in direction y for entire vol
        Slice-by-slice convolution between resized vol and $PSF(:,:,:,v)$
        $im(:,:,v)$ = Sum along depth of the convolution result
    end for
    return im
end function

function BP(im) ▷ Backward project the light field views of dimensions $N_x, s \cdot N_y, N_v$
    Initialize $vol = zeros(N_x, N_y, N_z)$
    for each v and d in N_v and N_z **do**
        $PSF^T(:,:,d,v)$ = rotate $PSF(:,:,d,v)$ at 180 degrees
    end for
    for each d in N_z **do**
        View-by-view convolution between im and $PSF^T(:,:,d,:)$
        Resize convoluted views at scale $1/s$ in direction y
        Apply in-plane rotation to each view at corresponding angle $-\theta(v)$
        $vol(:,:,d)$ = Sum along view of rotated results
    end for
    return vol
end function

Figure 39. Pseudocode of SLIM reconstruction algorithm.

imaging a sub-diffraction fluorescent bead. The point source is placed in the middle of FOV and axially scanned over a broad 600 μm depth range with a 2 μm step size using a motorized translation stage. The actual axial range and step size in the reconstruction depends on the specific experiments. The PSF is assumed to be spatially invariant within each sub-aperture image.

With our implementation, each measurement patch $I(x, y, v)$ has a resolution of 301 × 61 pixels. The numbers of channels (v) and axial slices (z) are configured based on the targeted framerate and depth range/step size. For example, with 19 sub-aperture measurements to reconstruct a volume of 305×305×151 pixels, the deconvolution takes around 30 seconds via eight iterations using a desktop computer with a modest GPU (Nvidia RTX 3070).

4.3 Imaging of flowing red blood cells in an embryonic zebrafish

Experimental characterization of blood flow in living organisms provides valuable insights into local metabolism, vascular development, and disease states. Using fluorescently labeled blood cells, various imaging methods have been demonstrated in single-cell velocimetry, such as in the larval zebrafish heart[30,32,40], tail[66,67], and mouse brain[15,42]. However, these methods are often limited to 2D imaging or restricted by a limited volumetric frame rate, which hinders the detection of fast flow and necessitates sedation of the animal to reduce motion artifacts. Here, we show that SLIM could be used to capture the fast-circulating red blood cells (RBCs) in a zebrafish at a kilohertz volumetric rate, both with and without sedation.

We imaged the transgenic zebrafish embryos expressing DsRed in RBCs at three days post fertilization (dpf). We excited the zebrafish brain using light-sheet-synthesized volumetric illumination and recorded fluorescence using SLIM with 19 sub-aperture images at 1,000 frames per second. The reconstruction reveals the 3D distribution of RBCs and allows for cell tracking

over time. **Figure 40a** shows two separate recordings from the dorsal and ventral view, each visualizing RBCs at representative time points (red) and the vasculature network by maximum intensity projection (MIP) throughout all frames (cyan). The flowing velocity is pulsatile temporally and varies spatially in the aorta and vein. The tracking reveals the velocity distribution in 3D and highlights vessels with a high-speed flow of up to 6 mm/s (**Figure 40a**). SLIM's kilohertz imaging seizes the transient motion at a millisecond time scale (**Figure 40b**), effectively eliminating the motion blur and enabling robust cell tracking compared to a lower imaging rate (**Figure 40c**).

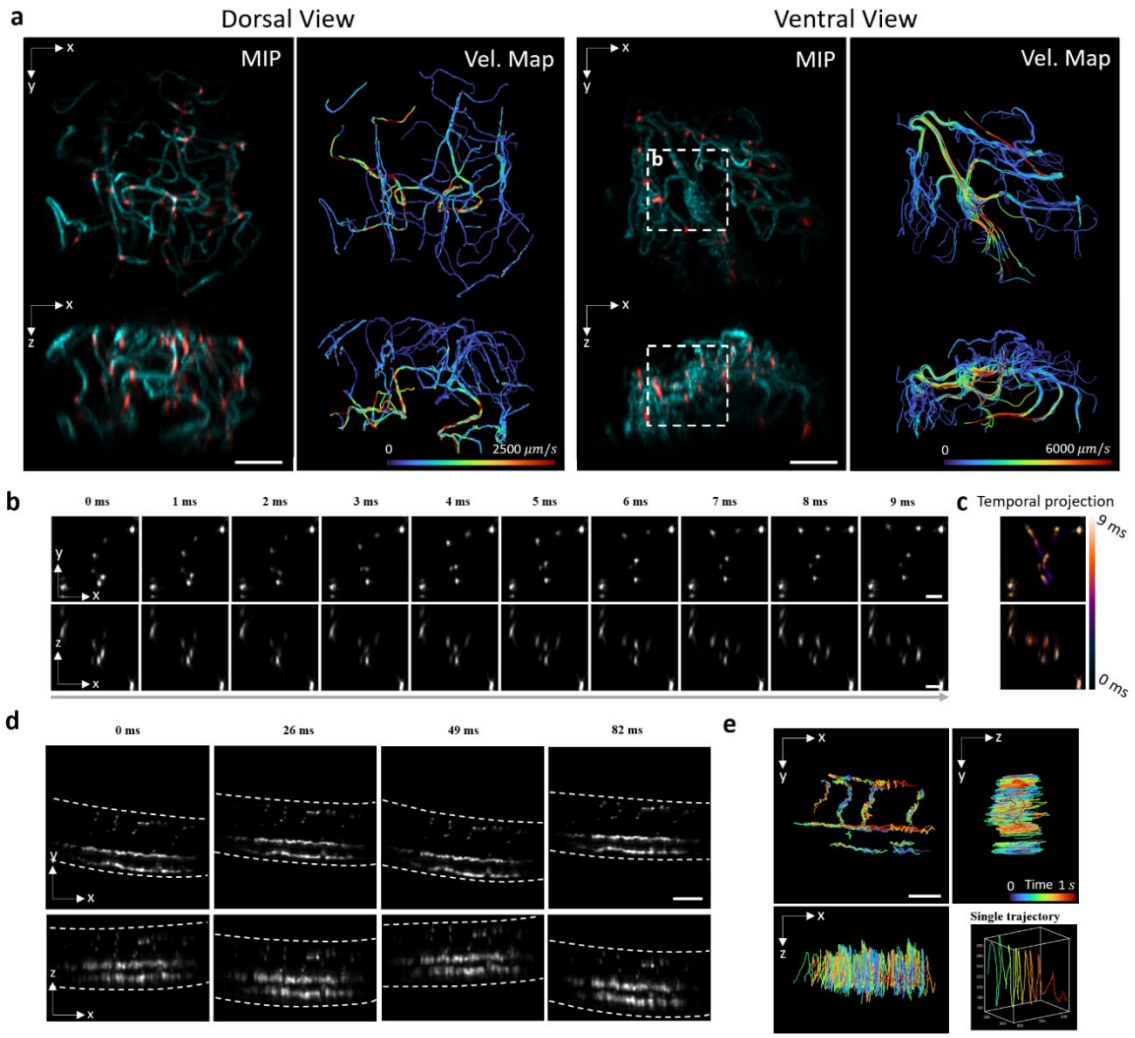

Figure 40. 3D imaging of hemodynamics in the embryonic zebrafish brain and tail at 1,000 vps. (**a**) MIPs of flowing blood cells at representative time points (red) and vascular network obtained by combining frames over time (cyan). The velocity maps show overlaid RBC trajectories color-coded by their instantaneous velocity. Two views, the dorsal view, and ventral view, are two datasets taken with differently-orientated embryos. Scale bar, 100 μm. (**b**) Zoom-in time-lapse of the region labeled by the white dotted box in (**a**) Scale bar, 30 μm. (**c**) Temporal projection color-coded by the time visualizes the motion blur with a lower imaging speed. (**d**) MIPs of a free-moving fish tail. (**e**) RBC trajectories from (**d**) are color-coded by time. The coordinate system has been rotated so that the x-y plane shows the RBC movement perpendicular to the tail swing direction. The single trajectory on the bottom right exhibits the compound motion of a single RBC during fish swimming. Scale bar, 100 μm.

We further demonstrated the speed advantage by imaging the free-moving tail of a zebrafish without sedation. The embryo was mounted on a cover glass with its head restrained using agarose while allowing the tail to move freely in the water. SLIM captured the high-frequency tail swings without any motion blur (**Figure 40d**), maintaining its capability to track individual RBCs and revealing the compound movement that combines oscillation vertical to the tail plane and normal progression along the vessels (**Figure 40e**). By combining with closed-loop tracking and a translational stage[10], SLIM's high-speed volumetric imaging holds promise for studying hemodynamics under natural conditions during locomotor behavior.

4.4 Optical recording of membrane action potentials in medicinal leech ganglions

The development of voltage imaging has enabled neuroscientists to examine neural dynamics with a high spatio-temporal resolution. However, it has long been a challenge to capture voltage signals *in vivo* across a large volume due to the extremely fast transients and low signal-to-noise ratio. With its millisecond temporal resolution, SLIM can precisely detect spike timings across a large 3D neural network, opening avenues for mapping the intricate interaction of neuronal components and elucidating the mechanisms underlying sensory processing and behavioral generation.

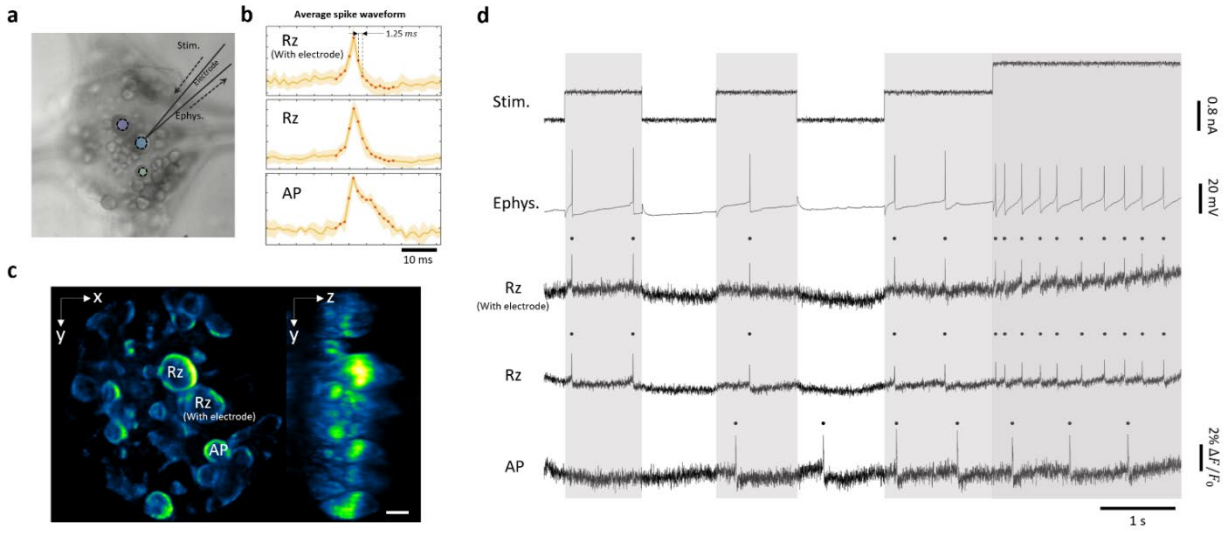

Figure 41. 3D imaging of membrane action potentials in medicinal leech ganglions at 800 vps. (**a**) Brightfield snapshot of leech ganglions. Microelectrode is denoted by a dotted line, which allows for simultaneous stimulation and electrophysiological recording. (**b**) Average spike waveform. The yellow area marks the standard deviation of signals. Orange dots represent temporally sampling points, with an interval of 1.25 ms. (**c**) MIP of SLIM reconstruction for voltage dye fluorescent signals. Scale bar, 50 μm. (**d**) Recording of stimulation current, electrophysiological readout, and optical measurements of ganglion cells: the impaled cell (top), its contralateral partner that is electrically coupled to it (middle), and an unconnected cell (bottom). Grey boxes represent the time window when stimulation is injected. Deeper color stands for larger stimulation. Spikes are detected and marked as black dots above the traces.

