Title

Computational Fluorescence Lifetime Imaging Microscopy for Biomedical Sciences

Author

Ma, Yayao

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Fluorescence Lifetime Imaging Microscopy for Biomedical Sciences

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Bioengineering

by

Yayao Ma

2023

ABSTRACT OF THE DISSERTATION

Computational Fluorescence Lifetime Imaging Microscopy

for Biomedical Sciences

by

Yayao Ma

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2023

Professor Liang Gao, Chair

Fluorescence lifetime imaging microscopy (FLIM) is a powerful imaging technique that enables the visualization of biological samples at the molecular level by measuring the fluorescence decay rate of fluorescent probes. This provides critical information about molecular interactions, environmental changes, and localization within biological systems. However, creating high-resolution lifetime maps using conventional FLIM systems can be challenging, as it often requires extensive scanning that can significantly lengthen acquisition times. This issue is further compounded in three-dimensional (3D) imaging because it demands additional scanning along the depth axis. To tackle this challenge, we developed two novel computational imaging techniques. The first technique is compressed FLIM based on a compressed sensing scheme. By leveraging the compressibility of biological scenes in a specific domain, we simultaneously record the time-lapse fluorescence decay upon pulsed

laser excitation within a large field of view. The resultant system can acquire a widefield fluorescence lifetime image within a single camera exposure, eliminating the motion artifact and minimizing the photobleaching and phototoxicity. The imaging speed, limited only by the readout speed of the camera, is up to 100 Hz. We demonstrated the utility of compressed FLIM in imaging various transient dynamics at the microscopic scale. The second technique is light field tomographic FLIM (LIFT-FLIM). This approach allows for the acquisition of volumetric fluorescence lifetime images in a highly data-efficient manner, significantly reducing the number of scanning steps required compared to conventional point-scanning or line-scanning FLIM imagers. Moreover, LIFT-FLIM enables the measurement of high-dimensional data using low-dimensional detectors, which are typically low-cost and feature a higher temporal bandwidth. We demonstrated LIFT-FLIM using a linear single-photon avalanche diode array on various biological systems, showcasing unparalleled single-photon detection sensitivity. Additionally, we expanded the functionality of our method to spectral FLIM and demonstrated its application in high-content multiplexed imaging of lung organoids. LIFT-FLIM has the potential to open up new avenues in both basic and translational biomedical research.

The dissertation of Yayao Ma is approved.

Jun Chen

Pei-Yu Chiou

Tzung Hsiai

Liang Gao, Committee Chair

University of California, Los Angeles

2023

Table of Contents

List of Figures

Acknowledgements

I would like to first acknowledge and express my deepest appreciation to Professor Liang Gao, my PhD advisor; his guidance, patience, and motivation have been always present throughout this journey. Secondly, I am very grateful to the rest of my dissertation committee: Prof. Jun Chen, Prof. Pei-Yu "Eric" Chiou, and Prof. Tzung Hsiai. Thank you for your time, effort, guidance, and all your constructive comments. Moreover, I would like to also express my deepest appreciation to my parents Huazhang Ma and Chunxia Zhao who supported me during all the ups and downs. I am also thankful and had the pleasure to work with and learn from my lab mates and project collaborators. Finally, I would like to thank TWICE Tzuyu for the inspiration and encouragement.

Chapter 1, Chapter 2.1 and 2.2 are a version of Ma, Yayao. "Compressed fluorescence lifetime imaging microscope." (2019) page 1-13. http://hdl.handle.net/2142/105865.

Chapters 2.3-2.5 are a version of Ma, Yayao, et al. "High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells." Proceedings of the National Academy of Sciences 118.3 (2021): e2004176118 from page 1-4. https://doi.org/10.1073/pnas.2004176118. Professor Liang Gao is the PI. Prof. Liang Gao and Prof. Catherine Best designed research analyzed data and wrote the paper. Youngjae Lee performed the research, analyzed data and wrote the paper.

Chapters 2.6 is a version of Yayao Ma, Xiaohua Feng, and Liang Gao, "Deep-learning-based image reconstruction for compressed ultrafast photography," Opt. Lett. 45, 4400-4403 (2020) page 1-4. Available at https://doi.org/10.1364/OL.397717. Professor Liang Gao is the PI. Prof. Liang Gao and Xiaohua Feng designed research analyzed data and wrote the paper.

Chapters 3 is a version of Feng, X., Ma, Y. & Gao, L. Compact light field photography

towards versatile three-dimensional vision. *Nature Commun*ications, **13**, 3333 (2022) page1-10. https://doi.org/10.1038/s41467-022-31087-9. Professor Liang Gao is the PI. Prof. Liang Gao and Xiaohua Feng designed research, analyzed data and wrote the paper.

Chapters 4 is a version of Light-field tomographic fluorescence lifetime imaging microscopy, submitted. Professor Liang Gao is the PI. Luzhe Huang, Xilin Yang and Aydogan Ozcan developed the deep learning network. Chandani Sen and Brigitte N. Gomperts prepared the lung organoid samples, while Samuel Burri, Claudio Bruschini and Edoardo Charbon developed the linear SPAD array. Robert B. Cameron and Gregory A. Fishbein contributed to the preparation of lung cancer slides and provided ground truth diagnosis. All authors analyzed the data and contributed to writing the manuscript.

The thesis work is supported by the National Institute of Health (NIH/NIGMS) under the grant number of R35GM128761(L.G.).

Biographical Sketch

EDUCATION

Master of Science in Electrical and Computer Engineering 2017-2019

University of Illinois at Urbana-Champaign

Champaign, IL

Bachelor of Science in Optoelectronics 2013-2017

Huazhong University of Science and Technology

Wuhan, China

PUBLICATIONS(* denotes equal contributions)

1. Feng, X.*., Ma, Y.*, & Gao, L. Compact light field photography toward versatile 3D-vision. Nature Communications, 13, 3333 (2022)

2. Ma, Y., Lee, Y., Best-Popescu, C., & Gao, L. (2021). High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells. Proceedings of the National Academy of Sciences, 118(3).

3. Ma, Y., Feng, X., & Gao, L. (2020). Deep-learning-based image reconstruction for compressed ultrafast photography. Optics letters, 45(16), 4400-4403.

4. Ma, Y., Ye, C., Ke, J., Zhang, J., Zhu, J., & Ling, Z. (2016). Array illumination of a Fresnel–Dammann zone plate. Applied optics, 55(26), 7218-7221.

5. Cui, Q., Park, J., Ma, Y., & Gao, L. (2021). Snapshot hyperspectral light field tomography. Optica, 8(12), 1552-1558

6. Xu, S., Ma, Y., Zhang, J., Zhou, S., & Zhu, J. (2019). Multiplanar imaging properties of Fermat-spiral Greek-ladder sieves with different point spread functions. Optics Communications, 434, 191-195.

7. Xu, S., Zhang, J., Zhou, S., Ma, Y., Wang, S., Zhang, Y., ... & Zhu, J. (2018). Multiplanar imaging properties of Theon sieves. Applied optics, 57(9), 1993-1997.

8. Zhang, Q., Wang, Y., Lin, F., Tang, Y., Cheng, P., Zhou, X., Zhu. Z., Ma, Y., ... & Bao, J. (2021). Laser-induced dynamic alignment and nonlinear-like optical transmission in liquid suspensions of 2D atomically thin nanomaterials. Optics Express, 29(22), 36389-36399.

Chapter 1 Introduction

1.1 Fluorescence lifetime

Fluorescence of fluorophores is a light-matter interaction process and contains three events: excitation, vibration relaxation (internal conversion) and fluorescence emission. A Jablonski energy diagram illustrates the process and the electronic transitions between ground states and excited states as is shown in Figure 1.1. Photons are absorbed by the molecules whose energy matches the energy difference between the discrete energy levels. These energy levels define the absorption and the emission band. Excitation refers to the absorption of a photon and can elevate an electron of a molecule from the ground state to an excited state. Excitation is caused by either mechanical or chemical mechanisms in nature and the energy source of the excitation can be a laser or a mercury lamp in fluorescence or chemical reactions in chemiluminescence. The fluorophores then transit to the excited electronic states. Excitation by a photon occurs in femtoseconds. Vibrational relaxation to the lowest vibrational energy level in the excited state lasts for picoseconds. The initial excited state is partially dissipated and transits to relaxed singlet excited states from which fluorescence emission results. Fluorescence emission lingers over a finite time and molecules return to the ground state. In addition to fluorescence emission, the excited states can be depopulated by other processes such as collisional quenching, fluorescence resonance energy transfer (FRET) and intersystem crossing. The ratio of the number of photons emitted to the number of photons absorbed is referred to as the quantum yield of fluorescence which measures the relative extent among these processes. Energy dissipation originates from the initial vibrational relaxations and the subsequent inclined fluorescence transitions, which derive from excited-state reactions, complex formations and FRET [1], to higher vibrational energy levels of the ground

state.

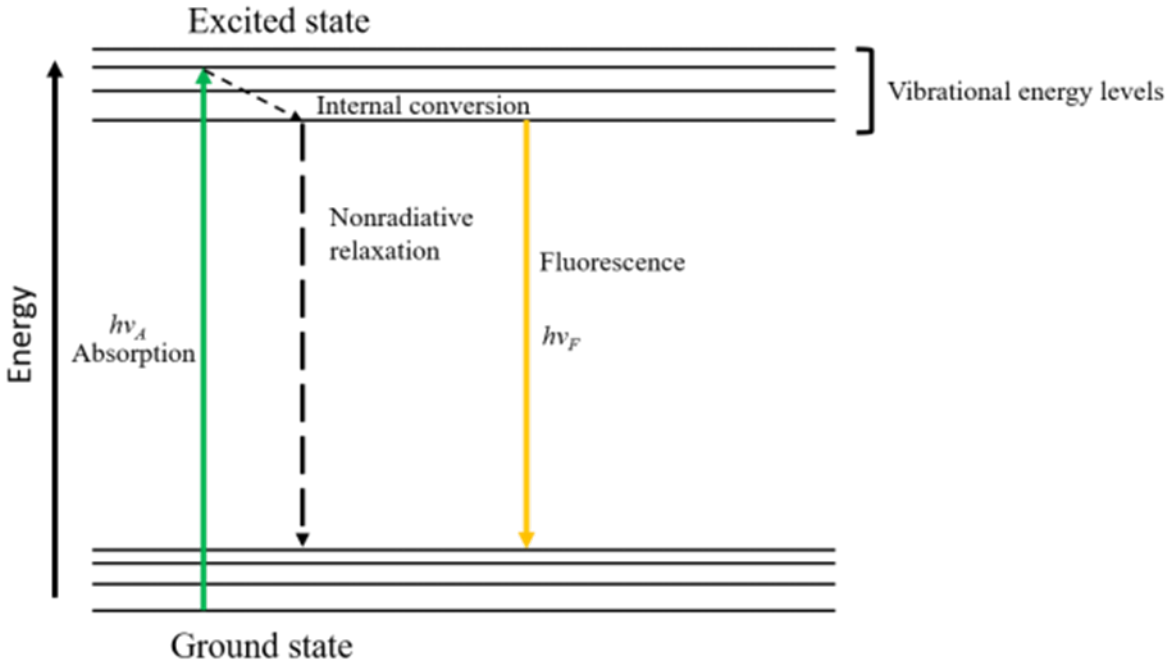

Figure 1.1. Jablonski energy diagram showing excitation and various possible relaxation mechanisms[6]. Each *hv* denotes the photon energy, where subscripts *A*, *F* denote absorption and fluorescence.

Due to energy dissipation, the energy of emitted photons is lower than of the excited photons and the energy difference represented by (hvA-hvF) defines the Stokes shift. The Stokes shift results in the red-shifted fluorescence emission relative to the excitation light. Emission wavelengths are independent of the excitation wavelength because of the vibrational relaxation to intermediate excited states. Fluorescence emission is incoherent because of the uncertain delays in the vibrational relaxations. Fluorescence emission spectrum generally is in mirror symmetry to the absorption spectrum. The characteristics of these spectra can be used to distinguish different fluorophores when they are simultaneously detected. Intensity of the excitation at the fluorescence excitation spectrum can affect fluorescence emission intensity. A single fluorophore can produce thousands of detectable photons given that the same

fluorophore can be repeatedly excited. Photobleaching refers to the phenomenon that one fluorophore is irreversibly destroyed and is permanently unable to fluorescence [2].

Fluorescence detection systems typically consist of four parts: 1) an excitation source, 2) filter sets, 3) a detector, 4) a fluorophore. These four parts should be compatible for the fluorescence detection fidelity. Fluorescence microscopes resolve fluorescence spatially for microscopic fluorescent samples. Fluorescence scanners resolve fluorescence spatially for macroscopic objects. Flow cytometers measure fluorescence characteristics in a large flowing particle stream to distinguish and isolate particles. Other fluorescence instrumentations include DNA sequencers and microfluidic devices. Different types of fluorescence instruments have different requirements on the fluorophores and measurements. For example, photobleaching should be avoided in fluorescence microscopy. However, it is not a significant problem in flow cytometry because of high flow speed and short exposure time under the excitation beam of the particles.

Fluorescence intensity depends on the same parameters as the absorbance and the fluorescence quantum yield of the target fluorophores, namely, on the excitation intensity and fluorescence collection efficiency of the detection instrument. In dilute solutions, fluorescence intensity is almost linearly proportional to the above parameters. In dense solutions, this relationship can be further distorted by self-absorption and the inner-filter effect. Background signals, which can result from autofluorescence or reagent background, seriously distort fluorescence detection sensitivity. Autofluorescence can be minimized by either appropriate filter sets or target fluorophores that absorb the autofluorescence. Reducing the detection bandwidth can increase the detection resolution but compromises the overall fluorescence intensity. Using fluorophores with > 500 nm excitation capability can mostly minimize the autofluorescence

from cells and tissues and other biological samples. Moreover, longer emission wavelengths reduce the light scattering by the dense media as well and contribute to greater penetration of the excitation light. Fluorescence Lifetime is defined as the average time that an excited molecule spends in the excited state before returning to the ground state. Lifetime decides the time available for the fluorophores to interact with their environment.

1.2 Fluorescence probes

In order to detect biological molecules such as proteins, polypeptides, antibodies and DNA/RNA and intracellular biological structures such as nuclei, mitochondria and collagen, a fluorescent probe is attached chemically or biologically to these molecules and structures. Generally, fluorescent probes use the fluorophores which are highly stable and sensitive and selectively bind to one specific region or functional group of the target molecular. A fluorophore can emit fluorescence upon light excitation. It absorbs photons at a specific energy level andre-emit photons at a lower energy level and longer wavelength. Maximum excitation and emission wavelength define the peak in the excitation and emission spectra, respectively. During fluorescence, the fluorophore exhibits a time-resolved decay pattern that can be easily characterized with a decay rate of $1/\tau$ and lifetime is τ equal to the time taken to decay to $1/e$ (≈ 0.368) of the original emission level and the decay process is mono-exponential in most cases. Lifetime of the fluorophores is independent of concentration, sample absorption, sample thickness, excitation intensity and photobleaching. So, it is highly robust and suitable to characterize a fluorophore and the correspondingly labelled protein. In the meanwhile, the fluorescence lifetime is relevant to environmental factors such as solution pH, ion concentration, molecular binding and proximity of energy acceptors. Consequently, the lifetime can be applied to functional imaging and reflects the interaction between labelled

cellular structures and the environment.

Fluorophores can be characterized and differentiated by the emission spectrum and fluorescence lifetime practically. In order to differentiate intracellular structures, multiple fluorophores are used to label different subcellular structures such as nuclei and mitochondria. Consequently, the ability to image them simultaneously (multiplexing) and separate them has been critical for understanding a variety of biological processes. Generally, two methods have been applied to achieve multiplexing. The first is the spectrum-based method and it is intensity-based, conventional and widely adopted and separates fluorophores with their emission spectra. The second method is based on the fluorescence lifetime, an identifying characteristic separate from the emission spectrum, which can be used to distinguish fluorophores. This approach relies on the fact that fluorophores exhibit a time-resolved decay pattern that can be easily characterized and is very often mono-exponential with a decay rate of $1/\tau$. Lifetime contrast is independent of fluorescence intensity and emission spectrum and is more robust than intensity-based methods.

1.3 Fluorescence lifetime imaging microscopy(FLIM)

Fluorescence lifetime imaging microscopy (FLIM) [4] has been extensively employed in a wide spectrum of biomedical applications, ranging from single cell studies [5] to medical diagnosis [6]. Rather than imaging the time-integrated fluorescent signals, FLIM measures the time-lapse fluorescent decay. Fluorescence lifetime imaging microscopy (FLIM) produces animage based on the differences in the excited state decay rate from a fluorescent sample as illustrated in Figure 1.2. Thus, FLIM is a fluorescence imaging technique where the contrast is based on the lifetime of individual fluorophores rather than their emission spectra. As an intrinsic characteristic of a fluorophore and its state, fluorescence lifetime does not depend on

concentration, absorption by the sample, sample thickness, photo-bleaching and excitation intensity; it is more robust than intensity-based methods. At the same time, the fluorescence lifetime depends on numerous environmental parameters such as pH, ion or oxygen concentration, molecular binding or the proximity of energy acceptors, making it the technique of choice for functional imaging because lifetime measurements are independent of intensity which can be significantly altered by tissue heterogeneities and depth location. FLIM enables a more quantitative study of molecular effects inside living organisms compared with conventional intensity-based approaches and is widely applied in various fields.

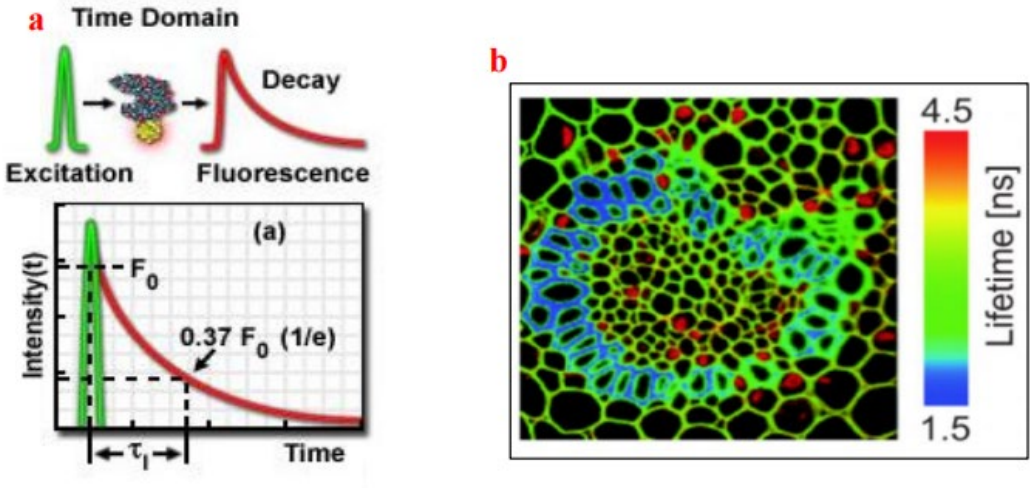

Figure 1.2 FLIM working principle[6]. (a) FLIM measures the time-lapse fluorescent decay. (b) FLIM produces an image based on the differences in the excited state decay rate from a fluorescent sample.

There are a variety of technical implementations of FLIM. Based on the detection mechanism, they are generally classified into two categories: time-domain FLIM and frequency-domain FLIM. Time-domain FLIM illuminates the sample with pulsed laser excitation, followed by measuring the fluorescent decay in sequential time channels using an ultrafast detector or detector array. Time-domain FLIM is performed using either scanning-based systems—such as a time-resolved confocal microscope [7] or a streak camera [8], or widefield-based

systems—such as a temporally- gated 2D camera or a single photon avalanche diode (SPAD) array. By contrast, frequency-domain FLIM illuminates the sample with high-frequency pulsed or modulated continuous-wave lasers, driving the fluorescence oscillating at the same frequency, however, with a reduced modulation degree and aphase shift due to the fluorophore's lifetime. To measure the modulation and phase of emitted fluorescence, the imager also modulates the gain of detector at the same frequency as the excitation light (homodyne methods) or a frequency slightly different from that of the excitation light (heterodyne methods). The measured data is then analyzed using the phasor approach.

Time-correlated single photon counting (TCSPC) is the most widely used method in time-domain FLIM to acquire fluorescence lifetime. TCSPC is a statistical method and relies on a large number of repetitive laser excitation and delay measurement cycles to accumulate enough photons for the delay histogram. In the case of intensive excitation, many electrons in the ground state of the fluorophore molecules are excited, resulting in multiple photon emission. TCSPC uses extremely low excitation intensity so that one single photon is emitted upon every laser excitation pulse. In TCSPC, the time interval between sample excitation laser pulse and the arrival of the emitted single photon at the detector is repeatedly measured to account for the statistical estimation of the fluorophore's emission decay profile at every laser excitation cycle. A large amount of data can be collected over many cycles. The delay time intervals are constructed into a histogram that summarizes the photon emission occurrence overtime upon excitation as shown in. The defined 'start' signal is realized by the electronics detecting the laser pulse which restricts the emitted photon counting rates, and the defined 'stop' signal is detected by a single-photon sensitive detector such as a photomultiplier tube (PMT), micro channel plate (MCP), single photon avalanche diode (SPAD) or hybrid PMT. In

application, a high number of emitted photons and correspondingly increased pixel dwell time are required to reach sufficient accuracy. However, due to the repetitive measurement for accuracy and mechanical scanning to the sample for large field of view (FOV),current TCSPC is limited by the slow frame rate.

Ultrafast biological dynamics are omniscient in life science due to the microcellular environment and rapid biological reaction. For example, protein folding, catalysis and ligand binding occur within from nanoseconds to milliseconds. Energy and charge transfer, isomerization and bond stretching, and twisting happen within picoseconds. Fluorescence probes are widely used to detect these dynamics. With a reference intensity camera, the intensity-based measurements infer only relative changes, and the results are easily confounded by variations in excitation fluence, signal decay by photobleaching, and background fluorescence, particularly when the imaging is performed in vivo. Compared with intensity-based approaches, FLIM provides a more quantitative assessment on the fluorescent probes because it is insensitive to the aforementioned factors. For example, FLIM has allowed precision detection of FRET in subcellular structures such as dendritic spines in neurons. As another example, confocal FLIM has been used to map the intracellular temperature distribution in live cells. Similarly, in endoscopy, lifetime imaging of tissue autofluorescence has shown great promise in demarcating malignant from normal tissue [9].

Chapter 2 Compressed fluorescence lifetime imaging microscopy

This chapter talks about compressed FLIM, a high-resolution, high-speed fluorescence lifetime imaging microscopy (FLIM) of live cells based on a compressed sensing scheme[187]. It first introduces compressed ultrafast photography (CUP) and then discusses compressed FLIM from its principle to applications. It also discusses a deep learning method to improve the image reconstruction quality of CUP.

2.1 Streak camera

A streak camera is an ultrafast optical imaging instrument that measures a wide range of the light intensity variation with time from a single-shot event. It is employed to measure the pulse duration of laser fusion lasers, free electron lasers and other pulsed lasers and for various time-crucial applications such as time-resolved Raman spectroscopy, fluorescence lifetime measurement,plasma light emission, laser ablation and LIDAR. Conventionally, a streak camera is a 1D ultrafast imaging device. Its narrow entrance slit (10- 50 μm wide) limits the imaging FOV to be a line. It operates by transforming the time profile of light pulse intensity into the spatial profile on a detector through the time-varying deflection of the light. Figure 2.1 illustrates the operating principle of the streak camera. Assume there are light pulses arriving sequentially at the slit. First, they pass through the slit and lenses and the light image is formed on the photocathode. Here at the photocathode, photons are converted into electrons in the time order and accelerated. Then they pass the sweep electrode sequentially and are deflected vertically at different angles because different sweeping voltages are applied at the different time slots. Then they hit the phosphor screen and are converted back to light, and the image is readout by a digital camera. Finally, in the streak image, the vertical axis records the time information and the horizontal axis captures the space information.

2.2 Compressed ultrafast photography (CUP)

Compressed ultrafast photography (CUP) is the worlds' fastest passive camera [12].

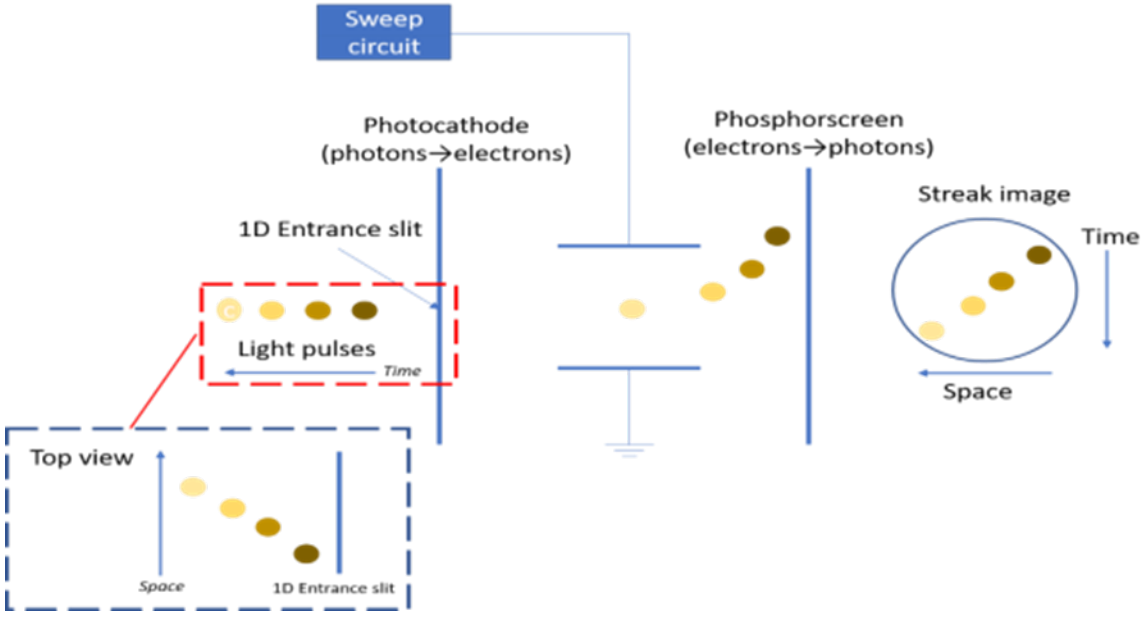

Figure 2.1 Streak camera working principle[6]

CUP uses conventional imaging optics and a streak camera but takes advantage of the ability of compressed sensing tools to recover images from sparse spatiotemporal data. The approach enables the information in the image to be encoded in the field of view of the streak camera, thereby allowing the system to read out images with full-frame capability at an extremely high speed (up to 100 billion fps). Figure 2.2 shows the schematic of CUP. Figure 2.3 shows representative images of CUP at macroscopic scales. The application of compressed sensing in ultrafast imaging has been demonstrated in compressed ultrafast photography (CUP) [12]. Compressed ultrafast photography (CUP) is a new computational ultrafast imaging technology that can capture transient dynamic events at 100 billion frames per second in a single camera exposure with a sequence depth of hundreds of frames. CUP synergistically combines two technologies—the streak camera and compressed sensing (CS) —and can perform 2D ultrafast passive imaging at 100 billion frames per second. Conventionally, a streak camera is a 1D ultrafast imaging device—its narrow entrance slit (10–50 μm wide) limits the imaging FOV to

be a line. To acquire a 2D high-speed video, one has to scan the imaging FOV along the direction perpendicular to the streak camera's entrance slit. However, this requirement poses the severe restriction that the event must be repetitive at each scanning position. Unlike other streak-camera-based ultrafast imagers, CUP overcomes this limitation with a fully opened entrance slit onto the streak camera to collect 2D image and uses compressed sensing to reconstruct the 2D image by spatially encoding an image with a random binary pattern using a digital micromirror device (DMD), followed by temporal shearing of the resultant signals using a streak camera. This spatially encoded and temporally sheared image is then detected by a CCD within a single exposure. The image reconstruction is the solution of the inverse problem of the above image formation processes. Given the spatiotemporal sparsity of the dynamic scene, which holds in many if not most natural scenes, a CS-based reconstruction algorithm can successfully decode the spatiotemporal mixing in the vertical axis of the streak camera and retrieve spatiotemporal information. Within a single exposure, CUP can capture up to 350 consecutive temporal frames at 100 billion frames per second Using CUP, many transient light-speed phenomena have been visualized, including the propagation, reflection, and refraction of a short laser pulse in space, faster-than-light propagation of non-information, and color-resolved fluorescent excitation and emission and dynamic volumetric imaging. To enable multicolor CUP, a spectral separation module in front of the streak camera can be added. As shown in Figure 2.4 a, a dichroic filter (562 nm cut-on wavelength) is mounted on a mirror at a small tilt angle (~5°). The light reflected from this module is divided into two beams according to the wavelength: green light (wavelength < 562 nm) is directly reflected from the dichroic filter, while red light (wavelength > 562 nm) passes through the dichroic filter and bounces from the mirror. Using the multicolor CUP system, a pulsed-laser-pumped

fluorescence emission process is imaged. A fluorophore, Rhodamine 6G, in water solution was

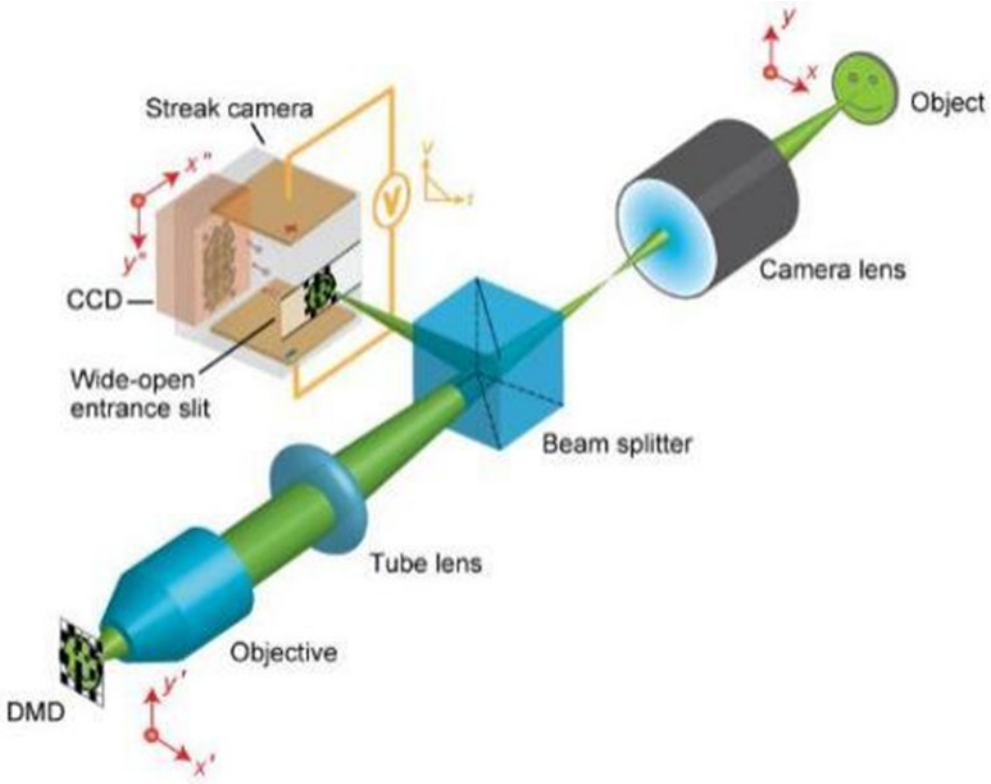

Figure 2.2 Optical schematics of CUP [12].

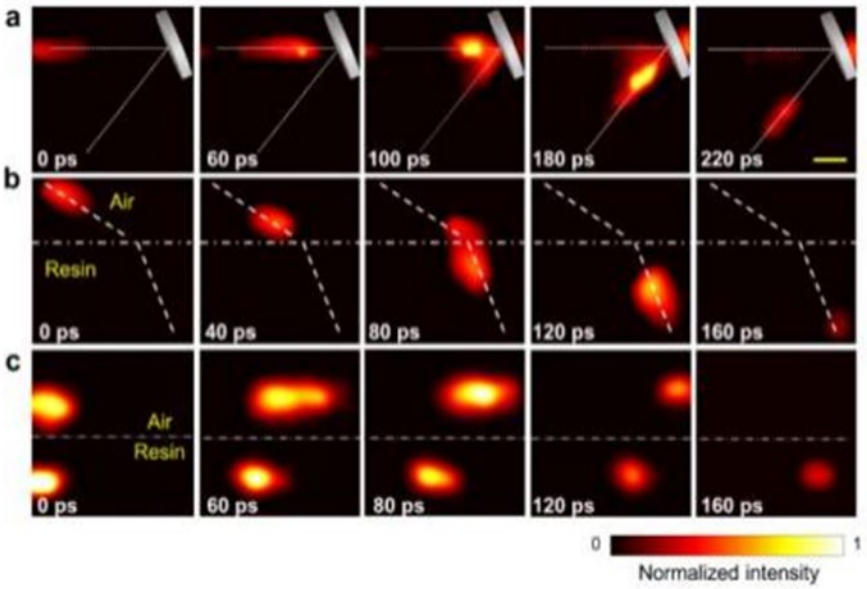

Figure 2.3 CUP of light propagation [12]. (a) Laser pulse reflected from a mirror. (b) Laser pulse refracted from an air-resin interface. (c) Laser pulses racing in air and resin. Scale bar, 10 mm.

excited by a single 7 ps laser pulse at 513 nm. Some representative temporal frames are shown in Figure 2.4 b. In addition, the time-lapse mean signal intensities within the dashed box in Figure 2.4 b for both the green and red channels (Figure 2.4 c) are calculated. Based on the measured fluorescence decay, the fluorescence lifetime was found to be 3.8 ns, closely matching a previously reported value. Currently, CUP relies on the unconstrained two-step iterative shrinkage/thresholding (TwIST) algorithm to reconstruct the event data cube. CUP is a two-dimensional ultrafast dynamic passive image platform and has three characteristics. First, it features high imaging speed and can reach picosecond time resolution. Second, it is a single-shot imagerand does not scan the sample, thereby minimizing the exposure to the sample, which is beneficial to live tissue imaging. Third, it is passive imaging and does not require active illumination. As a consequence, it is suitable for live cell imaging.

