Sustained 3D isotropic imaging of subcellular dynamics using adaptive VCD light-field microscopy 2.0

Lanxin Zhu[1,9], Jiahao Sun[1,9], Chengqiang Yi[1,9], Meng Zhang[3,9], Mian He[1], Liting Chen[4], Yicheng Zhang[4], Yanyi Huang[5], Chunhong Zheng[6], Jianbin Wang[7], Guangda Niu[8], Jiang Tang[8], Yuhui Zhang[3], Dongyu Li[1,2] and Peng Fei[1,2*]*

[1] *School of Optical and Electronic Information-Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, 430074, China.*

[2] *Advanced Biomedical Imaging Facility, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

[3] *Britton Chance Center and MoE Key Laboratory for Biomedical Photonics, Wuhan National Laboratory for Optoelectronic, Huazhong University of Science and Technology, Wuhan, 430074, China.*

[4] *Department of Hematology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.*

[5] *College of Chemistry, Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing, 100871, China*

[6] *International Cancer Institute, Peking University Cancer Hospital and Institute, Peking University, Beijing, China.*

[7] *School of Life Sciences, Tsinghua University, Beijing, China.*

[8] *Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China.*

[9] *These authors contributed equally: Lanxin Zhu, Jiahao Sun, Chengqiang Yi, Meng Zhang*

[*] *Correspondence：feipeng@hust.edu.cn; li_dongyu@hust.edu.cn*

Abstract

Long-term and high-spatiotemporal-resolution 3D imaging of living cells remains an unmet challenge for super-resolution microscopy, owing to the noticeable phototoxicity and limited scanning speed. While emerging light-field microscopy can mitigate this issue through three-dimensionally capturing biological dynamics with merely single snapshot, it suffers from suboptimal resolution insufficient for resolving subcellular structures. Here we propose an adaptive light-field microscopy (VCD-LFM 2.0) with new optics-aware and adaptive-tuning deep learning strategies (VCD 2.0) capable for highly-generalizable light-field reconstruction of diverse subcellular dynamics. VCD-LFM 2.0 accomplishes isotropic spatial resolution down to 120 nm without compromising superior temporal resolution and low phototoxicity. With the development of a compact light-field add-on and an open-source reconstruction software, we realize rapid (up to 333 volumes per second), sustained (up to 60 hours) 3D super-resolution imaging of diverse intracellular dynamics simply on a retrofitted inverted microscope. With VCD-LFM 2.0 approach, we quantitatively analyze both the instantaneous processes of lysosome-mitochondrial interaction and long-term evolution of mitochondrial across two complete cell cycles.

Introduction

In living cells, different organelles work together to execute diverse and intricate physiological functions. Elucidating these fast dynamics / interactions of organelles across the cell cycle needs the microscopes to have sufficiently high spatiotemporal resolution in four dimensions (time +3D space) and low phototoxicity rate for long-term observation [1-4]. This is highly challenging because current 3D microscopy techniques are constrained to the maximal photon budget a sample permits, which results in an inevitable trades-off among imaging speed, spatial resolution, and photon efficiency[5-12]. A series of 3D microscopy implementations including confocal microscopy[11, 13], 3D structured

illumination microscopy (3D SIM)[5, 7, 12] and light sheet microscopy (LSM)[9, 14] , have been intensively developed for live-cell imaging beyond diffraction limit, e.g., ~150 nm lateral and ~280-nm axial resolutions by SIM mode of lattice light-sheet microscope (LLSM-SIM)[9]. However, these scanning-based approaches often require recording hundreds of structured frames at different planes to reconstruct a super-resolved volume, thereby showing compromised temporal resolution limited to a few seconds and relatively high phototoxicity limited to hundreds of volumes acquisition[9]. Therefore, it's difficult for these approaches to observe either long-term evolution during the cell cycle or instantaneous subcellular process occurred at milliseconds timescale.

Unlike scanning-based microscopy, light-field microscopy (LFM) provides photon-efficient direct volumetric imaging by encoding both position and angular information of 3D signals on single 2D camera snapshots without time-consuming axial scanning [15-20]. Benefiting from the high-speed scanning-free 3D imaging, LFM has facilitated various biological studies on neural activities, cardiac hemodynamics and live cells[18, 19, 21]. In sharp contrast to the superior temporal resolution and low phototoxicity well suited for live imaging, the spatial resolution of LFM is often unsatisfactory, owing to the insufficient pixel sampling and limited sub numerical aperture existed in each simultaneously captured light-field view. Many efforts have been made to improve this limitation of LFM[19, 22, 23], for example, DAOSLIMIT can yield a spatial resolution of ~220 nm after a 9-times aperture scanning that slightly lower the imaging speed[19]. Meanwhile, the use of iterative light-field deconvolution for 2D-3D image reconstruction is vulnerable to artifacts, and incapable of surpassing the diffraction limit. There always exists an "impossible performance triangle" which greatly limits the design space of current 3D fluorescence microscopy techniques towards high speed, high spatial resolution and high photon efficiency imaging.

The recent advent of deep neural networks for image reconstruction enlarges the design space of microscopy by introducing prior knowledge of high-resolution data for learning and inference[24, 25]. We previously report view-channel-depth light-field microscopy (VCD-LFM), in which a VCD network model is trained to learn the nonlinear relationships between the 3D confocal ground truths and their 2D light-field projections and afterwards, can directly reconstruct high-resolution 3D volume from a single 2D light-field image through using the well-trained channels in the VCD model to transform the implicit features of the light-field views into depth information of a 3D stack. With deep-learning model combining the high-resolution advances of scanning microscopy into high-speed imaging of LFM, VCD-LFM holds the promise for high-speed and high-resolution 4D imaging of live samples[26-28]. However, current end-to-end supervised networks encounter constraints in term of both enhanced capabilities and generalization ability. When dealing with complex inverse problems, ensuring accuracy is challenging. For instance, reconstructing a 3D super-resolution (SR) image from a 2D under-sampled light-field image with spatial bandwidth compressed by around 500 times, requires the recovery of various degradations brought by noise, resolution, and dimensionality reduction. This presents a highly intricate ill-posed problem with a huge solution space. Traditional one-stage end-to-end networks, with limited amount of label data and convolutional layers, is difficult to find precise SR solutions in such a huge space, resulting in either limited resolution or reduced fidelity. Another typical challenge that supervised learning has to face is the generalization. This arises from the network's predominant focus on learning high-dimensional features of specific training samples, thereby limiting their applicability to the new

samples.