As a demonstration, we loaded a voltage-sensitive dye (FluoVolt, F10488, Thermo Fisher Scientific) into a dissected ganglion from a medicinal leech (*Hirudo verbana*)[68], obtained from leech.com and housed in artificial pond water maintained at 15°C. Detailed dissection procedures have been described before.[68,69] Briefly, an adult leech was anesthetized in ice-cold leech saline and an individual segmental ganglion (M10) was dissected out. The ganglion was pinned down ventral side up on a rectangular-shaped flat substrate made of Polydimethylsiloxane (PDMS) (Sylgard 184, Dow Corning). After removing the sheath that covers the ganglion, a voltage-sensitive dye[70] (FluoVolt, ThermoFisher) was bath-loaded using a peristaltic pump.

Fluorescent signals were recorded using SLIM with 29 sub-aperture images at 800 Hz, under illumination from an ultra-low-noise LED. Simultaneously, we introduced a microelectrode for electrophysiological stimulation and recording (**Figure 41a**). Glass microelectrodes (20–50 MΩ) filled with a recording solution (3 M potassium acetate and 60 mM potassium chloride) were used to penetrate the membrane of the target cell. Small negative holding currents were injected to maintain stability. Electrophysiological recordings were performed using Neuroprobe amplifiers (Model 1600; A-M Systems). Membrane voltage and electrode current were digitized at 10 kHz, synchronized with the camera trigger signal using a 16-bit data acquisition board (NI USB-6002; National Instruments). The camera triggers served as time stamps to align recorded frames with electrophysiological data.

After image reconstruction, we corrected for sample movement by running a 2D registration and demotion (**Figure 42**) between adjacent frames using a modified version of SWiFT-IR[71]. The 3D ROIs were then manually defined for each neuron. And the optical readout F_t was calculated by averaging the pixel intensities in the ROI and normalized by the temporal baseline: $F = (F_t - F_0)/F_0$, where F_0 is the temporal mean value. To detect spikes from the optical signal, we detrended the trace F by subtracting its median-filtered version (window size, 50 ms). It was then binarized by a Schmitt trigger, and a peak detection was performed to locate the voltage spikes.

Timing and waveforms (**Figure 41b**) of neuronal action potentials are adequately sampled from the reconstructed 3D image sequence (**Figure 41c**). The resultant time-lapse fluorescence intensities at selected neurons are shown in **Figure 41d**. SLIM measurements agreed with the electrophysiological record in quantitative detail, including the reduction of spike amplitude while strong depolarizing currents were applied.

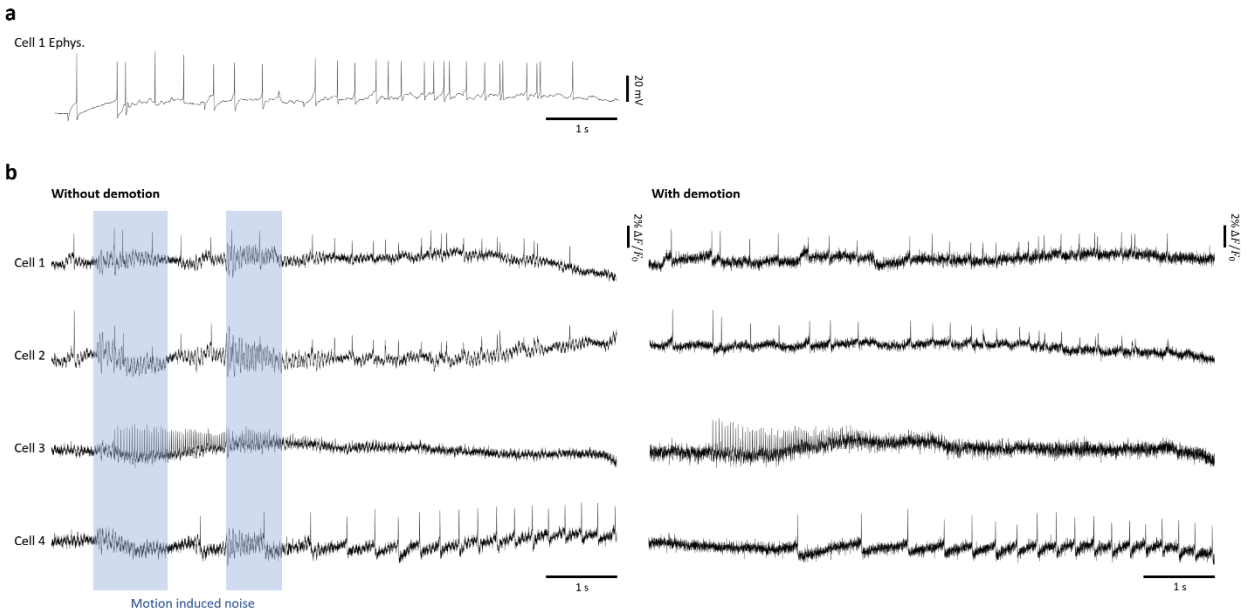

Figure 42. Image demotion for optical recording of membrane action potentials in leech. (**a**) The ground-truth electrophysiological signal provided by electrode on cell 1. (**b**) By assuming the sample being static during recording, motion correction has been applied to the image sequence by image registration between adjacent frames. With motion correction, the noise induced by sample/environment vibration can be suppressed. The blue boxes label the example time window when such noise appears severe and affects the detection of voltage spikes.

4.5 Imaging of free-swimming Vibrio cholerae bacteria

Bacterial chemotaxis is a behavior in which cells sense and respond to chemical gradients in order to swim using their flagellar appendages towards favorable chemoattractants and avoid harmful substances.[72,73] Chemotaxis and flagellar swimming motility are important components of virulence pathways and bacterial pathogenicity.[74] Fully capturing this behavior in 3D, however, is difficult due to the high swimming speeds of the cells and the 3D nature of their trajectories. The current standard for imaging bacterial swimming involves a computational approach by matching diffraction patterns in a 2D image to a reference library in order to reconstruct the 3D positional information of the cell for tracking.[75] A major limitation of this technique is the imaging rate, which is typically around 10-30 Hz. For a swimming bacterium (*Vibrio cholerae* average swim speeds are around 100 μm/s), this imaging rate will result in about 5-10 μm of displacement between

sequential images. Since this displacement is several cell body lengths long, much of the detailed behavior will not be captured.

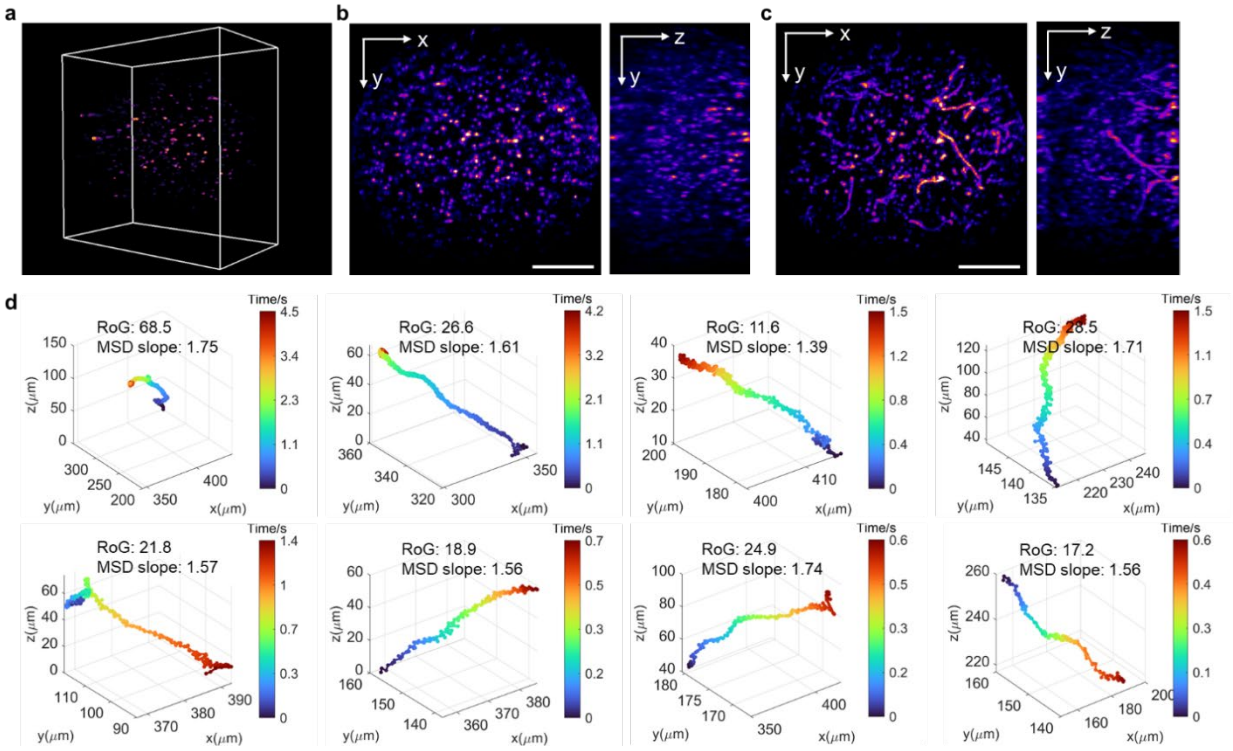

Figure 43. 3D imaging of free-swimming *Vibrio cholerae* bacteria at 200 vps. (**a**) 3D rendering volume of *Vibrio cholerae* bacteria. (**b**) MIP from x-y and y-z slices of *Vibrio cholerae* bacteria. *Vibrio cholerae* are stained with an external membrane stain. The total recording time was 5s. (**c**) MIPs of the swimming bacteria trajectory obtained by combining frames over time. (**d**) Representative trajectories of swimming bacteria with their respective Radius of Gyration (RoG) and Mean Squared Displacement (MSD) slope measurements labeled. Scale bar, 120 μm.