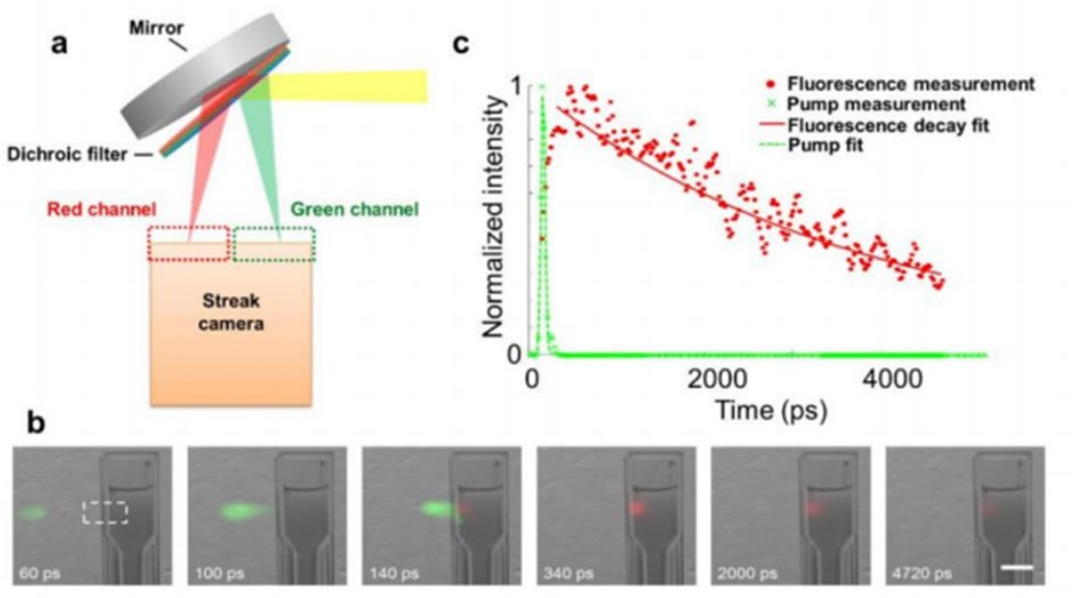

Figure 2.4 Multicolor CUP [12]. (a) Custom-built spectral separation unit. (b) Representative temporal frames of a pulsed laser-pumped fluorescence emission process. The pulsed pump laser and fluorescence emission are pseudo-colored based on their peak emission wavelengths. To explicitly indicate the spatiotemporal pattern of this event, the CUP-reconstructed frames are overlaid with a static background image captured by a monochromatic CCD camera. (c) Time-lapse pump laser and fluorescence emission intensities averaged within the dashed box in b. The temporal responses of pump laser excitation and fluorescence decay are fitted to a Gaussian function and an exponential function, respectively. Scale bar, 10mm..

2.3 Overview of compressed FLIM

To detect the fast fluorescence decay in FLIM, there are generally two strategies. Time-domain FLIM and frequency-domain FLIM as mentioned in Chapter 1. Despite being quantitative, the common drawback of FLIM is their dependence on scanning and/or repetitive measurements. For example, to acquire a two-dimensional (2D) image, a confocal FLIM imager must raster scan the entire field of view (FOV), resulting in a trade-off between the FOV and frame rate. To avoid motion artifacts, the sample must remain static during data acquisition. Alternatively, widefield FLIM systems acquire spatial data in parallel. Nonetheless, to achieve high temporal resolution, they still need to temporally scan either a gated window[36,37] or detection phase[38], or they must use TCSPC which requires a large number of repetitive measurements to construct a temporal histogram [39,40]. Limited by the scanning requirement, current FLIM systems operate at only a few frames per second when acquiring high-resolution images[41,42]. The slow frame rate thus prevents these imagers from capturing transient biological events, such as neural spiking[43] and calcium oscillation[44]. Therefore, there is an unmet need to develop new and efficient imaging strategies for high-speed, high-resolution FLIM.

To overcome the above limitations, herein we introduced the paradigm of compressed sensing into FLIM and developed a snapshot widefield FLIM system, termed compressed FLIM, which can image fluorescence lifetime at an unprecedented speed. Our method is made possible by a unique integration of (1) compressed ultrafast photography (CUP)[12] for data acquisition, (2) a dual-camera detection scheme for high-resolution image reconstruction[45], and (3) computer cluster hardware and graphic processing unit (GPU) parallel computing technologies for real-time data processing. The synergistic effort enables high-resolution (500

× 400) widefield lifetime imaging at 100 frames per second (fps). To demonstrate compressed FLIM, we performed experiments sequentially on fluorescent beads and live neurons, measuring the dynamics of bead diffusion and firing of action potential, respectively.

2.4 Results

2.4.1 Operaing principle of compressed FLIM

Compressed FLIM operates in two steps: data acquisition and image reconstruction. Briefly, the sample is first imaged by a high-resolution fluorescence microscope. The output image is then passed to the CUP camera for time-resolved measurement. Finally, we use a GPU-accelerated compressed sensing reconstruction algorithm—two-step iterative shrinkage/thresholding (TwIST)—to process the image in real time [46].

A compressed FLIM system (Fig. 2.5) consists of an epi-fluorescence microscope and a CUP camera.Upon excitation by a single laser pulse, the fluorescence is collected by an objective lens with a high numerical aperture (NA) and forms an intermediate image at the microscope's side image port. A beam splitter then divides the fluorescence into two beams. The reflected light is directly captured by a reference complementary metal–oxide–semiconductor (CMOS) camera, generating a time-integrated image:

$$E_I(x, y) = \int I(x, y, t)dt, \tag{1}$$

where $I(x, y, t)$ denotes the time-lapse fluorescence decay.

The transmitted light is relayed to a digital micromirror device (DMD) through a $4f$ imaging system consisting of a tube lens, a mirror, and a stereoscope objective. A static, pseudorandom, binary pattern is displayed on the DMD to encode the image. Each encoding pixel is turned either on (tilted -12° with respect to the DMD surface norm) or off (tilted +12° with respect to the DMD surface norm) and reflects the incident fluorescence in one of the two

directions. Two reflected fluorescence beams, masked with the complementary patterns, are both collected by the same stereoscope objective and enter corresponding sub-pupils at the objective's back focal plane. The fluorescence from these two sub-pupils are then imaged by two tube lenses, folded by right-angle prism mirrors (the lower right inset in Fig. 2.5), and form two complementary channel images at the entrance port of a streak camera. To accept the encoded 2D image, this entrance port is fully opened (~5 mm width) to its maximum, exposing the entire photocathode to the incident light. Inside the streak camera, the encoded fluorescence signals are temporally deflected along the vertical axis according to the time of arrival of incident photons. The final image is acquired by a CMOS camera—the photons are temporally integrated within the camera's exposure time and spatially integrated within the camera's pixel. The formation of complementary channel images can be written as:

$$\begin{aligned} E_{C1}(x,y) &= TSC_1I(x,y,t) \\ E_{C2}(x,y) &= TSC_2I(x,y,t). \end{aligned} \tag{2}$$

Here S is a temporal shearing operator, and T is a spatio-temporal integration operator. They describe the functions of the streak camera and CMOS camera, respectively. C_1 and C_2 are spatial encoding operators, depicting the complementary masks applied to two channel images, and $C_1 + C_2 = I$, where I is a matrix of ones. This complementary-encoding setup features a 100% light throughput, saving every photon in low-light conditions. Also, because there is no information lost, our encoding scheme enriches the observation and favors the compressed-sensing-based image reconstruction.

During data acquisition, we synchronize the streak camera with the reference camera in the transmission optical path. Therefore, each excitation event yields three images: one time-integrated fluorescence image, $E_I(x,y)$, and two spatially-encoded, temporally-sheared

channel images, $E_{C1}(x, y)$ and $E_{C2}(x, y)$.

The image reconstruction of compressed FLIM is the solution of the inverse problem of the above image formation process (Eq. 1-2). Because the two complementary channel images

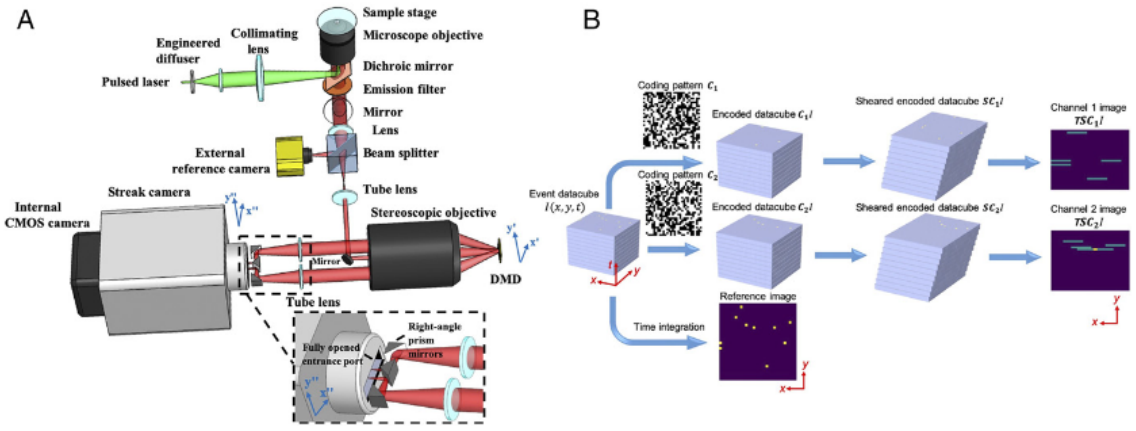

Fig.2.5 Schematic of compressed FLIM. Lower right inset[187]: Close-up of the configuration at the streak camera's entrance port. Light beams in two complementary encoding channels are folded by two right angle prisms before entering the fully opened entrance port of the streak camera. DMD: digital micromirror device. CMOS: complementary metal–oxide–semiconductor.

E_{C1}, E_{C2} are essentially associated with the same scene, the original fluorescence decay event can be reasonably estimated by applying a compressed sensing algorithm TwIST [46] to the concatenated data (E_{C1}, E_{C2}). Additionally, to further improve the resolution, we apply the time-integrated image recorded by the reference camera, E_I, as the spatial constraint. Finally, we fit a non-linear least-squares exponential curve to the reconstructed fluorescence decay at each spatial sampling location and produce the high-resolution fluorescence lifetime map. To reconstruct the image in real time, we implement our algorithm on GPU and computer cluster hardware.

The synergistic integration of hardware and algorithm innovations enables acquisition of high-quality microscopic fluorescence lifetime images. Operating in a snapshot format, the frame rate of compressed FLIM is limited by only the readout speed of the streak camera and up to

100 fps. The spatial resolution, determined by the numerical aperture of the objective lens, is in a submicron range, providing a resolving power to uncover the transient events inside a cell.

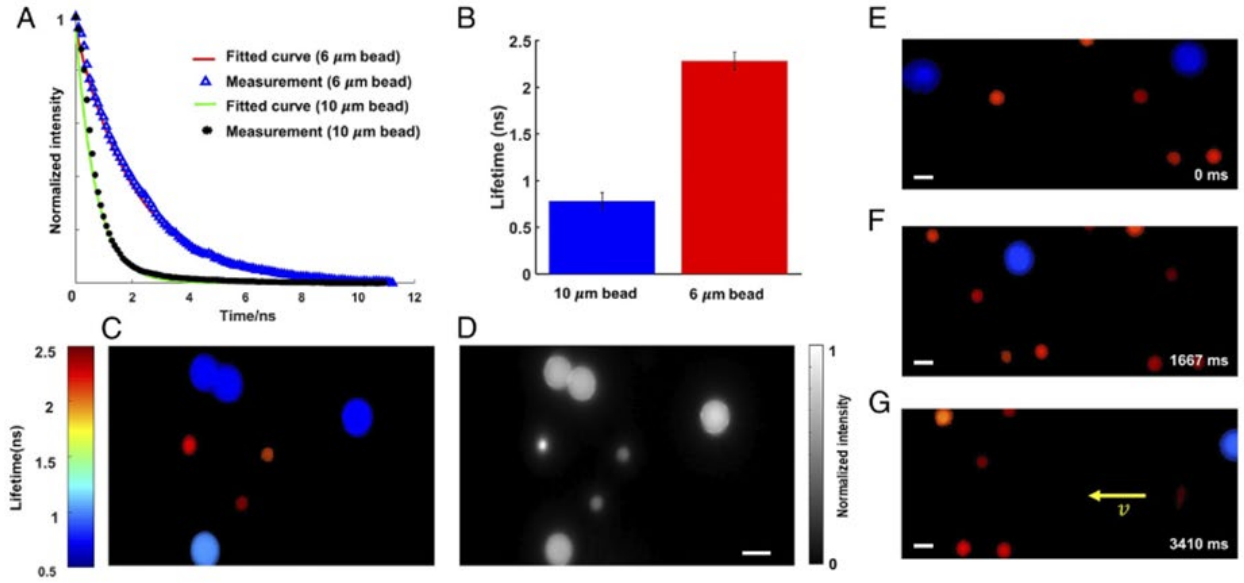

Fig.2.6 Lifetime imaging of fluorescent beads in flow[187]. a. Reconstructed fluorescence decays of two types of fluorescent beads. b. Mean lifetimes. The standard errors of the mean are shown as error bars. c. Reconstructed snapshot lifetime image. d. Reference intensity image. e-g. Lifetime images of fluorescent beads in flow at representative temporal frames. Scale bar: 10 μm.

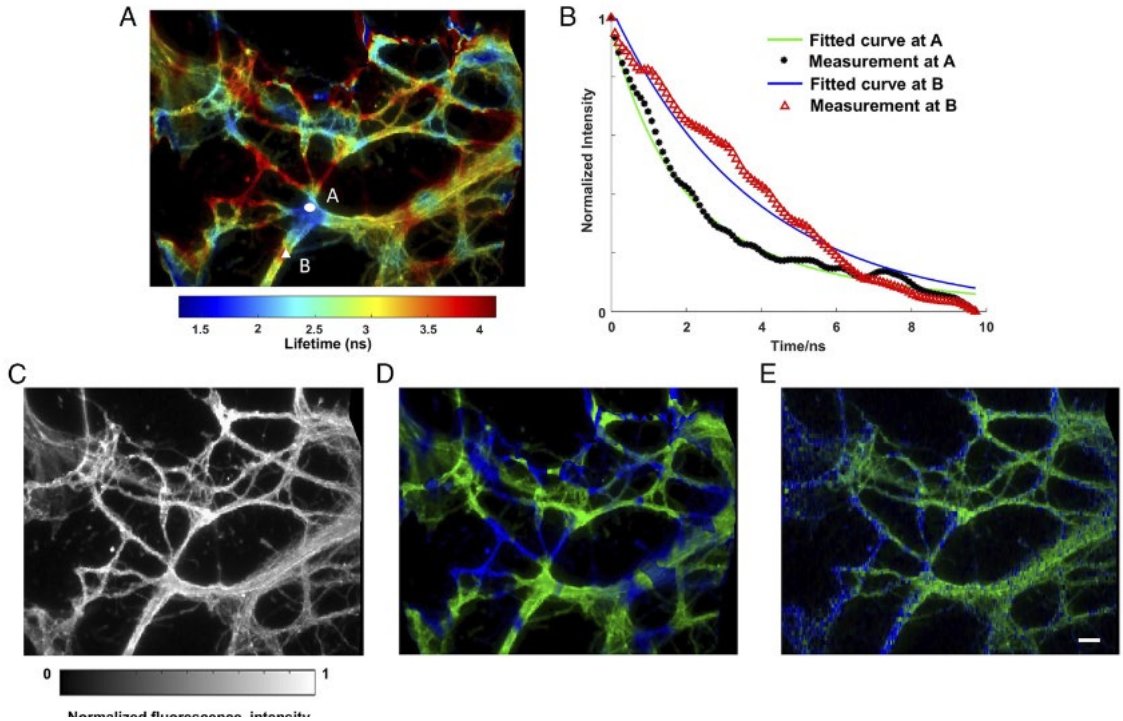

Fig.2.7 Lifetime imaging of neuronal cytoskeleton immunolabelled with two fluorophores. Alexa Fluor 555 immunolabels Vimentin. Alexa Fluor 546 immunolabels Tubulin[187]. a. Reconstructed lifetime image of immunofluorescently-stained neurons. b. Reconstructed fluorescence decays at two fluorophore locations. c. Reference intensity image. d. Lifetime unmixed image. Green channel, Vimentin; Blue channel, Tubulin. e. Ground-truth lifetime unmixed image captured by line-scanning streak imaging. Scale bar: 10 µm.

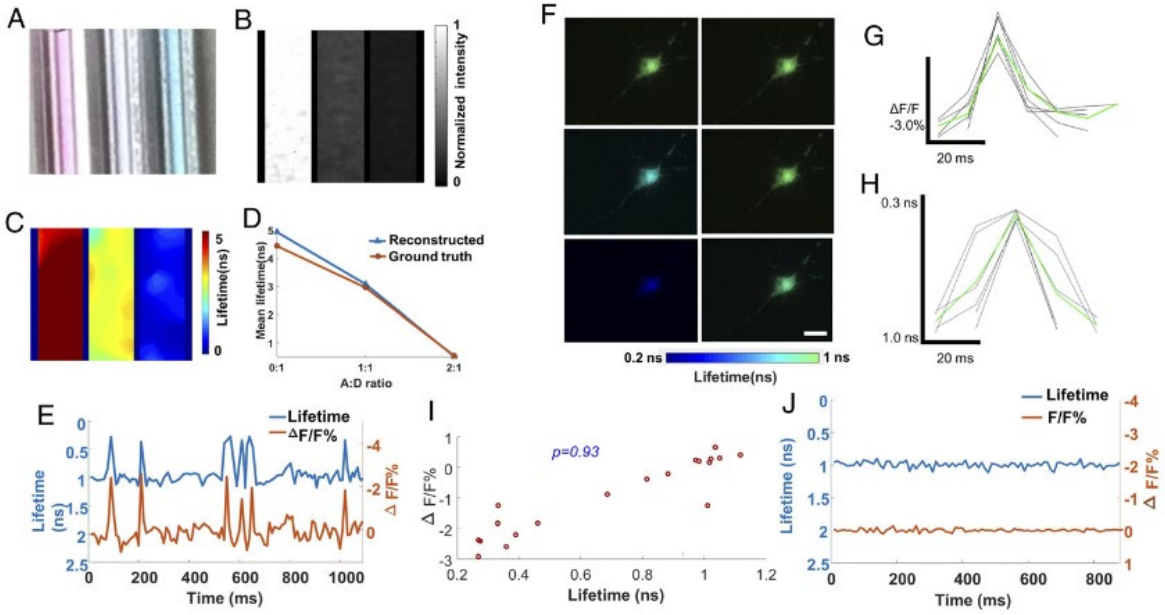

Fig.2.8 High-speed lifetime imaging of neural spiking in live neurons expressing MacQ-mOrange2. a-d. Lifetime imaging of FRET in phantoms. Acceptor Alexa 647 and donor Alexa 546 were mixed with varied concentration ratios (A:D ratio)[187]. a. Photograph of three mixed solutions with different A:D ratios (0:1, 1:1, 2:1). b. Reference intensity image. c. Reconstructed lifetime image. d. Comparison between measurement and ground truth. e-j. Lifetime imaging of FRET in live neurons. The FRET sensor MacQ-mOrange2 was expressed in a live neuron. e. Reconstructed lifetime images of the neuron at representative temporal frames upon potassium stimulation. f-g. Intensity and lifetime waveforms of neural spikes (black lines) and their mean (green line). h. Reconstructed time-lapse lifetime and intensity recording of neural spiking. The signals were averaged inside a cell. i. Scatter plot between lifetime and intensity of MacQ-mOrange2 measured at different times. The Pearson correlation coefficient is 0.93. j. Negative control. Scale bar: 10 µm.

2.4.2 Imaging fluorescent beads in flow

We demonstrated compressed FLIM in imaging fluorescent beads in flow. We mixed two types of fluorescent beads (diameters, 6 μm and 2 μm) in phosphate buffer solutions (PBS) and flowed them in a microtubing at a constant speed using a syringe pump. The ground-truth fluorescence lifetimes of these two types of fluorescent beads are 5.0 ns and 3.6 ns, respectively. We excited the beads at 532 nm and continuously imaged the fluorescence using compressed FLIM at 75 fps. As an example, the reconstructed time-lapse fluorescence decays at two beads' locations are shown in Fig. 2.6a. We pseudo-colored the bead image based on the fitted lifetimes (Fig. 2.6c). The result indicates that compressed FLIM can differentiate these two types of beads with very close lifetimes. For comparison, the corresponding time-integrated image captured by the reference camera is shown in Fig. 2.6d.

We further reconstructed the entire flow dynamics and show and representative temporal frames in Fig. 2.6e-g, respectively. Because the lifetime image was acquired in a snapshot format, no motion blur is observed. Moreover, we calculated the average fluorescence lifetimes of these two types of beads. The results 4.9 ns and 3.7 ns match well with the ground truth.

2.4.3 Lifetime unmixing of neural cytoskeletal proteins

Next, we applied compressed FLIM to cell imaging and demonstrated fluorescence lifetime unmixing of multiple fluorophores with highly overlapped emission spectra. Multi-target fluorescence labeling is commonly used to differentiate intracellular structures. Separation of multiple fluorophores can be accomplished by spectrally-resolved detection and multicolor analysis[47-49] or time-resolved detection by FLIM[50]. The spectral method fails when the emission spectra of the fluorophores strongly overlap. By contrast, FLIM has a unique edge in

this case provided a difference in fluorophores' lifetimes.

We imaged two protein structures in the cytoskeleton of neurons and unmixed them based on the lifetime. We immunolabelled Vimentin and Tubulin in the cytoskeleton with Alexa Fluor 555 and Alexa Fluor 546, respectively. The emission spectra of two fluorophores highly overlap but their fluorescence lifetimes differ (1.3 ns vs. 4.1 ns). Within a single snapshot, we captured a high-resolution lifetime image of immunofluorescently-stained neurons (Fig. 2.7a). Two representative decay curves associated with Alexa Fluor 555 and 546 are shown in Fig. 2.7b. Next, we applied a regularized unmixing algorithm to the lifetime data and separated Vimentin and Tubulin into two channels in Fig. 2.7d (green channel, Vimentin; blue channel, Tubulin). To acquire the ground-truth unmixing result, we operated our system in a slit-scanning mode and imaged the same field of view. The resultant unmixed result (Fig. 2.7e) matches well with compressed FLIM measurement. To further validate the distribution pattern of Vimentin and Tubulin in neuron cytoskeleton structures.

2.4.4 Imaging neural spikes in live cells

The complex functions of the brain depend on coordinated activity of multiple neurons and neuronal circuits. Therefore, visualizing the spatial and temporal patterns of neuronal activity in single neurons is essential to understand the operating principles of neural circuits. Recording neuronal activity using optical methods has been a long-standing quest for neuroscientists as it promises a noninvasive means to probe real-time dynamic neuronal function. Imaging neuronal calcium transients (somatic calcium concentration changes) with genetically encoded calcium indicators[51,52], fluorescent calcium indicator stains[53], and two-photon excitation methods using galvanometric[54] and target-path scanners[55] have been used to resolve suprathreshold spiking (electrical) activity.

However, calcium imaging is an indirect method of assessing neuronal activity, and spike number and firing rates using fluorescence recording is error prone, especially when used in cell populations that contain heterogeneous spike-evoked calcium signals. Thus, using functional calcium imaging to detect neuronal spiking in single-cell has been limited. Assessing neuronal function directly is invaluable for advancing our understanding of neurons and of the nervous system.

Genetically encoded voltage indicators (GEVIs) offer great promise for directly visualizing neural spike dynamics[56,57]. Compared with calcium imaging, GEVIs provide much faster kinetics that faithfully capture individual action potentials and sub-threshold voltage dynamics. Förster resonance energy transfer (FRET)-opsin fluorescent voltage sensors report neural spikes in brain tissue with superior detection fidelity compared with other GEVIs[58].

As a molecular ruler, FRET involves the nonradiative transfer of excited state energy from a fluorophore, the donor, to another nearby absorbing but not necessarily fluorescent molecule, the acceptor. When FRET occurs, both the fluorescence intensity and lifetime of the donor decrease. So far, most fluorescence voltage measurements using FRET-opsin-based GEVIs report relative fluorescence intensity changes ($\Delta F/F$) and fail to reveal the absolute membrane voltage because of illumination intensity variations, photobleaching, and background autofluorescence. By contrast, because FLIM is based on absolute lifetime measurement, it is insensitive to the environmental factors. Therefore, FLIM enables a more quantitative study with FRET-opsin-based GEVIs and provides a readout of the absolute voltage[59].

To demonstrate compressed FLIM can be used to detect FRET, we first imaged two fluorescence dyes (donor, Alexa Fluor 546; acceptor, Alexa Fluor 647) in phosphate buffer solutions (PBS) with varied mixed concentration ratios. The emission spectrum of the donor

overlaps considerably with the absorption spectrum of the acceptor, meeting the requirement for FRET. Acceptor bleed-through (ABT) contamination, i.e., the direct excitation and emission of the acceptor, is minimized by properly choosing the excitation wavelength and emission filter. We prepared three samples (Fig. 2.8a) with different concentration ratios (1:0, 1:1, 1:2) of donor and acceptor and imaged the fluorescence intensities and lifetimes using the reference time-integrated camera and compressed FLIM, respectively (Fig.2.8b-c). As expected, fluorescence emission intensity of the donor gradually diminishes with more acceptor presence and stronger fluorescence quenching. Also, as revealed by compressed FLIM, there is a decrease in the donor's fluorescence lifetime along with an increase in the acceptor's concentration. Furthermore, we performed the ground-truth measurement by switching the system to the line-scanning mode. The lifetime results acquired by compressed FLIM match well with the ground truth (Fig. 2.8d).

Next, we evaluated compressed FLIM in imaging a FRET-opsin-based GEVI, MacQ-mOrange249, to detect spiking in cultured neurons. During voltage depolarization, the optical readout is fluorescence quenching of the FRET donor mOrange2. We transfected neurons with plasmid DNA MacQ-mOrange2 and stimulated with high potassium treatment. We then used compressed FLIM to image the neural spikes. To determine fluorescence lifetime and intensity traces for individual cells, we extracted the pixels that rank in the top 50% of the SNR values, defined as $(\Delta F/F) \times \sqrt{\bar{F}}$, where ΔF/F is the voltage-dependent change in fluorescence intensity, and $\bar{F}$ is a pixel's mean baseline fluorescence intensity[58]. Average fluorescence lifetime and fluorescence intensity were calculated from these pixels in each frame. Fig. 2.8h shows the fluorescence intensity and lifetime traces of MacQ-mOrange2 sensor expressed in a cultured hippocampal neuron within 50 mM potassium environment imaged at 100 Hz.

Representative snapshots at 50 ms, 60 ms, 70 ms, 80 ms, 90 ms and 100 ms shown in Fig. 2.8e indicate occurrence of lifetime oscillations. The observed spiking irregularity may attribute to ion channel stochasticity[60], recurrent activity in the neuronal network[61], or modulation of neuronal excitability[62]. The average relative fluorescence intensity change ($\Delta F/F$) and absolute lifetime change ($\Delta\tau$) in response to one spiking event are -2.9% and -0.7 ns, respectively. Fig.2.8f-g present the experimentally determined fluorescence intensity and lifetime waveforms of single action potentials from MacQ-mOrange2 (black trace) and their mean (green trace, average over n=6 spikes). To further study the correlation between fluorescence intensities and lifetimes, we scatter plotted their relationship in Fig. 2.8i. The calculated Pearson coefficient is 0.93, indicating a high correlation between measured fluorescence intensities and lifetimes. Finally, to provide a negative control, we imaged MacQ-mOrange2 within a subthreshold non-activated 20μM potassium stimulation (Fig. 2.8j). Both the fluorescence intensities and lifetimes were stable during the entire time trace, and no oscillations were observed. Because advancement in imaging speed is crucial for resolving the dynamics of neural activity at the single cell and across neural networks, the results presented here demonstrated the utility of compressed FLIM in neuroimaging.

2.5 Discussion

Compared with conventional scanning-based FLIM imagers, compressed FLIM features three predominant advantages. First, based on a compressed-sensing architecture, compressed FLIM can produce high-resolution 2D fluorescence lifetime maps at 100 fps, allowing quantitative capture of transient biological dynamics. The gain in the imaging speed is attributed to the compressibility of a fluorescence scene in a specific domain. To show the dependence of reconstructed image quality on the compression ratio (CR) of a fluorescence event, we

calculated the CR when imaging a biological cell stained with a typical fluorophore with a lifetime of 4 ns. The observation time window on the streak camera was set as 20 ns. Here we define CR as the ratio of the total number of voxels ($N_x \times N_y \times N_z$; N_x, N_y, N_t, samplings along spatial axes x, y and temporal axis t, respectively) in the reconstructed event datacube to the total number of pixels ($N_{x\prime} \times N_{y\prime}$; $N_{x\prime}$, $N_{y\prime}$, samplings along spatial axes x', y' in the camera coordinate, respectively) in the raw image:

$$\mathrm{CR} = \frac{N_x \times N_y \times N_t}{N_{x\prime} \times N_{y\prime}}. \tag{3}$$

In compressed FLIM, the spatial information y and temporal information t occupy the same axis y' in the raw image. Therefore, their sum, $N_y + N_t - 1$, cannot exceed the total number of camera pixels along y' axis. Here we consider the equality case, $N_{y\prime} = N_y + N_t - 1$. Also, for simplicity, we assume the two complementary image channels fully occupy the entire x' axis on the camera, *i.e.*, $N_{x\prime} = 2N_x$. We then rewrite CR as:

$$\mathrm{CR} = \frac{N_y \times N_t}{2\left(N_y + N_t - 1\right)}\ . \tag{4}$$

Because of using a complementary encoding scheme, compressed FLIM possesses a 100% light throughput (ignoring reflection losses from the optical elements). The light throughput advantage can be characterized by the snapshot advantage factor, which is defined as the portion of datacube voxels that are continuously visible to the instrument[54]. When measuring a high-resolution image in the presented format (500 × 400 pixels), we gain a factor of 2×10^5 in light throughput compared with that in its point-scanning-based counterpart. Such a throughput advantage makes compressed FLIM particularly suitable for low light imaging applications.

Lastly, operating in a snapshot format, compressed FLIM eliminates the motion artifacts and

enables fast recording of stochastic biological events. We quantitatively computed the maximum blur-free motion allowed by our system. Assuming we image a typical fluorophore using a 1.4 NA objective lens and a 20 ns observation time window on the streak camera, the maximum blur-free movement during a single-shot acquisition equals the system's spatial resolution (~0.2 μm). The correspondent speed limit is 10 m/s. A lower NA objective lens or a shorter observation window will increase this threshold, however, at the expense of a reduced resolution and temporal sequence depth. The ability to capture rapid motion at the microscopic scale will be valuable to studying fast cellular events, such as protein folding[55] or ligand binding[56]. Moreover, compressed FLIM employs widefield illumination to excite the sample, presenting a condition that is favorable for live cell imaging. Cells exposed to radiant energy may experience physiological damages because of heating and/or the generation of reactive oxygen species (ROS) during extended fluorescence microscopy[57]. Because phototoxicity has a nonlinear relation with illumination radiance[58], widefield compressed FLIM prevails in preserving cellular viability compared with its scanning-based counterpart.

In compressed FLIM, the spatial and temporal information are multiplexed and measured by the same detector array. The system therefore faces two constrains. First, there is a trade-off between lifetime estimation accuracy and illumination intensity. Given ample photons, increasing the number of temporal samplings N_t will improve the lifetime estimation accuracy approximately in the manner of $\sqrt{N_t}$. Using a faster temporal shearing velocity in the streak camera will disperse the fluorescence signals to more detector array rows at the expense of a reduced signal-to-noise ratio. To maintain the photon budget at each temporal sampling bin, we must accordingly increase the illumination intensity at the sample, a fact which may introduce photobleaching and shorten the overall observation time. The second trade-off exists

between y-axis spatial sampling and t-axis temporal sampling. For a given format detector array, without introducing temporal shearing, the final raw image occupies N_y pixel rows. With temporal shearing, the image is sheared to $N_y + N_t - 1$ pixel rows with the constraint $N_y + N_t - 1 \leq N_{y\prime}$. Here $N_{y\prime}$ is the total number of detector rows. Increasing the image size N_y will decrease the N_t and thereby the lifetime estimation accuracy. Therefore, the image FOV and lifetime estimation accuracy must be balanced for a given application.

In summary, we have developed a high-speed, high-resolution fluorescent lifetime imaging method, compressed FLIM, and demonstrated its utility in imaging dynamics. Capable of capturing a 2D lifetime image within a snapshot, we expect compressed FLIM would have broad applications in blur-free observation of transient biological events, enabling new avenues of both basic and translational biomedical research.