Here we propose an adaptive VCD light-field microscopy (VCD-LFM 2.0), capable of accurately super-resolved and highly-generalizable reconstruction of diverse subcellular structures. To overcome the limitations in complex inverse SR challenges, a light-field-aware network have been devised based on the light-field propagation model. We design a physics-embedded hierarchical data synthesis pipeline to generate semi-synthetic multi-stage data priors, which allow progressive training of the network model. Also, we devised a decomposed-progressive optimization strategy to ensure the collaboration of the sub-networks when training the multi-stage data, thereby achieving precise SR solutions constrained by both optics and data. We prove that through the decoupling of the complex light-field degradations, our adaptive VCD 2.0 network can notably narrow down the inversion space for seeking the correct solution more efficiently. It also becomes highly generalizable to various structures with the addition of an adaptive tuning strategy, in which we use fresh 2D wide-field (WF) and light-field measurements to instantly tune the based model and make it always adaptive to the new samples. We further make the VCD-LFM 2.0 very easy by providing a cost-effective design of light-field add-on and an open-source program with GUI. Using an ordinary inverted microscope equipped with VCD-LFM 2.0, we can image the dynamics of intracellular structures in live cells at an isotropic spatial resolution of ~120 nm and a volume rate up to 333 Hz, enabling the analysis of the fast morphology changes and interactions of diverse organelles. With minimized photon invasion of VCD-LFM 2.0, we first achieve 3D live-cell imaging over 60 hours with recording over 40000 volumes for the tracking of mitochondrial evolution across two complete cell cycles.

Results

Principle and Implementation of VCD-LFM 2.0

Light-field microscopy encodes the spatial-angular patterns from 3D sample into a single 2D image[29, 30]. This procedure contains multiple imaging degradations including dimension compression from the diffraction-unlimited 3D objective to a diffraction-limited 2D LF projection, the frequency aliasing induced by the under-sampling of the microlens array (MLA) during the encoding of spatial-angular information and the noises mainly from the camera exposure (Fig. 1a). In our study, light-field reconstruction beyond diffraction limit needs to inverse this compression with a space-bandwidth product (SBP) expansion over 500 times (Method). This intricate and ill-posed inversion problem has resulted in an extensive solution space that maps the under-sampled LF measurement to the possible 3D SR solutions, thereby posing a big challenge to either analytic-model based approaches without data priors or standard end-to-end DL models constrained by limited label data and network parameters[26]. This leads to an unsatisfactory reconstruction performance especially when imaging the fine subcellular structures where both high spatial resolution beyond diffraction limit and high fidelity close to the ground truth are required for downstream tasks.

According to the physical process of LF imaging, VCD 2.0 disentangles the complex light field inverse problem into three subtasks: denoising, de-aliasing, and 3D reconstruction (Fig.1b, Supplementary Fig. 1, Supplementary Note 1, Method). The collaboration of these light-field-aware sub-networks with multi-stage data guiding thus facilitates the network convergence. During the model training phase, we introduced a physics-embedded hierarchical data synthesis pipeline based on the light-field model, to allow semi-synthetic multi-stage data priors to be conveniently generated

from the same 3D SR data and then progressively guide the sub-networks (Supplementary Fig. 2, Supplementary Note 2, Method). In this procedure, we first designed a sub-aperture shifted light-field projection (SAS LFP) strategy, in which we projected the 3D SR images containing sub-aperture shifts into a series of clean light-field images (Clean LF) and synthesized them into a single one without frequency aliasing (De-aliased LF). Then, we added various experiment-like noises and backgrounds to Clean LF to generate the corresponding Noisy LF which is highly similar to the experimental LF measurements. To achieve the SR light-field reconstruction that involves a 30-fold increase in sampling and transformation from 2D to 3D, we devised light-field-aware networks for each inversion task and a decomposed-progressive optimization strategy to efficiently facilitate the collaboration of multiple sub-networks, thus together ensuring the whole network can find an optimal solution with high resolution and high fidelity (Supplementary Note 3). Based on the multi-dimensional features of light-field images, we incorporated view-attention denoising modules, spatial-angular convolutional feature extraction operators and disparity constraints for high-fidelity (HiFi) denoising and de-aliasing of 4D (x, y, u, v) LF images. We also optimized VCD 3D reconstruction sub-network by incorporating multi-res blocks to extract features from more dimensions, ensuring the high quality of the reconstruction results. Considering the different difficulty in each task, the three sub-networks are jointly optimized using our decomposed-progressive optimization strategy, in which, each sub-network was optimized independently to ensure the high-quality solution for each sub-task at the beginning and then the sub-networks were grouped into "denoising", "denoising and de-aliasing" and "denoising, de-aliasing and reconstruction" for being optimized progressively.

Another inherent limitation of supervised learning is the generalization ability. Networks are often unable to generalize to unseen new signals because they primarily learn high-dimensional features of specific types of fluorescence signals during training, thereby limiting their applicability to other types of signals. We developed an adaptive-tuning strategy for VCD 2.0 to streamline its handling of diverse samples. When facing unseen samples, VCD 2.0 adopted few WF 2D images and corresponding LF measurements of new samples to rapidly tune the pre-trained VCD 2.0 model adaptive to the new samples (Fig.1c, Supplementary Fig. 3, Supplementary Note 4 and Method). In the fine-tuning phase, the WF images of the new samples were readily obtained using the regular port of the inverted microscope and served as the lateral constraint of 3D reconstructions, which were used for calculating the mean-square-errors (MSE) with sum projection of network's 3D inferences. To maintain the mapping function from LF to 3D reconstructions and prevent the over-fitting by the lateral constraint, a small amount of raw data used for the base model should be incorporated into the training process, acting as the volumetric constraint. The lateral and volumetric constraint were alternatively optimized during the fine-tuning phase to together contribute the optimized results in 3D. Through this adaptive-tuning strategy, we successfully transferred a VCD 2.0 model trained on the outer membrane of mitochondria to the outer membrane of lysosome and mitochondrial matrix, respectively (Supplementary Fig. 3). The finely-tuned network showed significantly reduced artifacts and enhanced fidelity, as compared with the version without fine-tuning yielding sharp but distorted reconstruction.