To demonstrate SLIM's capability in addressing this challenge, we imaged the dynamic behaviors of swimming *Vibrio cholerae*. Our setup involved capturing 29 sub-aperture images at a rate of 200 frames per second, facilitated by light-sheet-synthesized volumetric illumination (**Figure 43a**). This allows us to track each individual bacterium within imaging space simultaneously (**Figure 43b**) and obtain information about its swimming trajectory (**Figure 43c**) by MIP throughout all frames. This higher imaging volume rate reduces the spatial displacement in a cell's trajectory between sequential images allowing for observation of finer details in

swimming behaviors, such as reversal events occurring during the "run-reverse-flick" behavior in *Vibrio cholerae* (first picture of **Figure 43d**), where a change in flagellar rotation direction causes a buckling of the hook protein and subsequent reorientation of the swimming direction out of the conventional 2D plane of imaging.[76] The high volumetric imaging speed also allows us to quantify the diffusivity of swimming cells by calculating the Mean Squared Displacement (MSD) slope and Radius of Gyration of their trajectories, as labeled in **Figure 43d**[77], a task that is typically challenging with conventional 3D microscopy techniques.

4.6 Imaging of a beating embryonic zebrafish heart with scanning multi-sheet illumination

Although LFM techniques, including SLIM, offer the ability to numerically refocus to specific depths, they typically lack intrinsic optical sectioning capability. Its application is potentially hindered by the spatial resolution and reconstruction artifacts, and it favors objects with high sparseness.[78] Here, we demonstrated that SLIM can be combined with scanning multi-sheet illumination. The synergy enables high-contrast 3D imaging of densely-labeled fluorescent objects.

We constructed a dual-light-sheet illumination module and scanned the beams using a galvo-mirror driven by a sawtooth function. Rather than synchronizing the camera exposure with the entire scan range as in previous experiments, we operated the camera at a higher rate, allowing each frame to capture a subset of depth layers of the fluorescent object (**Figure 44a**). This approach significantly suppresses out-of-focus light and improves axial resolution in the reconstruction, as shown on fluorescent beads and zebrafish vasculature networks (*Tg(flk:mCherry)*) (**Figure 44b,c**). On the other hand, SLIM offers an ultra-high framerate and supports simultaneous multi-plane detection. These features enable SLIM to maintain a high volume rate even within this scanning scheme.

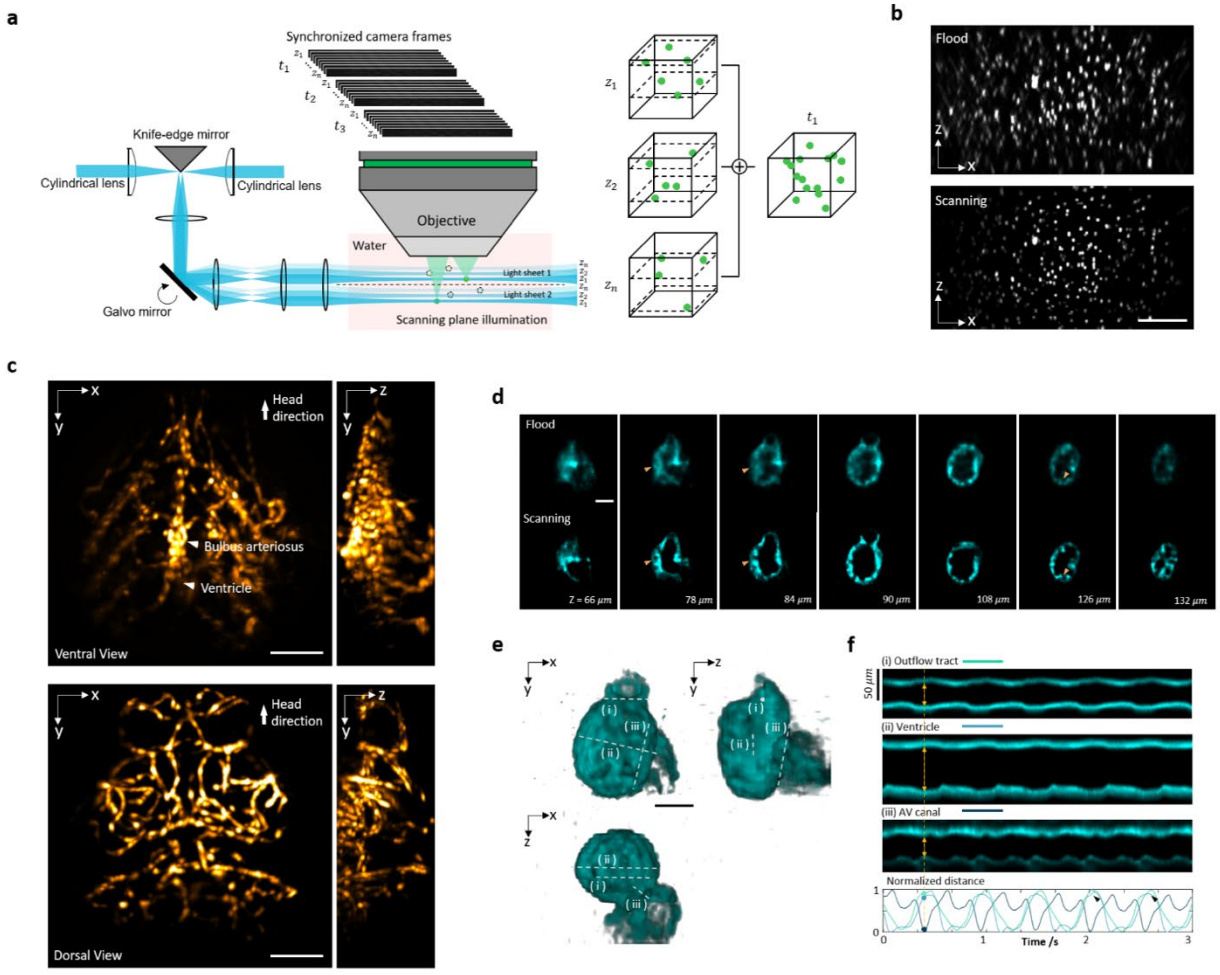

Figure 44. 3D imaging of a beating zebrafish heart with multi-plane scanning light sheet illumination at 300 vps. (**a**) Dual scanning light sheet replaces the flood illumination (*i.e.,* illuminating entire sample volume). In synchrony with the light sheet, the camera captures multiple frames at different scanning positions, each reconstructing two layers of the entire volume. By combining all measurements in one scan cycle, a 3D volume is synthesized for that time point. (**b**) The x-z MIPs of fluorescent beads show the enhancement of image contrast and axial resolution. Scale bar, 100 μm. (**c**) The structural images of the vasculature network in an embryonic zebrafish. Scale bar, 100 μm. (**d**) Comparison of x-y cross-section images between scanning light sheet illumination and flood illumination on cardiomyocytes in the zebrafish heart. The orange arrows mark the muscle structures that are clearly resolved in scanning mode but challenging to flood illumination. Scale bar, 50 μm. (**e**) The 3D rendering of myocardium at a representative time point. Scale bar, 50 μm. (**f**) Kymographs calculated by sampling the time-dependent distance of the cavity along the white dotted line in (**e**) The black arrows indicate the beat-to-beat variance of cardiac contraction. Scale bar, 50 μm.

We demonstrated this acquisition scheme by imaging a beating zebrafish heart (*Tg(cmlc2:GFP)*) at 300 vps (**Figure 44d**). This is achieved by scanning the dual light sheets at

300 Hz, synchronized with camera recording at 4,800 fps (8-bit speed mode). This setup allowed us to reconstruct the heart with 30 planes across 200 μm depth range with the microstructures like ventricular trabeculation clearly delineated (**Figure 44d,e**). The enhanced spatial resolution and contrast offer the potential for accurate segmentation of the heart chamber's geometry, facilitating cardiac studies, such as regional myocardial contractility analysis[32] and computational fluid dynamics (CFD) for hemodynamic forces simulation[79]. While current LFM cardiac imaging is mostly demonstrated on sparse markers like cardiomyocyte nuclei[26,30,80] and blood cells[30,55,81], SLIM with scanning multi-sheet illumination proves effective in resolving the densely labeled muscle tissue. It provides high 3D imaging speed to capture the beating heart in real time and outlines the time-dependent chamber dimension to detect beat-to-beat variations (**Figure 44f**).

4.6 Discussion

We presented SLIM as an innovative snapshot 3D detection method that addresses the pressing need for high-speed volumetric microscopy operating at kilohertz speeds. SLIM accomplishes this by capturing a condensed representation of the original light field using a compact ROI on the sensor. The sampling strategy is grounded in the principle that the inherent spatio-angular correlation in the light field can be exploited to recover signals from compressive measurement.[22,58] Validations across a range of applications, including hemodynamics, neural imaging, cardiovascular imaging, and bacterial dynamics, demonstrated SLIM's versatility and robustness.

SLIM's kilohertz 3D imaging speed, rarely provided by existing methods and often entailing significant design tradeoffs and demanding hardware requirements[57,82], presents new opportunities to investigate millisecond-scale dynamics in emerging fields like voltage imaging. It is universally adaptable to the vast majority of CMOS sensors, which generally allow higher frame rates at

reduced readout pixel rows. While we set kilohertz as a milestone for 3D fluorescence microscopy, SLIM has the potential to achieve tens and even hundreds of thousands of vps with current high-speed cameras[83]. It mitigates the trade-off between speed and other sensor characteristics, allowing us to prioritize cameras with higher sensitivity, lower noise, or other favorable attributes. It also presents great potential in imaging applications where limitations exist on camera pixel count due to fabrication complexity and cost, such as fluorescence lifetime imaging microscopy[84] and short-wave-infrared imaging[85].

SLIM offers a snapshot acquisition that effectively addresses the trade-off between pixel exposure time and volumetric frame rate encountered in conventional scanning-based 3D optical microscopy techniques. This unique approach gives SLIM distinct advantages in terms of photon efficiency and signal-to-noise ratio (SNR), especially beneficial for high-speed imaging of weak fluorescence.

The in-plane rotations and anisotropic scaling in SLIM are co-designed for reconstructing a FOV with a higher resolution than that of each sub-aperture image. This design is specifically tailored to utilize a low-format rectangular sensor, marking a fundamental departure from existing compressive light field photography.[58–60] The latter, stemming from coded aperture imaging, retrieves light fields at the same or lower resolutions than the multiplexed measurement. Moreover, SLIM does not require multiple shots[58] or learning on a sparse basis prior to reconstruction[60]. Our algorithm also bypasses the computational complexity associated with solving the 4D light field and directly reconstructs the 3D image stack for microscopy applications.