2.5 Methods in Compressed FLIM

2.5.1 Forward model

We describe compressed FLIM's image formation process using the forward model[5]. Compressed FLIM generates three projection channels: a time-unsheared channel, and two time-sheared channels with complementary encoding. Upon laser illumination, the fluorescence decay scene is first imaged by the microscope to the conjugated plane and $I(x, y, t)$ denotes the intensity distribution. A beam splitter then divides the conjugated dynamic scene into two beams. The reflected beam is directly captured by a reference CMOS camera, generating a time-integrated image. The optical energy, $E_I(m, n)$, measured at pixel *m, n,* on the reference CMOS camera is:

$$E_I(m,n) = \int dt \int dx \int dy I(x,y,t)\,\mathrm{rect}\left[\frac{x}{d} - \left(m + \frac{1}{2}\right), \frac{y}{d} - \left(n + \frac{1}{2}\right)\right], \tag{5}$$

where d is the pixel size of the reference camera.

The transmitted beam is then relayed to an intermediate plane (DMD plane) by an optical imaging system. Assuming unit magnification and ideal optical imaging, the intensity distribution of the resultant intermediate image is identical to that of the original scene. The intermediate image is then spatially encoded by a pair of complementary pseudorandom binary patterns displayed at the DMD plane. The two reflected spatially-encoded images have the following intensity distribution:

$$\begin{aligned} I_{C_1}(x',y',t) &= \sum_{i,j} I(x',y',t) C_{1,i,j}\,\mathrm{rect}\left[\frac{x'}{d'}-\left(i+\frac{1}{2}\right),\frac{y'}{d'}-\left(j+\frac{1}{2}\right)\right] \\ I_{C_2}(x',y',t) &= \sum_{i,j} I(x',y',t) C_{2,i,j}\,\mathrm{rect}\left[\frac{x'}{d'}-\left(i+\frac{1}{2}\right),\frac{y'}{d'}-\left(j+\frac{1}{2}\right)\right]. \end{aligned} \tag{6}$$

Here, $C_{1,i,j}$ and $C_{2,i,j}$ are elements of the matrix representing the complementary patterns with $C_{1,i,j} + C_{2,i,j} = 1$, i, j are matrix element indices, and d' is the binned DMD pixel size. For each dimension, the rectangular function (rect) is defined as:

$$\mathrm{rect}(x) = \begin{cases} 1, & \text{if } |x| \le \dfrac{1}{2} \\ 0, & \text{else} \end{cases}.$$

The two reflected light beams masked with complimentary patterns are then passed to the entrance port of a streak camera. By applying a voltage ramp, the streak camera acts as a shearing operator along the vertical axis on the input image. Assuming ideal optics with unit magnification again, the sheared images can be expressed as

$$\begin{aligned} I_{s_1}(x'',y'',t) &= I_{C_1}(x'',y''-vt,t) \\ I_{s_2}(x'',y'',t) &= I_{C_2}(x'',y''-vt,t), \end{aligned} \tag{7}$$

where v is the shearing velocity of the streak camera.

$I_{s_i}(x'', y'', t)$ (i = 1, 2) is then spatially integrated over each camera pixel and temporally integrated over the exposure time. The optical energy, $E_{C_i}(m, n)$ (i = 1, 2), measured at pixel *m, n,* is:

$$\begin{aligned} E_{C_1}(m,n) &= \int dt \int dx'' \int dy'' I_{s_1}(x'', y'', t)\, \mathrm{rect}\left[\frac{x''}{d''} - \left(m + \frac{1}{2}\right), \frac{y''}{d''} - \left(n + \frac{1}{2}\right)\right] \\ E_{C_2}(m,n) &= \int dt \int dx'' \int dy'' I_{s_2}(x'', y'', t)\, \mathrm{rect}\left[\frac{x''}{d''} - \left(m + \frac{1}{2}\right), \frac{y''}{d''} - \left(n + \frac{1}{2}\right)\right]. \end{aligned} \tag{8}$$

Here, d'' is the camera pixel size. Accordingly, we can voxelize the input scene, $I(x, y, t)$, into $I_{i,j,k}$ as follows:

$$I(x,y,t) \approx \sum_{i,j,k} I_{i,j,k}\, \mathrm{rect}\left[\frac{x'}{d'} - \left(i + \frac{1}{2}\right), \frac{y'}{d'} - \left(j + \frac{1}{2}\right), \frac{t}{\Delta_t} - \left(k + \frac{1}{2}\right)\right], \tag{9}$$

where $\Delta_t = d''/v$. If the pattern elements are mapped 1:1 to the camera pixels (that is, $d'=d''$) and perfectly registered, and the reference CMOS camera and the internal CMOS camera of the streak camera have the same pixel size (that is, $d = d''$), combining equations (5)-(9) yields:

$$\begin{aligned} E_I(m,n) &= \frac{d^3}{v} \sum_{k=0}^{n-1} I_{m,n,k} \\ E_{C_1}(m,n) &= \frac{d^3}{v} \sum_{k=0}^{n-1} C_{1,m,n-k} I_{m,n-k,k} \\ E_{C_2}(m,n) &= \frac{d^3}{v} \sum_{k=0}^{n-1} C_{2,m,n-k} I_{m,n-k,k}. \end{aligned} \tag{10}$$

Here $C_{i,m,n-k}\, I_{m,n-k,k}$ (i = 1, 2) represents the complimentary-coded, sheared scene, and the inverse problem of equation (10) can be solved using existing compressed-sensing algorithms[35,59,60].

2.5.2 Compressed FLIM image reconstruction algorithm

Given prior knowledge of the binary pattern, to estimate the original scene from the compressed FLIM measurement, we need to solve the inverse problem of equation (10). Because of the sparsity in the original scene, the image reconstruction can be realized by solving the following optimization problem

$$\hat{I} = \arg\min_{I}\left\{\frac{1}{2}\left\|E - \mathbf{O}I\right\|^2 + \beta\,\Phi(I)\right\}, \tag{11}$$

where $\boldsymbol{O}$ is the linear operator, $\Phi(I)$ is the regularization function and β is the weighing factor between the fidelity and sparsity. To further impose space and intensity constraint, we construct the new constrained solver:

$$\hat{I}_s = \arg\min_{I\in\mathbf{M},I>s}\left\{\frac{1}{2}\left\|E - \mathbf{O}I\right\|^2 + \beta\,\Phi(I)\right\}. \tag{12}$$

Here, **M** is a set of possible solutions confined by a spatial mask extracted from the reference intensity image and defines the zone of action in the reconstruction. This spatial constraint improves the image resolution and accelerates the reconstruction. *s* is the low intensity threshold constraint to reduce the low-intensity artifacts in the data cube. In compressed FLIM image reconstruction, we adopt an algorithm called two-step iterative shrinkage/thresholding(TwIST)[35], with $\Phi(I)$ in the form of total variation (TV):

$$\Phi(I) = \sum_{k=0}^{N_t-1}\sum_{i=1}^{N_x\times N_y}\sqrt{\left(\Delta_i^h I_k\right)^2 + \left(\Delta_i^v I_k\right)^2} + \sum_{m=1}^{N_x}\sum_{i=1}^{N_y\times N_t}\sqrt{\left(\Delta_i^h I_m\right)^2 + \left(\Delta_i^v I_m\right)^2} + \sum_{n=1}^{N_y}\sum_{i=1}^{N_x\times N_t}\sqrt{\left(\Delta_i^h I_n\right)^2 + \left(\Delta_i^v I_n\right)^2} \tag{13}$$

Here we assume that the discretized form of I has dimensions $N_x \times N_y \times N_t$ (N_x, N_y and N_t are respectively the numbers of voxels along *x, y* and *t*), and *m, n, k* are three indices. I_m, I_n, I_k denote the 2D lattices along the dimensions *m, n, k,* respectively. Δ_i^h and Δ_i^v are horizontal and

vertical first-order local difference operators on a 2D lattice. After the reconstruction, non-linear least squares exponential fitting is applied to data cube along the temporal dimension at each spatial sampling point to extract the lifemap.

2.5.3 Lifetime-based fluorophore un-mixing algorithm

For simplicity, here we consider only fluorophores with single-exponential decays. Provided that the sample consists of n mixed fluorophores with lifetimes $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$ and concentration $\boldsymbol{x} = (x_1, \ldots, x_n)$, upon a delta pulse excitation, the discretized time-lapse fluorescence decay is:

$$y_t = \sum_{k=0}^{n-1} A_{t,k} x_k . \tag{14}$$

Here, k is the fluorophore index, $A_{t,k}$ is an element of the fluorescence decay component matrix $\boldsymbol{A}$, and $A_{t,k} = \exp(-t/\tau_k)$. The inverse problem of equation (14) is a least squares problem with constraints. We choose 2-norm penalty and form the solvent for ℓ_2-regularized least squares problem:

$$\hat{x} = \arg\min_x \left\{ \frac{1}{2} \left\| y - \mathbf{A}x \right\|_2^2 + \lambda \left\| x \right\|_2^2 \right\}, \tag{15}$$

where the first term $\frac{1}{2}\|y - \boldsymbol{A}x\|_2^2$ represents the measurement fidelity, and the regularization term penalizes large norm of x. The regularization parameter λ adjusts the weight ratio between fidelity and 2-norm penalty.

In our experiment, to construct fluorescence decay component matrix $\boldsymbol{A}$, we first directly imaged Alexa Fluorophore 555 and Alexa Fluorophore 546 in solution and captured their time-lapse fluorescence decay. Then we computed their lifetimes by fitting the asymptotic portion of the decay data with single exponential curves. Finally, we applied the regularized

unmixing algorithm to the lifetime data and separated the fluorophores into two channels.

GPU assisted real-time reconstruction using computer cluster

Based on the iterative construction, compressed FLIM is computationally extensive. For example, to reconstruct a 500 × 400 × 617 (*x, y, t*) event datacube and compute a single lifetime image, it takes tens of minutes on a single PC. The time of constructing a dynamic lifetime movie is prohibitive. To accelerate this process, we (1) implemented the reconstruction algorithm using a parallel programming framework on two NVIDIA Tesla K40 GPUs and (2) performed all reconstructions simultaneously on a computer cluster (Illinois Campus Cluster). The synergistic effort significantly improved the reconstruction speed and reduced the movie reconstruction time to seconds. **Table 1** illustrates the improvement in reconstruction time when the computation is performed on a single PC vs. the GPU-assisted computer cluster.

2.5.4 Compressed FLIM: hardware

In the compressed FLIM system, we used an epi-fluorescence microscope (Olympus IX83) as the front-end optics. We excited the sample using a 515 nm picosecond pulse laser (Genki-XPC, NKT Photonics) and separated the fluorescence from excitation using a combination of a 532 nm dichroic mirror (ZT532rdc, Chroma) and a 590/50 nm band-pass emission filter (ET590/50m, Chroma). Upon excitation, an intermediate fluorescence image was formed at the side image port of the microscope. A beam splitter (BSX16, Thorlabs) transmitted 10% of light to a temporal-integration camera (CS2100M-USB, Thorlabs) and reflected 90% of light to the temporal-shearing channels. The reflected image was then relayed to a DMD (DLP LightCrafter 6500, Texas Instruments) through a 4f system consisting of a tube lens (AC508-100-A, Thorlabs) and a stereoscopic objective (MV PLAPO 2XC, Olympus; numerical

aperture, 0.50). At the DMD, we displayed a random, binary pattern to encode the image. The reflected light from both the "on" mirrors (tilted +12° with respect to the norm) and "off" mirrors (tilted -12° with respect to the norm) were collected by the same stereoscopic objective, forming two complementary channel images at the entrance port of a streak camera (C13410-01A, Hamamatsu). The streak camera deflected the image along the vertical axis depending on the time-of-arrival of incident photons. The resultant spatially-encoded, temporally-sheared images were acquired by an internal CMOS camera (ORCA-Flash 4.0 V3, C13440-20CU, Hamamatsu) with a sensor size of 1344(H) × 1016(V) pixels (1 × 1 binning; pixel size d = 6.5 μm). We synchronized the data acquisition of cameras using a digital delay generator (DG645, Stanford Research Systems).

2.5.5 Filter selection for FRET-FLIM imaging

To assure only fluorescence emission from the donor was collected during FRET-FLIM imaging, we chose a filter set (a 515 nm excitation filter and a 590/50 nm emission filter) to separate excitation and fluorescence emission. This filter combination suppressed the direct excitation of the acceptor to <3% and minimized the collection of the acceptor's fluorescence, thereby eliminating the acceptor bleed-through (ABT) contamination.

2.5.6 Spatial registration among three imaging channels

Because compressed FLIM imaged a scene in three channels (one temporal-integrated channel and two complementary temporal-shearing channels), the resultant images must be spatially registered. We calibrated the system using a point-scanning-based method. We placed an illuminated pinhole at the microscope's sample stage and scanned it across the FOV. At each point-scanning position, we operated the streak camera in the "focus" mode (i.e., without temporal shearing) and captured two impulse response images with all DMD's pixels turned

"on" and "off", respectively. Meanwhile, the reference CMOS camera captured another impulse response image in the temporal-integration channel. We then constructed a lookup table by extracting the pinhole locations in these three impulse response channel images. This lookup table was later used to register the three channel images $\left(E_I, E_{C_1}, E_{C_2}\right)$for concatenated image reconstruction.

2.5.7 Acquisition of encoding matrices $\mathbf{C}_1$ and $\mathbf{C}_2$

To acquire the encoding matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, we imaged a uniform scene and operated the streak camera in the "focus" mode. The streak camera directly captured the encoding patterns without temporal shearing. Additionally, we captured two background images with all DMD's pixels turned "on" and "off", respectively. To correct for the non-uniformity of illumination, we then divided the coded pattern images by the corresponding background images pixelwise.

2.5.8 Slit-scanning streak camera imaging

To form a ground-truth lifetime image, we employed the DMD as a line scanner and scanned the sample along the direction perpendicular to the streak camera entrance slit. We turned on the DMD's (binned) mirror rows sequentially and imaged the temporally-sheared line image in the correspondent imaging channel. Given no spatiotemporal mixing along the vertical axis, the fluorescence decay data along this line direction could be directly extracted from the streak image. Next, we computed the fluorescence lifetimes by fitting the decay data to single exponential curves. The resultant line lifetime images were stacked to form a 2D representation.

2.5.9 Confocal FLIM imaging

For confocal FLIM imaging, we used a bench-mark commercial system (ISS Alba FCS). The sample was excited by a Ti:Sapphire laser, and fluorescence was collected by a Nikon Eclipse

Ti inverted microscope. The time-lapse fluorescence decay was measured by a time-correlated single photon counting unit. To form a 2D lifetime image, the system raster scanned the sample. The typical time to capture a 2D lifetime image (256 × 256 pixels) was ~ 60 s.

2.5.10 Fluorescence beads

We used a mixture of 6 μm and 2 μm diameter fluorescence beads (C16509, Thermo Fisher; F8825 Thermo Fisher) in our experiment. To prepare the mixed beads solution, we first diluted the 6 μm and 2 μm diameter bead suspensions. After sonicating the two fluorescence bead suspensions, we pipetted 10 uL of the 6 μm bead suspension (~1.7×10^7 beads/mL) and 2 μm bead suspension (~4.5×10^9 beads/mL) into 1mL and 10mL phosphate buffer solutions (PBS), respectively. Next, we mixed 100 uL diluted 2 μm bead solution with 1 ml diluted 6 μm bead solution. The final mixed beads solution contained approximately 1.5×10^5 6 μm beads/mL and 4.1×10^5 2 μm beads/mL.

2.5.11 FRET phantom

We used Alexa Fluor 546 (A-11003, Thermo Fisher) as the donor and Alexa Fluor 647 (A-21235, Thermo Fisher) as the acceptor. We prepared the acceptor solutions with three different concentrations (0 mg/mL, 1 mg/mL, 2 mg/mL) and mixed them with the same donor solution(1 mg/mL). We then injected them into three glass tubes (14705-010, VWR) for imaging.

2.5.11 Primary cell culture:

Primary hippocampal neurons were cultured from dissected hippocampi of Sprague-Dawley rat embryos. Hippocampal neurons were then plated on 29mm glass bottom petri dishes that are pre-coated with poly-D-lysine (0.1 mg/ml; Sigma-Aldrich). To help the attachment of neurons (300 cells/mm^2) on to the glass bottom dish, neurons were initially incubated with a

plating medium containing 86.55% Minimum Essential Medium Eagle's with Earle's Balanced Salt Solution (MEM Eagle's with Earle's BSS, Lonza), 10% Fetal Bovine Serum (re-filtered, heat inactivated FBS; ThermoFisher), 0.45% of 20% (wt./vol.) glucose, 1× 100 mM sodium pyruvate (100×; Sigma-Aldrich), 1× 200 mM glutamine (100×; Sigma-Aldrich), and 1× penicillin/streptomycin (100×; Sigma-Aldrich). After three hours of incubation in the incubator (37°C and 5% CO_2), the plating media was aspirated and replaced with maintenance media containing Neurobasal growth medium supplemented with B-27 (Invitrogen), 1% 200 mM glutamine (Invitrogen) and 1% penicillin/streptomycin (Invitrogen) at 37 °C, in the presence of 5% CO_2. Half the media was aspirated once a week and replaced with fresh maintenance media. The hippocampal neurons were grown for 10 days before imaging.

2.5.12 Immunofluorescently stained neurons:

We immunolabelled the Vimentin (MA5-11883, Thermo Fisher) with Alexa Fluor 555 (A-21422, Thermo Fisher) and Tubulin (PA5-16891, Thermo Fisher) with Alexa Fluor 546 (A-11010, Thermo Fisher) in the neurons.

2.5.13 Transfect the neurons to express MacQ-mOrange2:

We used Mac mutant plasmid DNA Mac-mOrange2 (#48761, Addgene) to transfect hippocampal neurons. In DNA collection, as soon as we received the agar stab, we grew bacteria in Luria-Bertani (LB) broth with ampicillin in 1:1000 dilution for overnight in 37°C. Standard Miniprep (Qiagen) protocol was performed in order to collect DNA. DNA concentration was measured by Nanodrop 2000c (ThermoFisher). In neuron transfection, lipofectamine 2000 (Invitrogen) was used as transfecting reagent. In an Eppendorf tube, we stored 1mL of the conditioned culture media from 29mm petri plates neuron culture with 1mL of fresh media. We prepared two separate Eppendorf tubes and add 100uL of Neurobasal

medium to each tube. For one tube, 3ug of DNA was added while 4uL of lipofectamine 2000 was added to the other tube. After five minutes, two tubes were mixed together and incubated in room temperature for 20 minutes. This mixture was added to neuron culture in dish for 4 hours in the incubator (37°C and 5% CO_2). We took out the media containing lipofectamine 2000 reagent and added the stored conditioned and fresh culture media to neuron culture in dish. Hippocampal neurons were imaged 40 hours after transfection.

2.5.14 Potassium stimulation and imaging:

We used high potassium (50mM) treatment to stimulate neuron spike. The extracellular solution for cultured neurons (150 mM NaCl, 4 mM KCl, 10 mM glucose, 10 mM HEPES, 2 mM $CaCl_2$ and 2 mM $MgCl_2$) was adjusted to reach the desired final K^+ concentration (50mM) and maintain physiological osmolality at the same time. At each stimulation, we removed the media in the plate and pipetted high potassium extracellular solution at the same time.

2.6 Deep-learning-based reconstruction for compressed ultrafast photography

Currently, both compressed FIM and CUP relies on the two-step iterative shrinkage/thresholding (TwIST) algorithm to reconstruct the event datacube. The recovered image resolution is degraded by the temporal shearing operation of the streak camera [63]. Several improvements in the reconstruction algorithms have been made in the literature [63-65]. However, the resultant image quality is still non-optimal, and the optimization-based reconstruction methods typically need tens to hundreds of iterations to converge and often require fine-tuning of the hyper-parameter to obtain high-fidelity results, both of which are time-consuming. Also, the memory requirement is high due to the complex computation such as matrix inversion. Inspired by the recent advances in applying deep learning (DL) [66] to

computational imaging systems for faster and more accurate reconstruction [67-70], we present a deep-learning-based method for the CUP image reconstruction to improve the image quality and accelerate the reconstruction speed towards real-time display applications[188].

In CUP, as illustrated in Figure 2.9(a), the dynamic scene is first imaged by a lens to an intermediate image plane. Then a beam splitter divides the light towards two directions. The reflected light is directly imaged by an external CMOS camera (Thorlabs, CS2100M-USB). The transmitted light is passed to a DMD (Texas Instrument, LightCrafter 6500) by a 4f imaging system with a tube lens (Thorlabs, AC508-100-A) and a stereo objective (Olympus, MV PLAPO 2XC) . A static pseudorandom binary pattern is displayed onto the DMD to spatially encode the dynamic scene. Each encoding pixel is turned either on (tilted -12° with respect to the DMD surface norm) or off (tilted +12° with respect to the DMD surface norm) and reflects the incident light in one of the two directions. The reflected light masked with the pattern is collected by the same stereo objective and further relayed to the wide-open entrance slit of a streak camera (Hamamatsu, C13410-01A). Inside the streak camera, the incident light is temporally sheared by a sweeping voltage in the vertical axis according to the time of flight and imaged by an internal CMOS camera (Hamamatsu, ORCA-Flash 4.0) in a single 2D image. As discussed in 2.3, the reconstruction in CUP is a inherent large-sized 3D problem: the CUP captures 3D (x-y-t) data with a single 2D measurement (x-y). A key to reduce the complexity of applying DL for reconstruction is to recognize that, in the measurement operation, each 2D image slice (y-t) in the 3D datacube is independent of each other along the x axis: the T, S, C operator applies to the column of every instantaneous (x, y) image independently. As a result, the 2D image slice (y-t) corresponds to a 1D compressed line image (y) in the CUP measurement data. Therefore, the 3D image reconstruction can be

decomposed into massively-parallel 2D image reconstruction, as illustrated in Figure 1(b) and Figure 2. The measurements and the mask are first decomposed into independent line images (y, xi) and line mask Ci with i being the column index in the 2D measurement (x, y) and 2D mask C. The network input is then initialized to ATy with AT being the adjoint of A and function as an approximate inverse operator to reduce the learning burden. Such a setting is reliable for computational imaging problems [67] and has been widely used in recent works [71,72]. The network output is the 2D image slice (y, t). Network groups are constructed for each specific line masks, and image slices from the network groups are finally concatenated to the 3D datacube (x, y, t). Compared with the 3D mapping network that directly reconstructs the 3D datacube, this segmentation reconstruction method benefits from a smaller network, enabling faster training and requiring fewer training samples.

The deep learning network (Figure 2.10 inset) uses an encoder-decoder "U-net" architecture with modifications of replacing each convolution layer with a dense block (DB) to improve the training efficiency [73,74]. The encoder gradually condenses the spatiotemporal information into feature maps with increasing depths; the decoder recombines the information from the feature maps into the ultimate image. Specifically, first, the input goes through the "encoder" path, which consists of four dense blocks connected by max-pooling layers for downsampling. The intermediate output from the encoder encodes rich information along with the "depth" (1088 activation maps) with small lateral dimensions. Each dense block contains multiple layers, and each layer consists of batch normalization (BN), the rectified linear unit (ReLU) nonlinear activation, and convolution (conv) with 16 filters. Next, the low-resolution activation maps go through the "decoder" path, which consists of four additional dense blocks connected by upsampling convolutional (up conv) layers. Four skip connections are set across

different spatiotemporal scales along the encoder-decoder path to preserve high-frequency information. After the decoder path, an additional convolutional layer followed by the last layer produces the network output. For the loss function, we measure the l^2 distance between the network's prediction and the ground truth image. To provide strong restrictions on the forward model, we add extra constraints on the forward model as the second term in the loss function: the encoded streak camera image and the image captured by the reference camera. We denote x_i as the ground truth image, $\hat{x}_i$ as the network's prediction with i being the image index in the training image batch, and N as the training batch size. The resultant loss function $\ell(x, \hat{x})$ is:

$$\ell(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N} \|\hat{x}_i - x_i\|_2^2 + \lambda\|\boldsymbol{A}\hat{x}_i - \boldsymbol{A}x_i\|_2^2, \tag{16}$$

where λ is the parameter that controls the relative weights of each loss component. λ is set to 1 in the training. After training, the reconstructed (x, y, t) datacube is predicted from the streak camera measurement, CMOS camera measurement, and the mask.

To generate the dataset for training, we adopted two strategies. First, we assembled a collection of 3D image cube x by applying various dynamics (such as image shifting, reshuffling of the 2D images in the 3D cube) on the MNIST database and 1000 in-house experimental images of different objects. To obtain the corresponding measurement data set y, we applied the CUP forward operator A on the target image set x. To emulate experimental measurements, we encoded the dynamics by the mask captured in the real experiment and added the shot noises to the measurement data set y in the synthetic dataset. Second, we collected a small experimental dataset that contains the ground truth 3D image cube x. We obtained by the ground truth datacube x by a line-scanning operation in the CUP system. In line-scanning operation, we employed the DMD as a line scanner by turning on the DMD's

(binned) mirror rows sequentially and recording the temporally-sheared (x, t) image. By scanning the sample along the y direction, which is perpendicular to the entrance slit of the streak camera, we stacked all the (x,t) images to form the ground truth 3D datacube (x, y, t). Furthermore, to improve the network reconstruction accuracy, we performed image augmentation, including crop, shift, flip, and affine transformations on the training images to increase the training sample size. In total, 10,000 samples were used for training, which consists of MNIST (~5%), the in-house experimental objects (~70%), and experimental dynamics (~25%).

We trained the network using the Adam optimizer for 50: epochs. The leaning rate was initialized as 10^{-3} and scaled down by a factor of 0.5 every 5 epochs. For each epoch, we used measurements of the same dynamics with different added noise realizations to improve the robustness of the network. A total of 50 different noise realizations at the same noise level were generated for the synthetic datasets in the 50 epochs of one training. The training was performed on a campus cluster with two GPU (NVIDIA Tesla M2090) using Keras/Tensorflow. Once the network was trained, image reconstruction can be achieved in real time. To further show the speed improvement by the DL methods, we compared the reconstruction time when using DL, TwIST, and other algorithms—SALSA, FISTA, and GAP [74-76]. The DL reconstruction speed of a (256, 256, 32) datacube is at least 60 times faster than that enabled by the fastest GAP algorithm.

We first validated the deep learning (DL) method on simulated data and benchmarked it against the TwIST algorithm. A 256-by-256 Shepp-Logan (S-L) phantom was used as the base image. The simulated dynamic scene contained 9 frames, with the S-L phantom decaying exponentially. The intensity decay obeys $\boldsymbol{I}[n] = \exp(-0.25n)$, where $\boldsymbol{I}$ is the intensity trace,

n is the frame index, and the decay rate is 0.25. The streak camera measurement was generated according to the forward model, and both shot noise and 1% Gaussian white noise were added. The dynamic scene was then reconstructed using the TwIST-based constrained reconstruction method and our DL reconstruction method, respectively. Figures 2.11(a-c) show temporally integrated images of the ground truth, the reconstructed images by TwIST, and the DL method, respectively. For the large bright patch in Region 1 and the small bright spot in Region 2, the image produced by the DL reconstruction shows better contrast and resolution than that of TwIST reconstruction. The boundary between the dark and bright patches in Region 3 is also more prominent in the DL result than that in the TwIST result. To compare the reconstruction quality of the two methods across frames, we plotted normalized intensity changes against the frame index at the same circled pixels indicated in (a)-(c) as shown in Fig. 2.11(d). For the TwIST reconstructed intensity trace against the ground truth intensity trace is 0.06, and the reconstructed decay rate is 0.26 with 95% confidence bounds using a nonlinear least squares fitting method and single exponential model. For DL reconstruction, RMSE of the intensity trace is 0.07, and the reconstructed decay rate is 0.28 with 95% confidence bounds. The DL reconstruction thus provides a comparable temporal reconstruction accuracy to that of the TwisT algorithm.

We then benchmarked the DL method against the TwIST algorithm on experimental data. We imaged the fluorescence decay of a fluorescent tissue paper upon pulsed laser excitation. The 515 nm picosecond pulse laser (NKT Photonics, Genki-XPC, 7 ps pulse duration) first passed through an engineered diffuser and excited the fluorescent tissue paper. We separated the fluorescence from excitation using a combination of a 532 nm dichroic mirror (ZT532rdc, Chroma) and a 590/50 nm band-pass emission filter (ET590/50m, Chroma). We then used 10×

objective (Olympus, UPLFLN 10X2) and tube lens to relay the fluorescence to the intermediate image plane (shown in Fig. 2.9). Next, the CUP system collected the photons.

We reconstructed the dynamic scene using both the constrained TwIST and the DL reconstruction method. Figure 2.12(a) presents the reference image captured by the external reference camera. Figures 2.12(b-c) show temporally integrated images of the TwIST and DL reconstructed datacubes, respectively. In the spatial domain, the results of the DL method illustrate sharper boundaries and higher spatial resolution. Figure 4(d) shows the normalized intensity changes across time at the same circled pixels indicated in Figs. 4(b-c). For TwIST reconstruction, the reconstructed fluorescence lifetime (reciprocal of the decay rate) is 6.13 ns with 95% confidence bounds using the nonlinear least squares fitting method and a single exponential model. The DL reconstructed fluorescence lifetime is 6.29 ns with 95% confidence bounds. Figures 2.12(e) and 2.12(f) show the reconstructed frames at t= 0 ns, 1.3 ns, and 2.5 ns reconstructed by the TwIST and DL methods, respectively. The results indicate that, in the temporal domain, the reconstruction accuracy provided by DL and TwIST is similar.

Despite being able to produce a high quality image with much reduced reconstruction time, due to the segmentation method in deep learning reconstruction, the DL reconstruction results may have boundary artefects (stripes in each x-y temporal frame). This boundary artefects can be removed in post-processing [76]. The running time for this additional step is approximately 0.17 s, bringing the total DL reconstruction time (DL image reconstruction and image post-processing) to 0.27 s. This speed is still ~60 times faster than that enabled by the fastest algorithm GAP. In addition, the DL reconstruction is inferior to the iterative TwIST algorithm regarding the flexibility: the pre-trained network works only with specific mask patterns. A

potential solution to solve this problem is to use transfer learning, which can bypass the need to train the network from scratch[77]. Specifically, the pre-trained network for the old mask can be fine-tuned for a new mask, which can reduce training time and training sample size. Moreover, for larger data size in the reconstruction, deeper DL network and more training samples may be required for optimal performance.

In conclusion, we developed a DL reconstruction method for CUP. Compared with the conventional TwisT algorithm, the DL method can recover the dynamic scene with shaper boundaries, higher feature contrast, and fewer artifacts while maintaining a similar temporal reconstruction accuracy. Moreover, the DL method increases the reconstruction speed by a factor of over 500, thereby enabling real-time reconstruction of large-sized event datacubes.

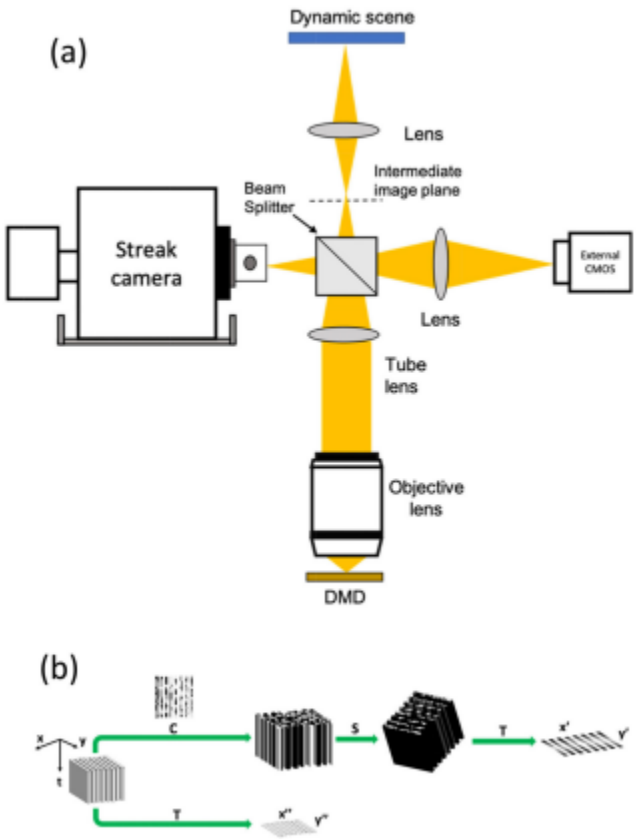

Fig.2.9 Schematic of the CUP system and CUP data acquisitions[188]. (a) Schematic of the CUP system. DMD, digital micromirror device. (b) Schematic of CUP data acquisition. t, time; x, y, spatial coordinates of the

dynamic scene; x', y', spatial coordinates at the streak camera; x'', y'', spatial coordinates at the external CMOS camera; C, spatial encoding operator; S, temporal shearing operator; T, spatiotemporal integration operator. The 3D image reconstruction can be decomposed into massively-parallel 2D image reconstruction.

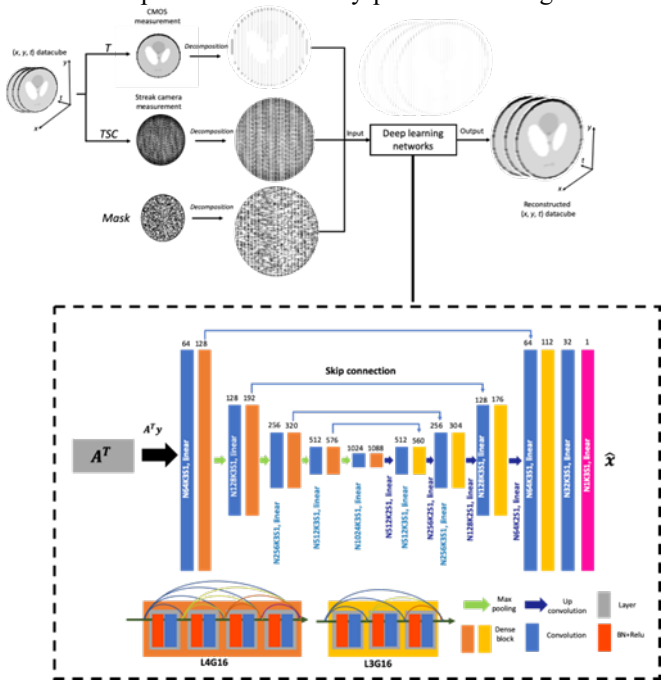

Fig. 2.10 Deep learning workflow and network architecture for CUP[188]. The measurements and the mask are first decomposed into independent line images (y, xi) and line mask Ci, i is the column index in the 2D measurement (y, x) and 2D mask C. Input to the network is then set to ATy and the network output is the 2D image slice (y, t). Network groups are constructed for each specific line mask, and image slices from the network groups are then concatenated to the 3D datacube (x, y, t). The deep learning network uses U-net structure. Notations: N: number of kernels, K: kernel size, S: stride, L: number of layers in the dense block, G: growth rate of dense block. N#K#S# denotes the number of kernels, kernel size and stride of the convolution layer, respectively. L#G# denotes the number of layers and growth rate inside the dense block, respectively.