With the abovementioned new designs, we demonstrated the performance advances of VCD 2.0 through a comparison with end-to-end VCD on the organelle data (Fig. 2a, Supplementary Fig. 4). To quantitatively illustrate the performance of the network in the process of solving inverse problems, we developed a new network comprehensive performance pyramid (NCPP) method

based on the comprehensive evaluation of the fidelity and resolution changes during the network training (Method). We quantified the fidelity and resolution of the reconstruction results during the network's convergence process by calculating the differences in structural similarity (Δ SSIM with 0 indicating the best similarity) and cut-off frequency (ΔKc with 0 indicating the highest resolution) between the inference results of 196 regions of interest (ROIs) and the ground truths (GTs). The position of data point on the coordinate axes reflects the network fitting ability while the concentration of the distribution represents the robustness of the network across different datasets. Through NCPP metric, we verified that it's indeed difficult for original VCD to find HiFi, SR solutions through a one-stage supervision, as evidenced by the premature convergence and halting at inferior resolution and fidelity (Fig. 2a, NPV map, left). In sharp contrast, VCD 2.0 under multi-stage data supervision achieved significant improvements in fidelity and resolution right after the initial task optimization (Initial stage, Fig 1b). The progressive optimization strategy further reduced the unnatural high-frequency artifacts caused by such local optimization through preventing the overfitting in the sub-networks. With the decomposed and progressive optimization of VCD 2.0, the network rapidly approached the global optimum (Fig. 2b, NCPP map, right), yielding HiFi and SR reconstructions in the last stage (Last stage, Fig 2b). In addition to the mitochondrial demonstrated in Fig. 2a, b, VCD 2.0 also exhibits better and more efficient reconstruction for other organelles (Supplementary Fig. 5).

Through a simple retrofit of commercial inverted microscope (Olympus IX73) using our compact light-field add-on (~220 mm×140 mm in size, full design in Supplementary Fig. 6 and Method), we performed LFM of the lysosomes in a fixed U2OS cell and reconstructed their 3D distributions using VCD 2.0 (Fig. 2c-e, Supplementary Fig. 7). The raw LF image of lysosome was effectively denoised, de-aliased and finally transformed into 3D volume (Fig. 2d). To validate the resolution enhancement and fidelity by VCD 2.0, we also obtained the *in-situ* Airyscan[13] images of lysosomes in the same fixed U2OS cell and enhanced them through an axial-to-lateral isotropic learning (Method). As verified by calculating the structure similarity (SSIM) using enhanced Airyscan microscope's results as references, the fine structures of lysosomes were accurately reconstructed throughout the 3D volume (Fig. 2f). The fidelity of VCD 2.0 reconstruction was quantified to be significantly higher when compared to current leading light-field reconstruction techniques, including LFD, virtually-scanning LFD (VS-LFD)[22] and VCD (Fig.2g, Method). Furthermore, while LFD, VS-LFD, VCD yielded average lateral resolutions of 780 nm, 315 nm, 261 nm, and average axial resolution of 1021 nm, 440 nm, 277 nm, respectively, VCD 2.0 achieved a near isotropic resolution of 120 nm, which is far superior to all the alternative approaches and close to the resolution of GT (Fig. 2h, Method).

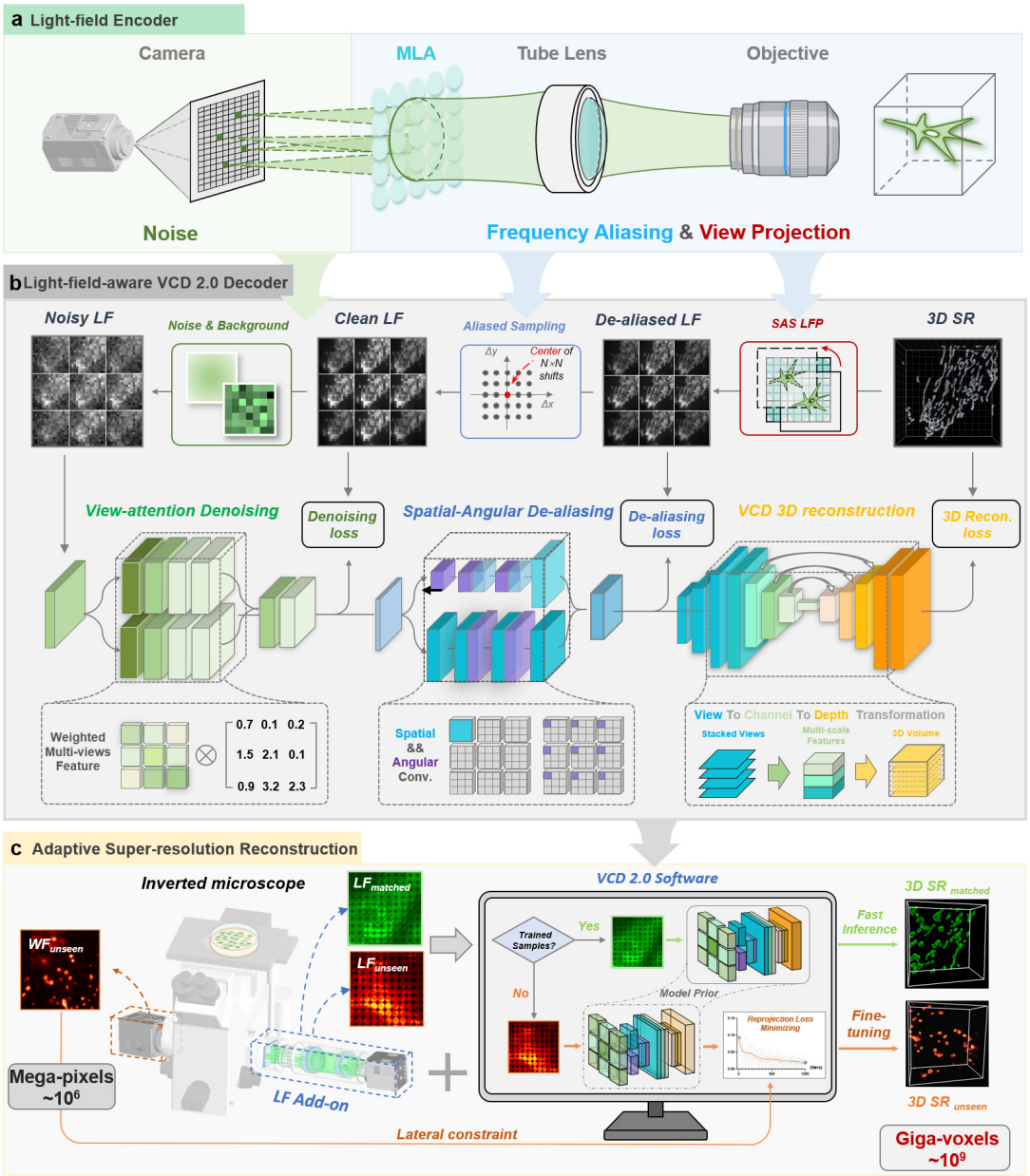

Figure 1. Design of VCD 2.0. The principle and architecture of VCD 2.0 and its training strategy. **a,** The stepwise network is designed based on the physical model of the light-field imaging process. The light-field imaging process encodes multidimensional degradation mainly including: (i) the frequency aliasing caused by the limited NA of objectives and sub-apertures of microlens, (ii) dimension compression during light field encoding and (ii) the noise addition by the camera recording. **b,** Three light-field-aware sub-networks with view-attention denoising, spatial-angular de-aliasing and VCD 3D reconstruction restoration tasks are designed to disentangles and sequentially solves the complex light-field inversion problem. Multi-stage training data including De-aliased LF, Clean LF and Noisy LF are synthesized from the same 3D SR data using an “physics embedded hierarchical data synthesis”, to guide the training of the sub-networks. The decomposed-progressive optimization strategy ensures the collaboration of the sub-networks when training the multi-stage data. **c,** The schematic illustrates the adaptive inference strategy of VCD 2.0. For sample types that have been trained using SR 3D data, users can directly infer 3D reconstructions (Task A with blue color). In the case of inferring unseen samples, the experimentally-