SLIM inherits the first-order parameters from FLFM, including magnification, numerical aperture (NA), and FOV. It can transform an existing FLFM[9,10,41,78] into a high-speed 3D imager with a significantly higher frame rate. However, like FLFM, SLIM faces challenges such as

compromised spatial resolution, depth-variant performance, and limited considerations for tissue scattering and lens aberrations. These factors constrain its applicability in complicated intravital environments. We have shown that multi-sheet scanning can extend SLIM's capability to densely-labeled tissue imaging. The refocusing capability within a largely extended depth of field makes SLIM compatible with various 3D illumination structures. Additionally, the literature presents several strategies to enhance SLIM's performance, such as background rejection by hardware[15] and computation[86], multi-focus optics for extended DoF[10,87], and sparsity-based resolution enhancement[78,88]. Furthermore, ongoing advancements in data-driven reconstruction algorithms, particularly physics-embedded deep learning models[11,30,55,80], hold great promise for addressing the ill-posed inverse problems associated with limited space-bandwidth and compressive detection in SLIM. These developments are expected to significantly enhance SLIM's capabilities and broaden its utility across diverse imaging scenarios.

Chapter 5
Light field microscopy with line sensors

5.1 Line sensor and light field imaging

The demand for single-shot high-speed imaging is ever-growing[89–91]. To circumvent the bandwidth limitation of electronic image sensors, strategies like compressed sensing encode high-dimensional spatiotemporal data and transform it into a low-dimensional format for fast readout: one single camera pixel thus carries multiplexed spatial or temporal information[92]. The scene can be numerically reconstructed provided that the signals are sparse in a specific domain. Such data compression has been implemented with encoders like a high-speed digital mirror device (DMD) [93–96], a piezoelectric stage[50], pixel shutters[97], and temporally shifting detectors such as a streak camera[93] and a time delay integration (TDI) camera[98].

Despite significant advances, most compressed imaging cameras can only capture two-dimensional (2-D) dynamics. To break this barrier, we recently developed light field tomography (LIFT)[62], which reformulates light field photography as a sparse-view CT problem and enables three-dimensional (3-D) imaging at an unprecedented frame rate. Rather than recording a 2-D image at each view angle, LIFT acquires *en-face* projections of perspective images and records the

data using a one-dimension (1-D) detector array (**Figure 45**) [22,62]. Such a measurement scheme significantly reduces the data load and allows the recording of light traveling at a sub-ten-picosecond temporal resolution by using streak camera[62]. Even with a regular 1D line sensor we still benefit from its excellent readout speed (hundreds of kilohertz), high sensitivity, high filling factor (~100%) and cost efficiency.

However, like other compressed imaging modalities, LIFT's image resolution is largely dependent on the data compression ratio. Additionally, sharing the same problem with sparse-view CT, the LIFT reconstruction bears artifacts and an anisotropic resolution as a result of a limited number of projection measurements[99–101]. Although these problems can be alleviated by filling the aperture of the main lens with more perspective image channels, this reduces each channel's aperture and, therefore, compromises the channel's diffraction-limited resolution (**Figure 46**).

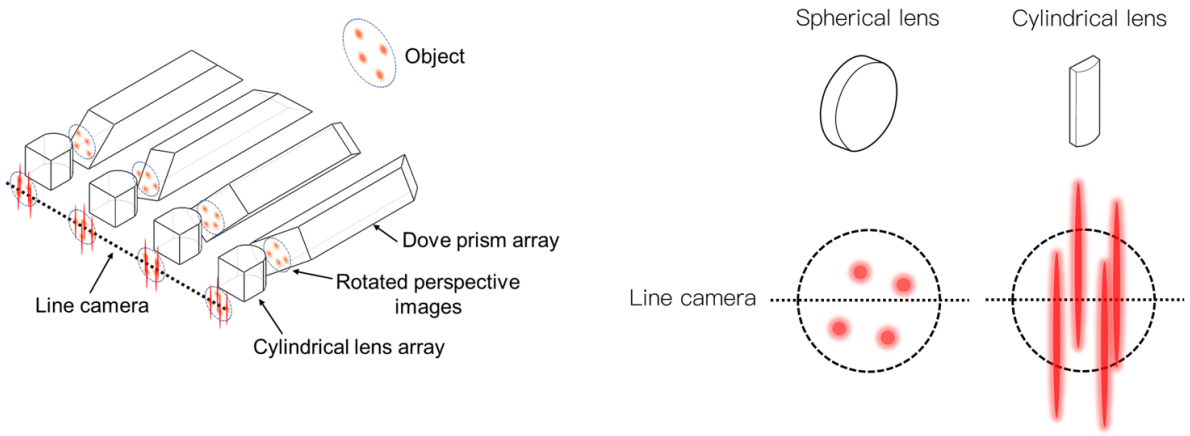

Figure 45. The principle of light field tomography (LIFT). Replacing spherical lens with cylindrical lens creates line projection of the object (which are four dots here). The combination of dove prism array and cylindrical lens array projects the image at different angles and thus allows a line sensor to record the light field information.

To solve this problem, we will present augmented LIFT through parallel spectral encoding in this chapter[63]. Our method is inspired by the fact that most imaging systems record only the spatial

coordinates of an optical field. The other dimensions of light, such as polarization[102], wavelengths[94,103,104], and angles[105], can be used for signal multiplexing. We previously showed that, by using a diffractive element like a grating, we could disperse the 1-D projections in LIFT and enable spectral imaging[53]. Here we utilize this spectral dimension and encode wavelengths with optical rotations for a significantly enriched projection angle set within a snapshot. This maintains the perspective image channel's aperture and provides LIFT with scalability to deal with signals with various sparsity.

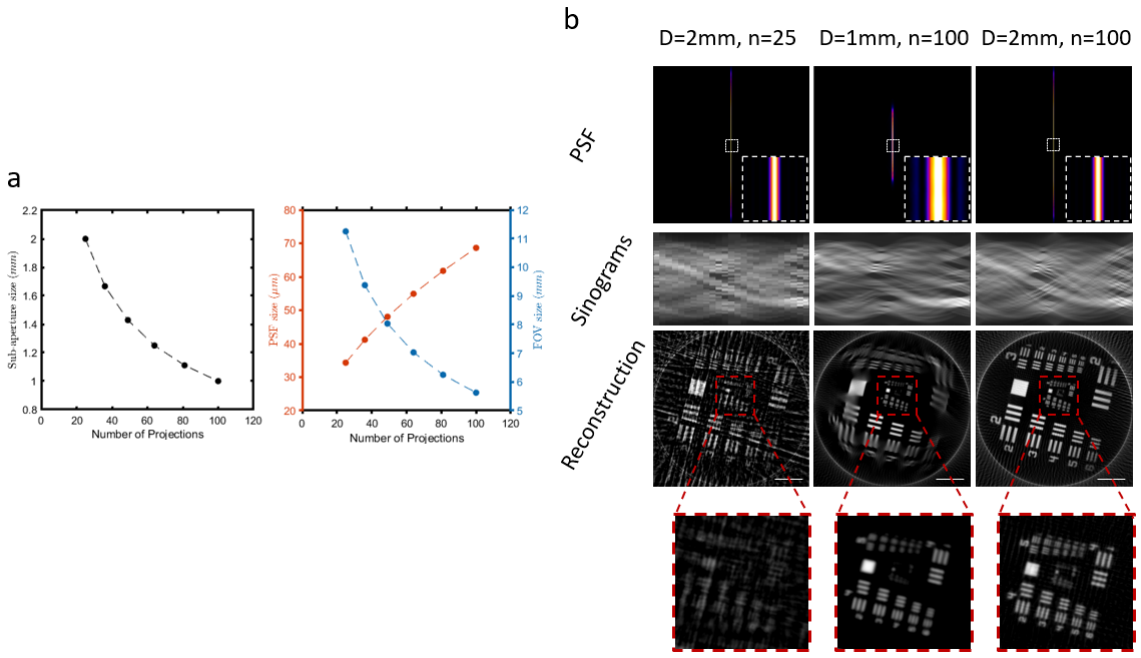

Figure 46. Simulation showing the trade-off between the number of projections and the sub-aperture size. (**a**) Without spectral encoding, more projections require smaller sized lenslets, resulting in a larger PSF and a smaller FOV as well as an increased difficulty in lens mounting and assembly. (**b**) PSFs, sinograms, and reconstructions with 5x5 lenslets (diameter, 2 mm, left column), 10x10 lenslets (diameter, 1mm, middle column) and 5x5 lenslets (diameter, 2mm, spectral encoding by four channels, right column). Multiplexing signals in wavelength maintains the PSF and FOV while increasing the number of projections. All simulations above use system parameters in Method sections (Objective, 4X0.13NA; Relay, 750mm/300mm, cylindrical lenslet, focal length 20mm). The USAF target ground truth has been scaled to occupy a large field of view. We consider ideal Objective and relay system under paraxial approximation. Scale bar, 2 mm.

5.2 Principle of spectral encoding in LIFT

We employ three Dove prisms to rotate the image with slightly different angles (Channel 1: 0; Channel 2: 2.3 degree; Channel 3: 5.3 degree). Each of them performs only in a given spectral bandwidth (Channel 1: 565 – 600nm; Channel 2: 550 – 565nm; Channel 3: 540 – 550nm), and three channels are merged as a single output (**Figure 47a**). The following detection system implements a Dove prism array and a cylindrical lens array to capture the light field in the form of 1-D projections (**Figure 47b**). Each lenslet occupies a sub-pupil area of the main lens, and the disparity enables synthetic refocusing in post-processing. Since the projection makes the information redundant in the non-power axis of the cylindrical lens, it reduces the necessary pixels readout to one row per sub-pupil image and thus boosts the imaging speed. In the end, to separate channels from the merged image, a diffractive grating disperses the 1-D projection and thus gives simultaneous acquisition of multiple channels of light field tomography.