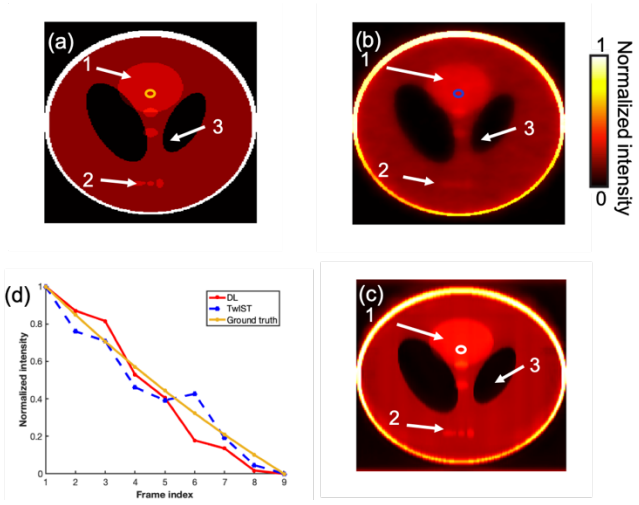

Fig. 2.11. Results of the numerical simulation[188]. (a-c) Temporally projected images of the ground truth, the TwIST reconstructed result, and the DL reconstructed datacubes. (d) Intensity trace against the frame index at the circled pixels in (a-c).

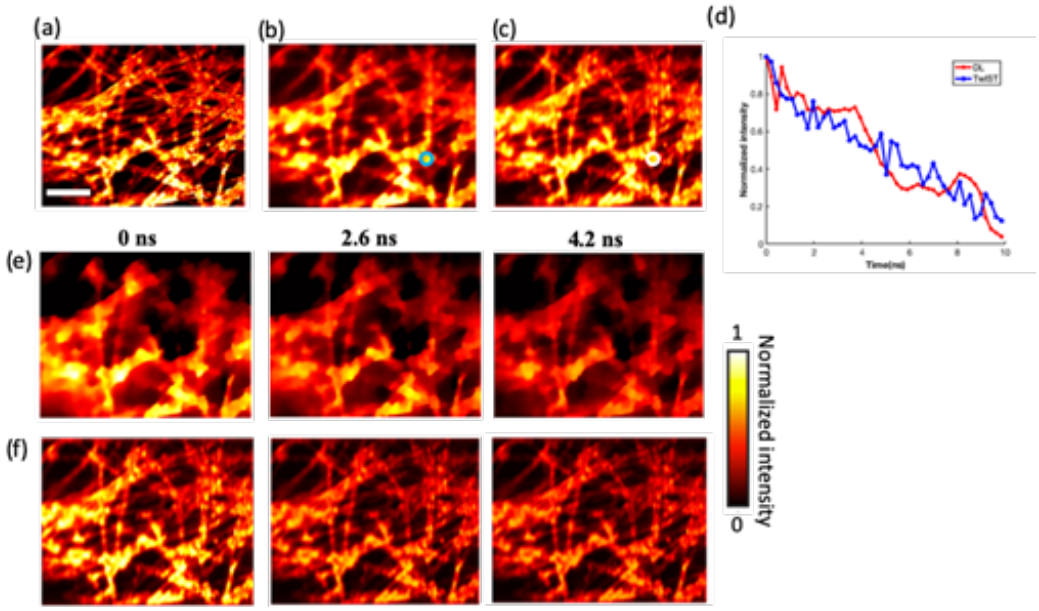

Fig. 2.12. Experimental results[188]. (a) Reference image captured by the external CMOS camera. (b-c) Temporally integrated images of TwIST and DL reconstructed datacubes. (d) Time-lapse intensity change at the same circled pixels in (b-c). (e-f) Reconstructed frames at t= 0 ns, 1.3 ns, 2.5 ns by the TwIST and DL methods, respectively. S]cale bar: 100 μm.

Chapter 3 Compact light field photography towards versatile three-dimensional vision

Although compressed FLIM can achieve high speed and high resolution FLIM, it does not have optical sectioning capability. To acquire 3D lifetime images, it still requires scanning along the depth axis. Moreover, it is established on compressed sensing and the image quality highly depends on the compression ratio and scene complexity. To break these limitations, we developed a imaging framework--referred to as compact light field photography(CLIP)—that explores the data redundancy in the conventional light field camera and acquires the light field data with high efficiency. This chapter discusses CLIP from its principle to its applications[133].

3.1 Overview of CLIP

Three-dimensional (3D) imaging is vital for sensing, modeling, and understanding the physical world[78,79], with a broad range of applications in navigation, robotics, and medical imaging[80,81]. However, there is an inherent dimensionality gap between a 3D scene and the recording sensors, which can at most be arranged on a two-dimensional (2D) surface as in curved sensor arrays[82]. As a result, only a 2D projection of the scene can be captured from a given perspective. To recover depths, one must perform additional measurements along an extra axis of light: an angular axis in multi-view measurements or a temporal axis in time-of-flight sensing. While multi-view methods, including stereo[83], structured-light[84], and light field cameras[85], can attain exceptional depth accuracy (<100 μm) at near distances and operate at a relatively high speed, their accuracies degrade quadratically with distance and ultimately fail at a long range[83]. Except for structured light that employs active illuminations, multi-view methods also rely heavily on object texture for effective depth extraction. On the other hand, time-of-flight techniques are agonistic of textures and can maintain the depth resolution over a large detection range[86-88]. However, high-speed and

dense depth mapping robust against stochastic motions is still challenging for time-of-flight cameras. The distinct strengths and limitations of multi-view methods and time-of-flight techniques have long divided the design of 3D imaging cameras, confining the capability and application scopes of incumbent 3D vision solutions.

Performing multi-view time-of-flight measurements has potentially disruptive benefits. Besides bringing 3D imaging to an ultrafast time scale[89] and fuelling the development of new 3D vision capabilities[90,91], a dense multi-view measurement can substantially bolster the sensing range of a time-of-flight camera15 and make it possible to see through occlusions[92,93], a recurring challenge for visual tracking in computer vision[94]. However, current methods for acquiring multi-view time-of-flight signals suffer from either a limited number of views or being too time-consuming that hinders dynamic imaging. More problematically, such direct multi-view acquisition of time-of-flight signals exacerbates the "big data" issue. Even with a low spatial resolution, sensing along the extra angular or temporal dimension yields a large amount of data—multi-view measurements generate a plethora of images from different views, and a single time-of-flight (temporal) trace involves thousands of time points. Adding yet another dimension will increase the system complexity and data load so fast that makes real-time image processing and streaming impracticle16. Driven by the growing need of high-resolution 3D imaging with large-format detectors like the megapixel SPAD sensors[95], designing efficient multi-view measurements becomes increasingly more relevant. Moreover, sensors for infrared wavelengths, sub-picosecond measurements[96], and other specialized applications[97] are still limited in element counts, which prevent dense 2D image sampling[98] and consequently hamper 3D imaging via conventional multi-view methods.

To address these challenges, we present compact light field photography (CLIP) to sample dense light fields8 with a significantly improved efficiency and flexibility. Unlike previous compressive light field cameras[99-101] that require densely sampled 2D images for recovering a 4D light field, CLIP is a systematic framework to design and transform any imaging model that employs nonlocal data acquisition into a highly efficient light field imaging approach: by distributing the designed or existing nonlocal image acquisition process into different views and modelling the correlations inherent in 4D light fields, CLIP can recover the 4D light field or directly retrieve refocused images from a measurement dataset even smaller than a single sub-aperture image. Under the CLIP framework, sensors of arbitrary formats—a single pixel, a linear array, or a sparse 2D area detector—can be employed for efficient light field imaging by transforming the imaging models of a single pixel camera, x-ray computed tomography and a diffuser camera[102], to name a few. Additionally, CLIP is natively applicable to camera array systems, and promotes robustness against defective sensor measurements and severe scene occlusions. With CLIP, we seamlessly synergized multi-view with time-of-flight techniques, and demonstrated single-shot 3D imaging of texture-less scenes in an extended depth range, robust 3D vision through severe occlusions, and real-time non-line-of-sight (NLOS) imaging with curved and disconnected walls, a critical task for field applications not yet fulfilled[103–108]

3.2 Results

3.2.1 Principle of compact light field photography

In linear systems, the image acquisition process can be written in a general matrix formalism as:

$$\mathbf{f} = \mathbf{A}\mathbf{h} + \boldsymbol{\sigma}, \qquad (17)$$

where $\boldsymbol{\sigma}$ is the measurement noise, $\mathbf{h}$ and $\mathbf{f}$ are the vectorized image and measurement, respectively. $\mathbf{A}$ is the $m \times N^2$ system matrix (for an image resolution of $N \times N$ throughout the manuscript), which is square ($m=N^2$) for full-rate sampling and rectangular ($m<N^2$) under compressive sensing. This formulation, though general, typically assumes all m measurements of the scene are obtained from a single view and hence possessing no light field capability. To record light fields with an angular resolution of l, the measurement procedure must be repeated either parallelly (via a lens array[109]) or sequentially at l different views[110], leading to a measurement dataset of size $m \times l$. In CLIP, we break this convention by employing nonlocal acquisition for the row entries of matrix $\mathbf{A}$ and splitting the m measurements into l different views, thereby compactly recording light field data with an angular resolution of l.This transforms the image model into:

$$\mathbf{f} = \begin{bmatrix} \mathbf{A_1} & \cdots & \mathbf{0} \\ \vdots & \mathbf{A_2} & \vdots \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{A}_l \end{bmatrix} \begin{bmatrix} \mathbf{P_1} \\ \mathbf{P_2} \\ \vdots \\ \mathbf{P}_l \end{bmatrix} + \boldsymbol{\sigma} = \mathbf{A}' \begin{bmatrix} \mathbf{P_1} \\ \mathbf{P_2} \\ \vdots \\ \mathbf{P}_l \end{bmatrix} + \boldsymbol{\sigma} = \mathbf{A}'\mathbf{P} + \boldsymbol{\sigma}, \quad (18)$$

where $\mathbf{A_k}$ is the *k-th* sub-matrix such that $\mathbf{A} = [\mathbf{A_1};\ \mathbf{A_2};\ \cdots, \mathbf{A_l}]$, $\mathbf{A}'$ is the transformed block-diagonal matrix and $\mathbf{P} = [\mathbf{P_1};\ \mathbf{P_2};\ \cdots, \mathbf{P_l}]$ is the 4D light field. While one can exploit the sparisity prior to compressively recover a 4D light field at this stage, CLIP can further retrieve a refocused image directly by explicitly modelling the correlations in the 4D light field to better cope with complex scenes.

This is inspired by the observation that images of the same scene acquired from different views share the same content in photographic applications. There is only a depth-dependent disparity between any two sub-aperture images, as illustrated in Fig. 3.1a. Therefore, one can

explicitly model the correlations among the sub-aperture images by digitally propagating the light field, which relates the sub-aperture image at view k (denoted as $\mathbf{P_k}$) to a reference sub-aperture image $\mathbf{h}$ via an invertible shearing operator $\mathbf{B_k}$ as $\mathbf{P_k} = \mathbf{B_k h}$, and the m measurement data acquired from l views now becomes:

$$\mathbf{f} = \begin{bmatrix} \mathbf{f_1} \\ \mathbf{f_2} \\ \vdots \\ \mathbf{f_l} \end{bmatrix} = \begin{bmatrix} \mathbf{A_1} & \cdots & \mathbf{0} \\ \vdots & \mathbf{A_2} & \vdots \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{A}_l \end{bmatrix} \begin{bmatrix} \mathbf{B_1 h} \\ \mathbf{B_2 h} \\ \vdots \\ \mathbf{B}_l \mathbf{h} \end{bmatrix} + \boldsymbol{\sigma} = \begin{bmatrix} \mathbf{A_1 B_1} \\ \mathbf{A_2 B_2} \\ \vdots \\ \mathbf{A_l B_l} \end{bmatrix} \mathbf{h} + \boldsymbol{\sigma} = \mathbf{F(d)h} + \boldsymbol{\sigma}, \quad (19)$$

where $\mathbf{f_k}$ is a vector that contains m_k measurements at view k, and the total number of measurements in $\mathbf{f}$ is $m = \sum_{k=1}^{l} m_k$. The whole system model $\mathbf{F(d)}$ becomes a function of the depth d, which is the key to recover the image h with different focal settings—by applying the shearing matrix $\mathbf{B_k}$ (hence $\mathbf{F(d)}$) to depth d as in light field cameras, the reconstructed image will be correspondingly focused thereon. CLIP thus can attain light field imaging (conventionally of a data size $m \times l$) with a measurement data size of only m. It is worth noting that further reduction of the measurement data is possible by multiplexing the measurement from all the views onto a single measurement vector: $\mathbf{f^c} = \mathbf{Tf}$, with $\mathbf{T}$ being the integration operator.

The nonlocal acquisition strategy is pivotal to encode all scene points of the image into each view's smaller sub-measurement vector $\mathbf{f_k}$ (i.e., $m_k \ll N^2$) for attaining an effective angular resolution of l. This is similar to the incoherent multiplexing requirement in compressive sensing, where a rich pool of nonlocal acquisition schemes has been developed for a range of applications during the past decades, benefiting CLIP. Such a nonlocal acquisition also endows CLIP with imaging robustness against defective pixels or scene occlusions. Because the complete scene is encoded in any subset of the measurements, image recovery is not

substantially affected by a fraction of defective pixel readings, despite that the conditioning of image reconstruction might deteriorate. Similarly, an object that is completely blocked in certain views by its surrounding objects, as in Fig. 3.1b, could be partially visible to the remaining views, which contain incomplete but complementarily global information of the object to enable its retrieval. Furthermore, CLIP's nonlocal acquisition can take advantage of the compressibility of natural photographs for compressive imaging ($m<N^2$) to minimize data load, particularly when coupled with time-of-flight imaging. As illustrated in Fig. 3.1c for a point-scanning-based LiDAR imaging of a crowded office scene, the camera captures only a thin slice of the 3D volume at each time bin under an ultrashort illumination. As a result, the crowded 3D scene is decomposed into a sequence of instantaneous 2D images that are far simpler than its continuously-wave-illuminated photograph. Such instantaneous compressibility also holds for NLOS imaging, albeit in a different representation basis.

Three exemplar CLIP embodiments utilizing a single pixel element (0D), a linear array (1D), and a 2D area detector are illustrated in Fig. 3.1d-f, respectively. The single-pixel camera[111,112] (Fig. 3.1d) sequentially encrypts the scene with different random codes and measures light intensities with a bucket detector. To sample light fields without redundant data, CLIP splits the measurements by scanning the detector during code update along the *u*-*v* direction (*u, v*, angular axis behind the collection lens) into *l* positions. With random binary codes, each measurement integrates ~50% of the image pixels, and $m_k>=7$ measurements in each view cover every pixel with a high probability ($p = 1 - 0.5^7 > 99\%$). For CLIP with 1D sensors, the *x*-ray CT imaging model is transformed by using a cylindrical lens to cast along the lens' invariant axis (the axis without optical power) a line-integral of the image onto an individual pixel as in Fig. 3.1e, allowing a 1D detector array to parallelly records $m_k=N$

measurements to cover all image pixels. Light fields of the scene are then acquired with an array of cylindrical lenses, each being oriented at a distinct angle with respect to the 1D sensor.

For CLIP imaging with 2D detectors (of various sparsity), one can design a complex-valued mask to produce a wide-field, depth-dependent point spread function (PSF) (Fig. 3.1f) to multiplex sub-aperture measurements (i.e., $\mathbf{f^c = Tf}$). Moreover, we can unify wavefront coding[113,114], coded-aperture[115] techniques, and diffuser cameras into the CLIP framework, where the full recovery of 4D light fields is unnecessary. Adaption to camera array systems can be readily accomplished by making each camera (of any dimension) record a few nonlocal coefficients of the scene and sufficiently overlapping individual cameras' fields of view.

3.2.2 3D imaging through occlusions

Seeing through occlusions has been previously achieved by dense camera array systems16, which apply synthetic aperture processing to blur down the occluder while keeping the object of interest coherently enhanced. However, a clear separation of the object and occluder in 3D is difficult due to the defocused background and the limited depth sectioning capacity of camera array systems. We show here background-free 3D vision through severe occlusion with time-of-flight (ToF) CLIP imaging. For the proof-of-concept demonstration, we built a ToF-CLIP system with a streak camera as the 1D ultrafast sensor for snapshot acquisition of large scale 3D time-of-flight data. The streak camera is spatially multiplexed by seven customized plano-convex cylindrical lenslets (diameter of 2 mm and a focal length of 4 mm) along its entrance slit at distinct orientations to mimic a camera array system (Methods). The baseline of the camera is 15 mm, and the field of view is 30 mm at a distance of 60 mm. With

~1,000 pixels, CLIP implicitly recorded a 125×125×7 light field dataset and streamed a temporal sequence of 1,016 points (i.e., 1,000 spatial × 1,016 temporal) at a 100 Hz repetition rate for high-speed imaging. A femtosecond laser was modulated by a motorized assembly consisting of a concave lens and a diffuser for programmable illumination between a diverged and a collimated light. The diverged laser shines from an oblique angle to cover both the object and occluder.

Background-free 3D imaging through severe occlusions for three different scenes is shown in Fig. 3.2a-c. In all cases, the objects that are completely blocked by their preceding items, as rendered in the front view images emulating what a conventional camera would capture, can be well retrieved by the ToF-CLIP camera with correct 3D locations and geometric shapes. For larger objects such as the letter V in Fig. 3.2b that remain partially visible, its occluded parts are recovered with a weaker intensity. This is because the occluded parts contribute less effective measurement signals for the image reconstruction, equivalent to imaging with a smaller synthetic aperture. Trackings through occlusions is demonstrated in Fig. 3.2d, where a 2×2 grid pattern made of white foam was mounted on a translation stage behind a rectangular obscurer and moved back and forth across the camera field of view. Motion of the grid pattern varied the severity of occlusion smoothly from none to a complete obscurance as shown in the representative frames. Except for a weaker intensity caused by occlusion, the grid pattern is adequately recovered at all the time instances

It is noteworthy that a clear separation between the objects and occluder is consistently achieved in all the scenes. No defocused background signals from the occluder are discernible on the blocked object, highlighting the benefits of merging dense multi-view measurement with ToF by CLIP. Because an occluder reduces the number of measurements for the blocked

object, and the current CLIP camera has a compression factor of ~20 (with respect to a single sub-aperture image), the occluded objects that can be well recovered are restricted to be relatively simple in geometry. Nonetheless, the imaging outcomes are still remarkable, considering the reduction of measurements against camera array systems of the same resolution is more than 100 folds, and conventional compressive imaging using the same amount of data shows similar imaging characteristics but lacks light field capability to see through occlusions.

3.2.3 Flash LiDAR within an extended depth range

For high-quality 3D imaging of indoor scenes, multi-view methods require an unwieldy system baseline apart from their dependence on object texture. By contrast, flash LiDAR imaging can maintain a high precision at longer distances in a compact form but suffers from a stringent tradeoff between the sensing range and light throughput[116]. To demonstrate that CLIP is well-posed to lift such a tradeoff, we tuned the camera's field of view to 1.5 m × 1.5 m at a nominal distance of 3.0 m by moving the lenslet array closer to the slit and aligned the laser to be approximately confocal with the camera while providing a diverged illumination after passing through the motorized assembly.

An example of single-shot flash LiDAR imaging with an extended depth range is shown in Fig. 3.3a, where several texture-less letters were placed at different depths spanning a range of ~2 meters. The extended depth of field is highlighted in Fig. 3.33b by computationally refocusing the camera from far to near, as indicated in the top view image. The resultant LiDAR projection photograph clearly renders the defocusing blur for objects that deviate from their actual focal settings, whereas an all-in-focus image generated by CLIP in Fig. 3.3c allows a sharper portrait of the entire 3D scene. The flash LiDAR imaging resolution was estimated to

be about 30 mm laterally and ~10 mm axially (depth direction). While this example features a relatively simple scene to facilitate the comparison between the reference photograph and LiDAR images

We demonstrated ToF-CLIP in dynamic imaging of a 3D scene by mounting a letter V on a rotation stage and manually rotating it at an irregular speed against a simple and cluttered background, respectively. The resultant motions were filmed by the CLIP-ToF camera at a 100 Hz frame rate and a reference video camera at 60 Hz. The two videos were then numerically synchronized after temporally downsampling the LiDAR results to 60 Hz for comparison. Representative frames of the dynamic results are shown in Fig. 3.3d and e for the simple and cluttered background, respectively.

3.2.4 NLOS imaging with curved and disconnected surfaces

Unlike LiDAR that detects the directly scattered photons, NLOS imaging analyzes multiply scattered light from a diffusive surface to reveal objects hidden from direct line-of-sight. A key ingredient for such an analysis is the precise knowledge of the relaying surface's 3D geometry, which was previously obtained via nontrivial calibrations by a stereo camera or scanning-based LiDAR[117] ranger, hampering applications in the field where the relay surface evolves with the camera's viewpoint and 3D surroundings. The ToF-CLIP camera addresses this critical need for real-time mapping of the relay surface via built-in flash LiDAR imaging. More importantly, it can accommodate a non-planar surface geometry for NLOS imaging using array detectors with its light field capability. Paired with a proposed hybrid time-frequency domain reconstruction algorithm, which can handle general surfaces with a computational complexity of o(N4) (Methods), ToF-CLIP can attain real-time NLOS imaging with arbitrary curved surfaces. While NLOS imaging with a dynamic and curved surface has

been demonstrated by Manna40 et. al., its reception point was fixed at a stationary point rather than being on the dynamic surface, making it inapplicable for real-time imaging with array detectors. Similarly, the preprocessing step[118] proposed by Lindell et.al. that adapts the f-k migration reconstruction algorithm to deal with slightly curved surfaces in confocal NLOS imaging has a computational complexity of o(N5logN), which is higher than the time-domain phasor field method and thus inefficient for real-time reconstruction.

To demonstrate our approach, we directed the CLIP-ToF camera towards a scattering wall with a fixed focus. The field of view is tuned to be approximately 0.5 m × 0.5 m at a standoff distance of ~1 m. The geometry of the wall was mapped by the flash LiDAR, and NLOS signal reception was delayed accordingly to avoid the strong reflections from the collimated laser spot on the wall. The hidden scene was then reconstructed in real-time by the hybrid time-frequency domain algorithm.

We demonstrated NLOS imaging with planar, disconnected, and curved walls in Fig. 3.4a-c. For all the relay walls, the hidden scenes were placed over 1 m away from the laser spot on the wall and then imaged with a single laser shot at an average laser power of 700 mW. The 3D flash LiDAR measurement of the walls are shown in the first column, and the NLOS imaging results for two example objects in each category were rendered in a 2D front view (from the wall's perspective) and a 3D point cloud format in the following columns. Both the 3D positions and morphological structures of the hidden objects were decently recovered for NLOS imaging with all the relay walls. The importance of an extended depth-of-field to cope with disconnected and curved surfaces is illustrated in Fig. 3.4d-e respectively, where the camera's extended depth of field is disabled by computationally refocusing the camera onto different planes (from rear to the front) before reconstructing the hidden scenes. Due to

defocus effects that blur the spatiotemporal data on the walls, the reconstruction quality degrades noticeably compared with the images reconstructed with an extended depth of field (highlighted by the green boxes). It is worth noting that NLOS imaging with the curved surface suffered from secondary laser inter-reflections (i.e., laser reflections between the surface parts before incident onto the hidden objects) during the experiments, which caused the imaging artifacts in Fig. 3.4c, despite that the phasor field method is robust against multiple inter-reflections. This is primarily because the secondary laser reflection is much stronger than the inter-reflections of the weaker NLOS photons. Still, CLIP's capability to handle disconnected and curved surfaces is an important step to achieve point-and-shoot NLOS imaging in the field.

3.2.5 Discussion

The angular and temporal axes are two basic elements in the plenoptic function P(spatial: x, y, angular: u, v, wavelength: λ, time: t), which completely characterizes the fundamental properties of light. With decades' investigations and developments, current 3D imagers have reached the practical sensing limit about what is offered by measuring along the angular or temporal axis. A more complete acquisition of the plenoptic function (or the light fields), such as both angular and temporal dimension, towards versatile and more capable 3D vision remains a largely untapped area. This is mainly attributed to the lack of efficient schemes for sampling the resultant high-dimensional data with a 2D image sensor. However, for most applications, recording the entire high-dimensional light field is not the ultimate goal but an intermediate step to gain versatile image processing abilities, such as digital refocusing, extending the depth of field, or depth extraction. At the same time, expanding to the high-dimensional space promotes data sparsity[89], making full-rate acquisition unnecessary and

inefficient. Previous endeavors to sample light fields in an implicit and more efficient manner include the coded aperture[119] and wavefront-coding techniques. Unfortunately, their acquisition schemes for 3D imaging[114] or depth of field extension[115] require a densely sampled 2D image that precludes their applications in ultrafast, infrared[111,112], or Terahertz[120] imaging applications, where the detector resolution is severely limited. The CLIP framework encompasses and goes far beyond these methods to utilize sensors of arbitrary formats for efficient light field imaging (compressive or not), with a flexible nonlocal sampling strategy that promotes imaging robustness and better exploitation of the sparsity characteristic of high-dimensional data. An important application of canonical light field camera that remains to be explored by CLIP is to measure, and consequently correct for, optical aberrations as it can recover the full 4D light fields. Still, even without recovering the 4D light field, CLIP can readily correct for the Petzval field curvature owing to its refocusing capability, which can facilitate the coupling of planar sensors with monocentric systems[121,122] for wide-field or panoramic imaging.

CLIP's recording of a 4D light field is essentially an efficient dimensionality reduction in the optical domain, allowing high-dimensional information to be acquired with sensors of lower dimensionality such as the ubiquitous 1D or 0D (single pixel) detectors, which are still the dominant sensor format for imaging at the ultrafast time scale or the infrared, terahertz spectral regime. This feature facilitates CLIP to be deployed as a universal platform for snapshot multidimensional imaging[123] that samples the plenoptic function in a massively parallel manner. For instance, by extending the 1D ultrafast sensor in CLIP to 2D area detectors such as a megapixel SPAD, the extra spatial dimension could be readily used to measure the objects' spectra. Quantitative polarization information can also be extracted by overlaying the

pixels with a layer of polarizers, similar to utilizing a color filter array for color photography. The polarization cues have been routinely used for depth sensing and refinement[124,125], which can potentially further bolster CLIP's 3D vision performance.

3.3 Methods in CLIP

3.3.1 Experimental setup

A high dynamic range streak camera (C13410-01A, Hamamatsu Photonics) is used as the 1D array of time-of-flight sensors, which parallelly record the temporal signal within a single snapshot and a temporal resolution around 60 ps (for an observation window of 10 ns). For CLIP implementation with 1D sensor, seven customized plano-convex cylindrical lenslets (diameter of 2 mm and a focal length of 4 mm) are secured on a 3D printed holder and mounted on 3-axis translation stage to align them with the streak camera's entrance slit, which is effectively the virtual sensor plane. The cylindrical lenslets are rotated to different angles that are approximately uniform in the range of [-45o, 45o] to obtain incoherent measurements among different views. For optimal imaging through occlusions, the cylindrical lenslet angles are further randomly distributed along the lenslet array direction, as the effective measurement entries for the occluded objects are reduced to a subset of the measurement entry in the imaging model. Such random distribution maximizes statistically the incoherence among any subset of the measurements to ensure consistent image recovery performance. The imaging field of view is adjusted by tuning the translation stage that changes the distance between the lenslets and streak camera's slit. For transient illumintion, a femtosecond laser (808 nm, 7 mJ per pulse, ~100 fs pulse width, Astrella-F-1K, Coherent Inc.) is synchronized with the streak camera and trigged at a repetition rate of 100 Hz and an average power of 700 mW. An assembly comprised with a concave lens (LD1464-B, Thorlabs) and a diffuser (EDC-20-

14175-A, RPC Photonics) is switched by a step motor to diverge the laser beam for flooded illumination in flash LiDAR imaging. For NLOS imaging, the assembly is displaced for a collimated radiation.

3.3.2 Image reconstruction

Recovering a 4D light field or a refocused image from Eq. 18 and 19 can be inverted by solving a corresponding optimization problem:

$$argmin_{\hat{\mathbf{h}}} \ \|\mathbf{f} - \mathbf{F}(\mathbf{d})\mathbf{h}\|_2^2 + \ \mu\|\varphi(\mathbf{h})\|_1, \tag{20}$$

$$argmin_{\hat{\mathbf{P}}} \ \|\mathbf{f} - \mathbf{A}'\mathbf{P}\|_2^2 + \ \mu\|\varphi(\mathbf{P})\|_1, \tag{21}$$

where $\|.\|_1$ is the l_1 norm and $\varphi(\cdot)$ is a domain in which the 4D light field $\mathbf{P}$ (or image) is sparse. μ is a hyperparameter that balances the data fidelity and regularization term. In the framework of regularization by denoising algorithm[126], the representation domain $\varphi(\cdot)$ is not explicitly specified and the regularization step is implemented by a state-of-the-art image denoising algorithm such as BM3D or even a neural network. We adopted the BM3D and total variation (TV) denoisers for the regularization, owing to the existence of efficient algorithms[127]. Also, to minimize light field processing time that involves numerous refocusing steps (depth retrieval, extending depth of field), the reconstruction process at each step is initialized with the previous solution to exploit the proximity of the solutions.

It is worth noting that while recovering the 4D light field is always compressive in CLIP, directly retrieving a refocused image is not necessarily the same. Still, a major appeal of CLIP is to use a small number of sensors for recording a large-scale light field, which typically falls into the compressive sampling regime

3.3.3 Camera Calibration

To obtain quantitative and absolute 3D positions for flash LiDAR imaging, the CLIP camera is

metric-calibrated to extract its intrinsic matrix. A calibration pattern is made of a planar blackboard with a grid of 3×4 diffusing plates (circular white foam with a 10 mm diameter). With a femtosecond laser illumination, the CLIP camera captures the calibration pattern at ~ten different orientations, and the grid positions are automatically extracted in each image. The intrinsic matrix of the camera is then obtained by Zhang's calibration method[128]

3.3.4 Flash LiDAR and NLOS experiments

Room light was turned on during all experiments and the gating functionality of streak camera was enabled to makes it robust against ambient illuminations for time-of-flight imaging.

(1) Flash LiDAR coordinate transformation. With the camera being the origin (assuming roughly confocal illumination and detection), the flash LiDAR produces distance measurements in a polar coordinate, not the direct *z* components. The radial distance along different imaging pixels needs to be transformed into a rectilinear coordinate for correct 3D modelling. Given the camera's intrinsic matrix *K* and the homogeneous pixel coordinate [*u, v, 1*], each pixel's projection angle with respect to the camera's optical axis can be derived as:

$$\begin{bmatrix} tan\theta_x \\ tan\theta_y \\ 1 \end{bmatrix} = \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix} = K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \qquad (22)$$

where the absolute position of a point object is denoted as [x, y, z] in the rectilinear coordinate system.The LiDAR measurement distance r is related to the rectilinear coordinate as:

$$r = \sqrt{x^2 + y^2 + z^2} = z\sqrt{(x/z)^2 + (y/z)^2 + 1}\,. \qquad (23)$$

Combing Eq. 6 and 7, the absolute 3D position in the rectilinear coordinate is obtained as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix} = \frac{r}{\sqrt{(x/z)^2+(y/z)^2+1}} K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}. \qquad (24)$$

(2) Time-gain compensation of flash LiDAR signal. Unlike conventional scanning-based LiDAR that employs collimated laser illumination, the diverging light in flash LiDAR has a r^2 intensity decay, which makes distant objects to receive (and reflect) much fewer photons. We employed a time-gain compensation $(t\text{-}d)^2$ to partially equalize the signal intensity for objects positioned at a distance larger than d but shorter than d_2, after which the signals are too noisy for non-reflective objects. A customized compensation curve could also be used for the best visualization of the 3D scene, a common practice in medical ultrasound imaging.

(3) NLOS geometric calibration. The absolute 3D coordinates of the relay surface in NLOS imaging are measured via flash LiDAR in real-time. To extract the laser spot position on the wall and to cope with different walls in field applications that will modify the laser position thereon, the fixed propagation line of the laser is characterized in the absolute 3D space. The laser illumination spot on the surface is then dynamically calculated by intersecting the propagation line with the relay wall's 3D point cloud. To parameterize the propagation line, two thin microscopic slides are placed along the collimated beam path, with the tiny Fresnel reflection from the slides encoding two points that the laser passed through. The absolute positions of the two points are then measured using the LiDAR technique.