obtained 2D WF and corresponding LF images are used to instantly tune the pre-trained model adaptive to the new samples and then yield improved reconstruction with notably higher fidelity (3D SR_{unseen}).

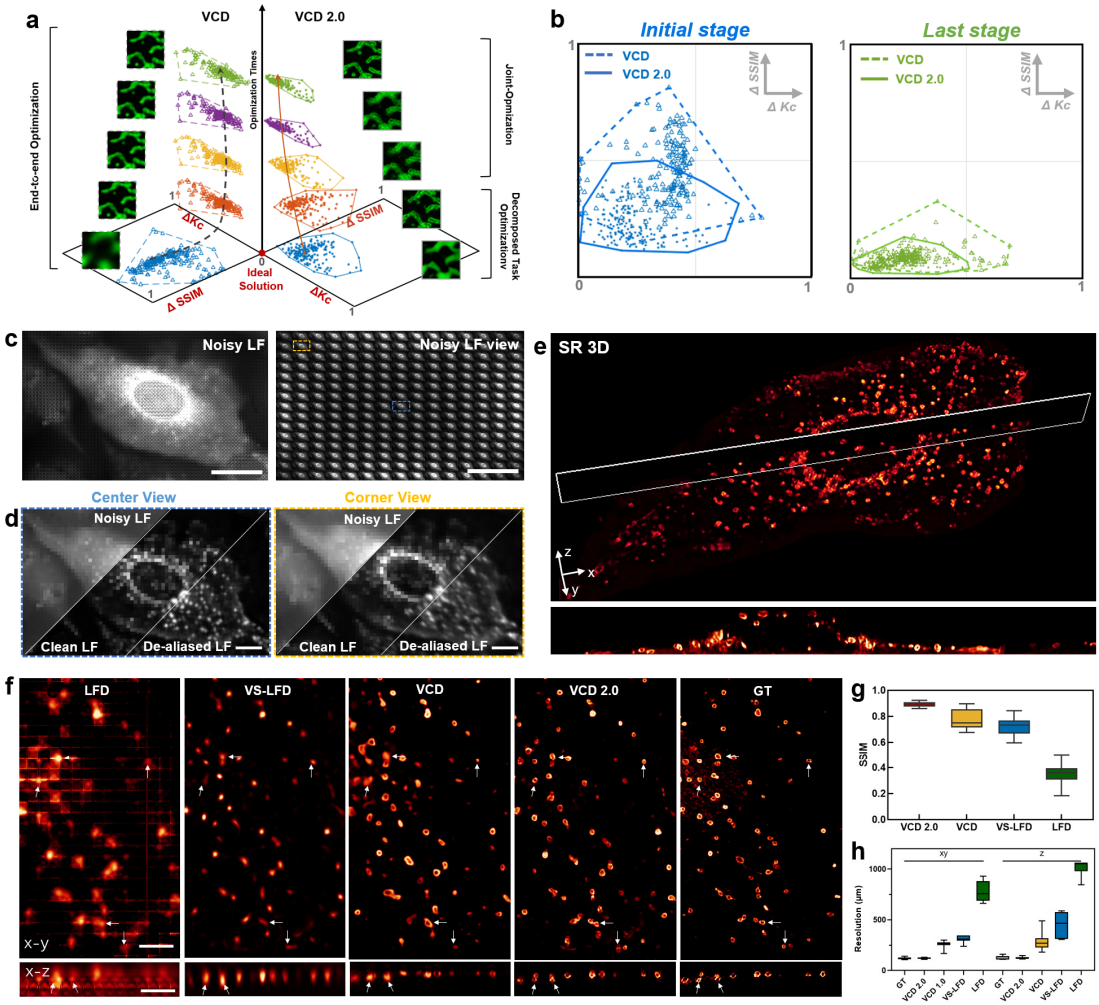

Figure 2. The network performance of VCD 2.0. a, The comparative NCPP map evaluating the fitting capability of previous end-to-end VCD network and VCD 2.0 network during the optimization process. The fitting ability of the network is comprehensively quantified by the fidelity (ΔSSIM, with 0 indicating the best fidelity) and spatial resolution (ΔKc, with 0 indicating the best resolution) of the 3D reconstructions throughout the entire training process, where ΔKc and ΔSSIM are the difference values of cut-frequency and multiscale structural similarity between the 3D reconstructions and ground truths, respectively. The ideal solution, reflecting a lossless 3D reconstruction, is positioned at (0,0). 196 ROIs are calculated for statistical analysis of K_c and SSIM (hollow triangle: VCD, solid circle: VCD 2.0). The shorter distance to the center axis represents the less quality difference as compared with ground truth. **b**, The 2 charts show the network performance at initial optimization stage (10th epoch) and final stage (well-convergence at 250th epoch). **c,** The experimental Noisy LF image of lysosome outer membranes in fixed U2OS cell. Its separated views are shown in the right panel. **d,** The Noisy, Denoised, De-aliased LF views (indicated by dotted boxes in **c**) and **e,** the 3D SR results from three sub-networks of VCD 2.0. Scale bar, 5 μm. **f,** The comparison of reconstructions by VCD 2.0 and current leading light-field reconstruction approaches including LFD, VS-LFD and VCD. The white arrows indicate the noticeable errors in LFD, VS-LFD and

VCD results whereas being accurately resolved by VCD 2.0. Scale bar, 2 μm. **g,** Structure similarity (SSIM) metric quantitatively comparing the fidelity of VCD 2.0 and other approaches with using enhanced Airyscan data (GT) as reference (n= 15 volumes). **h,** Decorrelation analysis quantifying the lateral and axial resolution of reconstructions by VCD 2.0, LFD, VS-LFD, VCD and GT. n=40 volumes were analyzed at both xy and xz planes.