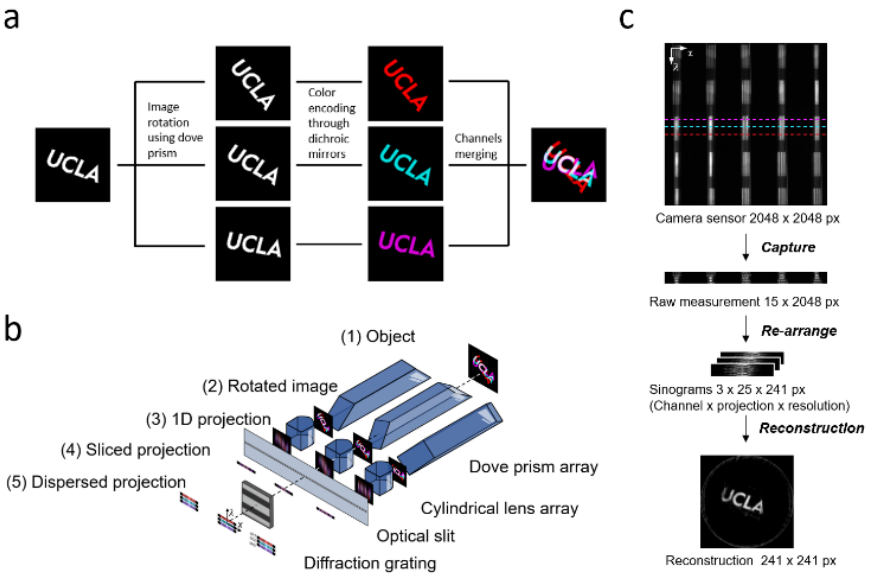

Figure 47. Schematic of the optical system and image formation. (**a**) Dove prisms with different angles followed by dichroic mirrors encode rotation into three color channels. Images are merged at the intermediate image plane. The pseudo color in the schematic does not indicate the actual wavelength used in the system. (**b**) Dove prism and cylindrical lens array are used to capture light field of the object. Each lenslet is located at a different position of the Objective's aperture and forms a 1-D projection of the original 2-D light field

sub-image. A transmissive grating disperses the 1-D projection so that we can separate three channels previously encoded with different image rotations. The slit is perpendicular to the non-power axis of cylindrical lens and the dispersing direction of the grating. Only three lenslets in the array are shown for simplicity. (**c**) The pipeline of our imaging method. We read out only a few rows of pixels and rearrange them to get the sinogram within each channel. The color dotted lines show example sampling positions for three channels. By combining the computational tomography and light field imaging, we can reconstruct the object in 3D with much fewer pixel readout and a higher imaging speed.

The schematic of the system is illustrated in **Figure 48**. We use a microscopic objective (4X0.13NA, RMS4X-PF, Olympus). In the infinity space of the objective, beamsplitters (BS1: 30(R):70(T), BS019; BS2: 50(R):50(T), BS013, Thorlabs) split three channels, and dichroic mirrors (DM1: AT565dc; DM2: T550lpxr, Chroma) merge them after Dove prisms (PS995M, Thorlabs).

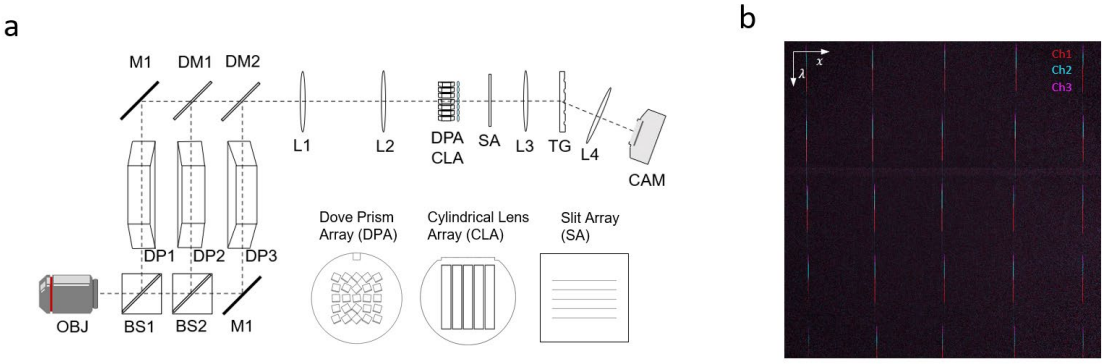

Figure 48. (**a**) The schematic of the entire system. OBJ, Objective; BS, beamsplitter; M, mirror; DP, Dove prism; DM, dichroic mirror; L, achromatic doublet; DPA, Dove prism array; CLA, cylindrical lens array; SA, slit array; TG, transmissive grating; CAM, sCMOS camera sensor. (**b**) The image at the camera sensor plane of a pinhole (50 μm). Three channels are separated in the vertical direction and encoded in pseudo colors for visualization. By sampling in the horizontal direction, we acquire 1-D projections of the pinhole.

The pupil plane of the objective is then relayed by a 4f system (ACT508-300-A, ACT508-750-A, Thorlabs) to an array of custom Dove prisms and cylindrical lenses (**Figure 49**). Each Dove prism in the array has a clear aperture of 2 mm, while each cylindrical lens (focal length = 20 mm) covers five Dove prisms with an extended length of 17 mm. The 1D projections are focused onto

a slit array (slit width = 50 μm), followed by being dispersed by a grating (300 groves/mm, GT50-03, Thorlabs) in the Fourier plane of another relay system (f = 100mm). The raw image is captured by an sCMOS sensor (Prime BSI or Kinetix, Teledyne). We selected multiple ROIs for exposure and readout, which allows us to convert a 2-D sensor into an array of fast 1-D detectors. This configuration enables a higher speed because of a lighter data load. Compared to a streak camera in the original LIFT implementation, an sCMOS camera is less costly and easier to maintain. More importantly, it allows configurable 2-D sensor areas where the encoded signals can be demultiplexed. For illumination source, we used a Halogen lamp (HL250-AY, AmScope) with diffuser (DG10-120, Thorlabs) for static scenes and a LED (UHP-F-5-560, Prizmatix) with collimator (LLG5-CM1, Prizmatix) for dynamic scenes. Overall, the systematic magnification is 0.178x with a FOV around 4.2 $mm \times 4.2\ mm$.

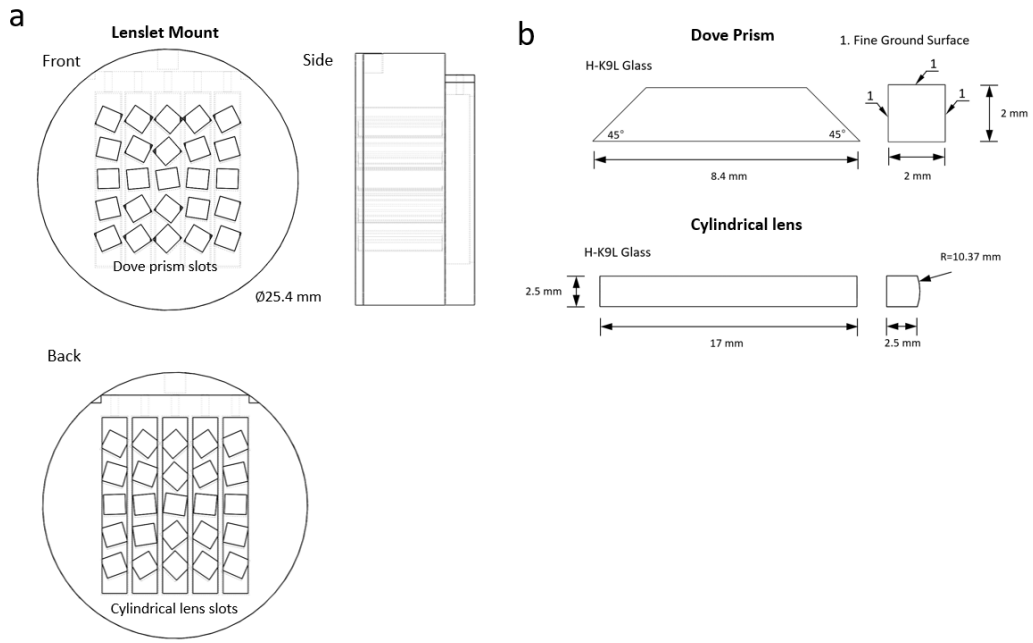

Figure 49. **Lens and mount designs.** (**a**) 3D printed mount for the lenslet array. The angles of dove prism slots are uniformly distributed between -45 to 45 degrees. (**b**) The parameters for custom Dove prisms and cylindrical lens lenslets.

Figure 47c shows the pipeline of our imaging process. The vertical direction spans the spectrum of the projection measurement associated with the perspective image. By sampling corresponding wavelength for each channel, we read out 15 rows of pixels (five rows per channel) to acquire three sinograms (25 projections for each) from the same snapshot. Each channel differs in the global rotation applied during the spectrum encoding stage, thus complementing each other to mitigate the sparse-view problem in computed tomographic (CT).

For the channel i, we model the formation of the vectorized sinogram b_i as

$$b_i = \boldsymbol{TRF_i}g = \boldsymbol{A_i}g \tag{10}$$

where g is the vectorized 2-D image, $\boldsymbol{F_i}$ is the geometrical transformation between channels during spectral encoding, $\boldsymbol{R}$ is the rotation operator applied by the Dove prism array, and T denotes the signal integration by the cylindrical lens.

$\boldsymbol{R}$ is the collection of n (n=25) rotation operators, which can be expanded as

$$\boldsymbol{R} = \begin{bmatrix} R_{\theta 1} \\ R_{\theta 2} \\ \vdots \\ R_{\theta n} \end{bmatrix}$$

$\boldsymbol{A_i}$ will be the forward operator of the channel i, and the image reconstruction of a 2D slice can be achieved by iteratively solving the following optimization problem:

$$\underset{\hat{g}}{argmin} \sum_i \|b_i - \boldsymbol{A_i}g\|_2^2 + \rho\varphi(g) \tag{11}$$

where $\varphi(g)$ is a transform function sparsifying the image. We chose total variation while other functions like $l1$ and wavelet transform can also be applicable. ρ is a hyperparameter that weighs the regularization term. Equation (11) is solved using the FISTA algorithm[106].

In conventional light field imaging, the synthetic refocusing is performed by shifting and adding the sub-pupil images[107]. To extend our method to a 3D scene, we shift each sub-pupil image with respect to its pupil location before passing it to the iterative reconstruction [53,62]. If we define the spatial coordinate of each dove prism in the array as (u, v), we translate the sub-pupil image in the direction and distance of vector $s * (u, v)$, where s is a parameter depending on the axial location of synthetic focal plane. Since we acquire only a 1-D projection of a sub-pupil image, we need to find the amount of translation perpendicular to the projection axis. Assuming that one Dove prism lenslet rotates the image by θ counterclockwise and the non-power direction of cylindrical lens aligns with the sub-pupil coordinate axis v, we shift each 1-D sub-pupil projection by[53]

$$s * u * \cos(\theta) + s * v * \sin(\theta) \tag{12}$$

During the system calibration, we measure the rotation angles of each channel by aligning images of the same object acquired at the intermediate image plane within different spectra. We place a pinhole at the nominal focal plane of the objective lens. From the full-frame sensor image, we can locate and store the center positions of each projection. Then we translate the pinhole axially with a known distance and find the shifting parameter s that gives the sharpest refocused image. By this means, we map s with the depth position in experiments. We further capture a standard USAF resolution target image at each depth and compute the geometrical transformation among channels using image registration. This step is crucial because the transformation is depth-dependent. After the system calibration, we only readout limited pixels at projection positions for a higher framerate. The pixel rearrangement and reconstruction will be using the same calibration data for all following experimental measurements.