(4) Hybrid frequency-time domain NLOS reconstruction. Real-time NLOS imaging with arbitrarily-curved surfaces lacks an efficient solver. Recent developments of fast reconstruction algorithms[129] significantly reduced the computational cost to make real-time imaging feasible, but are exclusively limited to the paradigm of approximately planar surfaces. The pre-processing method presented by Lindell et. al. can transform the measurement data from a curved surface into a format suitable for fast frequency-domain solvers, but has a computational complexity of $o(N^5 logN)$, which is slightly more expensive than the universal

filtered backprojection algorithm and thus less appealing. Time-domain methods are more flexible in accommodating arbitrary wall geometries but more time-consuming. The hybrid frequency-time domain reconstruction method proposed here first converts the spatiotemporal measurement on a curved surface $y_r(\boldsymbol{r_p}, t)$ onto a virtual plane via wave propagation in time domain and then reconstruct the hidden scenes with existing efficient frequency-domain phasor field method[129]. Under the phasor-field framework, the spatiotemporal waveform on the virtual plane can be calculated in time-domain as:

$$f(r_v, t) = \int_{-w}^{w} y_r(\boldsymbol{r_p}, t) * p(t-\tau) d\boldsymbol{r_p}, \qquad (25)$$

where $\tau = \frac{\boldsymbol{r_v} - \boldsymbol{r_p}}{c}$ is the travel time from the point $\boldsymbol{r_p}$ on the curved surface to a point $\boldsymbol{r_v}$ on the virtual planar surface, and $p(t)$ is the convolutional kernel in the phase-field method. Wave migration in the time domain has two major advantages. First, it is more efficient than frequency domain migration: with both the curved and virtual plane being sampled with a spatial resolution of N^2, time-domain migration has a computational complexity of $o(N^4)$ instead of $o(N^5 logN)$. Second, it does not restrict the sampling pattern on either the curved surface or the virtual plane. While the frequency-domain method needs a spatial interpolation operation to deal with nonuniform sampling on a planar surface due to a 2D camera's projective distortion, time-domain migration can readily achieve a regular sampling on the virtual plane for the subsequent frequency-domain reconstruction[129]. Combined with the complexity of $o(N^3 logN)$ for the frequency-domain phasor field reconstruction, the total complexity for the hybrid time-frequency domain reconstruction is $o(N^4)$, still orders of magnitude faster than time domain methods. The memory complexity for the frequency-domain phasor field and relevant *f-k* migration reconstruction have been analyzed in the literature to be $o(N^3)$ and $\sim o(50N^3)$[52], respectively

We further accelerate the reconstruction on a GPU (Nvidia RTX3080Ti) using CUDA. For a 128×128×128 imaging volume with a spatiotemporal data cube of 125×125×1016, the NLOS reconstruction time is ~0.03 seconds, which can reach a 30 Hz video rate. The actual bottleneck lies in the iterative CLIP reconstruction of the spatiotemporal data cube on the wall, which takes about 2.0 seconds.

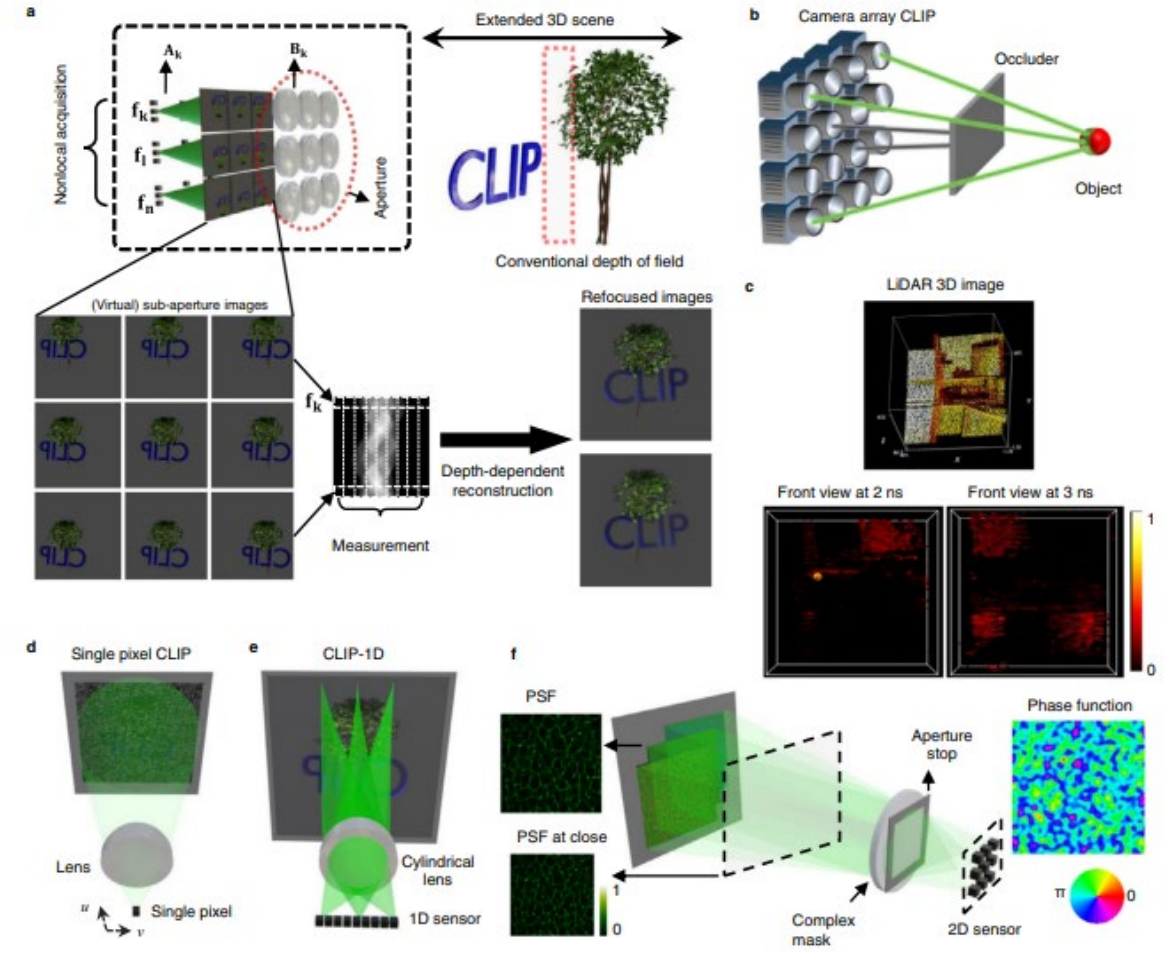

Fig. 3.1. Principle of compact light field photography[133]. a A conventional light field camera captures the scene from different views with a lens array and records all sub-aperture images. In contrast, CLIP records (operator A_k) only a few nonlocal measurements (f_k to f_n) from each sub-aperture image and exploits the depth-dependent disparity (modeled by B_k) to relate the sub-aperture images for gathering enough information to reconstruct the scene computationally. Refocusing is achieved by varying the depth-dependent disparity model B_k. b Seeing through severe occlusions by CLIP as a camera array, with each camera only recording partial nonlocal information of the scene. A obscured object (from the camera with black rays) remains partially visible to some other views (with green rays), whose nonlocal and complementary information enables compressive retrieval of the object. c Illustration of instantaneous compressibility of the time-of-flight measurements for a 3D scene in a flash LiDAR setup, where a transient illumination and measurement slice the crowded 3D scene along the depth (time) direction into a sequence of simpler instantaneous 2D images. d-f CLIP embodiments that directly perform nonlocal image acquisitions with a single-pixel, a linear array, and 2D area detectors,

respectively. A single pixel utilizes a defocused spherical lens to integrate a coded image, with u and v behind the lnes being the angular dimension. A cylindrical lens yields along its invariant axis a radon transformation of the en-face image onto a 1D sensor. The complex-valued mask such as a random lens produces a random, wide-field PSF that varies with object depth to allow light field imaging. PSF: point spread function; CLIP: compact light field photography; LiDAR: light detection and ranging. 1D,2D,3D: one, two, and three-dimensional.

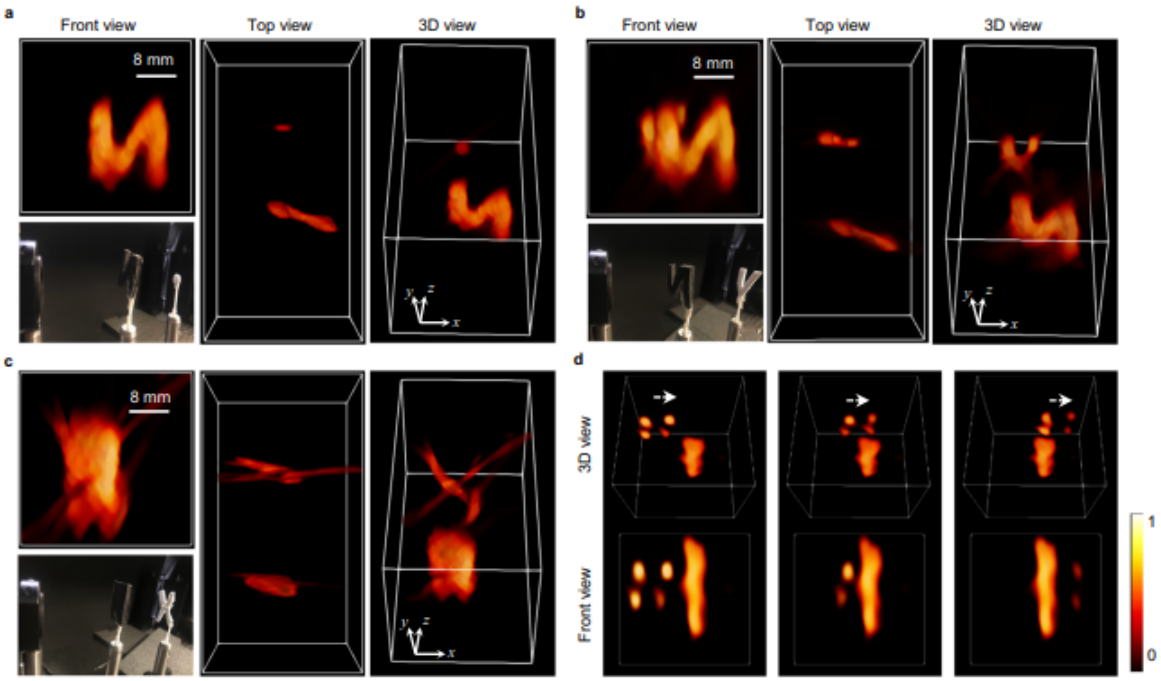

Fig. 3.2. Three-dimensional imaging (3D) through occlusions[133]. a-c Reconstructed 3D images rendered in different perspective for three scenes: circular plate (a) and letter V (b) behind the letter N, and letter X (c) blocked by a rectangular plate. The severe occlusions are evident from the front view images, with the larger objects in the front completely blocked the object right behind them. In contrast, CLIP is able to unambiguously reconstruct the obstructed objects in 3D without any defocusing signals from the preceding occluder. d Three representative frames of imaging a 2×2 grid pattern moving across the CLIP camera FOV behind a rectangular occluder. Note that signals from the black occluders are enhanced relative to the objects for better visualization.

Fig. 3.3. Snapshot flash LiDAR imaging over an extended depth range[133]. a Flash LiDAR imaging of a letter scene. From left to right are the reference photographs, a projected two-dimensional LiDAR images along the depth direction, and the 3D (three-dimensional) point-cloud representation of the scene. b flash LiDAR of the same 3D scene without extending the imaging depth of field, obtained by refocusing the camera onto a single focal plane. Note the defocus blur in the near and far objects. c Computational all-in-focus image. d-e Two representative frames for the dynamic imaging of a manually rotated letter V in a simple and cluttered scene, respectivel

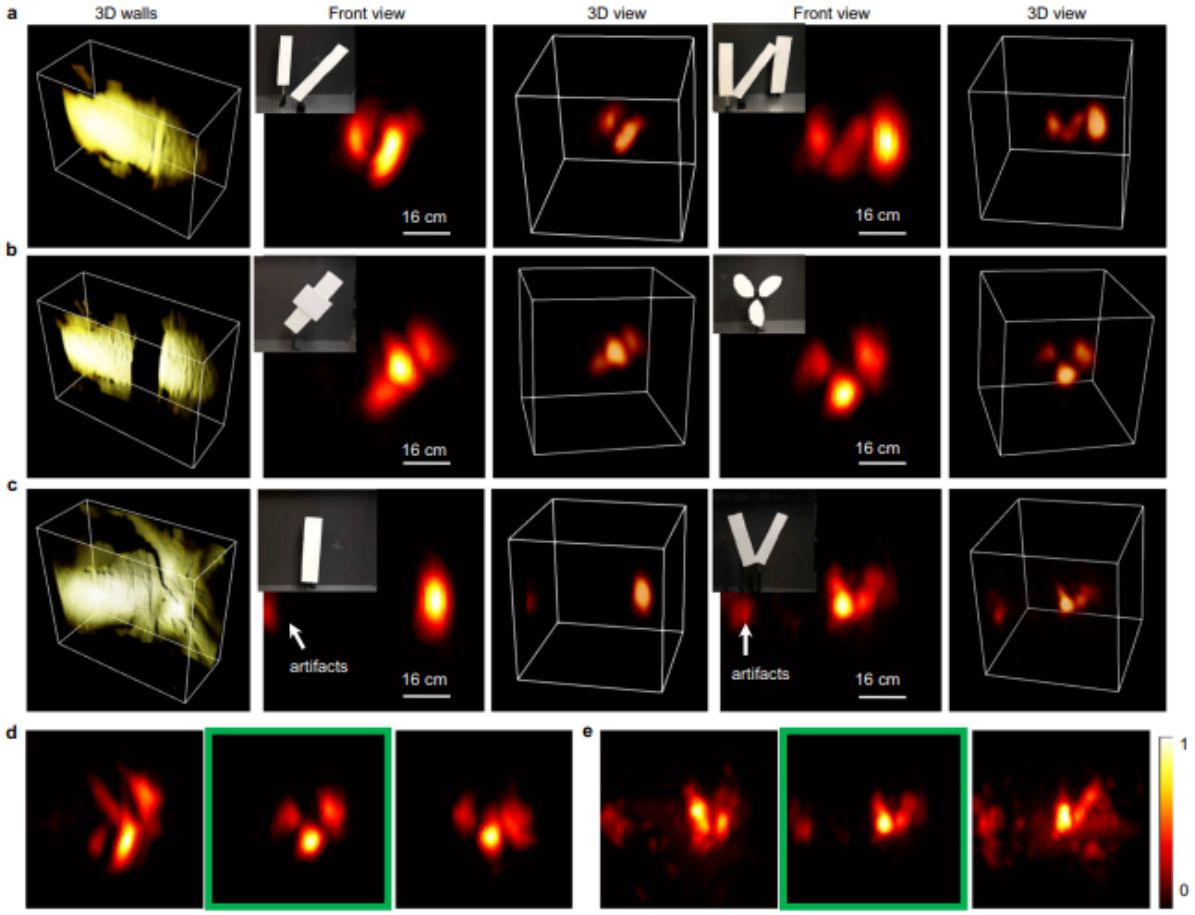

Fig. 3.4. NLOS imaging by CLIP-ToF.[133] a-c Imaging with planar, disconnected, and curved surfaces, respectively. From left to right are the flash LiDAR imaging of the relay surfaces, and two example hidden objects rendered as a projection image in the front view, and a 3D (three-dimensional) point cloud. Ground truth photographs of the object are shown in the inset of the front view image. d-e Reconstructed NLOS images for the disconnected and curved surfaces, respectively, with defocus errors on the relay wall, and those recovered with extended depth of field (highlighted by the green box). The quality of reconstruction degrades when the camera's extended depth of field is disabled.

Chapter 4 Light-field tomographic fluorescence lifetime imaging microscopy

Based on CLIP, light-field tomographic fluorescence lifetime imaging microscopy(LIFT-FLIM) is developed to solve the problems of compressed FLIM: lack of 3D imaging capability and high compression ratio. This approach allows for the acquisition of volumetric fluorescence lifetime images in a highly data-efficient manner, significantly reducing the number of scanning steps required compared to conventional point-scanning or line-scanning FLIM imagers. Moreover, LIFT-FLIM enables the measurement of high-dimensional data using low-dimensional detectors, which are typically low-cost and feature a higher temporal bandwidth. We demonstrated LIFT-FLIM using a linear single-photon avalanche diode array on various biological systems, showcasing unparalleled single-photon detection sensitivity. Additionally, we expanded the functionality of our method to spectral FLIM and demonstrated its application in high-content multiplexed imaging of lung organoids. LIFT-FLIM has the potential to open up new avenues in both basic and translational biomedical research.

4.1 Overview of LIFT-FLIM

The use of single-photon avalanche diode (SPAD) arrays in time-domain FLIM provides a solution to this long-standing problem by enabling parallel measurement of fluorescence decays at multiple image pixels. Moreover, SPAD arrays offer considerably greater sensitivity than conventional gated cameras, making them an excellent choice for low-light imaging applications. A SPAD imager can operate in either time gate or time-correlated single photon counting (TCSPC) mode, with TCSPC being the preferred detection method for its higher precision, faster speed, and greater sensitivity. However, the native fill factor of 2D SPAD arrays operating in the TCSPC mode is generally low (<10%) [130] due to the physical limitations posed by the inclusion of intricate timing electronics for each pixel. Although the addition of microlenses can recover some of the fill factor loss, this method is effective only

for collimated impinging light. In contrast, a linear SPAD array offers a significantly higher fill factor close to 50% [131], which results in a substantially increased light throughput. Moreover, the fabrication cost of a linear SPAD array is much lower than its 2D counterpart, making it more accessible for general labs. Nonetheless, when it comes to high-resolution imaging of a 2D or 3D scene using a linear SPAD array system, the challenge is much like that of point-scanning FLIM—the conventional approach involves scanning the entire field of view or volume using a vast number of steps, which can result in a protracted imaging duration.

To tackle the aforementioned challenges and streamline the acquisition of 3D FLIM data using a linear SPAD array, we devised a novel computational imaging technique called LIght-Field Tomographic Fluorescence Lifetime Imaging Microscopy (LIFT-FLIM). Our approach has been only recently made possible by an emerging technique, light field tomography (LIFT) [132,133], which is highly efficient in acquiring light field data for 3D imaging. Sharing its roots with light field photography [134-136], LIFT acquires multiple views of a 3D object and determines depth information through disparity analysis. However, rather than directly capturing a 2D perspective image, LIFT measures only the en-face projections of the image, thereby transforming 2D perspective images into lines. This allows us to map high-dimensional optical information to a low-dimensional space through pure optical operations. In LIFT-FLIM, we take advantage of this transformation by directly capturing 1D projection images using a linear SPAD array, allowing for 3D fluorescence lifetime imaging with exceptional single-photon sensitivity.

Aside from its impressive 3D imaging capability, LIFT also possesses inherent compatibility with spectral imaging [137]. Capitalizing on this feature, we showcased the system's versatility by extending its functionality to include spectral FLIM (sFLIM). We achieved this

by dispersing the 1D projection images utilizing a diffraction grating, and then feeding the resulting image into a time-gated camera for precise lifetime measurement. This allows for the simultaneous acquisition of 3D FLIM images at multiple wavelengths, making our system a versatile tool for analyzing both lifetime and spectral information. We demonstrated LIFT-FLIM and LIFT-sFLIM on various biological systems and showed their potential for high-content multiplexed imaging.

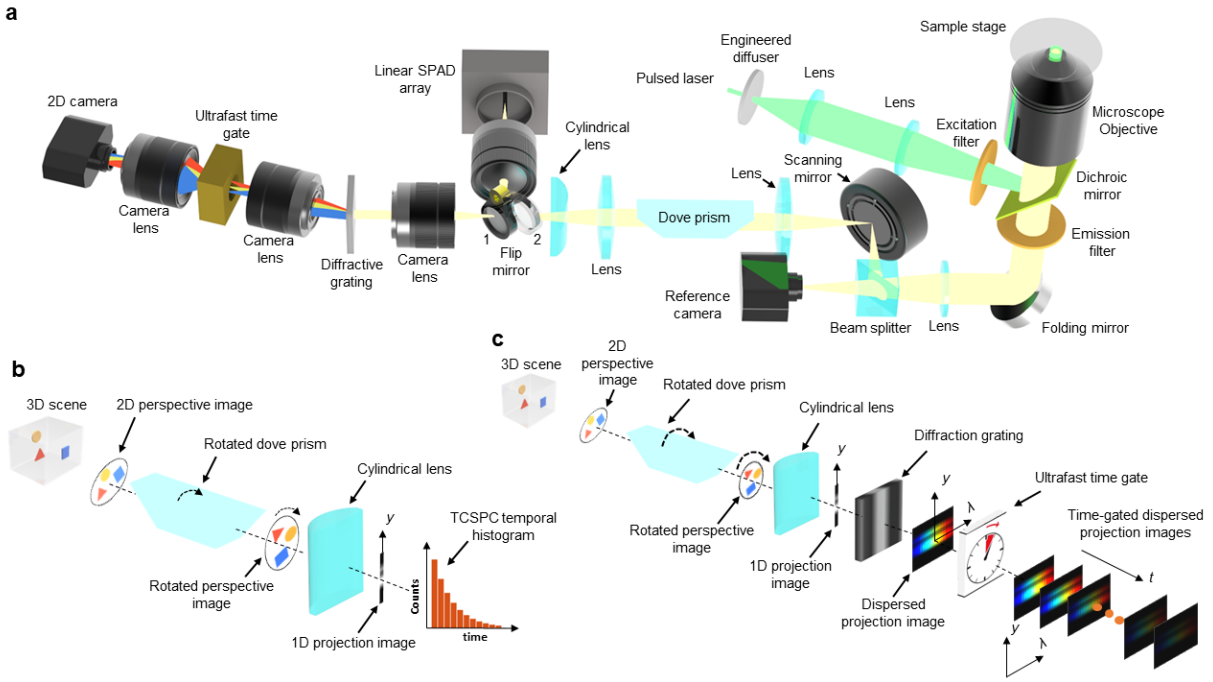

Fig.4.1. Optical setup and image formation models. a. System schematics. b. Image formation model of LIFT-FLIM. c. Image formation model of LIFT-sFLIM. SPAD, single-photon avalanche diode; TCSPC, time-correlated single photon counting.

4.2 Results

4.2.1 Operating principle and characterization

Based on computational imaging, LIFT-FLIM operates in two steps: data acquisition and image reconstruction. We show a LIFT-FLIM system in **Fig. 4.1a**. Upon pulsed laser excitation, the fluorescence is collected by a microscope objective lens with a high numerical aperture (NA) and forms an intermediate image at the microscope's side image port. A beam

splitter then divides the fluorescence into two beams. The transmitted light is recorded directly by a complementary metal-oxide-semiconductor (CMOS) camera, resulting in a reference intensity image. On the other hand, the reflected light creates an intermediate image on a scanning mirror. As the mirror tilts, it imparts twice its angle of tilt onto the outgoing rays. The reflected light from the scanning mirror is then collimated by a lens and forms a pupil image at a plane, where we position a dove prism. By adjusting the tilt angle of the mirror, we can shift the position of the pupil image on this plane. This enables us to selectively direct the light rays corresponding to a specific view angle through the dove prism, which will rotate the light rays and produce a rotated perspective image. Next, we use a cylindrical lens to compress the rotated perspective image into a line, which is essentially an en-face projection of the original perspective image along an orientation twice the rotation angle of the dove prism. This transformed line image can be either directly measured by a linear SPAD camera (**Fig. 4.1b**) or further spectrally dispersed and measured by a time-gated camera (**Fig. 4.1c**).

To compute a 3D image, conventional light field cameras require the acquisition of a comprehensive set of spatial and angular information of a light field, resulting in a significant data load. However, our previous work demonstrated that this acquisition method is inefficient and generates a substantial amount of redundant data [133], which can be reduced by distributing a nonlocal image acquisition process, such as en-face projection measurement, into different views. Moreover, this allows for the measurement of a high-dimensional light field using low-dimensional detectors, which are typically low-cost and feature a higher temporal bandwidth. In LIFT-FLIM, we leverage this advantage and capture only en-face 1D projection images at each scanned sub-pupil location. Furthermore, we can create an arbitrarily shaped off-focus point spread function (PSF) by strategically shuffling the orientation angles

of our projection measurements across various views in a programmed manner

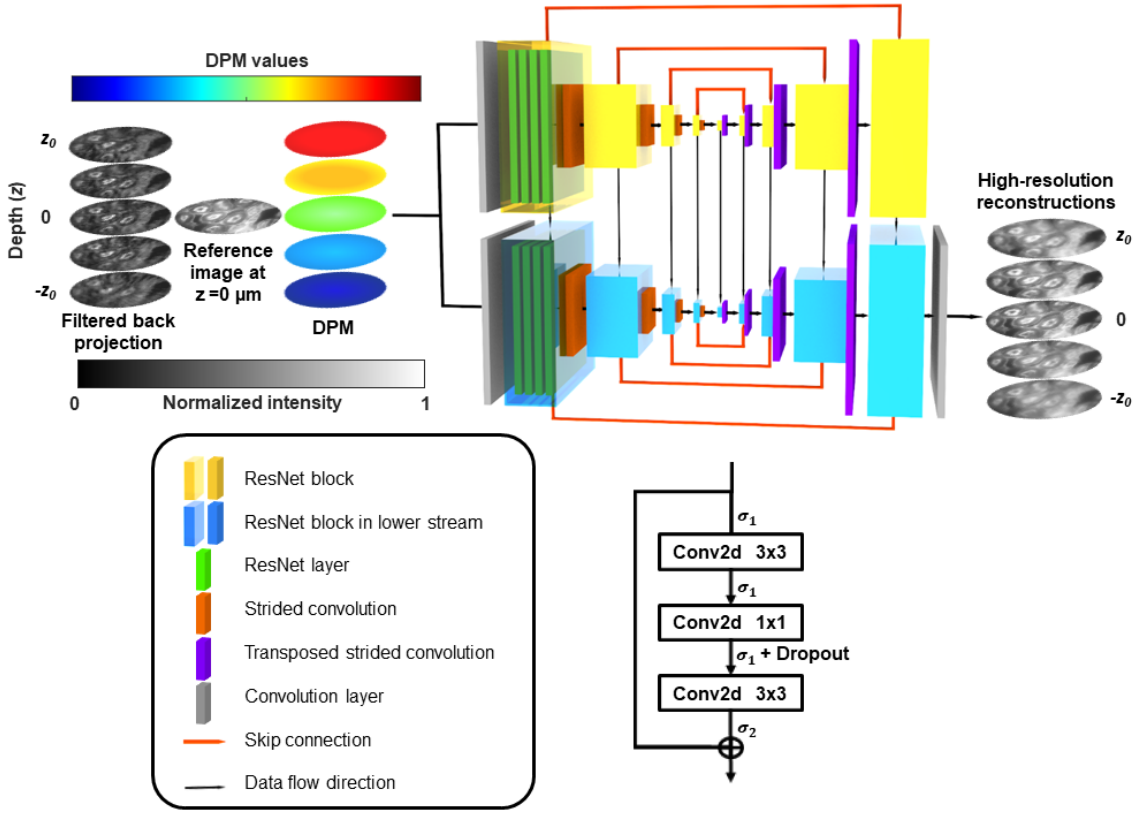

Fig. 4.2. Deep-learning-based image enhancement neural network. The network consists of two down- and up-sampling streams. Each stream has five ResNet blocks in both down-sampling and up-sampling paths. Each ResNet block contains four ResNet layers, and each ResNet layer has two 3×3 convolutional layers and one 1×1 convolutional layer, as indicated in the bottom right panel. Strided convolutional layers were added between the two adjacent ResNet blocks to halve the spatial dimensions in the down-sampling path, and conversely transposed strided convolutional layers were utilized to implement up-sampling in the up-sampling path. The spatial dimensions of the ResNet blocks in the sampling streams from left to right are 256×256, 128×128, 64×64, 32×32, 16×16, 32×32, 64×64, 128×128, 256×256. The central 16×16 ResNet blocks are shared by the down- and up-sampling streams. Skip connections connect each ResNet block in the down-sampling path with its counterpart block in the up-sampling path. The inputs to the network include LIFT refocused depth image stack using filtered back projection from depth $-z_0$ to depth z_0, reference image captured at depth zero, and a DPM stack. The output is a high-resolution image stack at the corresponding depths. DPM: digital propagation matrix. σ_1, σ_2: activation functions. Conv2d: convolution 2D.

We formulate the image formation of LIFT-FLIM using a linear model. For a given perspective image P_k at view k ($k = 1, 2, \ldots, K$), the projection measurement along angle θ is

$$\mathrm{f}_k^{\theta} = \boldsymbol{T}\boldsymbol{R}^{\boldsymbol{\theta}}\mathrm{P}_k, \tag{26}$$

where P_k has a dimension of N^2 (N is the image dimension in pixels), $\boldsymbol{T}$ is the en-face projection operator, and $\boldsymbol{R}^{\boldsymbol{\theta}}$ is the image rotation operator, which describes the function of the dove prism rotated at $\theta/2$.

Rather than capturing a complete set of $N_\theta = N$ projection angles at each view, we acquire only a subset of n_k projection angles at view k. This process can be explicitly written as:

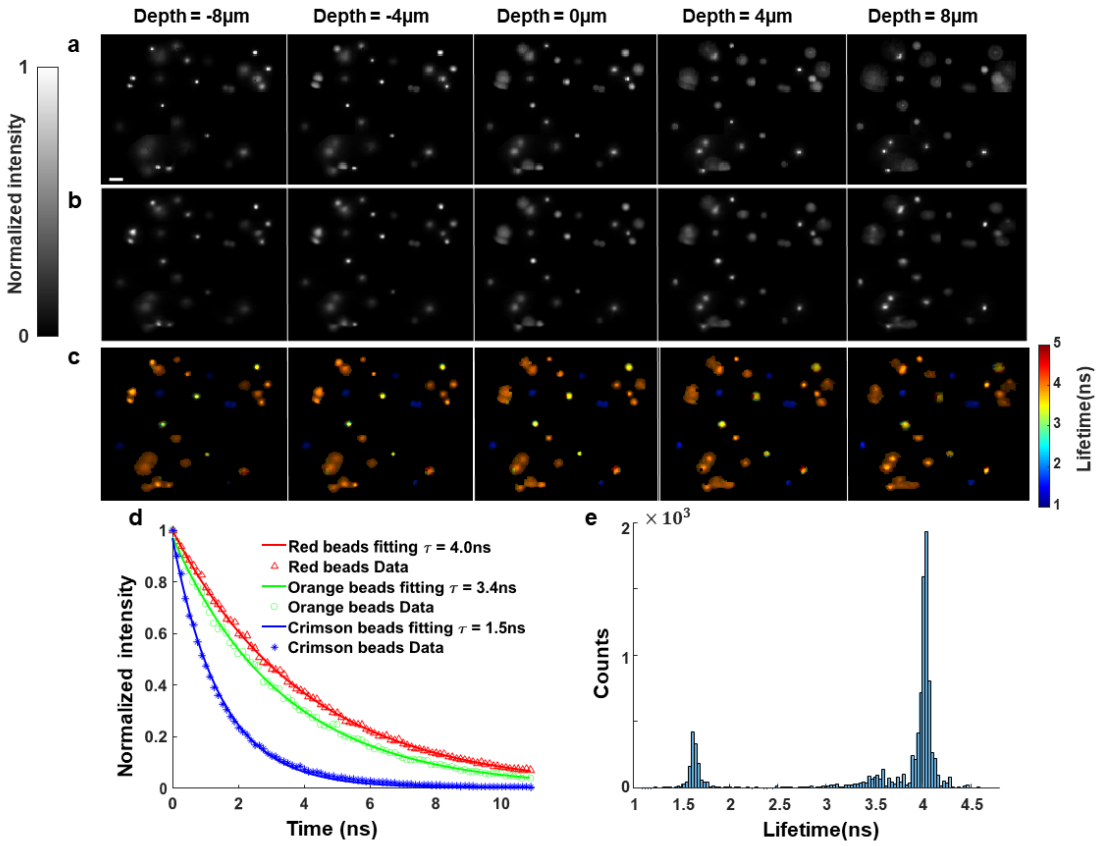

Fig. 4.3. LIFT-FLIM of mixed fluorescent beads. a. Reference intensity images at depths of -8μm, -4μm, 0μm, 4μm, and 8μm. b. Time-integrated LIFT-FLIM images at the corresponding depths. c. Lifetime images at the corresponding depths. d. Fluorescence decay curves at representative beads' locations. e. Histogram of pixel lifetimes at depth zero. Scale bar: 20μm.

$$
\mathrm{f} = \begin{bmatrix}
view\_1 \begin{Bmatrix} \mathrm{f}_1^{\theta_1} \\ \mathrm{f}_1^{\theta_2} \\ \vdots \\ \mathrm{f}_1^{\theta_{n_1}} \end{Bmatrix} \\
\vdots \\
view\_k \begin{Bmatrix} \mathrm{f}_k^{\theta_{s_{k-1}+1}} \\ \mathrm{f}_k^{\theta_{s_{k-1}+2}} \\ \vdots \\ \mathrm{f}_k^{\theta_{s_{k-1}+n_k}} \end{Bmatrix} \\
\vdots \\
view\_K \begin{Bmatrix} \mathrm{f}_K^{\theta_{s_{K-1}+1}} \\ \mathrm{f}_K^{\theta_{s_{K-1}+2}} \\ \vdots \\ \mathrm{f}_K^{\theta_{s_{K-1}+n_K}} \end{Bmatrix}
\end{bmatrix}
$$

$$
= \mathbf{T} \begin{bmatrix}
view\_1 \begin{Bmatrix} \boldsymbol{R}^{\boldsymbol{\theta_1}} \\ \boldsymbol{R}^{\boldsymbol{\theta_2}} \\ \vdots \\ \boldsymbol{R}^{\boldsymbol{\theta_{n_1}}} \end{Bmatrix} & \cdots & \mathbf{0} \\
\vdots & view\_k \begin{Bmatrix} \boldsymbol{R}^{\boldsymbol{\theta_{s_{k-1}+1}}} \\ \boldsymbol{R}^{\boldsymbol{\theta_{s_{k-1}+2}}} \\ \vdots \\ \boldsymbol{R}^{\boldsymbol{\theta_{s_{k-1}+n_k}}} \end{Bmatrix} & \vdots \\
\mathbf{0} & \ddots & \mathbf{0} \\
\mathbf{0} & \cdots & view\_K \begin{Bmatrix} \boldsymbol{R}^{\boldsymbol{\theta_{s_{K-1}+1}}} \\ \boldsymbol{R}^{\boldsymbol{\theta_{s_{K-1}+2}}} \\ \vdots \\ \boldsymbol{R}^{\boldsymbol{\theta_{s_{K-1}+n_K}}} \end{Bmatrix}
\end{bmatrix} \begin{bmatrix} \mathrm{P}_1 \\ \mathrm{P}_2 \\ \vdots \\ \mathrm{P}_K \end{bmatrix}
+ \sigma \qquad (27)
$$

$$
= \begin{bmatrix} \boldsymbol{A_1} & \cdots & \mathbf{0} \\ \vdots & \boldsymbol{A_2} & \vdots \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \boldsymbol{A}_K \end{bmatrix} \begin{bmatrix} \mathrm{P}_1 \\ \mathrm{P}_2 \\ \vdots \\ \mathrm{P}_K \end{bmatrix} + \sigma.
$$

Here f is a stack of projection measurements (also referred to as a sinogram), and it has a dimension of $N \times N_\theta$. $s_{k-1} = \sum_{i=1}^{i=k-1} n_i$, and $s_{K-1} + n_K = N_\theta . \sigma$ denotes the measurement noise. $\boldsymbol{A_k}$ is a combined function of en-face projection and image rotation operators on perspective image P_k. Because the images from different views capture the same scene, they share a common underlying content with only a depth-dependent disparity between any two sub-aperture images. Therefore, the correlation between sub-aperture images can be modeled by digitally propagating the light field, *i.e.* the sub-aperture image P_k at view k can be related to a depth-dependent image feature kernel $h(d)$ through an invertible shearing operator $\boldsymbol{B_k}$ as $P_k = \boldsymbol{B_k}(d)h(d)$, where $\boldsymbol{B_k}$ is also a function of depth d Accordingly, we transform **Eq. 27** to:

$$\begin{aligned} f &= \begin{bmatrix} \boldsymbol{A_1} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{A_2} & \vdots \\ \boldsymbol{0} & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{A_K} \end{bmatrix} \begin{bmatrix} \boldsymbol{B_1}(d)h(d) \\ \boldsymbol{B_2}(d)h(d) \\ \vdots \\ \boldsymbol{B_K}(d)h(d) \end{bmatrix} + \sigma = \begin{bmatrix} \boldsymbol{A_1 B_1}(d) \\ \boldsymbol{A_2 B_2}(d) \\ \vdots \\ \boldsymbol{A_K B_K}(d) \end{bmatrix} h(d) + \boldsymbol{\sigma} \\ &= \boldsymbol{F}(d)h(d) + \sigma \end{aligned} \tag{28}$$

The overall image forward operator $\boldsymbol{F}(d)$ becomes a function of depth d, which is essential for recovering image $h(d)$ with various focal settings. Noteworthily, although individual P_k is measured at only a subset of projection angles, the underlying image feature kernel $h(d)$ is measured on a complete angular basis, as $\boldsymbol{F}(d)$ concatenates image rotation operators across all views.