Comparative performances of VCD-LFM 2.0 and other approaches

VCD-LFM 2.0 enabled 4D imaging of large-scale mitochondrial dynamics in dozens of cells with a volumetric imaging rate up to 200 Hz and a field of view (FOV) of ~ 220 × 220 × 10 μm^3 (Fig. 3a). We compared the reconstruction of VCD 2.0 with those from the alternative LFM reconstruction approaches including VS-LFD [22] and VCD[26] based on end-to-end network model (Fig. 3b). While VCD 2.0 reconstruction clearly showed the outer membranes of mitochondria, these structures remained unresolvable in LFD, VS-LFD and VCD reconstructions, owing to their suboptimal resolutions and noticeable artifacts. VCD 2.0 yielded a near isotropic lateral and axial resolutions of ~120 nm, which were compared with 230 nm and 370 nm by VS-LFD, 180 nm and 350 nm by VCD, respectively (Fig. 3c). The hundred-hertz light-field imaging rate together with subcellular-resolution reconstruction thus allowed the visualization of the fast morphology changes of the mitochondria, such as fission and fusion, occurred in three dimensions and at milliseconds time scale (Fig. 3d). Since LFM only requires light exposure once per volume, its phototoxicity is notably lower than the scanning-based 3D microscopy (Fig. 3e). Meanwhile, the inclusion of denoising module in VCD 2.0 ensures stable light-field reconstruction from low-exposure, noisy measurements (Supplementary Fig. 8). As a result, low-phototoxicity imaging combined with low-exposure reconstruction capability together permitted high spatiotemporal resolution live-cell imaging in long term, yielding over 40000 SR volumes with less than 50% photobleaching. In sharp contrast, plane-scanning-based light-sheet fluorescence microscopy (LSFM)[31] and point-scanning-based Airyscan microscopes[13] suffered from noticeable photobleaching after imaging merely 500 and 50 volumes, respectively (Fig. 3e, f, Method). We compared the resolution and volumetric speed of our VCD-LFM 2.0 with current leading 3D fluorescence microscopy techniques for live-cell imaging, including scanning-based SR microscopes such as Airyscan microscope, Instant SIM (iSIM)[8] and LLSM-SIM[9] and LFM modalities such as sLFM[19] , Fourier light-field microscopes (FLFM)[21] and VCD-LFM[26] (Fig. 3g). While scanning-based SR microscopes exhibit trades-off among spatial resolution and imaging speed, VCD-LFM 2.0 has apparently enlarges this limitation, showing spatial resolution as well as volumetric speed far superior to not only SR microscopes, but also other LFM modalities (Fig. 3g).

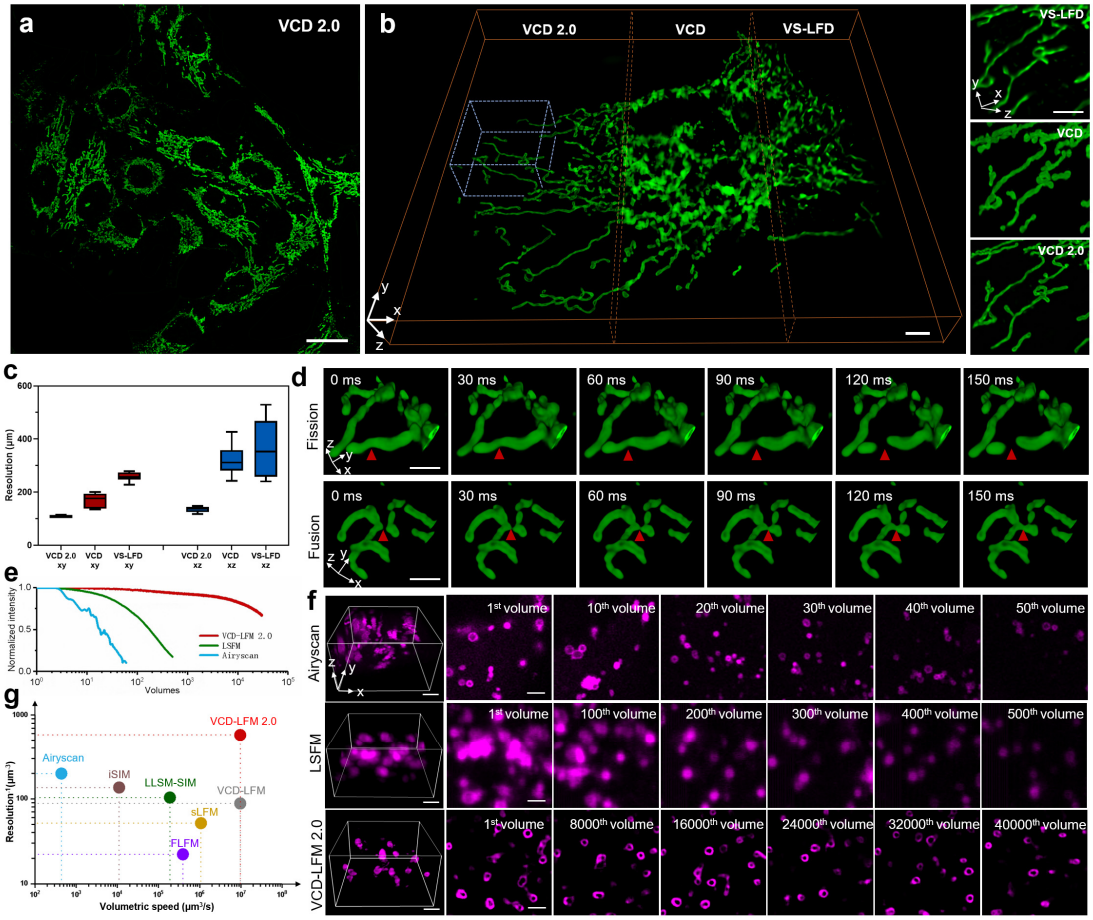

Figure 3. The performance of VCD-LFM 2.0. a, The volume rendering of the large-scale mitochondrial dynamics in dozens of live U2OS cells reconstructed by VCD 2.0 within a large FOV of 220×220×10 μm^3 using a 60×/1.3 NA objective. Scale bar, 20 μm. **b,** The volume rendering of the reconstructed results of mitochondrial outer membranes in live U2OS cells by VCD 2.0, VCD, VS-LFD. Insets show the magnified volume renderings of the ROI indicated by blue dotted box. Scale bar, 5 μm. **c,** Decorrelation analysis quantifying spatial and axial resolution of VS-LFD, VCD and VCD 2.0. n= 20 volumes were analyzed at both xy and xz planes. **d,** Time-lapse 3D visualization captures the rapid morphological transformations of the mitochondrial outer membrane occurred at milliseconds time scale, illustrating both the processes of mitochondrial fission and fusion. Scale bar, 2 μm. **e,** Comparisons of the photobleaching rates between VCD-LFM 2.0, LSFM and Airyscan. **f,** Corresponding images of lysosomes in live U2OS cells imaged via Airyscan (4 minutes every volume for a whole cell), LSFM (4 seconds every volume for a whole cell) and VCD-LFM 2.0 (5 ms every volume for at least a whole cell). Scale bar, 2 μm. **g,** The comparisons of the performance in resolution and volumetric imaging speed between Airyscan, iSIM, LLSM-SIM, FLFM, sLFM, VCD-LFM, and VCD-LFM 2.0.