5.3 Results of augmented LIFT

Sparse-view CT is prone to noise and structural artifacts. When testing our system on the USAF resolution target, a single channel measurement can frequently fail by generating deteriorated line pairs and stripe-like structures (**Figure 50a**). With an insufficient number of projection angles, the resolving capability is also subject to the object's orientation and its relative location in the FOV. Since each channel observes the sample with slightly different rotations, reconstruction varies given the same group of line pairs (**Figure 50**). This underlies the misinterpretation of the sample structures. Combining all three channels, the augmented measurement provides twice more projections than a single channel, leading to higher contrast and lower noise. The resolving capability also gains robustness in dealing with samples of different orientations (**Figure 50**).

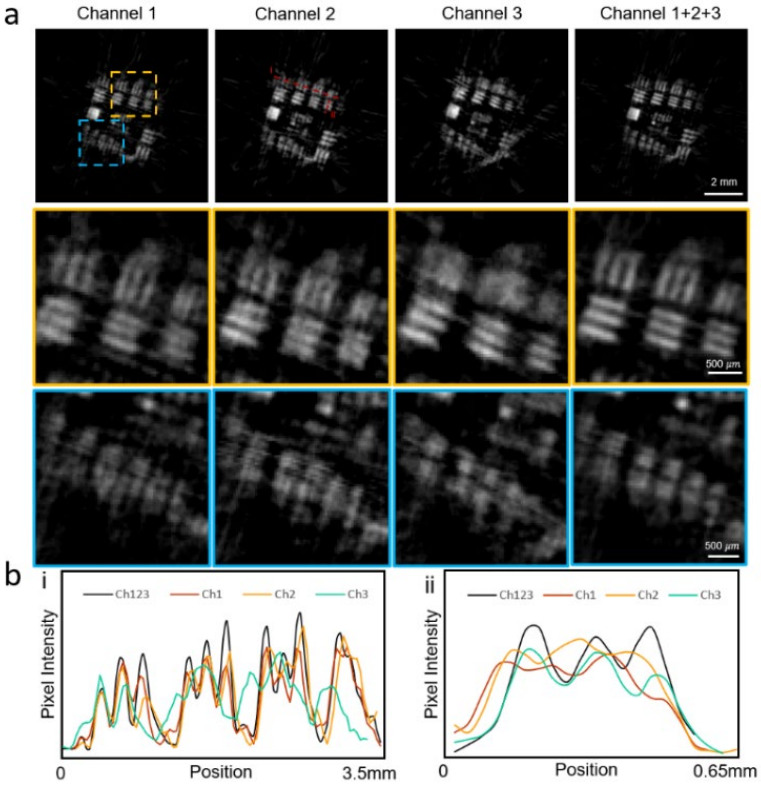

Figure 50. Enhancement of image quality by combining three channels measurement. (**a**) Comparison of reconstruction results of multi-channel vs. single channel when imaging a USAF resolution target. By combining all three channels information, the image gains higher image quality, a higher contrast, reduced

artifacts, and stable resolving capability on differently oriented objects. Yellow and blue box mark the area enlarged for a close-up comparison. (**b**) The pixel intensity profile of sampling line **i** and **ii**, which are labeled in a. Our proposed method delivers a more reliable spatial resolution, while single channel suffers from fewer number of projections, and it is dependent on the object orientation. For multi-channel results, contrast levels $(I_{max} - I_{min})/(I_{max} + I_{min})$ are calculated for line pair groups 2-2 to 2-5 as 0.60, 0.43, 0.41 and 0.28. The resolvable line pair 2-5 indicates a lateral resolution of 157 μm.

To further validate the enhancement, we tested our method on synthetic images displayed on an LCD panel and used structure similarity index measure (SSIM) as a metric to quantify the reconstruction quality. We found that the images with complex structures like retinal vessels (**Figure 51**) benefit the most from the multichannel reconstruction. This is because objects with fine details and textures are less compressible, and thus it requires more projection measurements (*i.e.*, a lower compression ratio) for CT reconstruction.

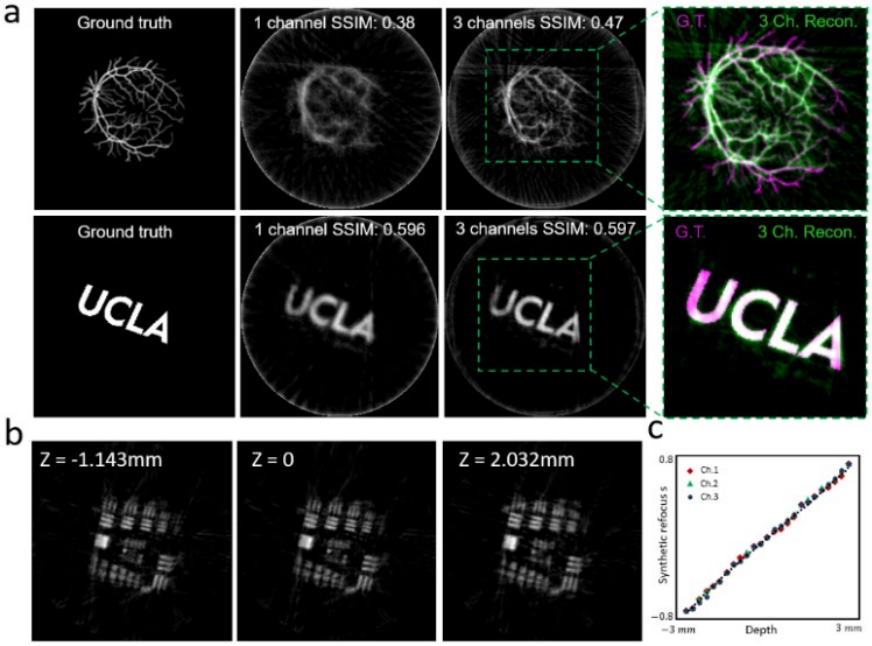

Figure 51. Validation of reconstruction fidelity and synthetic refocus. (**a**) We display various samples on an LCD display panel in front of our system and quantify the reconstruction fidelity and quality using SSIM (structural similarity index measure). The color-coded merged image shows the similarity of our reconstruction to the ground truth. We observe a more significant advantage of our multi-channel over single-channel measurement when imaging samples with more complex structure. (**b**) The USAF resolution target is placed at different defocus positions to show the synthetic refocusing ability of our imaging method. **c.** We adopt the classical 'shifting and add' algorithm for image refocusing. By shifting each sub-image of the light field with regard to their position at the aperture, we can bring sharp focus back to a defocused object. The relation between the shifting and object depth is plotted. Red dots denote each measurement, and the dotted line is the fitting.

LIFT captures the light field of the scene, which permits 3-D reconstruction. To demonstrate the improvement in image quality through spectral encoding also applies to refocused depths, we translated the resolution target axially with various defocus. By shifting each projection accordingly in post-processing, the augmented LIFT can focus on different depths with the same enhancement from multichannel reconstruction (**Figure 51b, c**).

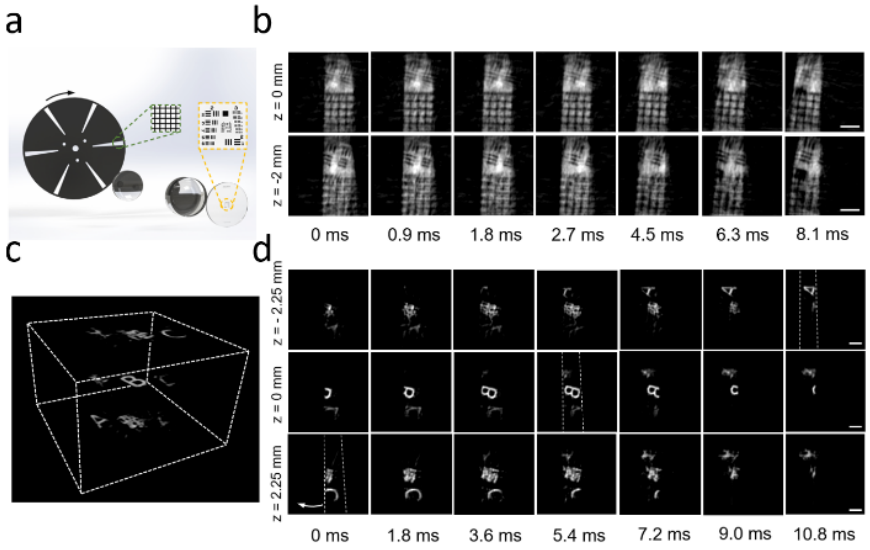

Figure 52. High-speed 3-D imaging experiments. (**a**) We designed an experiment where a rotating optical chopper is relayed to the sample to generate a mask in high-speed motion. The chopper and the sample are displaced at different depths. To represent the chopper wheel plane, we attached a small grid. The data was acquired at 1111 Hz. (**b**) Representative time-lapse frames of the grid (upper row) and the USAF resolution target (lower row). Through synthetic refocus, we can reconstruct sharp images of dynamics at different depths in post-processing. Scale bar, 1 mm. (**c**) Displaced alphabet characters are used to replace USAF resolution target in (a). (**d**) Representative time-lapse frames of each character masked by the rotating optical chopper. White dotted line delineates the shape of chopper, and the arrow is the rotating direction. Scale bar, 1 mm.

The parallel spectral encoding preserves the LIFT's snapshot 3-D imaging ability, which allows us to capture fast dynamics at different depths simultaneously. The measurement of 5 x 5 sub-aperture images, which would occupy the entire sensor frame in a conventional light field camera, is now compressed in a format of 15 x 3200 pixels. By reducing the full frame to a few rows of

pixels, we can record the light field at a kilohertz frame rate by reading out only selected ROIs in a scientific CMOS (3200 x 3200 px, 16-bit mode, full frame rate, 83 Hz). We demonstrated the speed by imaging a rotating optical chopper wheel (MC1F6P10, Thorlabs), which was relayed to the sample to generate a mask in motion (**Figure 52a**). The slit moved at an approximate speed of 200 mm/s and crossed the entire FOV within 20 ms. And our system could seize the transient movement without motion blur. We mounted a grid pattern on the chopper wheel and positioned a USAF resolution target at a defocused distance. We reconstructed a sharp image sequence for both objects via digitally refocusing in post-processing (**Figure 52b**). We performed a similar experiment using axially displaced alphabet characters (**Figure 52c, d**). While each frame benefits from the augmented sinogram via spectral encoding, it's flexible to adjust the frame rate by configuring the camera ROIs for the actual number of channels read out. For example, when dealing with simple objects where a high compression ratio is permissible, we can switch to a single channel for higher imaging speed.

5.4 Discussion

Conventional light field acquisition is redundant for 3-D imaging because the sub-aperture images duplicate each other except for a disparity cue[62]. LIFT solves this problem by reducing the data dimension and thus provides an efficient way for high-speed volumetric imaging. In this work, we further augmented LIFT with spectral encoding to mitigate the common drawbacks of sparse-view CT reconstruction. Kilohertz 3-D microscopic imaging was demonstrated over a large volume (~4.2 mm x 4.2 mm x 4.5 mm) with reduced reconstruction errors and aliasing artifacts.