For direct fluorescence lifetime measurement using a linear SPAD array with TCSPC (**Fig. 4.1b**), the image formation model is a time-lapse version of **Eq. 28**, which can be expressed

as:

$$\mathrm{f}(t) = \boldsymbol{F}(d)h(d,t) + \sigma(t), \tag{29}$$

where $\mathrm{f}(t)$ is a time-lapse sinogram constructed by the projection measurements at the time bin t of a TCSPC temporal histogram.

For spectral FLIM measurement using a gated ultrafast camera (**Fig. 4.1c**), the image forward model is a function of both time t and wavelength λ:

$$\mathrm{f}(t,\lambda) = \boldsymbol{F}(d)h(d,t,\lambda) + \sigma(t,\lambda), \tag{30}$$

where $\mathrm{f}(t,\lambda)$ is a spectrally resolved, time-lapse sinogram constructed by the projection measurements at the gated time t and wavelength λ.

The image reconstruction of LIFT-FLIM and -sFLIM involves solving the inverse problems of **Eq. 29** and **30**, respectively. Like standard computed tomography, this can be accomplished through simple inverse Radon transform or more advanced optimization algorithms like a Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [138,139]. We depict the workflow for processing the light field data, such as image refocusing, extending the depth of field, and rendering a 3D image in **Methods**.

Akin to conventional light field cameras, LIFT-FLIM and -sFLIM divide the aperture to extract the depth information. Therefore, they have a reduced lateral resolution (~1.8 μm) compared with the native diffraction-limited resolution of the objective lens. To improve the quality of the reconstructed images, we developed a deep-learning-based image enhancement neural network [140-142] (**Fig. 4.2**). The input to the neural network consists of reconstructed LIFT depth images, a diffraction-limited reference image captured at the depth zero, and digital propagation matrices (DPMs), which represent the axial distance from the reference image plane to the target plane on a per-pixel basis [143]. The neural network then uses a

PixelCNN++ architecture [144] to generate high-resolution outputs at corresponding depths. The axial resolution, determined by the NA of the objective lens and the number of views acquired, was measured to be ~3.0 μm for point objects. The temporal resolutions of LIFT-FLIM and -sFLIM depend on the characteristics of the image sensor. For example, using a linear SPAD array provides a temporal resolution of 50 ps with TCSPC [145], whereas using a gated ultrafast camera yields a temporal resolution of 70 ps [146].

4.2.2 LIFT-FLIM of mixed fluorescent beads

We validated the 3D lifetime imaging performance of LIFT-FLIM on fluorescent beads. We mixed three types of fluorescent beads with lifetimes of 1.5 ns, 3.4 ns, and 4.0 ns, respectively, in an agarose gel. We simultaneously excited the beads using a filtered supercontinuum laser and imaged the fluorescence using LIFT-FLIM with a linear SPAD array. Moreover, we captured the ground-truth intensity images (**Fig. 4.3a**) at depths from -8 μm to 8 μm using a reference camera by mechanically scanning the microscope's focus.

To compare LIFT-FLIM images with the ground-truth images obtained, we summed the signals at all time bins in the TCSPC temporal histogram and reconstructed the time-integrated images at the corresponding depths (**Fig. 4.3b**). The resulting images exhibit a high degree of similarity to the ground-truth images, demonstrating the system's numerical refocusing ability. We further generated the time-lapse LIFT-FLIM images and computed the average lifetime at each image pixel using mono-exponential curve fitting (**Fig. 4.3c**). Three representative fluorescence decays reconstructed at beads' locations are shown in **Fig. 4.3d**. The derived fluorescence lifetimes are consistent with the beads' specifications. Furthermore, we generated a histogram of the lifetimes of all pixels at depth zero (**Fig. 4.3e**). This histogram displays three distinct peaks that correspond to the lifetimes of three different types of fluorescence

beads. This observation reinforces the reliability and accuracy of our lifetime measurement.

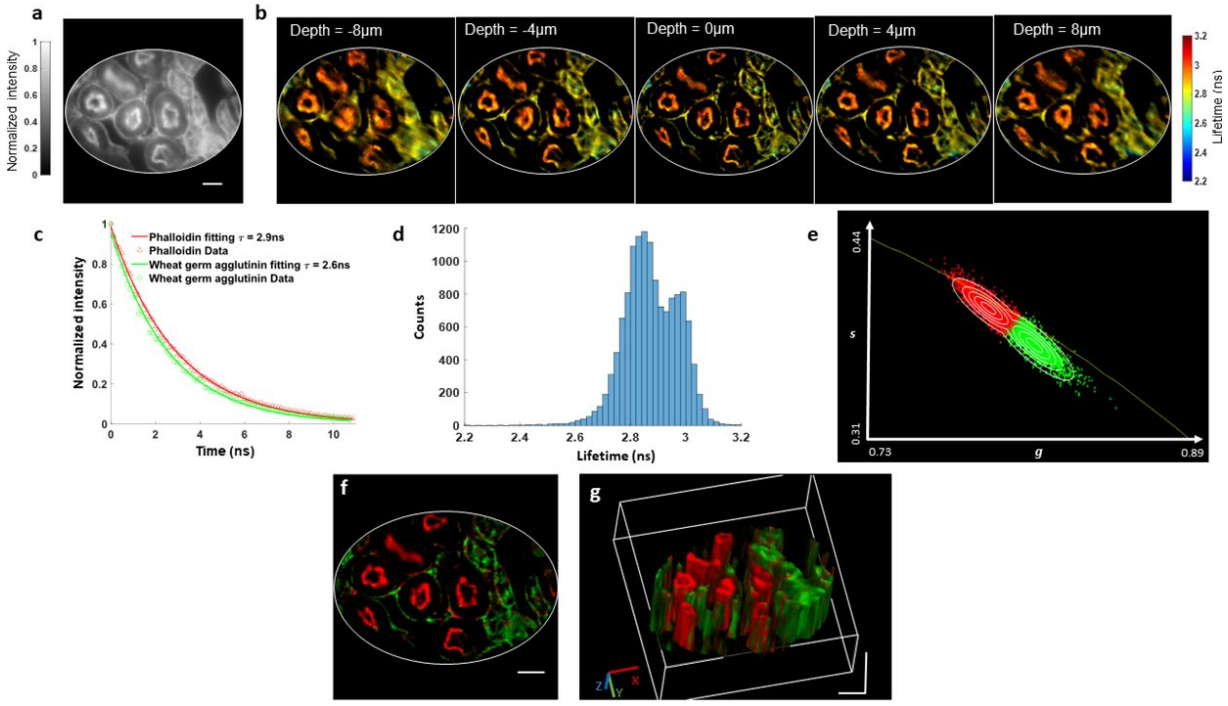

Fig. 4.4. LIFT-FLIM of a mouse kidney tissue section. a. Reference intensity image at depth zero. b. Reconstructed lifetime images at depths of -8μm, -4μm, 0μm, 4μm, and 8μm. The refocusing to continuous depths is visualized in Movie 2. c. Fluorescence decay curves at two representative fluorophore locations. d. Histogram of pixel lifetimes at depth zero. e. Phasor plot. The data points were pseudocolored based on its probability belonging to a specific cluster (Red, Phalloidin; Green, WGA). The probability contour lines ranging from outer to inner space correspond to values of 0.1, 0.3, 0.5, 0.7, and 0.9. f. Unmixed fluorophore image at depth zero. Red channel, Phalloidin. Green channel, WGA. g. 3D visualization of unmixed fluorophores' distribution. Scale bars in all figures: 20 μm.

4.2.3 LIFT-FLIM of a mouse kidney tissue section

We tested LIFT-FLIM on a standard biological sample (a mouse kidney section, FluoCells™ Prepared Slide from ThermoFisher) and demonstrated its ability in lifetime unmixing. The sample was stained with Alexa Fluor 488 wheat germ agglutinin (WGA) for labeling cell membrane and Alexa Fluor 568 phalloidin for labeling filamentous actin (F-actin). These two fluorophores have distinct but close fluorescence lifetimes (Alexa Fluor 488, 2.6 ns *vs.* Alexa

Fluor 568, 2.9 ns). The fluorescence intensity image captured at the focal plane by the reference camera is shown in **Fig. 4.4a**. The refocused LIFT-FLIM fluorescence lifetime images at representative depths are displayed in **Fig. 4.4b**. **Figure 4.4c** shows the fluorescence decay curves measured at two fluorophore locations, and **Fig. 4.4d** shows the histogram of all pixels' lifetimes at depth zero, where the two peaks indicate the two underlying lifetime components.

Next, we applied an unsupervised phasor approach [147,148] to the fluorescence lifetime data and calculated the probability of each pixel belonging to a specific lifetime component cluster. **Figure 4.4e** displays the phasor plot for the fluorescence lifetime image at depth zero, with each data point color-coded to represent its corresponding probability and overlaid with probability contour lines. We then classified the image pixels in the time-integrated LIFT-FLIM image based on this probability and unmixed the fluorophores into pseudo-colored channels. **Figure 4.4f** shows a representative unmixed image at depth zero (red channel, phalloidin; green channel, WGA). Repeating this procedure for all depths yields a 3D unmixed image, as shown in **Fig. 4.4g**.

4.2.4 LIFT-FLIM of a human lung cancer pathology slide

We demonstrated LIFT-FLIM in autofluorescence imaging of an unstained human lung cancer pathology slide. Previous studies show that FLIM can access tumor metabolism by imaging endogenous chromophores such as NAD(P)H and FAD, enabling its application in cancer diagnosis and intraoperative surgical guidance [149,150]. Particularly in pathological imaging, FLIM holds great promise as an alternative approach for label-free detection of tissue lesions [151,152]. However, conventional FLIM microscopes with a high collecting NA suffer from a shallow depth of field. When imaging a panoramic FOV through multiple captures and

stitching, the system must mechanically adjust its focus at each position to correct for potential focal drift that can occur during extensive scanning, complicating the imaging procedure. Here we show that, by using numerical refocusing, LIFT-FLIM enables an extended depth of field and allows for capturing an all-in-focus image without the need for accounting for the focal drift.

We excited the sample at 450nm and collected the autofluorescence in the range of 490-700 nm. The primary endogenous fluorophore that accounts for the fluorescence emission at this wavelength is flavin adenine dinucleotide (FAD). To image a large FOV, we scanned the sample and stitched the images. The resultant fluorescence intensity image captured by the reference camera is shown in **Fig. 4.5a**, where certain parts of the FOV are blurred due to the focal drift. In contrast, LIFT-FLIM can numerically correct for this defocus error in post-processing and form an all-in-focus image, as shown in **Fig. 4.5b**. For quantitative comparison, we plotted signal intensities along a dashed line in **Fig. 4.5a-b** and show the results in **Fig. 4.5c**. The image features appear to have much sharper edges in LIFT-FLIM compared to those captured by the reference camera (~36% reduction in full-width at half maximum). Next, we computed the lifetimes for the stitched all-in-focus LIFT image and presented a lifetime map in **Fig. 4.5d**. A zoom-in area (**Fig. 4.5e**) shows a significant level of lifetime heterogeneity. To correlate this observation to the tissue state, we stained an adjacent slide from the same tissue sample using standard hematoxylin and eosin (H&E) and imaged it under a widefield microscope. After the histological image was obtained, a pathologist

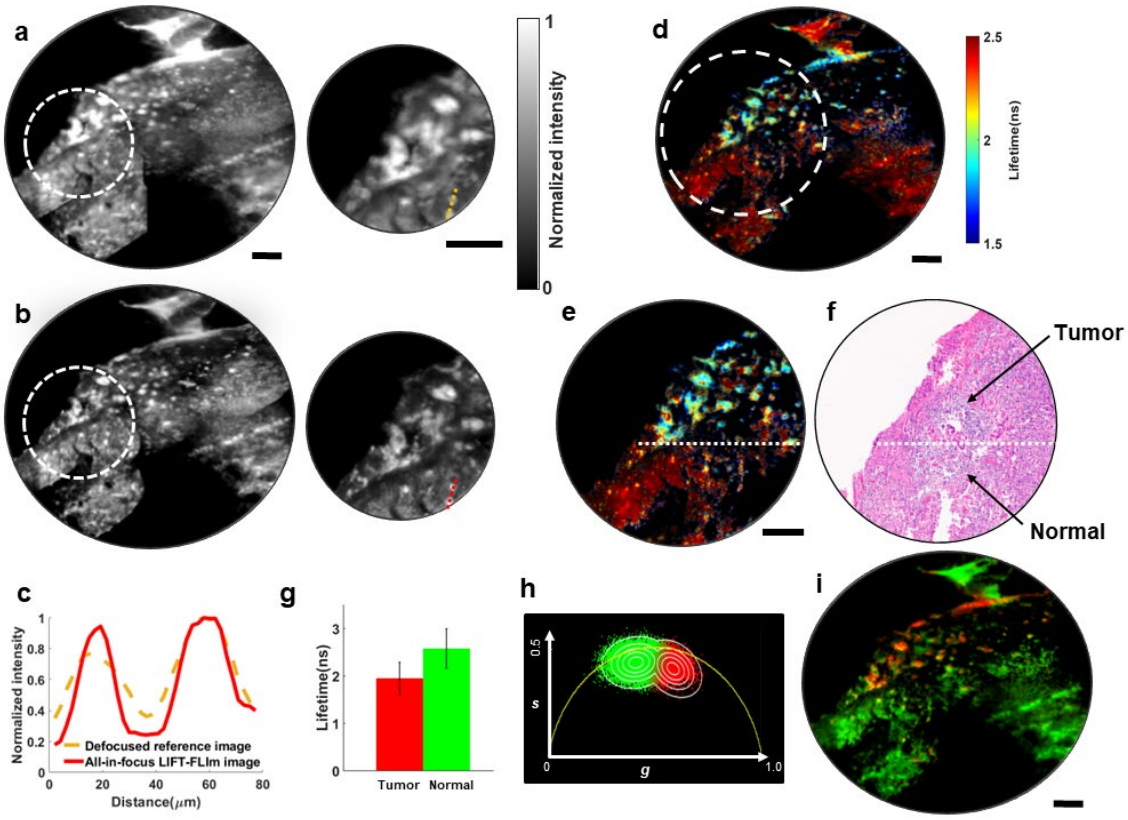

Fig. 4.5. LIFT-FLIM of a human lung cancer pathology slide. a. Left panel: stitched reference intensity image. Right panel: zoom-in image of the circled area in the left panel. The image is blurred due to focal drift during extensive scanning. b. Left panel: stitched all-in-focus time-integrated LIFT-FLIm image. Right panel: zoom-in image of the circled area in the left panel. c. Intensity profiles of dashed lines in a and b. d. Stitched all-in-focus lifetime image. The lifetime image is masked with an intensity threshold e. Zoom-in image of the circled area in d. f. Hematoxylin and eosin (H&E) stained image from an adjacent tissue slice. The tumor/normal tissue boundary was identified by a pathologist and annotated with a white dashed line. g. Average pixel lifetimes in the tumor and normal tissues areas in e. The standard deviation (SD) is shown as error bars. h. Phasor plot. The data points were pseudocolored based on its probability belonging to a specific cluster (Red, tumor; Green, normal). The probability contour lines ranging from outer to inner space correspond to values of 0.1, 0.3, 0.5, 0.7, and 0.9. i. Classified tissue map. Red channel, tumor; Green channel, normal. Scale bars in all figures: 100 μm.

reviewed it to identify the boundary between the tumor and normal tissue, as illustrated in **Fig. 4.5f**. Comparing the average pixel lifetimes above (1.9±0.3 ns) and below (2.6±0.4 ns) the annotated boundary (**Fig. 4.5f**) reveals a significant difference (**Fig. 4.5g**). The observed reduction in autofluorescence lifetimes in the tumor areas compared to that in the normal tissue is consistent with previous reports [153-155] and may indicate a shift towards glycolysis and cancer metabolism [156]. To classify the tissue based on the lifetime, we again applied an unsupervised phasor approach to the fluorescence lifetime data. The resultant phasor plot and

classified tissue map are shown in **Fig. 4.5h** and **4.5i**, respectively (red channel, tumor; green channel, normal tissue).

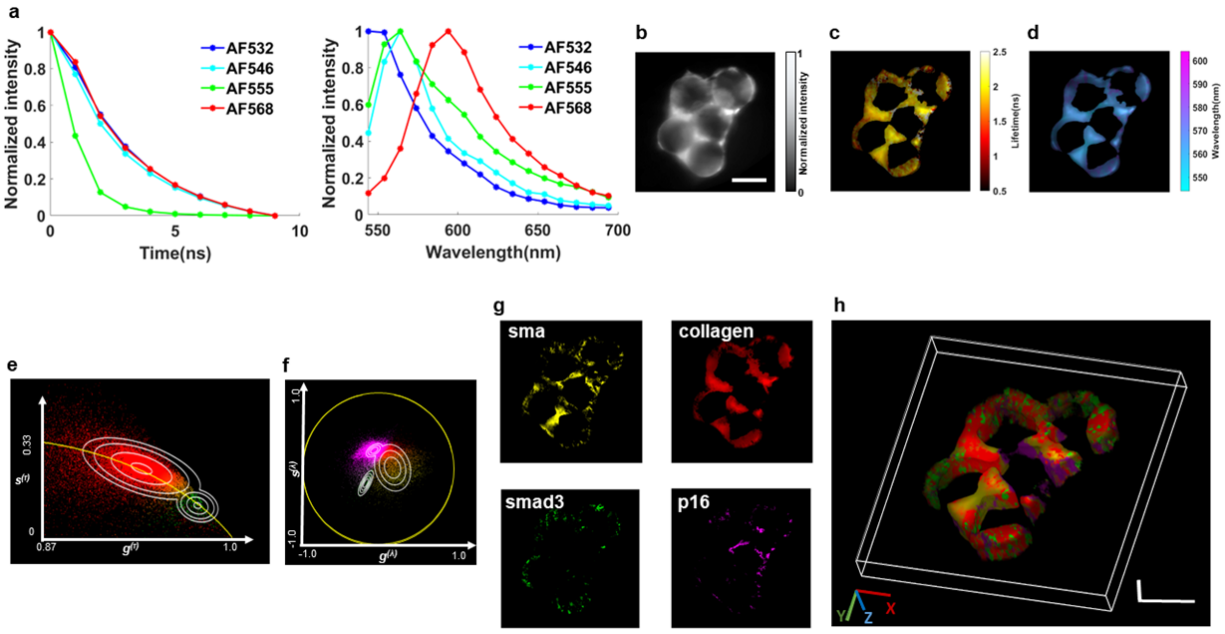

Fig. 4.6. LIFT-FLIM of lung organoids. **a**. Fluorescence decay curves (left panel) and emission spectra (right panel) of the fluorophores used. **b-d.** Reconstructed LIFT-sFLIM (**b**) intensity image, (**c**) wavelength-integrated lifetime image, and (**d**) time-integrated wavelength image at depth zero**. e.** Lifetime phasor plot. The data points were pseudocolored based on its probability belonging to a specific cluster (Red, sma, collagen, p16; Green: smad3). The probability contour lines ranging from outer to inner space correspond to values of 0.2, 0.35, 0.65, and 0.95. **f.** Spectral phasor plot. The data points were pseudocolored based on its probability belonging to a specific cluster (Magenta: collagen, smad3; Yellow: sma; Green, p16). The probability contour lines ranging from outer to inner space in the magenta and yellow clusters correspond to values of 0.45, 0.65, and 0.85, while the contour lines in the green cluster correspond to values of 0.25, 0.45, 0.65, and 0.85. **g.** Unmixed component images at depth zero. **h.** 3D visualization of unmixed fluorophores' distribution in the organoid.. Scale bar: 100μm in all figures.

4.2.4 LIFT-sFLIM of lung organoids

We demonstrated LIFT-sFLIM in 3D multiplex imaging of lung organoids. Organoids, 3D multicellular stem-cell-derived constructs that mimic *in vivo* tissue, have gained growing interest for modeling tissue development and disease [157-159]. Particularly, organoids hold great promise for high-content phenotypic screening because they recapitulate many aspects of parent tissues and can be derived from patient material, rendering them ideal model systems for personalized medicine and drug discovery [160-164].

One primary challenge for high-content phenotypic screening of organoids is extraction of multivariate information from organoids labeled with multiple biomarkers [165-167]. Here we show that, by acquiring both the spectral and lifetime information, LIFT-sFLIM provides a powerful solution to overcome this challenge. We cultured lung alveolar organoids with different combinations of primary healthy human lung fibroblasts and epithelial cells grown on alginate scaffolds that mimic the alveolar micro-architecture [168]. We used the antibodies and labeled epithelial-mesenchymal transition by α smooth muscle actin (α-sma) expression, ECM deposition by collagen (collagen I) expression, cell apoptosis by SMAD signaling pathway (smad3), and cellular senescence by $P16^{INK4A}$ (p16) expression.

Figure 4.6a depicts the fluorescence emission decay curves and spectra of the four fluorophores that were utilized in the secondary antibodies. While the fluorophores AF 532, 546, and 568 have close fluorescence lifetimes, their spectral emission peaks are well separated. On the other hand, AF 546 and AF 555 exhibit significant spectral overlaps but differ in fluorescence lifetimes. The combination of four fluorophores used in this study presents a challenge for conventional imaging techniques. Specifically, neither FLIM nor spectral imaging alone can simultaneously capture and distinguish all four fluorophores. This

limitation underscores the need for innovative imaging approaches, such as LIFT-sFLIM, which can integrate both spectral and temporal information to enable reliable separation and quantification of multiple fluorophores in complex biological samples.

Using LIFT-sFLIM, we acquired a five-dimensional (5D) dataset (x, y, z, t, λ) (x, y, z, spatial coordinates; t, fluorescence decay time; λ, wavelength). **Figure 4.6b** shows the LIFT-sFLIM reconstructed intensity image at depth zero. The wavelength-integrated lifetime image and time-integrated wavelength image at depth zero are shown in **Fig. 4.6c-d**, respectively. To unmix the fluorophores, we applied a spectral-lifetime phasor approach to the 5D dataset. The resultant color-coded phasor plots in the lifetime and spectral domains are shown in **Fig. 4.6e-f**, respectively. Consistent with the spectral and lifetime data presented in **Fig. 4.6a**, our analysis revealed two distinct clusters in the lifetime phasor plot and three distinct clusters in the spectral phasor plot. By combining spectral and temporal information, we separated the fluorophores into four color-coded channels. Representative images at depth zero are shown in **Fig. 4.6g**. By repeating this procedure at all depths, we generated a 3D color-coded image that depicts the distribution of each fluorophore in the organoid, as shown in **Fig. 4.6h**.

4.3 Discussion

Using LIFT-FLIM for 3D lifetime imaging offers a crucial benefit of reducing the number of scanning steps required compared to traditional point- or line-scanning time-domain FLIM techniques. To produce a 3D image of $N_x \times N_y \times N_z$ voxels, a FLIM system that uses point- or line-scanning requires a total of $N_x \times N_y \times N_z$ or $N_y \times N_z$ (if line scans are done along the y axis) scanning steps, respectively. Here, N_x, N_y, and N_z denote the number of spatial samplings in a 3D space. For simplicity, we consider $N_x = N_y = N$. In contrast, because

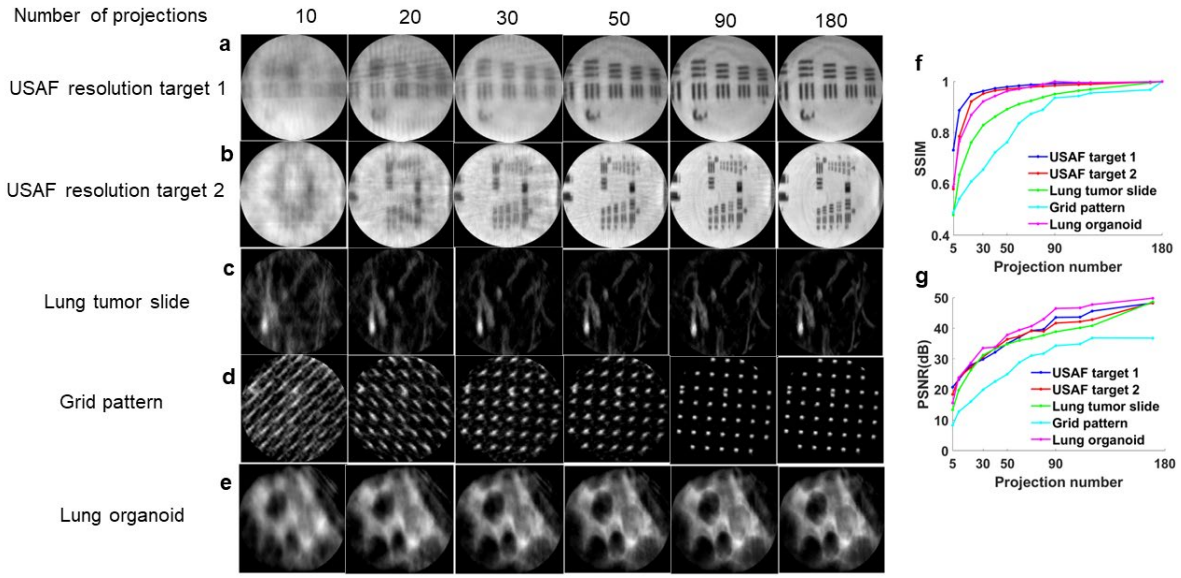

Fig.4.7. LIFT reconstruction under different compression ratio values. a-e. Reconstructions of various objects under different number of projections (i.e., compression ratio). f. Structure similarity index measure (SSIM) *vs.* number of projections. g. Peak signal-to-noise ratio (PSNR) *vs.* number of projections.

LIFT-FLIM distributes projection measurements into different views, it demands only N_θ scanning steps, where N_θ is a total number of projection angles. Therefore, LIFT-FLIM reduces the scanning steps required by a factor of $N^2 \times N_z/N_\theta$ or $N \times N_z/N_\theta$ compared to point- or line-scanning systems.

For non-compressive measurement, we set N_θ equal to N. Our findings indicate that, in the light field imaging, the effective number of depth samplings, N_z, equals the number of angular samplings, K. As a result, the scanning reduction factor is either $N \times K$ or K when compared to point- or line-scanning systems. With our current N and K values set at 180 and 15, respectively, the resulting scanning reduction factors are 2,700 and 15 in comparison to point- or line-scanning systems.

Alternatively, like sparse-view computed tomography [169], we can choose an N_θ less than N for compressive measurement. We define a compression ratio (CR) as

$$\mathrm{CR} = N/N_\theta. \tag{31}$$

To quantify the dependence of the reconstructed image quality on the CR, we adopted the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) as evaluation metrics. We varied the CR by increasing N_θ and calculated the corresponding PSNRs and SSIMs. **Figure 4.7** illustrates reconstructions of sparse and complex objects at different CR values. In general, reducing the CR improves both PSNR and SSIM in the reconstructed image. Additionally, our findings suggest that the quality of the reconstructed image is highly dependent on the CR for complex objects, like the lung tumor image displayed in **Fig. 4.7c**. In such cases, a lower CR value, less than 4.5, is necessary to achieve high-quality image reconstruction (SSIM≥0.9). Conversely, when imaging a sparse object, such as an USAF resolution target, a CR of 9 is sufficient to recover a high-quality image. Therefore, by adjusting N_θ, LIFT-FLIM can tailor the CR to the complexity of a sample, resulting in effective measurements for a given object.

The imaging speed of LIFT-FLIM is determined by the total number of projections N_θ acquired and the time duration at each projection. For LIFT-FLIM using a linear SPAD array, the duration at each projection includes both the pixel exposure time and temporal histogram readout time. For LIFT-sFLIM using a gated ultrafast camera, the duration at each projection equals the product of the number of time gates and the camera frame time. Importantly, when imaging simple objects, the system can be operated in the compressive measurement mode, where a reduced N_θ can be acquired to accelerate the imaging speed without compromising the image quality.

The spatial resolution of LIFT-FLIM is fundamentally limited by optical diffraction when

performing non-compressive measurements. Due to the division of the aperture, LIFT-FLIM has a lateral resolution of $\lambda f/D$, where λ is the wavelength, f is the focal length of the objective lens, and D is the sub-aperture diameter associated with a perspective image. Given K views, $D = D_0/\sqrt{K}$, where D_0 is the original aperture of the objective lens. Therefore, the lateral resolution is $\sqrt{K}$ times greater than the native resolution of the objective lens. Although this is a common issue encountered by all light field cameras, we can mitigate it by acquiring fewer views and increasing the sub-aperture size to enhance the resolution at the expense of reduced depth accuracy. On the other hand, when performing compressive measurements, the spatial resolution of LIFT-FLIM is practically limited by the CR. While a higher CR is favored in terms of imaging speed, it deteriorates the reconstructed image quality and resolution for complex objects. Hence, selecting an appropriate CR value for a given object involves striking a balance between imaging speed and resolution.

LIFT-FLIM images can be reconstructed and analyzed in real-time. For instance, when processing uncompressed measurement data, a simple inverse Radon transform takes about 0.13 seconds per time bin on an Nvidia RTX3080Ti GPU with CUDA. Subsequently, deep learning enhancement and phasor analysis require 0.079 and 0.024 seconds, respectively. Parallel computing reduces the total post-processing time to less than 0.3 seconds.

The light throughout of LIFT-FLIM depends on the sub-aperture size of a perspective image, the ratio of projection line image width to the detector pixel's size, and the fill factor of the image sensor. LIFT-FLIM is built on an unfocused light field imaging configuration, where the projection line width at the image sensor equals to the sub-aperture diameter, D, multiplying with a pupil demagnification ratio, r. Given the pixel pitch, p, and fill factor, κ, the percentage of light measured by the image sensor pixel is

$$\zeta_{\mathrm{FILM}} = \frac{D}{D_0} \times \frac{p\kappa}{Dr} = \frac{p\kappa}{D_0 r}. \tag{32}$$

Here D/D_0 describes the light loss due to the view selection during pupil scanning, where D_0 is the original aperture of the objective lens. Therefore, a lower pupil demagnification ratio (*i.e.,* a shorter focal length of the cylindrical lens in **Fig. 4.1a**) can lead to a higher system light throughput. In our current system, due to use of only off-the-shelf optics and a SPAD array, we have r equal to 0.1, resulting in an overall light throughput of 0.01. To further enhance the system performance, one possible approach is to utilize custom optics that feature a lower pupil demagnification ratio, r, together with a rectangularly shaped SPAD pixel that has a longer pixel pitch, p, in the direction of the projection line width. Alternatively, instead of scanning the pupil to choose the views, it is possible to simultaneously capture all perspective images by employing an array of dove prisms with different orientations, as we have previously demonstrated. However, this setup necessitates the use of multiple linear SPAD arrays, each of which measures a projection line image in a synchronized fashion.