Dual-color VCD-LFM 2.0 for 5D *intoto* imaging and quantification of lysosomes-mitochondria interactions

With the high spatiotemporal resolution of VCD-LFM 2.0, we conducted simultaneous 5D (3D space + time + spectrum) SR imaging of the outer membranes of lysosomes (tagged with EGFP-rab7) and mitochondria (tagged with Tomm20-EGFP) in live U2OS cells at a volume rate of 50 Hz (Fig. 4a). The isotropic subcellular resolution presented in five dimensions then permitted *intoto*

visualization of the Lysosome-Mitochondria (Lyso-Mito) interactions (Fig. 4b, c). While 2D microscopes merely provided projection images of Lyso-Mito contacts, our ability to reconstruct 3D processes eliminated false judgments and yielded more accurate analytic results (Fig. 4d). By measuring the distance between lysosome and mitochondria in 2D and 3D (Method), respectively, we identified a 24 % margin of inaccuracy in 2D results (Fig. 4e, 10 false cases in 41 events from 17 cells), proving the significance of VCD-LFM 2.0 for investigating organelle interactions in 4D. Recent research has revealed that lysosome-mitochondria contact sites may serve as a markers for mitochondrial fission[32]. We scrutinized 35 fission cases and 31 fusion cases from 17 cells reconstructed by VCD 2.0 to measure the proportion of Lyso-Mito contact in the mitochondrial fission / fusion events. The proportions of Lyso-Mito contacts at mitochondrial fission and fusion sites were 48 % and 41% in our statistics, respectively. These rates are found significantly lower than those previously identified from 2D images (Fig. 4f). With the high spatiotemporal resolution of VCD-LFM 2.0, we also analyzed whether such Lyso-Mito contact would affect the speed of mitochondrial fission and fusion. The results showed a strong positive correlation between the fission / fusion speed and lysosome contacts, indicating that the lysosomes could indeed accelerate the mitochondrial fission / fusion through their interactions (Fig. 4g).

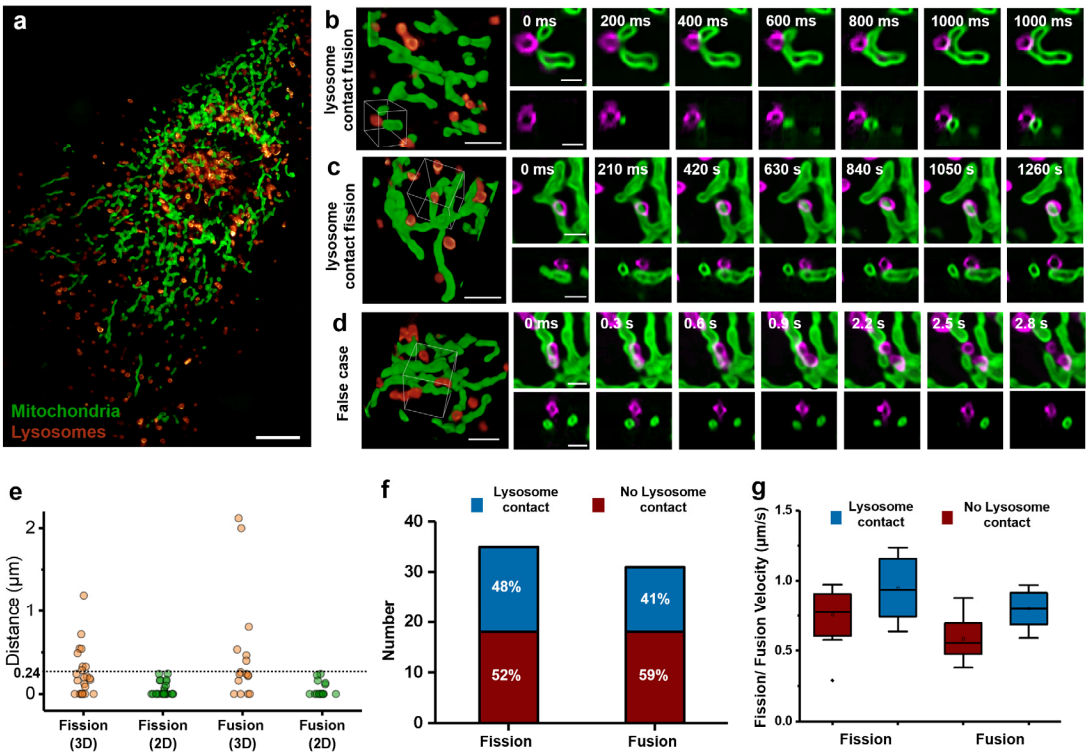

Figure 4. VCD-LFM 2.0 enabling 5D *intoto* imaging and quantitative analysis of mitochondria-lysosomes interactions at high spatiotemporal resolution. a, A dual-color volume rendering of the outer membranes of mitochondria (tagged with Tomm20-EGFP) and lysosomes (tagged with EGFP-rab7) in a live U2OS cell. Scale bar, 10 μm. **b, c,** Time-lapse images of typical lysosomes-mediated mitochondrial fusion and fission events, respectively. Volume rendering shows the first volume of the events in 3D. The xy and xz slices of the ROI indicated by white boxes in the volume rendering accurately show the rapid interactions between the lysosome and the fission / fusion sites of mitochondrial. **d,** An example of a false positive case where the lysosome appears to be in contact with the fission sites in 2D

projection, but in fact there is a significant distance between them in 3D observation. **e,** The 3D distances between the lysosomes and mitochondrial contact sites reveal the inaccuracies in the identification of Lyso-Mito contacts in 2D. The dashed line indicates the close proximity (<240 nm) between lysosomes and mitochondrial constriction sites that can be identified as Lyso-Mito contacts. **f,** The proportion of mitochondrial fission and fusion events occurred with and without Lyso-Mito contacts identified in 3D. **g,** The velocity of mitochondrial fission and fusion is significantly higher when in contact with lysosomes.