In the current system, the sCMOS sensor is the bottleneck for the imaging speed. Because the ROIs are sequentially exposed and there exists an extra overhead time for each readout, the achievable framerate (1111 Hz) is far below the theoretical limit (>17 kHz). This problem can be

alleviated by optimizing the camera's firmware. The current method is not directly compatible with the 1-D sensors like the streak camera, because spectral multiplexing requires an extra sensor area for dispersion. Despite the technical challenge, we could seek to stack synchronized 1-D sensor arrays as a solution. Additionally, because the encoding system introduces aberrations (mainly astigmatism), we can't directly concatenate sinograms from different channels. Instead, our algorithm approaches it by iteratively merging 2-D reconstructions in the image space, which relies on the prior knowledge of the geometrical transformation between the channels. Lastly, an underlying assumption of our method is the sample has a similar appearance across wavelengths, a condition that holds if the image contrast is dominated by a single chromophore. Although beyond the scope of current work, fluorescence imaging can also be made possible by tailoring the channel wavelength to the fluorophore's emission. We expect our augmented acquisition method will expand the application realm of LIFT by enabling a higher image quality while maintaining its snapshot advantage.

Chapter 6
Conclusion

The speed advantage of LFM arises from the significantly reduced digital readout. A 3D image stack is estimated from single 2D or even 1D measurement of far fewer pixels. Despite being inherently ill-posed, this approach performs remarkable well in capturing critical dynamic signals across various animal models. These successful demonstrations suggest that many transient processes in organism exhibit high spatiotemporal sparsity which can be effectively exploited by compressive reconstruction algorithms[22]. When designing microscopy for high-speed functional imaging, the key may not be to capture more pixels in short time but to employ a more efficient sampling strategy, as exemplified by LFM.

The recording duration and imaging speed often conflict in high-speed imaging due to the limited bandwidth of cameras and storage devices, which constrains the total number of frames that can be saved within a given timeframe. For similar reasons, high-speed cameras used in slow-motion cinematography are equipped with massive on-board storage yet still support only a few seconds of maximum recording time (e.g., Phantom TMX 7510 with 512 GB RAM can record for 4.4 seconds at 76,000 fps). However, extended recording durations are crucial in fields such as

brain function study, where mapping neural activities to animal behaviors requires longer observation periods. LFM offers a cost-efficient solution for long-term, high-speed recording.

This thesis introduces several LFM systems with volumetric imaging speeds ranging from 100 Hz to 3,333 Hz. Although each system is optimized for different types of processes, they collectively represent a trade-off between imaging speed (data compression rate) and spatial resolution. This inherent limitation is common to all LFM systems and poses a challenge to their broader adoption within the biological research community. How can we build a microscope with both high temporal and spatial resolution?

Data-driven algorithms[55,80] like VCD-LFM can recover high-frequency features from aliased measurements by training on large, relevant datasets. These algorithms work in tandem with compressive imaging, shifting the information burden from optical hardware to post-acquisition software. However, challenges such as the "black box" nature of these models and the risk of hallucinations still need to be addressed before they can be widely adopted by biologists without deep learning expertise.

From a hardware perspective, advancements in camera technology are expected to yield sensors with more sensitive, finer, and lower-noise pixels, alongside larger array formats and higher frame rates. However, while these improvements enhance the sampling of the spatioangular components of the light field, optical diffraction inevitably becomes a limiting factor. For example, the lenslet array in Fourier LFM divides the pupil of the microscope objective into multiple sub-aperture images, resulting in each image having a lower numerical aperture (NA) than the objective and, consequently, reduced resolving power. Additionally, each lenslet must balance depth of field and spatial resolution, with higher spatial resolution typically compromising depth range. Overcoming these optical constraints requires moving beyond traditional optical components and

design methodologies, exploring alternatives like diffusers[23,61,108], metasurfaces[109], and end-to-end optics optimizations[110,111].

Bibliography

1. Adelson, E. H. & Bergen, J. R. The Plenoptic Function and the Elements of Early Vision. in *Computational Models of Visual Processing* (The MIT Press, 1991).
2. Ng, R. *et al. Light Field Photography with a Hand-Held Plenoptic Camera*. Stanford University Computer Science Tech Report https://hal.science/hal-02551481 (2005).
3. Levoy, M., Ng, R., Adams, A., Footer, M. & Horowitz, M. Light field microscopy. *ACM Trans. Graph.* **25**, 924–934 (2006).
4. Ng, R. Fourier slice photography. *ACM Trans. Graph.* **24**, 735–744 (2005).
5. Broxton, M. *et al.* Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express, OE* **21**, 25418–25439 (2013).
6. Prevedel, R. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat Methods* **11**, 727–730 (2014).
7. Llavador, A., Sola-Pikabea, J., Saavedra, G., Javidi, B. & Martínez-Corral, M. Resolution improvements in integral microscopy with Fourier plane recording. *Opt. Express, OE* **24**, 20792–20798 (2016).
8. Scrofani, G. *et al.* FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples. *Biomed. Opt. Express, BOE* **9**, 335–346 (2018).
9. Guo, C. *et al.* Fourier light-field microscopy. *Opt. Express, OE* **27**, 25573–25594 (2019).

10. Cong, L. *et al.* Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio). *eLife* **6**, e28158 (2017).
11. Zhu, L. *et al.* Sustained 3D isotropic imaging of subcellular dynamics using adaptive VCD light-field microscopy 2.0. 2023.03.15.532876 Preprint at https://doi.org/10.1101/2023.03.15.532876 (2023).
12. Yi, C. *et al.* High-fidelity, generalizable light-field reconstruction of biological dynamics with physics-informed meta neural representation. 2023.11.25.568636 Preprint at https://doi.org/10.1101/2023.11.25.568636 (2024).
13. Han, K. *et al.* 3D super-resolution live-cell imaging with radial symmetry and Fourier light-field microscopy. *Biomed. Opt. Express, BOE* **13**, 5574–5584 (2022).
14. Hua, X., Liu, W. & Jia, S. High-resolution Fourier light-field microscopy for volumetric multi-color live-cell imaging. *Optica, OPTICA* **8**, 614–620 (2021).
15. Zhang, Z. *et al.* Imaging volumetric dynamics at high speed in mouse and zebrafish brain with confocal light field microscopy. *Nat Biotechnol* **39**, 74–83 (2021).
16. Lu, Z. *et al.* Long-term intravital subcellular imaging with confocal scanning light-field microscopy. *Nat Biotechnol* 1–12 (2024) doi:10.1038/s41587-024-02249-5.
17. Wu, J. *et al.* Iterative tomography with digital adaptive optics permits hour-long intravital observation of 3D subcellular dynamics at millisecond scale. *Cell* **184**, 3318-3332.e17 (2021).
18. Gu, M. *Advanced Optical Imaging Theory*. (Springer Berlin, Berlin, 2013).
19. Voelz, D. G. *Computational Fourier Optics: A MATLAB Tutorial*. (SPIE Press, Bellingham, Wash., 2011).
20. Goodman, J. W. *Introduction to Fourier Optics*. (Roberts & Co, Englewood, Colo, 2005).

21. Liu, W., Kim, G.-A. R., Takayama, S. & Jia, S. Fourier light-field imaging of human organoids with a hybrid point-spread function. *Biosensors and Bioelectronics* **208**, 114201 (2022).

22. Feng, X., Ma, Y. & Gao, L. Compact light field photography towards versatile three-dimensional vision. *Nat Commun* **13**, 3333 (2022).

23. Antipa, N. *et al.* DiffuserCam: lensless single-exposure 3D imaging. *Optica, Vol. 5, Issue 1, pp. 1-9* (2018) doi:10.1364/OPTICA.5.000001.

24. Xue, Y., Davison, I. G., Boas, D. A. & Tian, L. Single-shot 3D wide-field fluorescence imaging with a Computational Miniature Mesoscope. *Science Advances* **6**, eabb7508 (2020).

25. Yanny, K. *et al.* Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy. *Light Sci Appl* **9**, 171 (2020).

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

27. Lu, Z. *et al.* Phase-space deconvolution for light field microscopy. *Opt. Express, OE* **27**, 18131–18145 (2019).

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

29. Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555**, 487–492 (2018).

30. Wang, Z. *et al.* Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nat Methods* **18**, 551–556 (2021).

31. Truong, T. V. *et al.* High-contrast, synchronous volumetric imaging with selective volume illumination microscopy. *Commun Biol* **3**, 1–8 (2020).

32. Wang, Z. *et al.* A hybrid of light-field and light-sheet imaging to study myocardial function and intracardiac blood flow during zebrafish development. *PLOS Computational Biology* **17**, e1009175 (2021).

33. Lee, J. *et al.* 4-Dimensional light-sheet microscopy to elucidate shear stress modulation of cardiac trabeculation. *J Clin Invest* **126**, 1679–1690 (2016).

34. Mickoleit, M. *et al.* High-resolution reconstruction of the beating zebrafish heart. *Nat Methods* **11**, 919–922 (2014).

35. Weber, M. *et al.* Cell-accurate optical mapping across the entire developing heart. *eLife* **6**, e28307 (2017).

36. Liebling, M., Forouhar, A. S., Gharib, M., Fraser, S. E. & Dickinson, M. E. Four-dimensional cardiac imaging in living embryos via postacquisition synchronization of nongated slice sequences. *JBO* **10**, 054001 (2005).

37. Chen, J. *et al.* Displacement analysis of myocardial mechanical deformation (DIAMOND) reveals segmental susceptibility to doxorubicin-induced injury and regeneration. *JCI Insight* **4**, (2019).

38. Hsu, J. J. *et al.* Contractile and hemodynamic forces coordinate *Notch1b*-mediated outflow tract valve formation. *JCI Insight* **4**, (2021).

39. Vermot, J. *et al.* Reversing Blood Flows Act through klf2a to Ensure Normal Valvulogenesis in the Developing Heart. *PLOS Biology* **7**, e1000246 (2009).

40. Voleti, V. *et al.* Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nat Methods* **16**, 1054–1062 (2019).

41. Yi, C. *et al.* Video-rate 3D imaging of living cells using Fourier view-channel-depth light field microscopy. *Commun Biol* **6**, 1–8 (2023).

42. Meng, G. *et al.* Ultrafast two-photon fluorescence imaging of cerebral blood circulation in the mouse brain in vivo. *Proceedings of the National Academy of Sciences* **119**, e2117346119 (2022).

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

44. Abdelfattah, A. S. *et al.* Sensitivity optimization of a rhodopsin-based fluorescent voltage indicator. *Neuron* **111**, 1547-1563.e9 (2023).

45. Hochbaum, D. R. *et al.* All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat Methods* **11**, 825–833 (2014).

46. Wu, J. *et al.* Kilohertz two-photon fluorescence microscopy imaging of neural activity in vivo. *Nat Methods* **17**, 287–290 (2020).