On the other hand, for LIFT-sFLIM using a gated ultrafast camera, the light throughput is determined by the sub-aperture size of a perspective image, the diffraction efficiency of the grating, χ, and quantum efficiency of the gated ultrafast camera, η.

$$\zeta_{\mathrm{sFILM}} = \frac{D}{D_0} \times \chi \times \eta. \tag{33}$$

Since $D/D_0 = 1/\sqrt{K}$, where K is the total number of views acquired, **Eq. 33** can be rewritten as $\zeta_{\mathrm{sFILM}} = \chi\eta/\sqrt{K}$. Hence, reducing the number of angular samplings can boost the light throughput, but this comes at the cost of decreased depth accuracy. Noteworthily, here the pupil magnification ratio, r, has no effect on the light throughput. Rather, it governs the

spectral resolution of the system like in a conventional pushbroom imaging spectrometer [170,171].

To investigate how the number of photons received at a pixel affects the quality of the reconstructed image, we conducted simulations under a shot-noise-limited condition. Provided that the pixel with the maximum count in the image collects M photons, the corresponding shot noise is $\sqrt{M}$ photons. We introduced photon noise to all pixels in the projection images and reconstructed the images with various values of M, while maintaining a constant number of projections across all data points in the plot. **Figure 4.8** presents the reconstruction results of a Shepp-Logan phantom under different M values. The results indicate that a larger M (*i.e.,* more photons) can lead to a higher PSNR. For high-quality image reconstruction (PSNR$\geqslant$20 dB), M must be greater than 64 photons.

To sum up, we have created a highly data-efficient 3D FLIM technique that relies on light field tomography and extended its capabilities to 3D sFLIM. We believe that LIFT-FLIM and -sFLIM will find broad applications in high-throughput and high-content imaging of biological cells and tissues, opening up new avenues for both fundamental and translational biomedical research.

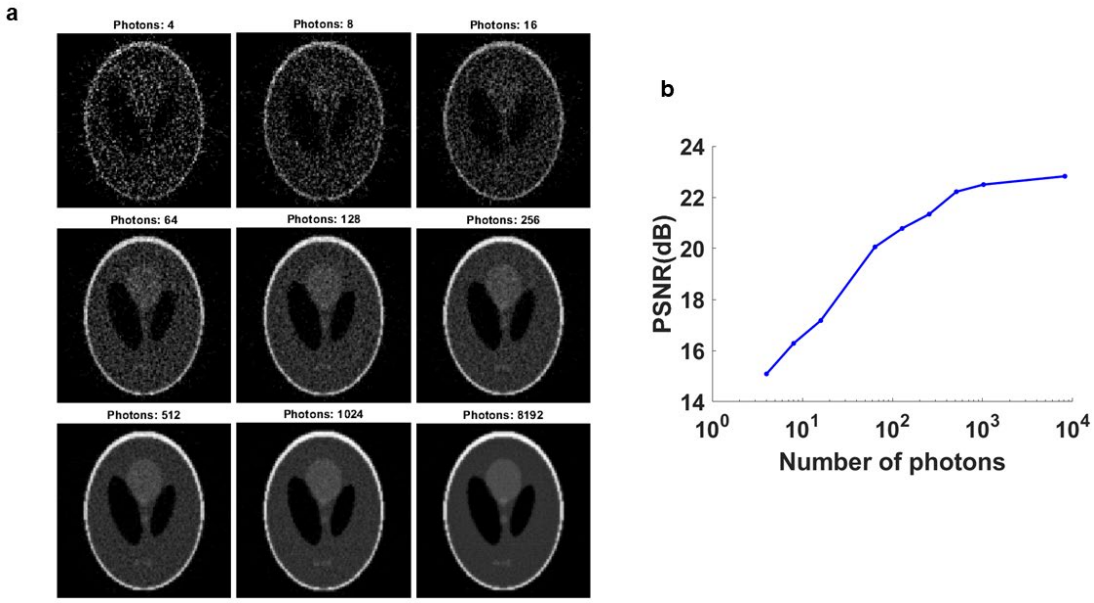

Fig.4.8. Reconstruction under different noise levels. a. Reconstructions of a Shepp-Logan phantom with different photons numbers. b. Peak signal-to-noise ratio (PSNR) *vs.* number of photons.

4.4 Methods in LIFT-FLIM

4.4.1 Experimental setup

In a LIFT-FLIM and -sFLIM system, we used an epi-fluorescence microscope (IX83, Olympus) as the front-end optics and excited the sample with a pulsed laser source (SuperK FIANIUM, FIU-15, NKT Photonics, for LIFT-FLIM; SIRIUS GR-2, Spark Laser, for LIFT-sFLIM). The emitted fluorescence is collected by a microscope objective lens (UPLXAPO60XO, Olympus; UPLXAPO20X, Olympus), and an intermediate fluorescence image is formed at the side image port of the microscope.

To split the light, we employed a beam splitter (BSX16, Thorlabs), which transmits 10% of the light to a reference camera (CS2100M-USB, Thorlabs) and reflects 90% of the light to the LIFT-FLIM camera. We placed a scanning mirror (MR-10-30, Optotune) at the intermediate image plane to shift the pupil image.

The fluorescence is then directed through a 4f system, which consists of two lenses (ACT508-250-A and AC254-150-A, Thorlabs) with a focal length of 250 mm and 150 mm, respectively. To rotate the perspective image, we mounted a Dove prism (PS990M, Thorlabs) on a motorized rotation stage (PRM1Z8, Thorlabs) and positioned the assembly at the Fourier plane of the 4f system. We also positioned a cylindrical lens (LJ1095L1-A, Thorlabs, invariant axis along the y-axis) 131mm after the second lens in the 4f system, which generates a 1D en-face projection of a perspective image along the y-axis. To locate the projection line image, we identified the line with the smallest width. Additionally, we compensated for the focal shift and spherical aberration introduced by the cylindrical lens by defocusing.

The subsequent system is split into two arms, namely the LIFT-FLIM and LIFT-sFLIM arms. The former employs a linear SPAD array [34], while the latter utilizes a 2D ultrafast time-gated camera (High rate image intensifier, LaVision). To switch the light path between the two arms, we placed a flip mirror (TRF90, Thorlabs; PF10-03-G01, Thorlabs) at the line image plane.

When the mirror is positioned at 1, the fluorescence is directed towards the LIFT-FLIM sub-system through a camera lens (YN100mm F2, YONGNUO) and directly measured by the linear SPAD camera. The linear SPAD camera comprises 256 effective CMOS SPAD pixels with a pitch of 26.2μm. Operating in the TCPSC mode, the SPAD camera provides a temporal resolution of 50 picoseconds [34]. It is connected to a FPGA (Spartan 6, Xilinx) with 64 time-to-digital converters (TDCs) and histogram engines, enabling it to process up to 8.5 giga-photons per second. By rotating the dove prisms in a set of angles at assigned views, we sequentially acquired the 1D en-face projections and constructed a sinogram.

In position 2 of the flip mirror, the emitted fluorescence is directed to the LIFT-sFLIM sub-

system. The line image is relayed to the image sensor plane by a pair of camera lenses (YN100mm F2, YONGNUO). To disperse the line image along the x-axis, we positioned a transmission diffraction grating (GT50-03, Thorlabs) at the Fourier plane of the relay system. The resultant dispersed projection image is then sampled in time by an ultrafast time gate and further relayed to a 2D camera (CS2100M-USB, Thorlabs) by a camera lens (YN100mm F2, YONGNUO). By varying the delay between the time gate and the laser reference signal, we acquired a series of time-resolved dispersed projection images. To synchronize the scanning mirror, the dove prism rotation stage, the camera, and the laser, we employed a digital delay generator (DG645, Stanford Research Systems). To maximize information content for image reconstruction, we chose the dove prism rotation angles from a set of angles that are evenly spaced in the range of [0, 90°].

To tune the illumination wavelength from the supercontinuum laser, we built a wavelength-selecting module using a digital micromirror device (DMD). The collimated white laser beam is first dispersed by a transmission diffraction grating (GT50-03, Thorlabs) and line focused onto the surface of the DMD (DLP LightCrafter 6500, Texas Instruments) through a cylindrical lens (LJ1125L1-A, Thorlabs). The broadband illumination has a line dispersion of 54.4 nm/mm on the DMD surface. The DMD has 1920×1080 micromirrors, each of which can be individually tilted ±12^{o} relative to the norm. Each column of the DMD corresponds to a different wavelength with a 0.4nm/column wavelength resolution. By adjusting the mirror pattern, we can select any desired illumination wavelengths. The laser light of selected wavelengths is then spatially recombined by another identical set of cylindrical lens and diffraction grating and directed towards the LIFT-FLIM sub-system.

When imaging the mixed fluorescence beads and mouse kidney tissue section, we used

multiband excitation with two different wavelengths (488 nm and 561 nm) and separated fluorescence from excitation using the combination of a multiband dichroic mirror (ZT405/488/551/647rpc, Chroma) and a multiband emission filter (ZET405/488/561/647m, Chroma). For imaging the human lung cancer pathology slide, we used 450 nm laser excitation, a 495 nm dichroic mirror (T495lpxr, Chroma), and a long-pass emission filter (ET500lp, Chroma). In the case of lung organoids, we used 532 nm laser excitation, a 532 nm dichroic mirror (ZT532rdc, Chroma), and a long pass emission filter (ET542lp, Chroma). The laser fluence at the sample focal plane was approximately 9.9 x10^{-7} J/cm^2, 1.1x10^{-7} J/cm^2, and 2.9 x10^{-3} J/cm^2 for the mouse kidney section, lung cancer pathology slide, and lung organoid imaging experiments, respectively. These laser fluences were well below the cell damage threshold of 4 J/cm^2 [172,173].

4.4.2 Image reconstruction

To obtain an image of a monochromatic scene at a specific time point and depth from the measurement described by Eq. 3, we iteratively solve an optimization problem:

$$\text{argmin } \|\mathrm{f} - \boldsymbol{F}(d)h(d)\|_2^2 + \mu\|\varphi(h(d)\|_1 , \tag{9}$$

where $\|.\|_2$ denotes the l_2 norm, $\|.\|_1$ denotes the l_1 norm, and $\varphi(\cdot)$ is a data regularization term. μ is a hyperparameter that balances the data fidelity and regularization term. In the framework of regularization by denoising [174], $\varphi(\cdot)$ is not explicitly specified, and the regularization can be implemented by a state-of-the-art image denoising algorithm such as BM3D or a neural network. We adopted the BM3D and total variation (TV) denoisers for the regularization due to the availability of efficient algorithms [175]. Besides the iterative

method, inverse Radon transformation is an alternative approach that could have been used for image reconstruction with lower computational cost.

4.4.3 Refocusing and extending the depth of field

A light field acquired by conventional light field cameras can be parameterized by the aperture plane (u, v) and the image plane (x, y). Indexing view k as (u_k, v_k), the image $\mathrm{P}_k(x, y)$ observed from view k can be related to a reference image feature kernel $h(x, y)$ by

$$p_k(x, y) = h(x - su_k, y - sv_k), \tag{10}$$

where s is a depth-dependent shearing parameter. In conventional light field imaging, refocusing is performed by shifting and adding the sub-aperture images [176]. Unlike conventional light field cameras, LIFT-FLIM first rotates a perspective image, followed by transforming the rotated image into a line. Therefore, the depth-dependent shearing must be performed parallel to the projection axis.

For sub-aperture (u_k, v_k) at the projection angle of θ, the shearing of 1D sub-aperture projection is given by

$$s \cdot u_k \cdot \sin\theta - s \cdot v_k \cdot \cos\theta. \tag{11}$$

For numerical refocusing, we applied the correspondent shearing factor to each projection image and updated sinogram for reconstructing the depth image.

Extending the depth of field can be achieved through a similar approach to conventional light field imaging, which involves refocusing onto different depths, extracting the sharpest feature

for each pixel, and assembling an all-in-focus image [177].

4.4.4 System calibration and resolution

Scanning mirror calibration

To ensure that the scanning range of sub-apertures fully utilizes the entire aperture of the objective lens, we calibrated the scanning mirror's horizontal and vertical tilt angles. Additionally, to optimize the light throughput of each sub-aperture, we employed a rectangular iris instead of a round one at the aperture stop. The rectangular shape reduces the gaps between adjacent scanned pupil positions and allows more light to pass through a sub-aperture.

Projection center calibration

To calibrate the central position of each projection line image at the sensor plane, we imaged a pinhole (P10D, Thorlabs) positioned at the center of the FOV on the sample stage. We captured images of the pinhole at every projection angle θ and view k and directly localized the center of each line image as $y_{p0}^{(\theta,k)}$. Subsequently, we extracted the projection data based on the center location to form a sinogram.

Spectrum calibration and resolution

To calibrate the spectral response, we positioned a pinhole (P10D, Thorlabs) at the sample stage and illuminated it with monochromatic light at varied wavelengths. The resulting pixel locations of the projections were recorded and fitted with a linear polynomial, as illustrated in **Fig. 4.9a**. The slope of the line determines the spectral sampling of the system, which was calculated to be 0.14 nm. The spectral resolution is defined as the full-width at half maximum (FWHM) of the spectral response. A 1 nm bandpass filter (FL532-1, Thorlabs) was used to

limit the source wavelength for this measurement, and the raw spectral response is displayed in **Fig. 4.9b**, where the FWHM was approximately 9.2 nm. However, this width was a convolution of the geometrical image of the pinhole on the camera (approximately 7 pixels), the bandwidth of the light source (approximately 7 pixels), and the system spectral resolution. The width of a convoluted function (in pixels) can be computed as [178]:

$$w(f_1 * f_2 * f_3) = w(f_1) + w(f_2) + w(f_3) - 2, \tag{12}$$

where w denotes the width of the function, $*$ denotes the convolution operator, and f_i $(i = 1, 2, 3)$ denotes the individual function in a discrete form. Based on this equation, the width of the spectral resolution on the camera was estimated to be 48 pixels. Given a 0.14 nm spectral sampling, the spectral resolution is 6.6 nm.

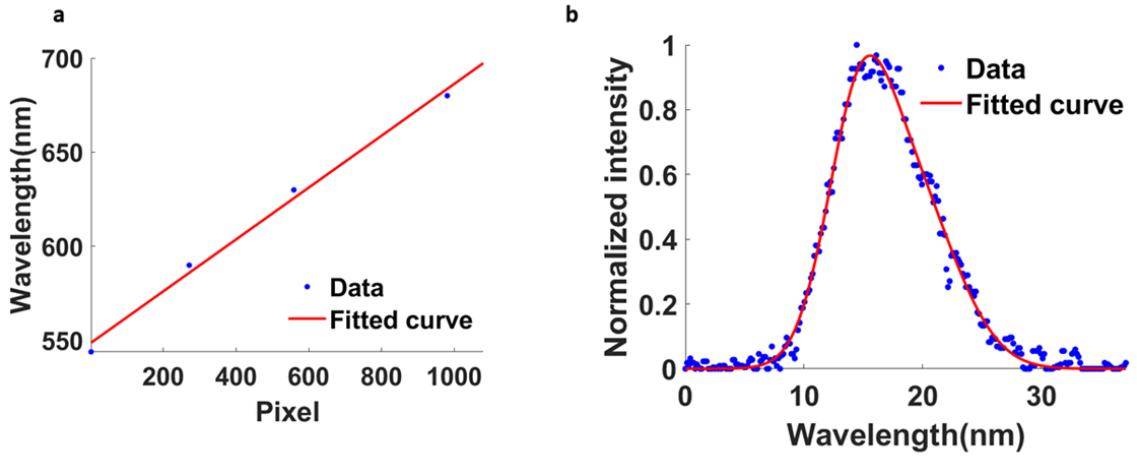

Fig.4.9 Spectral calibration. a. Spectral response locations in the pixels of different wavelengths on the camera. b. Raw spectral response under monochromatic 532 nm illumination (bandwidth = 1 nm).

Spatial resolution and field-of-view(FOV)

To quantify the spatial resolution, we imaged a fluorescence bead with a diameter of 4 μm (F8858, Thermo Fisher), where the lateral and axial full-width at half maximum (FWHM) are approximately 4.6 μm and 5.7 μm, respectively. However, the lateral FWHM width was a convolution of the geometric image of the bead on the camera (around 4 pixels) and the system lateral resolution. The width of a convoluted function can be calculated as [178]:

$$w(f_1 * f_2) = w(f_1) + w(f_2) - 1 \tag{13}$$

where w denotes the width of the function, $*$ denotes the convolution operator, and f_i $(i = 1, 2)$ denotes the individual function in a discrete form. Using this formula, the width of the lateral resolution on the camera was calculated to be 2 pixels. Given a 0.9 μm spatial sampling (camera pixel pitch of 26.2 μm divided by system magnification ratio of 29), the lateral resolution was estimated to be 1.8 μm. Similarly, for axial FWHM, the width was a convolution of the bead size and the system lateral resolution. As a result, the axial resolution was estimated to be 3 μm.

We also imaged a group of bars from a USAF resolution target (Group 7 element 3-6) along both horizontal and vertical directions and plotted the intensities along the dashed line. The image visibility, defined as $(I_{max} - I_{min})/(I_{max} + I_{min})$, where I is the intensity, was calculated for each group of bars using the peaks and valleys of the intensity. With a visibility threshold of 0.2, the spatial resolution of the bars was determined to be 2.2 μm along both vertical and horizontal directions, indicating an isotropic resolution. The FOV of our system was measured to be 227 μm x 143 μm when using a 60x microscope objective lens

(UPLXAPO60XO, Olympus).

Depth calibration

To calibrate the depth, we utilized a fluorescence bead (C16509, Thermo Fisher) and translated it along the depth axis from -16 μm to 16 μm with a 2 μm step. At each depth, we captured an image and then performed digital refocusing by adjusting the shearing parameter, as described in **Methods: Refocusing and extending the depth of field**. The goal was to identify the shearing parameter that would bring each image into the sharpest focus, which was determined by maximizing a focus measure (*e.g.*, sum of modified Laplacian) for each pixel in the image. The best focus shearing parameter at each physical depth was then recorded. The resultant shearing parameter to depth curve was fitted with linear models. With this calibration curve, we can digitally refocus a 3D object to a specific depth using the correspondent shearing parameter. To validate the accuracy of our shearing parameters, we imaged 3D fluorescence beads in an agarose gel and compared LIFT refocusing against the ground-truth depth images captured by a reference camera, as shown in **Fig. 4.3**.

Camera registration

To register the LIFT-FLIM and -sFLIM image with the reference camera, we imaged a 7 by 13 grid pattern. We then extracted the point locations from the reconstructed LIFT-FLIM and -sFLIM images and the reference image and calculated the homography matrix to establish a pixel-to-pixel correspondence between the two cameras. The reprojection error using the homography is less than one pixel, ensuring an accurate registration. This homography matrix is then used to register the LIFT-FLIM and -sFLIM images with reference images for deep learning reconstruction.

4.4.5 Preparation of training data for deep learning

To generate the ground truth depth image stack, we translated the sample stage along the depth axis while capturing images using the reference camera. The resulting image stack was then transformed to the LIFT-FLIM and -sFLIM camera coordinates using a homography matrix obtained through camera registration. We subtracted a background image from the warped image stack.

The LIFT-FLIM and -sFLIM images were captured at depth = 0, and a depth image stack was computed by numerical refocusing. We calculated a uniform DPM at each depth and appended them to the reconstructed image stack. We created a mask by setting a threshold to the reconstructed image to identify regions with sample fluorescence signals above the background. This mask was then applied to both the ground truth image stack and the LIFT-FLIM and -sFLIM image stack. For training, we constructed a total of 200 image stack pairs per task in the training dataset, each comprising an input image collection (LIFT-FLIM/sFLIM reconstruction stack, DPM stack, and a wide-field image at depth=0) and a ground-truth image stack.

4.4.6 Deep learning network architecture and training

The PixelCNN++ architecture[144] was adopted for LIFT-FLIM and -sFLIM refocusing. As illustrated in **Fig. 4.2**, our model consists of the down- and up-sampling streams and the lower down- and up-sampling streams. Each stream has 5 ResNet blocks [179] in both the down-sampling and up-sampling paths. Each ResNet block contains 4 ResNet layers, and each ResNet layer has two 3×3 convolutional layers and one 1×1 convolutional layer. The ResNet layer utilized two activation functions σ_1, σ_2 defined below:

$$\sigma_1(x) = ELU(x \oplus (-x))$$

$$\sigma_2(x_1 \oplus x_2) = x_1 \odot Sigmoid(x_2).$$

Here $\oplus$ means concatenation along the channel axis and $\odot$ is element-wise multiplication. Exponential Linear Unit (ELU) [69] and Sigmoid function are defined as

$$ELU(x) = \begin{cases} x, & for\ x > 0 \\ \alpha(\exp(x) - 1), & for\ x \leq 0 \end{cases},$$

$$Sigmoid(x) = \frac{1}{1 + \exp(-x)},$$

where α is a hyperparameter that controls the value to which an ELU saturates for negative net inputs. Strided convolutional layers were added between two sequential ResNet blocks to halve the spatial dimensions in the down-sampling path, and conversely transposed strided convolutional layers were utilized to implement up-sampling in the up-sampling path. Skip connections connect each ResNet block in the down-sampling path with its counterpart block in the up-sampling path such that relatively higher-frequency image features can flow through the model.

The training loss of our model is a linear combination of Fourier domain mean absolute error (FDMAE) [180,181], mean square error (MSE) and the perceptual loss:

$$L(y, \hat{y}) = \alpha L_{FDMAE}(y, \hat{y}) + \beta L_{MSE}(y, \hat{y}) + \gamma L_p(y, \hat{y}). \quad (14)$$

Here α, β and γ are weights of each loss term, and were empirically set as 0.1, 0.1 and 1.0, respectively. $y, \hat{y} \in R^{N^2}$ are the vectorized ground truth and predicted images, respectively. The FDMAE loss is defined as

$$L_{FDMAE}(y, \hat{y}) = \|Fy - F\hat{y}\|_1, \quad (15)$$

where $F \in R^{N^2}$ is the Fourier transform matrix. The MSE loss is defined as

$$L_{MSE}(y, \hat{y}) = \|y - \hat{y}\|_2. \quad (16)$$

The perceptual loss is defined as the sum of MSE losses between the feature maps of y and $\hat{y}$ generated by a Visual Geometry Group 16 (VGG16) network [183]:

$$L_p(y, \hat{y}) = \sum_{k=1}^{K} w_k \cdot \|VGG_k(y) - VGG_k(\hat{y})\|_2, \quad (17)$$

where $VGG_k(\cdot)$ represent the feature map of the input image after the k_{th} block of VGG16, and w_k is the weight for the corresponding feature maps. In this work we used the first three blocks of VGG16 for image feature extraction, i.e., K = 3, and empirically set w_1 = 0.5, w_2 = 0.15, w_3 = 0.1. An Adam optimizer with exponentially decaying learning rate was utilized for parameter optimization. The initial learning rate was set as 10^{-4} and the decay rate was 0.999995 per epoch.

Our models were implemented using PyTorch framework [184] on a machine with Intel Xeon W-2195 processor and four RTX 2080Ti graphic cards. All models converged after around 5000 epochs, which took approximately 2 to 3 days.

4.4.7 Image stitching

We used a feature-based image stitching algorithm to create a panorama view of the human lung cancer pathology slide from multiple scanned FOVs with overlapping regions. This process involved detecting and matching image features, estimating the geometric

transformation between images, and computing the transformation mapped each image onto the panorama. Moreover, to correct the artifacts in the stitched image caused by connecting the individual images, we applied an intensity averaging technique to the neighboring pixels at the artifact's coordinates.

4.4.8 Phasor Analysis

To facilitate fast and accurate analysis of LIFT-FLIM and LIFT-sFLIM data, we utilized a phasor approach to unmix the underlying chromophores. In our experiments involving mouse kidney tissue sections and human lung cancer pathology slides, we reconstructed a multidimensional array XYT at each depth Z, where X, Y and Z are the spatial dimensions, and T represents fluorescence decay time. After phasor transformation, we fed the resulting phasor coordinates into an unsupervised unmixing algorithm [147,148] to determine the probability of each pixel belonging to a specific cluster. It is important to note that the number of clusters present was assumed to be known a prior, which is generally the case as we label our samples with fluorescence probes or make assumptions regarding the sample composition [148]. For instance, we modeled our data using two clusters for the mouse kidney tissue sections (Alexa Fluor 488 WGA and Alexa Fluor 568 phalloidin) and human lung cancer pathology slide (normal and tumor). Using the probabilities obtained from the unmixing algorithm, we assigned colors to the unmixed image pixels: each cluster was assigned a unique color, and the RGB coordinates of the colors were combined using the probabilities as coefficients. This process results in a color code for each pixel, as shown in **Fig. 4.4f**.and **Fig. 4.5i**.

In LIFT-sFLIM, the image is represented by a multidimensional array $XYTS$ at each depth Z, where S denotes the spectral dimension. To simplify the analysis, we focus on the $T \times S$ matrix $I(t,\ \lambda)$ at one pixel location, where fluorescence decay is sampled at m points ($t =$

$t_1, \dots, t_m$), and the spectrum is sampled at n points ($\lambda = \lambda_1, \dots, \lambda_n$). We also consider wavelength-integrated time decay $I^{\Lambda}(t) = \sum_{\lambda=1}^{n} I(t, \lambda)$ and time-integrated emission spectrum $I^{T}(\lambda) = \sum_{t=1}^{m} I(t, \lambda)$ to streamline the process. In our lung organoid experiment, $I^{\Lambda}(t)$ and $I^{T}(\lambda)$ can be expressed as linear combinations of four components:

$$I^{\Lambda}(t) = f_1 I_1^{\Lambda}(t) + f_2 I_2^{\Lambda}(t) + f_3 I_3^{\Lambda}(t) + f_4 I_4^{\Lambda}(t)$$
$$I^{\mathrm{T}}(t) = f_1 I_1^{\mathrm{T}}(t) + f_2 I_2^{\mathrm{T}}(t) + f_3 I_3^{T}(t) + f_4 I_4^{T}(t), \quad (18)$$

where f_i ($i = 1,2,3,4$) are the amplitude fractions of each component with the constraint that their sum equals unity ($\sum_{i=1}^{4} f_i = 1$). Once transformed into the phasor space, $I^{\Lambda}(t)$ and $I^{\mathrm{T}}(t)$ are represented by points constituted by the linear combinations of phasors of its pure components $I_i^{\Lambda}(t)$ $(i = 1, 2, 3, 4)$ and $I_i^{\mathrm{T}}(t)$ $(i = 1, 2, 3, 4)$. Based on the temporal and spectral profiles of the pure fluorophores in **Fig. 4.6a**, the optimal number of clusters to model the data was set to be two and three for temporal and spectral unmixing, respectively. Similar to the experiments conducted on mouse kidney tissue sections and lung cancer pathology slides, we computed the probability estimate of pixels belonging to a particular cluster as p_i^{Λ} $(i = a, b)$ and p_i^{T} $(i = a, b, c)$, such that $\sum_i p_i^{\Lambda} = 1$ and $\sum_i p_i^{\mathrm{T}} = 1$, where a, b, c are indices of classified clusters. Note that the following assumptions are made regarding probability estimates for clusters: p_a^{Λ} for smad3, p_b^{Λ} for sma, collagen and p16, p_a^{T} for sma, p_b^{T} for collagen and smad3, and p_c^{T} for p16. By combining the probability estimate with **Eq. 18**, we can obtain:

$$p_a^{\Lambda} = f_3$$
$$p_b^{\Lambda} = f_1 + f_2 + f_4$$

$$p_a^{\mathrm{T}} = f_1$$
$$p_b^{\mathrm{T}} = f_2 + f_3$$
$$p_c^{T} = f_4. \quad (19)$$

Using this equation, we can calculate the amplitude fractions of the components in the mixture.

4.4.9 SPAD histograms post-processing

Processing the raw histograms from SPAD involves three steps: background subtraction, delay correction, and nonlinear correction (**Fig. 4.10a**). First, in the background subtraction step, we collected the background signal under the same condition as experiments, which results from the dark count of the SPAD camera and stray light from the environment. The resulting background histogram was then subtracted from the raw histogram. **Figure 4.10b** shows the histograms after subtraction at two representative pixel locations, indicating that the zero references of histograms of the pixels are not aligned due to the delays in the FPGA from the input to the delay line not being matched [185]. Delay calibration was then performed to correct for the misalignment of histograms. In the delay calibration step, to measure the zero-reference bins of each pixel, we shined a picosecond laser beam onto the pixels and registered the start of the event at each pixel as the zero-reference bin location. **Figure 4.10c** presents the histograms at the pixels in **Fig. 4.10b**, indicating the zero reference at the maximum bins. The aligned histograms are shown in **Fig. 4.10d** after shifting by the zero reference bin values. After the delay correction, a nonlinear correction was conducted to smooth the non-linearities inherent in the delay chains [186]. Using non-time-correlated uniform illumination to the linear SPAD array, we collected a sufficient amount of histograms (~100) and processed the

resultant averaged histogram using histogram equalization to create a uniform histogram [186]. A correction matrix was computed and stored for each pixel during the histogram equalization. Multiplying the correction matrix with raw histograms yields smoothed histograms. **Figure 4.10e** shows the raw histogram and smoothed histogram for a 2.5 ns period under uniform illumination, and **Fig. 4.10f** shows the smoothed histogram after multiplication with the correction matrix.

4.4.10 Ground truth lung organoid imaging using a confocal fluorescence microscope

To show the ground truth locations of individual biomarkers in our lung organoid experiment, we cultured another four sets of lung organoids under the same condition and labeled them with individual fluorophores. We then imaged the organoids using a standard confocal fluorescence microscope (Zeiss LSM 880 Confocal). **Figure 4.11** presents the color-coded images, which display a similar appearance to our LIFT-sFLIM results.

4.4.11 Sample preparation

Mixed fluorescent beads

To create a mixed fluorescent bead sample, we embedded three types of fluorescent beads (F8858, C16509, F8831, Thermo Fisher) into agarose gels. First, we diluted the bead suspensions and sonicated them. Then, we pipetted 10uL of the 4um bead suspension ($\sim 5.7\times10^7$ beads/mL), 10μL of the 6μm bead suspension ($\sim 1.7\times10^7$ beads/mL), and 10uL of the 10μm bead suspension ($\sim 3.6\times10^6$ beads/mL) into 10 mL of PBS (10010023, Thermo Fisher) for each type of bead. We mixed 100 μL of the diluted 4 μm bead solution, 100 μL of the diluted 6 μm bead solution, and 100 μL of the diluted 10 μm bead solution to create the final mixed beads solution, which contained approximately 1.9×10^4 4 μm beads/mL, 5.6×10^3 6 μm beads/mL, and 1.2×10^3 10 μm beads/mL.

We prepared the agarose gel by making a 1% [weight/volume] solution of low melting point agarose (A6013, Sigma-Aldrich) in PBS, heating it until it completely dissolved, and cooling it down to approximately 40°C. We added 2.5 μL of the mixed beads solution to 400 μL of the agarose solution. After sonication, we added a 50 μL drop of the mixture onto a glass bottom dish (P35G-1.5-14-C, Mattek) and allowed it to solidify for a few minutes. Finally, we imaged the ~1 mm thick gel, which contained immobilized fluorescent beads, using the methods described in the main text.

Distal lung organoid preparation

We used a hydrostatic droplet generator to fabricate alginate microbead scaffolds with an average diameter of 100 μm, which mimics the size of pulmonary alveoli. After generating the microbeads, we coated them with collagen I (354249, Corning) and dopamine (H8502, Sigma) in a two-step process to functionalize them for cell culture. The detailed protocol for alginate bead generation and functionalization can be found in [57].

Human primary adult normal lung fibroblasts were isolated from distal lung tissue from a de-identified healthy donor (65-year-old, male, Caucasian, non-smoker, non-alcoholic) procured from the International Institute for the Advancement of Medicine (IIAM). Human lung tissue was procured under the UCLA-approved IRB protocol #16-000742. The fibroblast (crawled out population) and epithelial (MACS sorted $EpCAM^+$ population) were isolated from the distal tissue and used in this study.

To develop the 3D model, we used a high aspect ratio vessel (HARV) bioreactor vessel (model: RCCS-4H; Synthecon, Houston, Texas) of 2 mL volume and added 0.5 mL of functionalized microbeads and 1.5 mL of media containing a total of 1 million cells (epithelial:fibroblast=1:1). The vessel was screwed into the bioreactor base and rotated for 48h

to allow optimum cell-bead adherence. After 48h, the cell-coated bead solution was aliquoted 100µL per well in a glass-bottom 96-well plate (P96-1.5H-N, Cellvis) and the plate was briefly centrifuged (1000g X 2 min) to settle the cells/ beads at the bottom of the plate. An additional 150µL media was added to each well. The plate was then kept inside an incubator (37°C,5%CO_2,95%RH) and monitored for the formation of self-organized 3D structures. Within the next 72h, the fully-formed 3D co-culture organoids with micro-alveolar structures were observed in each well.