Long-term live imaging and quantitative analysis of mitochondrial fates across cell cycles

The low phototoxicity and robust denoising capability of VCD-LFM 2.0 enable 3D imaging of mitochondria evolution (tagged with Cox4-EGFP) in live *U2OS* cells across a long timescale up to 60 hours. Considering both rapid mitochondrial dynamics and their long-term fates need to be studied, we specifically designed an automatic imaging strategy to track the high-speed fission and fusion process throughout the entire cell cycle with minimal phototoxicity (Fig. 5a, Method). During each cycle of 4 hours, we continuously imaged the mitochondria for 30 minutes across 8 FOVs, with exposure time of 500 ms and a temporal resolution of 10 seconds (Fig.5b). During this 30-min mid-term observation window, the imaging pipeline included the following steps: (i) Implementing a VCD-based focusing strategy to correct sample drift along z axis. Mitochondrial were imaged and quickly reconstructed by a lite-version VCD network to calculate the z-drift distance, which would be further corrected by moving the objective. This process typically takes merely 1 second. (ii) Imaging the mitochondria dynamics at fixed depth of interest for 30 minutes, with an exposure time of 500 ms and interval time of 10 seconds. Meanwhile, we imaged *in-situ* chromosomes of the same FOV for identifying the related cell stage with an exposure time of 500 ms and interval time of 100 seconds. After 36-hours observation containing 9 cycles of 30-minture continuous recording, we discovered diverse viabilities of the cells within one batch, identifying both inertia without cell division and active 2-generation divisions across one entire cell cycles (Fig. 5c).

Smart VCD-LFM 2.0 of live cell as such included capture of instantaneous mitochondrial fission / fusion events (500 ms), lineage tracing of single mitochondrion's change in middle term (30 minutes), and evolution mapping of all mitochondrial in long term (60 hours), thereby enabling *intoto* investigation of the mitochondrial fates during entire cell cycle. For example, up to seven generations of the fissions / fusions of an individual mitochondrion were recorded during the 2nd 30-minute observation. Each fissions / fusions activities of the selected mitochondrion and their resulting later generations were successfully visualized in three dimensions (Fig. 5d). We casted lineage tracing of these mitochondrion activities (Fig. 5e), showing that the later generations of the source mitochondrion (the Magenta ones in Fig. 5e) tended to marry with their relatives with only one exception being fused with a foreign mitochondrion (the yellow one in Fig. 5e). Furthermore, we followed the fates of the daughter mitochondria from peripheral and midzone fissions at whole-cell scale. After analyzing 108 / 113 mitochondrial at G1 / G2 stages, we validated that while most of the small peripheral daughter mitochondria were excluded from further fusions or divisions, a conclusion which is consistent with previous study[33], the rates of no event occurred in these small peripheral daughter are significantly different between G1 (92%) and G2 (80.7%) phases (Fig. 5f, Supplementary Fig. 9). This is probably because the G2 phase is the phase before the cell division, in which the mitochondria are more active to prepare for the cell division[34]. Furthermore, VCD-LFM 2.0 results also allowed us to study the morphology changes of mitochondria, including variations in major and minor axis length, throughout two entire cell cycles (Fig. 5g, Supplementary

Fig. 10). The substantial variations in both morphology and event rates of mitochondria underscore the risk of drawing unreliable conclusions relying on the images obtained in a short time window.