47. Zhang, T. *et al.* Kilohertz two-photon brain imaging in awake mice. *Nat Methods* **16**, 1119–1122 (2019).

48. Wang, Z. *et al.* Imaging the voltage of neurons distributed across entire brains of larval zebrafish. 2023.12.15.571964 Preprint at https://doi.org/10.1101/2023.12.15.571964 (2023).

49. Pavani, S. R. P. *et al.* Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences* **106**, 2995–2999 (2009).

50. Llull, P. *et al.* Coded aperture compressive temporal imaging. *Opt. Express, OE* **21**, 10526–10545 (2013).

51. Wagadarikar, A. A., Pitsianis, N. P., Sun, X. & Brady, D. J. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Opt. Express, OE* **17**, 6368–6388 (2009).

52. Zhao, R., Cui, Q., Wang, Z. & Gao, L. Coded aperture snapshot hyperspectral light field tomography. *Opt. Express, OE* **31**, 37336–37347 (2023).

53. Cui, Q., Park, J., Ma, Y. & Gao, L. Snapshot hyperspectral light field tomography. *Optica, OPTICA* **8**, 1552–1558 (2021).

54. Skocek, O. *et al.* High-speed volumetric imaging of neuronal activity in freely moving rodents. *Nat Methods* **15**, 429–432 (2018).

55. Lu, Z. *et al.* Virtual-scanning light-field microscopy for robust snapshot high-resolution volumetric imaging. *Nat Methods* **20**, 735–746 (2023).

56. Tian, T., Yuan, Y., Mitra, S., Gyongy, I. & Nolan, M. F. Single Photon Kilohertz Frame Rate Imaging of Neural Activity. *Advanced Science* **9**, 2203018 (2022).

57. Guo, R. *et al.* EventLFM: Event Camera integrated Fourier Light Field Microscopy for Ultrafast 3D imaging. Preprint at https://doi.org/10.48550/arXiv.2310.00730 (2023).

58. Ashok, A. & Neifeld, M. A. Compressive light field imaging. in *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV* vol. 7690 221–232 (SPIE, 2010).

59. Babacan, S. D. *et al.* Compressive Light Field Sensing. *IEEE Transactions on Image Processing* **21**, 4746–4757 (2012).

60. Marwah, K., Wetzstein, G., Bando, Y. & Raskar, R. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph.* **32**, 46:1-46:12 (2013).

61. Antipa, N., Necula, S., Ng, R. & Waller, L. Single-shot diffuser-encoded light field imaging. in *2016 IEEE International Conference on Computational Photography (ICCP)* 1–11 (2016). doi:10.1109/ICCPHOT.2016.7492880.

62. Feng, X. & Gao, L. Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging. *Nat Commun* **12**, 2179 (2021).

63. Wang, Z., Hsiai, T. K. & Gao, L. Augmented light field tomography through parallel spectral encoding. *Optica, OPTICA* **10**, 62–65 (2023).

64. Wang, Z. *et al.* Kilohertz volumetric imaging of in-vivo dynamics using squeezed light field microscopy. 2024.03.23.586416 Preprint at https://doi.org/10.1101/2024.03.23.586416 (2024).

65. Biggs, D. S. C. & Andrews, M. Acceleration of iterative image restoration algorithms. *Appl. Opt., AO* **36**, 1766–1775 (1997).

66. Roustaei, M. *et al.* Computational simulations of the 4D micro-circulatory network in zebrafish tail amputation and regeneration. *Journal of The Royal Society Interface* **19**, 20210898 (2022).

67. Zhou, Y., Zickus, V., Zammit, P., Taylor, J. M. & Harvey, A. R. High-speed extended-volume blood flow measurement using engineered point-spread function. *Biomed. Opt. Express, BOE* **9**, 6444–6454 (2018).

68. Tomina, Y. & Wagenaar, D. A. A double-sided microscope to realize whole-ganglion imaging of membrane potential in the medicinal leech. *eLife* **6**, e29839 (2017).

69. Tomina, Y. & Wagenaar, D. Dual-sided Voltage-sensitive Dye Imaging of Leech Ganglia. *BIO-PROTOCOL* **8**, (2018).

70. Miller, E. W. *et al.* Optically monitoring voltage in neurons by photo-induced electron transfer through molecular wires. *Proceedings of the National Academy of Sciences* **109**, 2114–2119 (2012).

71. Wetzel, A. W. *et al.* Registering large volume serial-section electron microscopy image sets for neural circuit reconstruction using FFT signal whitening. in *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* 1–10 (2016). doi:10.1109/AIPR.2016.8010595.

72. Berg, H. C. Chemotaxis in Bacteria. *Annual Review of Biophysics* **4**, 119–136 (1975).

73. Keegstra, J. M., Carrara, F. & Stocker, R. The ecological roles of bacterial chemotaxis. *Nat Rev Microbiol* **20**, 491–504 (2022).

74. Butler, S. M. & Camilli, A. Going against the grain: chemotaxis and infection in Vibrio cholerae. *Nat Rev Microbiol* **3**, 611–620 (2005).

75. Grognot, M., Mittal, A., Mah'moud, M. & Taute, K. M. Vibrio cholerae Motility in Aquatic and Mucus-Mimicking Environments. *Applied and Environmental Microbiology* **87**, e01293-21 (2021).

76. Son, K., Guasto, J. S. & Stocker, R. Bacteria can exploit a flagellar buckling instability to change direction. *Nature Phys* **9**, 494–498 (2013).

77. Conrad, J. C. *et al.* Flagella and Pili-Mediated Near-Surface Single-Cell Motility Mechanisms in *P. aeruginosa*. *Biophysical Journal* **100**, 1608–1616 (2011).

78. Yoon, Y.-G. *et al.* Sparse decomposition light-field microscopy for high speed imaging of neuronal activity. *Optica, OPTICA* **7**, 1457–1468 (2020).

79. Vedula, V. *et al.* A method to quantify mechanobiologic forces during zebrafish cardiac development using 4-D light sheet imaging and computational modeling. *PLOS Computational Biology* **13**, e1005828 (2017).

80. Wagner, N. *et al.* Deep learning-enhanced light-field imaging with continuous validation. *Nat Methods* **18**, 557–563 (2021).

81. Zhang, Y. *et al.* DiLFM: an artifact-suppressed and noise-robust light-field microscopy through dictionary learning. *Light Sci Appl* **10**, 152 (2021).

82. Sacconi, L. *et al.* KHz-rate volumetric voltage imaging of the whole Zebrafish heart. *Biophysical Reports* **2**, 100046 (2022).

83. HiCAM: High-Speed, High-Sensitivity Imaging Camera - Lambert. *lambertinstruments.com* https://lambertinstruments.com/products/hicam (2020).

84. Datta, R., Heaster, T. M., Sharick, J. T., Gillette, A. A. & Skala, M. C. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. *JBO* **25**, 071203 (2020).

85. Wang, F. *et al.* Light-sheet microscopy in the near-infrared II window. *Nat Methods* **16**, 545–552 (2019).

86. Zhang, Y. *et al.* Computational optical sectioning with an incoherent multiscale scattering model for light-field microscopy. *Nat Commun* **12**, 6391 (2021).

87. Zhang, Y. *et al.* Multi-focus light-field microscopy for high-speed large-volume imaging. *PhotoniX* **3**, 30 (2022).

88. Pégard, N. C. *et al.* Compressive light-field microscopy for 3D neural activity recording. *Optica, OPTICA* **3**, 517–524 (2016).

89. Liang, J. & Wang, L. V. Single-shot ultrafast optical imaging. *Optica, OPTICA* **5**, 1113–1127 (2018).

90. Gao, L. & Wang, L. V. A review of snapshot multidimensional optical imaging: Measuring photon tags in parallel. *Physics Reports* **616**, 1–37 (2016).

91. Mikami, H., Gao, L. & Goda, K. Ultrafast optical imaging technology: principles and applications of emerging methods. *Nanophotonics* **5**, 497–509 (2016).

92. Park, J., Feng, X., Liang, R. & Gao, L. Snapshot multidimensional photography through active optical mapping. *Nat Commun* **11**, 5602 (2020).

93. Gao, L., Liang, J., Li, C. & Wang, L. V. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature* **516**, 74–77 (2014).

94. Wang, P., Liang, J. & Wang, L. V. Single-shot ultrafast imaging attaining 70 trillion frames per second. *Nat Commun* **11**, 2091 (2020).

95. Liang, J., Wang, P., Zhu, L. & Wang, L. V. Single-shot stereo-polarimetric compressed ultrafast photography for light-speed observation of high-dimensional optical transients with picosecond resolution. *Nat Commun* **11**, 5252 (2020).

96. Lu, Y., Wong, T. T. W., Chen, F. & Wang, L. Compressed Ultrafast Spectral-Temporal Photography. *Phys. Rev. Lett.* **122**, 193904 (2019).

97. Mochizuki, F. *et al.* Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor. *Opt. Express, OE* **24**, 4155–4176 (2016).

98. Park, J. & Gao, L. Continuously streaming compressed high-speed photography using time delay integration. *Optica, OPTICA* **8**, 1620–1623 (2021).

99. Davison, M. E. The Ill-Conditioned Nature of the Limited Angle Tomography Problem. *SIAM J. Appl. Math.* **43**, 428–448 (1983).

100. Willemink, M. J. & Noël, P. B. The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. *Eur Radiol* **29**, 2185–2195 (2019).

101. Han, X. *et al.* Algorithm-Enabled Low-Dose Micro-CT Imaging. *IEEE Transactions on Medical Imaging* **30**, 606–620 (2011).

102. Hai, N., Kumar, R. & Rosen, J. Single-shot TIE using polarization multiplexing (STIEP) for quantitative phase imaging. *Optics and Lasers in Engineering* **151**, 106912 (2022).

103. Nakagawa, K. *et al.* Sequentially timed all-optical mapping photography (STAMP). *Nature Photon* **8**, 695–700 (2014).

104. Zang, Z. *et al.* Ultrafast parallel single-pixel LiDAR with all-optical spectro-temporal encoding. *APL Photonics* **7**, 046102 (2022).

105. Li, Z., Zgadzaj, R., Wang, X., Chang, Y.-Y. & Downer, M. C. Single-shot tomographic movies of evolving light-velocity objects. *Nat Commun* **5**, 3085 (2014).

106. Beck, A. & Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009).

107. Ng, R. *et al.* Light field photography with a hand-held plenoptic camera. (Stanford University, 2005).

108. Liu, F. L., Kuo, G., Antipa, N., Yanny, K. & Waller, L. Fourier DiffuserScope: single-shot 3D Fourier light field microscopy with a diffuser. *Opt. Express, OE* **28**, 28969–28986 (2020).

109. Hua, X. *et al.* Ultra-compact snapshot spectral light-field imaging. *Nat Commun* **13**, 2732 (2022).

110. Sitzmann, V. *et al.* End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph.* **37**, 114:1-114:13 (2018).

111. Zhang, B. *et al.* End-to-end snapshot compressed super-resolution imaging with deep optics. *Optica, OPTICA* **9**, 451–454 (2022).