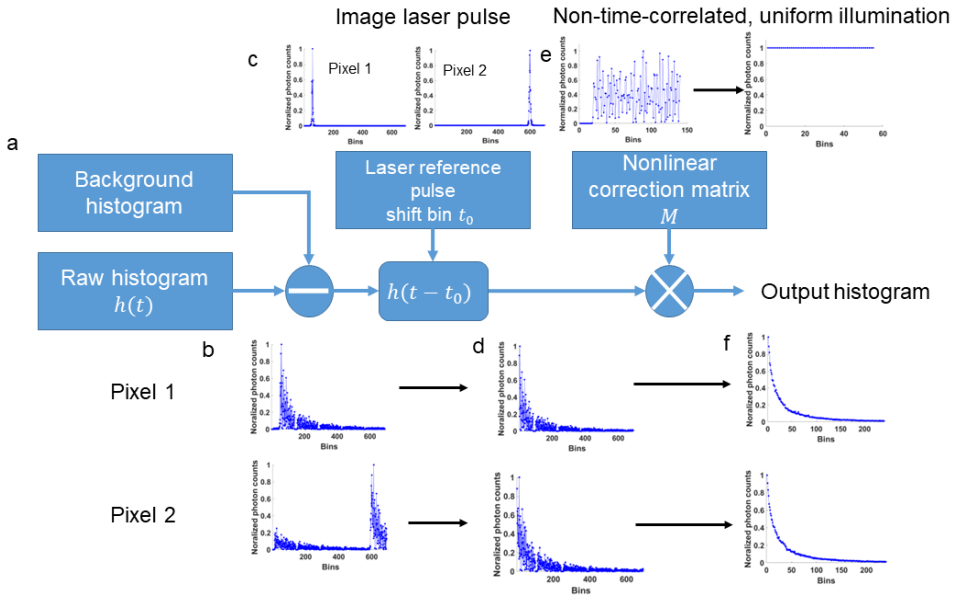

Fig.4.10 SPAD histogram post-processing. a. SPAD data post-processing pipeline. b. Histograms after subtraction at two representative pixel locations. c. Histograms of the pixels when measuring the same laser pulse. d. Aligned histograms after shifting b by the zero reference bin values. e. Raw histogram and smoothed histogram for a 2.5 ns period under uniform illumination. f. Smoothed histograms after multiplication with the correction matrix. SPAD: single-photon avalanche diode.

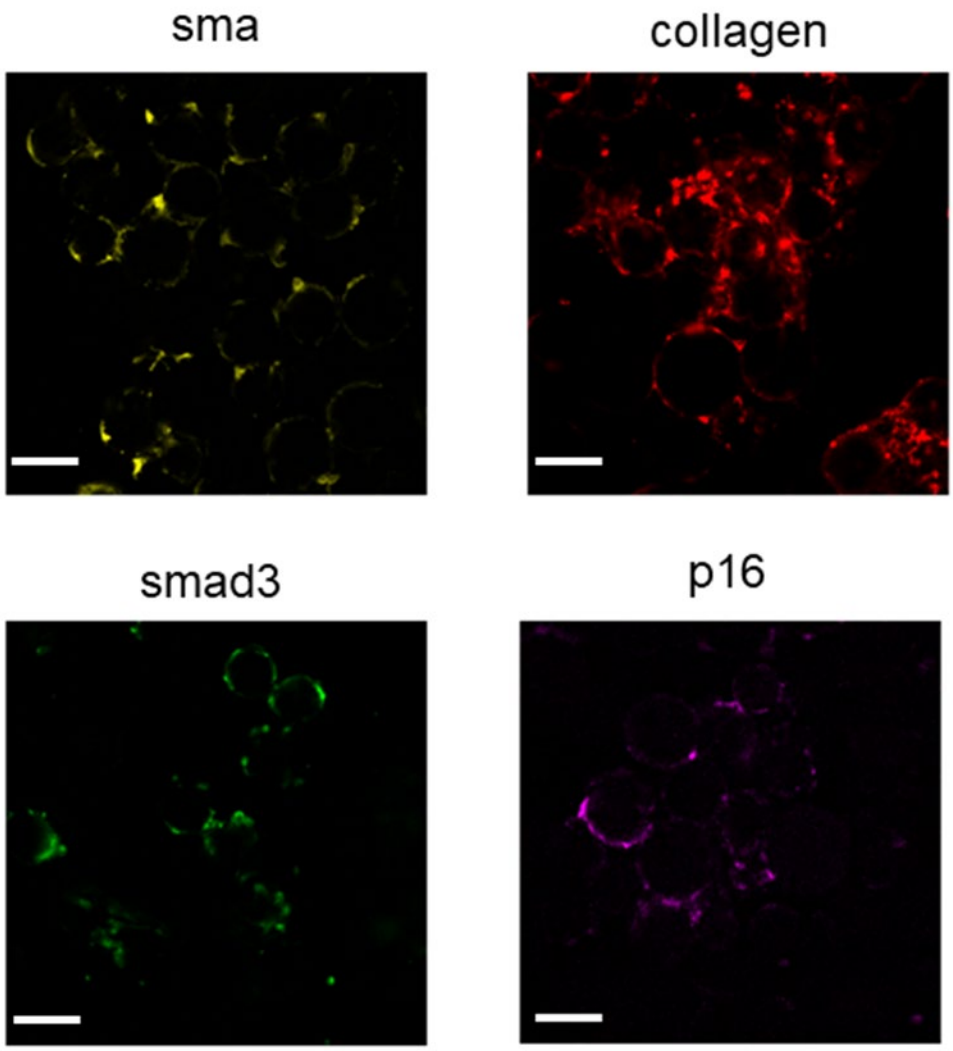

Fig.4.11. Ground-truth imaging of individually labeled lung organoids using a confocal fluorescence microscope. Scale bars in all figures: 100μm.

Chapter 5 Bibliography

1. L. V. Wang and H.-I. Wu, Biomedical optics: principles and imaging, Hoboken: John Wiley & Sons, 2012.

2. M. T. Spence and I. D. Johnson, The molecular probes handbook: A guide to fluorescent probes and labeling technologies, Carlsbad: Live Technologies Corporation, 2010.

3. J. R. Lakowicz, Principles of fluorescence spectroscopy, Springer Science & BusinessMedia, 2013.

4. J. W. Borst and J. V. Antonie, "Fluorescence lifetime imaging microscopy in life sciences," Measurement Science and Technology, vol. 21,no. 10,p. 102002, 2010.

5. P. J. Verveer, F. S. Wouters, A. R. Reynolds and P. I. H. Bastiaens, "Quantitative imaging of lateral ErbB1 receptor signal propagation in the plasma membrane," Science, vol. 290, no. 5496,pp. 1567- 1570, 2000.

6. Ma, Yayao. "Compressed fluorescence lifetime imaging microscope." (2019).

7. W. Becker, "DCS- 120 confocal scanning FILM systems," 2015. [Online]. Available: http://www.tokyoinst.co.jp/product_file/file/BH03_tec01_ja.pdf.

8. R. V. Krishnan, A. Masuda, V. E. Centonze and B. Herman, "Quantitative imaging of protein-protein interactions by multiphoton fluorescence lifetime imaging microscopy using a streak camera," Journal of Biomedical Optics, vol. 8,no. 3,pp. 362-367, 2003.

9. A. J. Bower, J. Li, E. J. Chaney, M. Marjanovic, D. R. Spillman and S. A. Boppart, "High-speed imaging of transient metabolic dynamics using two-photon fluorescence lifetime imaging microscopy," Optica, vol. 5,no. 10,p. 1290- 1296, 2018.

10. "Guide to streak cameras," Photonics, Hamamatsu, Hamamatsu City, Japan, 2008.

11. Hamamatsu Photonics, "Guide to streak cameras," 2008. [Online]. Available:

https://www.hamamatsu.com/resources/pdf/sys/SHSS0006E_STREAK.pdf.

12. L. Gao, J. Liang, C. Li and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature, vol. 516,no. 7529,pp. 74-77, 2014.

36. K. Dowling, S.C.W. Hyde, J.C. Dainty, P.M.W. French, J.D. Hares, 2-D fluorescence lifetime imaging using a time-gated image intensifier, Opt Commun, 135 (1997) 27-31.

37. M.J. Cole, J. Siegel, S.E.D. Webb, R. Jones, K. Dowling, M.J. Dayel, D. Parsons-Karavassilis, P.M.W. French, M.J. Lever, L.O.D. Sucharov, M.A.A. Neil, R. Juskaitis, T. Wilson, Time-domain whole-field fluorescence lifetime imaging with optical sectioning, J Microsc-Oxford, 203 (2001) 246-257.

38. J.P. Eichorst, K.W. Teng, R.M. Clegg, Fluorescence lifetime imaging techniques: Frequency-domain FLIM, in: P.F. L Marcu, and DS Elson (Ed.) Fluorescence Lifetime Spectroscopy and Imaging: Principles and Applications in Biomedical Diagnostics, CRC Press, 2014, pp. 165-186.

39. M. Gersbach, R. Trimananda, Y. Maruyama, M. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. Henderson, E. Charbon, High frame-rate TCSPC-FLIM using a novel SPAD-based image sensor, in: SPIE NanoScience+ Engineering, International Society for Optics and Photonics, 2010, pp. 77801H-77801H77813.

40. L. Parmesan, N.A. Dutton, N.J. Calder, N. Krstajic, A.J. Holmes, L.A. Grant, R.K. Henderson, A 256 x 256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy, Memory, 900 M4-M5.

41. C.J. van der Oord, K.W. Stoop, L.K. van Geest, Fluorescence lifetime attachment LIFA, in: BiOS 2001 The International Symposium on Biomedical Optics, International Society for Optics and Photonics, 2001, pp. 115-118.

42. B. Krämer, V. Buschmann, U. Ortmann, M. Wahl, A. Bülter, F. Koberling, R. Erdmann, Compact FLIM and FCS Upgrade kit for laser scanning microscopes (LSMs), Technical Note, (2007).

43. A. Grinvald, R.D. Frostig, E. Lieke, R. Hildesheim, Optical imaging of neuronal activity, Physiological Reviews, 68 (1988) 1285-1366.

44. T.J. Herron, P. Lee, J. Jalife, Optical imaging of voltage and calcium in cardiac cells & tissues, Circ Res, 110 (2012) 609-623.

45. Wang, Lizhi, et al. "Dual-camera design for coded aperture snapshot spectral imaging." Applied optics 54.4 (2015): 848-858.

46. J. M. Bioucas-Dias, M. A. T. Figueiredo, A new TwIST: Two-step iterative shrinkage thresholding algorithms for image restoration. IEEE Trans. Image Process. 16, 2992–3004 (2007)

47. Liyanage, M. et al. Multicolour spectral karyotyping of mouse chromosomes. Nat. Genet. 14, 312–315 (1996).

48. Garini, Y., Gil, A., Bar-Am, I., Cabib, D. & Katzir, N. Signal to noise analysis of multiple color fluorescence imaging microscopy. Cytometry 35, 214–226 (1999).

47. Tsurui, H. et al. Seven-color fluorescence imaging of tissue samples based on Fourier spectroscopy and singular value decomposition. J. Histochem. Cytochem. 48, 653–662 (2000).

48. Bastiaens, P.I. & Squire, A. Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell. Trends Cell Biol. 9, 48–52 (1999).

49. Chang, C.W., Sud, D. & Mycek, M.A. Fluorescence lifetime imaging microscopy. Methods Cell Biol. 81, 495–524 (2007).

50. Wahl, M., Koberling, F., Patting, M., Rahn, H. & Erdmann, R. Time resolved confocal

fluorescence imaging and spectroscopy system with single molecule sensitivity and sub-micrometer resolution. Curr. Pharm. Biotechnol. 5, 299–308 (2004).

51. Wallace DJ, Meyer zum Alten Borgloh S, Astori S, Yang Y, Bausen M, Kügler S, Palmer AE, Tsien RY, Sprengel R, Kerr JN, et al. 2008. Single-spike detection in vitro and in vivo with a genetic Ca2+ sensor. Nat Methods 5: 797–804.

52. Stosiek C, Garaschuk O, Holthoff K, Konnerth A. 2003. In vivo two-photon calcium imaging of neuronal networks. Proc Natl Acad Sci 100: 7319– 7324.

53. Greenberg DS, Houweling AR, Kerr JND. 2008. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nat Neurosci 11: 749–751

54. Lillis, Kyle P., et al. "Two-photon imaging of spatially extended neuronal network dynamics with high temporal resolution." Journal of neuroscience methods 172.2 (2008): 178-184.

55. Zanin, Natacha, et al. "STAM Interaction with Hrs Controls JAK/STAT Activation by Interferon-α at the Early Endosome." bioRxiv (2019): 509968.

56. Y. Gong, The evolving capabilities of rhodopsin-based genetically encoded voltage indicators, Curr Opin Chem Biol, 27 (2015) 84–89.

57. Mollinedo-Gajate, Irene, Chenchen Song, and Thomas Knöpfel. "Genetically Encoded Fluorescent Calcium and Voltage Indicators." (2019): 1-21.

58. Y. Gong, M.J. Wagner, J.Z. Li, M.J. Schnitzer, Imaging neural spike in brain tissue using FRET-opsin protein voltage sensors, Nature communications, 5 (2014).

59. D. Brinks, A.J. Klein, A.E. Cohen, Two-Photon Lifetime Imaging of Voltage Indicating Proteins as a Probe of Absolute Membrane Voltage, Biophys J, 109 (2015) 914–921.

60. Schneidman E, Freedman B, Segev I. Ion channel stochasticity may be critical in

determining the reliability and precision of spike timing[J]. Neural computation, 1998, 10(7): 1679-1703.

61. Van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity[J]. Science, 1996, 274(5293): 1724-1726.

62. Destexhe, Alain, et al. "Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons." Neuroscience 107.1 (2001): 13-24.

63. Zhu, Liren, et al. "Space-and intensity-constrained reconstruction for compressed ultrafast photography." Optica 3.7 (2016): 694-697.

64. Yang, Chengshuai, et al. "Optimizing codes for compressed ultrafast photography by the genetic algorithm." Optica 5.2 (2018): 147-151.

65. Yang, Chengshuai, et al. "Compressed ultrafast photography by multi-encoding imaging." Laser Physics Letters 15.11 (2018): 116202.

66. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, http://www.deeplearningbook.org.

67. G. Barbastathis, A. Ozcan, and G. Situ, “On the use of deep learning for computational imaging,” Optica, vol. 6, no. 8, pp. 921–943, Aug 2019.

68. M. Qiao, Z. Meng, J. Ma, and X. Yuan, “Deep learning for video compressive sensing,” APL Photonics, vol. 5, no. 3, p. 030801, 2020.

69. X. Yuan, Y. Liu, J. Suo, and Q. Dai, “Plug-and-play algorithms for largescale snapshot compressive imaging,” in CVPR, June 2020.

70. Elias Nehme, Lucien E. Weiss, Tomer Michaeli, and Yoav Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," Optica 5, 458-464 (2018)

71. A. Goy, K. Arthur, S. Li, and G. Barbastathis, “Low photon count phase retrieval using deep learning,” Physical review letters, vol. 121, no. 24, p. 243902, 2018.

72. M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, “Deep-learning-based ghost imaging,” Scientific reports, vol. 7, no. 1, p. 17865, 2017.

73. G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2017), pp. 2261–2269.

74. Li, Yunzhe, Yujia Xue, and Lei Tian. "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media." Optica 5.10 (2018): 1181-1190. Beck, Amir, and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." SIAM journal on imaging sciences 2.1 (2009): 183-202.

75. X. Yuan, “Generalized alternating projection based total variation minimization for compressive sensing,” IEEE International Conference on Image Proc Conference on Image Processing (ICIP), 2539-2543(2016).

76. Guan, Juntao, Rui Lai, and Ai Xiong. "Wavelet deep neural network for stripe noise removal." IEEE Access 7 (2019): 44544-44554. essing (ICIP), 2539-2543(2016).

77. Thanh, Nguyen, et al. "Deep learning approach to Fourier ptychographic microscopy." Optics Express (2018).

78. Rogers, C. et al. A universal 3D imaging sensor on a silicon photonics platform. Nature 590, 256–261 (2021).

79. Yokoyama, A. 3D imaging sensing technology. in ACM SIGGRAPH 97 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH ’97 114–115 (ACM, 1997).

80. Faccio, D., Velten, A. & Wetzstein, G. Non-line-of-sight imaging. Nat. Rev. Phys. 2, 318–327 (2020).

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

82. Brady, D. J. et al. Multiscale gigapixel photography. Nature 486, 386–389 (2012).

83. Schechner, Y. Y. & Kiryati, N. Depth from Defocus vs. Stereo: How Different Really Are They? Int. J. Comput. Vis. 39, 141–162 (2000).

84. Geng, J. Structured-light 3D surface imaging: a tutorial. Adv. Opt. Photonics 3, 128–160 (2011).

85. Levoy, M. Light Fields and Computational Imaging. Computer 39, 46–55 (2006).

86. Jiang, Y., Karpf, S. & Jalali, B. Time-stretch LiDAR as a spectrally scanned time-of-flight ranging camera. Nat. Photonics 14, 14–18 (2020).

87. Riemensberger, J. et al. Massively parallel coherent laser ranging using a soliton microcomb. Nature 581, 164–170 (2020).

88. Gyongy, I. et al. High-speed 3D sensing via hybrid-mode imaging and guided upsampling. Optica 7, 1253–1260 (2020).

89. Feng, X. & Gao, L. Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging. Nat. Commun. 12, 1–9 (2021).

90. Shrestha, S., Heide, F., Heidrich, W. & Wetzstein, G. Computational imaging with multi-camera time-of-flight systems. ACM Trans. Graph. 35, 33:1-33:11 (2016).

91. Kim, Y. M. et al. Multi-view image and ToF sensor fusion for dense 3D reconstruction. in 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 1542–1549 (IEEE 2009).

92. Jayasuriya, S., Pediredla, A., Sivaramakrishnan, S., Molnar, A. & Veeraraghavan, A. Depth Fields: Extending Light Field Techniques to Time-of-Flight Imaging. Preprint at https://arxiv.org/abs/1509.00816 (2015).

93. Wilburn, B. et al. High performance imaging using large camera arrays. ACM Trans. Graph. 24, 765–776 (2005).

94. Huang, Y. & Essa, I. Tracking multiple objects through occlusions. in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 1051–1058 (IEEE 2005).

95. Morimoto, K. et al. Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications. Optica 7, 346–354 (2020).

96. Liang, J., Zhu, L. & Wang, L. V. Single-shot real-time femtosecond imaging of temporal focusing. Light Sci. Appl. 7, 1–10 (2018).

97. Donati, S. & Tambosso, T. Single-Photon Detectors: From Traditional PMT to Solid-State SPAD-Based Technology. IEEE J. Sel. Top. Quantum Electron. 20, 204–211 (2014).

98. Bruschini, C., Homulle, H., Antolovic, I. M., Burri, S. & Charbon, E. Single-photon avalanche diode imagers in biophotonics: review and outlook. Light Sci. Appl. 8, 1–28 (2019).

99. 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).

100. 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 (IEEE 2016).

101. Cai, Z. et al. Lensless light-field imaging through diffuser encoding. Light Sci. Appl. 9, 143 (2020).

102. Antipa, N. et al. DiffuserCam: lensless single-exposure 3D imaging. Optica 5, 1–9 (2018).

103. Liu, X. et al. Non-line-of-sight imaging using phasor-field virtual wave optics. Nature 572, 620–623 (2019).

104. O'Toole, M., Lindell, D. B. & Wetzstein, G. Confocal non-line-of-sight imaging based on the light-cone transform. Nature 555, 338–341 (2018).

105. Gariepy, G., Tonolini, F., Henderson, R., Leach, J. & Faccio, D. Detection and tracking of moving objects hidden from view. Nat. Photonics 10, 23–26 (2016).

106. Saunders, C., Murray-Bruce, J. & Goyal, V. K. Computational periscopy with an ordinary digital camera. Nature 565, 472–475 (2019).

107. Rapp, J. et al. Seeing around corners with edge-resolved transient imaging. Nat. Commun. 11, 5929 (2020).

108. Wu, C. et al. Non–line-of-sight imaging over 1.43 km. Proc. Natl. Acad. Sci. 118, (2021).

109. Lam, E. Y. Computational photography with plenoptic camera and light field capture: tutorial. J. Opt. Soc. Am. A 32, 2021 (2015).

110. Chang, J., Kauvar, I., Hu, X. & Wetzstein, G. Variable Aperture Light Field Photography: Overcoming the Diffraction-Limited Spatio-Angular Resolution Tradeoff. in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 3737–3745 (IEEE 2016).

111. Edgar, M. P., Gibson, G. M. & Padgett, M. J. Principles and prospects for single-pixel imaging. Nat. Photonics 13, 13–20 (2019).

112. Gibson, G. M. et al. Single-pixel imaging 12 years on: a review. Opt. Express 28, 28190–28208 (2020).

113. Dowski, E. R. & Cathey, W. T. Extended depth of field through wave-front coding. Appl. Opt. 34, 1859–1866 (1995).

114. Wang, Z. et al. Single shot, three-dimensional fluorescence microscopy with a spatially rotating point spread function. Biomed. Opt. Express 8, 5493–5506 (2017).

115. Levin, A., Fergus, R., Durand, F. & Freeman, W. T. Image and depth from a conventional camera with a coded aperture. ACM Trans. Graph. 26, 70:1-70:9 (2007).

116. Godbaz, J. P., Cree, M. J. & Dorrington, A. A. Extending AMCW Lidar Depth-of-Field Using a Coded Aperture. in Computer Vision – ACCV 397–409 (Springer, 2011).

117. Manna, M. L., Nam, J.-H., Reza, S. A., Velten, A. & Velten, A. Non-line-of-sight-imaging using dynamic relay surfaces. Opt. Express 28, 5331–5339 (2020).

118. Lindell, D. B., Wetzstein, G. & O'Toole, M. Wave-based Non-line-of-sight Imaging Using Fast F-k Migration. ACM Trans. Graph. 38, 116:1-116:13 (2019).

119. Levin, A., Hasinoff, S. W., Green, P., Durand, F. & Freeman, W. T. 4D frequency analysis of computational cameras for depth of field extension. ACM Trans. Graph. 28, 97:1-97:14 (2009).

120. Stantchev, R. I., Yu, X., Blu, T. & Pickwell-MacPherson, E. Real-time terahertz imaging with a single-pixel detector. Nat. Commun. 11, 2535 (2020).

121. Dansereau, D. G., Schuster, G., Ford, J. & Wetzstein, G. A Wide-Field-of-View Monocentric Light Field Camera. in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 3757–3766 (IEEE, 2017).

122. Kim, M. S. et al. An aquatic-vision-inspired camera based on a monocentric lens and a silicon nanorod photodiode array. Nat. Electron. 3, 546–553 (2020).

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

124. Kadambi, A., Taamazyan, V., Shi, B. & Raskar, R. Polarized 3D: High-Quality Depth Sensing with Polarization Cues. in 2015 IEEE International Conference on Computer Vision (ICCV) 3370–3378 (IEEE 2015).

125. Zhu, D. & Smith, W. A. P. Depth From a Polarisation + RGB Stereo Pair. in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 7578–7587 (IEEE 2019).

126. Romano, Y., Elad, M. & Milanfar, P. The Little Engine That Could: Regularization by Denoising (RED). SIAM J. Imaging Sci. 10, 1804–1844 (2017).

127. Kamilov, U. S. A Parallel Proximal Algorithm for Anisotropic Total Variation Minimization. IEEE Trans. Image Process. 26, 539–548 (2017).

128. Zhang, Z. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1330–1334 (2000).

129. Liu, X., Bauer, S. & Velten, A. Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. Nat. Commun. 11, 1–13 (2020).

130. Bruschini, Claudio, et al. "Single-photon avalanche diode imagers in biophotonics: review and outlook." Light: Science & Applications 8.1 (2019): 87.

131. Krstajić, N., Levitt, J., Poland, S., Ameer-Beg, S. & Henderson, R. 256 × 2 SPAD line sensor for time resolved fluorescence spectroscopy. Opt. Express 23, 5653–5669 (2015).

132. Feng, X. & Gao, L. Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging. Nature Communications 12, 2179,

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

134. Ng, Ren. Digital light field photography. stanford university, 2006.

135. Wu, Gaochang, et al. "Light field image processing: An overview." IEEE Journal of Selected Topics in Signal Processing 11.7 (2017): 926-954.

136. Ihrke, Ivo, John Restrepo, and Lois Mignard-Debise. "Principles of light field imaging: Briefly revisiting 25 years of research." IEEE Signal Processing Magazine 33.5 (2016): 59-69.

137. Qi Cui, Jongchan Park, Yayao Ma, and Liang Gao, "Snapshot hyperspectral light field tomography," Optica 8, 1552-1558 (2021)

138. Beck, Amir, and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." SIAM journal on imaging sciences 2.1 (2009): 183-202.

139. Bioucas-Dias, José M., and Mário AT Figueiredo. "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration." IEEE Transactions on Image processing 16.12 (2007): 2992-3004.

140. Barbastathis, George, Aydogan Ozcan, and Guohai Situ. "On the use of deep learning for computational imaging." Optica 6.8 (2019): 921-943.

141. Wang, H., et al. "H. Günayd in, LA Bentolila, C. Kural, and A. Ozcan,“." Deep learning enables cross-modality super-resolution in fluorescence microscopy,” Nat. Methods 16 (2019):

103-110.29 Wu, Y., Rivenson, Y., Wang, H. et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. Nat Methods 16, 1323–1331 (2019).

142. Rivenson, Y., et al. "Deep learning microscopy. Optica 4, 1437–1443." (2017).

143. Wu, Y., Rivenson, Y., Wang, H. et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. Nat Methods 16, 1323–1331 (2019).

144. Salimans, Tim, et al. "Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications." arXiv preprint arXiv:1701.05517 (2017).

145. Bruschini, Claudio, et al. "LinoSPAD2: a 512x1 linear SPAD camera with system-level 135-ps SPTR and a reconfigurable computational engine for time-resolved single-photon imaging." Quantum Sensing and Nano Electronics and Photonics XIX. Vol. 12430. SPIE, 2023.

146. http://www.kentech.co.uk/index.html?/&2

147. Scipioni, L., Rossetta, A., Tedeschi, G. et al. Phasor S-FLIM: a new paradigm for fast and robust spectral fluorescence lifetime imaging. Nat Methods 18, 542–550 (2021).

148. Alex Vallmitjana, Belén Torrado, and Enrico Gratton, "Phasor-based image segmentation: machine learning clustering techniques," Biomed. Opt. Express 12, 3410-3422 (2021)

149. Skala, Melissa C., et al. "In vivo multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia." Proceedings of the National Academy of Sciences 104.49 (2007): 19494-19499.

150. Walsh, Alex J., et al. "Optical metabolic imaging identifies glycolytic levels, subtypes, and early-treatment response in breast cancer." Cancer research 73.20 (2013): 6164-6174.

151. Datta, Rupsa, et al. "Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications." Journal of biomedical optics 25.7 (2020): 071203-071203.

152. Marcu, Laura, Paul MW French, and Daniel S. Elson, eds. Fluorescence lifetime spectroscopy and imaging: principles and applications in biomedical diagnostics. CRC press, 2014.

153. Wang M, Tang F, Pan X, Yao L, Wang X, Jing Y, Ma J, Wang G, Mi L. Rapid diagnosis and intraoperative margin assessment of human lung cancer with fluorescence lifetime imaging microscopy. BBA Clin. 2017 Apr 27; 8:7-13.

154. Wang, Q., Hopgood, J. R., Fernandes, S., Finlayson, N., Williams, G. O., Akram, A. R., Dhaliwal, K. & Vallejo, M. A layer-level multi-scale architecture for lung cancer classification with fluorescence lifetime imaging endomicroscopy. Neural Computing and Applications 34, 18881-18894 (2022)

155. Fernandes, S., Williams, G., Williams, E., Finlayson, N., Wang, Q., Dorward, D., Dhaliwal, C., Wallace, W., Akram, A. & Dhaliwal, K. S99 Fluorescence-lifetime imaging: a novel diagnostic tool for suspected lung cancer. Thorax 76, A63-A64 (2021)

156. Vander Heiden M.G., Cantley L.C., Thompson C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009; 324:1029–1033.

157. Hofer, M. & Lutolf, M. P. Engineering organoids. Nature Reviews Materials 6, 402-420 (2021)

158. Drost, J. & Clevers, H. Organoids in cancer research. Nature Reviews Cancer 18, 407-418 (2018)

159. Schutgens, F. & Clevers, H. Human organoids: tools for understanding biology and treating diseases. Annual Review of Pathology: Mechanisms of Disease 15, 211-234 (2020)

160. Lukonin, I., Zinner, M. & Liberali, P. Organoids in image-based phenotypic chemical screens. Experimental & Molecular Medicine 53, 1495-1502 (2021)

161. Renner, H., Grabos, M., Becker, K. J., Kagermeier, T. E., Wu, J., Otto, M., Peischard, S., Zeuschner, D., TsyTsyura, Y. & Disse, P. A fully automated high-throughput workflow for 3D-based chemical screening in human midbrain organoids. eLife 9, e52904 (2020)

162. Driehuis, E., Kretzschmar, K. & Clevers, H. Establishment of patient-derived cancer organoids for drug-screening applications. Nat Protoc 15, 3380-3409 (2020)

163. Du, Y., Li, X., Niu, Q., Mo, X., Qui, M., Ma, T., Kuo, C. J. & Fu, H. Development of a miniaturized 3D organoid culture platform for ultra-high-throughput screening. Journal of Molecular Cell Biology 12, 630-643 (2020)\

164. Takahashi, T. Organoids for drug discovery and personalized medicine. Annual review of pharmacology and toxicology 59, 447-462 (2019)

165. Czerniecki, S. M., Cruz, N. M., Harder, J. L., Menon, R., Annis, J., Otto, E. A., Gulieva, R. E., Islas, L. V., Kim, Y. K. & Tran, L. M. High-throughput screening enhances kidney organoid differentiation from human pluripotent stem cells and enables automated multidimensional phenotyping. Cell stem cell 22, 929-940. e924 (2018)

166. Lukonin, I., Serra, D., Challet Meylan, L., Volkmann, K., Baaten, J., Zhao, R., Meeusen, S., Colman, K., Maurer, F. & Stadler, M. B. Phenotypic landscape of intestinal organoid regeneration. Nature 586, 275-280 (2020)

167. Zhou, T., Tan, L., Cederquist, G. Y., Fan, Y., Hartley, B. J., Mukherjee, S., Tomishima, M., Brennand, K. J., Zhang, Q. & Schwartz, R. E. High-content screening in hPSC-neural

progenitors identifies drug candidates that inhibit Zika virus infection in fetal-like organoids and adult brain. Cell stem cell 21, 274-283. e275 (2017)

168. Dan C. Wilkinson, Jackelyn A. Alva-Ornelas, Jennifer M.S. Sucre, Preethi Vijayaraj, Abdo Durra, Wade Richardson, Steven J. Jonas, Manash K. Paul, Saravanan Karumbayaram, Bruce Dunn, Brigitte N. Gomperts, Development of a Three-Dimensional Bioengineering Technology to Generate Lung Tissue for Personalized Disease Modeling, Stem Cells Translational Medicine, Volume 6, Issue 2, February 2017, Pages 622–633,

169. H. Kudo, T. Suzuki, and E. A. Rashed, “Image reconstruction for sparse-view CT and interior CT-introduction to compressed sensing and differentiated backprojection,” Quantum Imaging Med. Surg. 3(3), 147–161 (2013).

170. Mouroulis, Pantazis, Robert O. Green, and Thomas G. Chrien. "Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information." Applied Optics 39.13 (2000): 2210-2220.

171. Mouroulis, Pantazis Z., and Michael M. McKerns. "Pushbroom imaging spectrometer with high spectroscopic data fidelity: experimental demonstration." Optical Engineering 39.3 (2000): 808-816.

172. R. Dixit, R. Cyr, Cell damage and reactive oxygen species production induced by fluorescence microscopy: Effect on mitosis and guidelines for non-invasive fluorescence microscopy. Plant J. 36, 280–290 (2003).

173. X. Yuan et al., Effect of laser irradiation on cell function and its implications in Raman spectroscopy. Appl. Environ. Microbiol. 84, 8 (2018).

174. Romano, Y., Elad, M. & Milanfar, P. The little engine that could regularization by denoising (RED). SIAM J. Imaging Sci. 10, 1804–1844(2017).

175. Kamilov, U. S. A parallel proximal algorithm for anisotropic total variation minimization. IEEE Trans. Image Process. 26, 539–548 (2017)

176. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, Light Field Photography with a Hand-Held Plenoptic Camera (Stanford University, 2005).

177. Boykov, Y., Veksler, O. & Zabih, R. Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23, 18 (2001).

178. A. V. Oppenheim, J. Buck, M. Daniel, A. S. Willsky, S. H. Nawab, and A. Singer, Signals & Systems (Pearson Educación, 1997).

179. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

180. Clevert, Djork-Arné, Thomas Unterthiner, and Sepp Hochreiter. "Fast and accurate deep network learning by exponential linear units (elus)." arXiv preprint arXiv:1511.07289 (2015).

181. Huang, Luzhe, et al. "GedankenNet: Self-supervised learning of hologram reconstruction using physics consistency." arXiv preprint arXiv:2209.08288 (2022).

182. Chen, Hanlong, et al. "eFIN: Enhanced Fourier Imager Network for generalizable autofocusing and pixel super-resolution in holographic imaging." arXiv preprint arXiv:2301.03162 (2023).

183. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

184. Paszke, A. et al., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035.

185. https://linospad.ch/LinoSPAD_software_manual.pdf

186. Samuel Burri, Harald Homulle, Claudio Bruschini, Edoardo Charbon, "LinoSPAD: a time-resolved 256x1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second," Proc. SPIE 9899, Optical Sensing and Detection IV, 98990D (29 April 2016)

187. Ma, Y., Lee, Y., Best-Popescu, C., & Gao, L. (2021). High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells. Proceedings of the National Academy of Sciences, 118(3).

188. Ma, Y., Feng, X., & Gao, L. (2020). Deep-learning-based image reconstruction for compressed ultrafast photography. Optics letters, 45(16), 4400-4403.