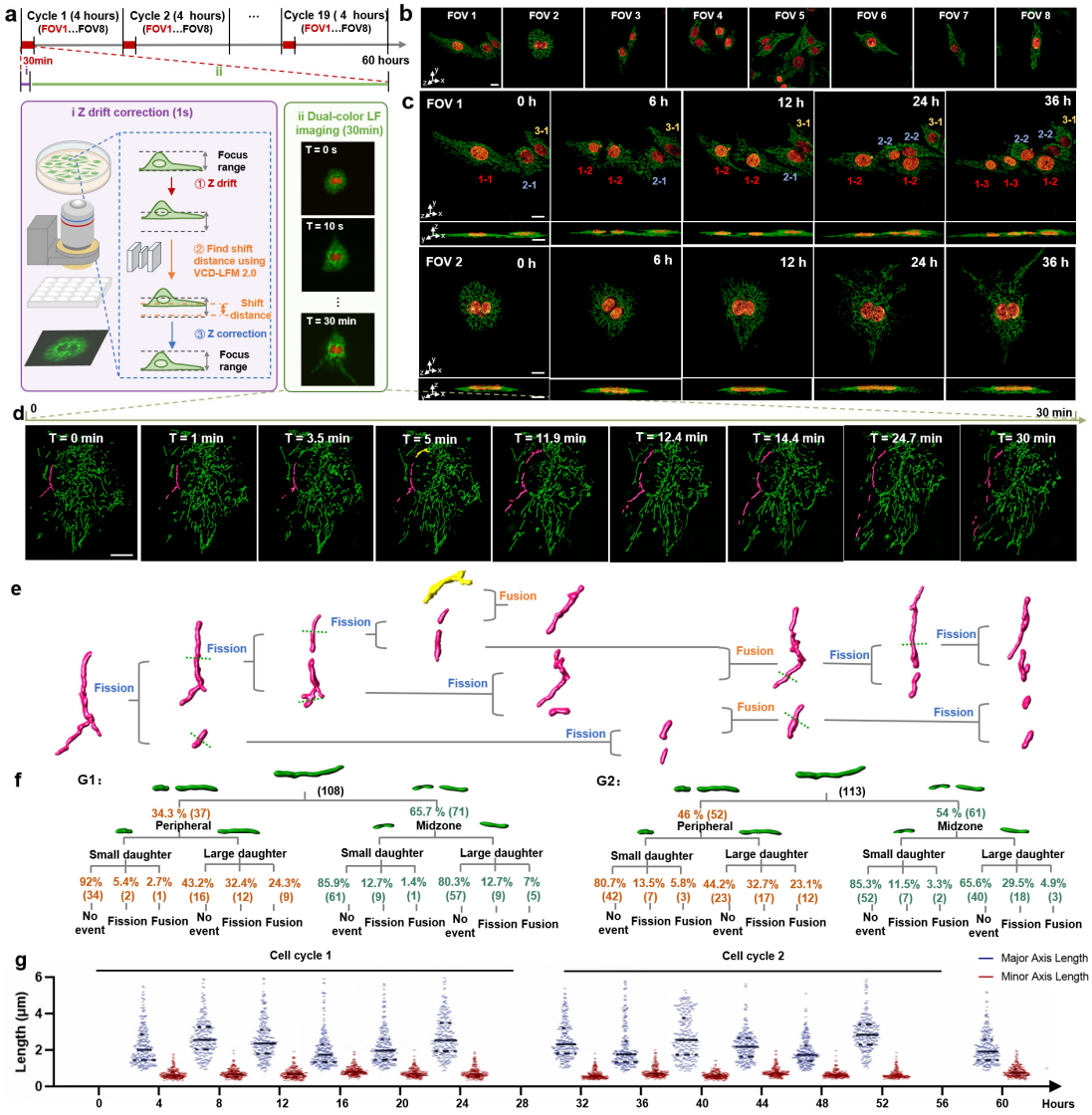

Figure 5. Long-term imaging of mitochondrial fates and morphology changes across cell cycles using VCD-LFM 2.0. a, Automatic VCD-LFM 2.0 imaging pipeline designed for the long-term imaging (up to 60 hours) of mitochondria (tagged with Cox4-EGFP) and chromosomes (tagged with H2B). In each cycle of 4 hours, the cells were continuously imaged for 30 minutes. The 30-min observation majorly includes the following steps: (i) An instant VCD-based auto-focusing strategy to correct sample drift along the z axis. Quick light-field imaging and reconstruction of mitochondria were performed to calibrate the z-axis drift of the system. The drift was then corrected by moving the objective using piezo scanner. (ii) VCD-LFM 2.0 imaging of the mitochondria dynamics and chromosomes in cell nuclei for 30 minutes. 8 FOVs were sequentially imaged in each cycle with an interval time of 10 seconds for mitohchondria and 100 seconds for chromosomes and an exposure time of 500 ms for both. **b,** The MIPs of 8 FOV imaged in the first cycle. **c,** The volume renderings of time-lapse images visualizing diverse cell viabilities with one undergoing 2-generation divisions (the first FOV in b) and another showing inertia without cell division (the second FOV in b) during the same 60-hours imaging. The inset numbers indicate three cells and their later generations. **d,** The 3D visualization of a representative mitochondrion during 30 minutes. The

mitochondria are highlighted in the 3D rendering of whole cell. **e,** Detailed lineage tracing of the selected single mitochondrion with 7 generations traced. The mitochondria daughters rendered by magenta were from the same source mitochondrion (leftmost one) while the yellow rendering was a foreign mitochondrion. **f,** Schema depicting the different fates of the daughter mitochondria in different stages of interphase G1 and G2 in the cell cycle. Mitochondria were tracked for 30 minutes. Both the quantities and portions of the specific mitochondria events were calculated. **g,** The morphology changes (Major axis length and minor axis length) of mitochondria during two entire cell cycles throughout 60 hours. Scale bar, 10 μm.

Discussion

Sustained observation of the subcellular biological dynamics at their physiological status is essential to investigating diverse organelle functions and their interactions in cell biology. This is difficult for either scanning-based SR microscopes with suboptimal speed and photon invasion, or high-speed light-field microscopes with limited single-cell resolution. VCD-LFM 2.0 circumvents this compromise between speed, resolution and photon efficiency based on the new development of deep-learning VCD pipeline now with efficient solution seeking and strong generalization abilities. We'd like to summarize the following technical advances of VCD-LFM 2.0 enabled by our new developments of light-field-aware VCD 2.0 model and adaptive tuning strategy. First, VCD 2.0 includes new light-field-aware model, data generation procedure and training optimization strategy to maximize the network's solving ability to light-field inversion problem with huge solution space, thereby efficiently pushing the resolution of LFM beyond diffraction limit. It decouples the highly complex inversion problem in LFM imaging, achieving a SBP increase over 500 fold. Meanwhile, the strong denoising by VCD 2.0 in conjunction with photon-efficient LFM modality together enables ultra long-term cell observation under physiological status, wiping off the phototoxicity issue existed commonly in scanning-based SRM with allowing over 10-fold more measurements. Furthermore, the intrinsically weak generalization ability of a supervised network is also improved in VCD 2.0 by using readily-accessible WF and LF measurements to instantly tune the model adaptive to new types of signals. It is worth noting that VCD 2.0 has been wrapped into a fully opensource program with a user-friendly GUI provided. Together with the compact LFM add-on we devised, VCD-LFM 2.0 allows the accomplishment of 3D SR imaging of live cells through a simple retrofit on inverted fluorescence microscopes widely-used in those laboratories without much optics or computation skills. With VCD-LFM 2.0, we four-dimensionally imaged highly dynamic interactions between lysosomes and mitochondrial membranes at near-isotropic120-nm spatial resolution while maintained a video-rate temporal resolution. The high-quality 4D spatiotemporal patterns revealed by VCD-LFM 2.0 are quite different with those previous results obtained by 2D-SIM, indicating an actually weaker interaction of lysosomes at the moment of mitochondrial fission / fusion. We also demonstrated SR fluorescent imaging of live cells with probably longest observation time of 60 hours and up to 40000 volumes recorded. The image results support lineage tracing of both short-term changes and long-term fates of mitochondria for the first time at single-mitochondrion resolution and across an entire cell cycle.

In addition to standard VCD-LFM 2.0, we envision that the stepwise and fine-tuning strategies of VCD 2.0 are also helpful to the improvement of other LFM modalities, such as Fourier LFM and sLFM[18, 19, 21], without the need of extra hardware. While the fine-tuning strategy notably enhances the generalization ability of the supervised network, it is envisaged that the incorporation of more

data, especially those with higher resolution and volumetric prior, will further improve the performance of fine-tuning. Meanwhile, it's noted that VCD 2.0 still requires a considerable amount of high-resolution label data to initiate the training of base model. We anticipate that this dependency on external data will be circumvented by a full-cycle self-supervision reconstruction strategies, in which the *in-situ* WF measurements of the samples can be used as internal training labels to further improve the resolution of LF views and lead to reconstruction with higher quality. A completely unsupervised deep-learning light-field reconstruction could be also made possible by using implicit network representation (INR)[35, 36]. It's promising because the intrinsic view synthesis capabilities in INR is indeed helpful to the enhancement of axial resolution and reduction of reconstruction artifacts. But the learning-and-representing mode for each single LF image is very time consuming and computationally demanding. INR-based reconstruction needs to solve this efficiency problem before it becomes as practical as DL and deconvolution approaches. Given that a lot of *in vivo* biological dynamics occur in the deep tissues, we also expect the combination of VCD 2.0 with more advanced LFM techniques, e.g., LFM with AO or two photon excitations, or even a system working at NIR-II window, to *intoto* observe the subcellular dynamics inside living animals. Taken together, we believe VCD-LFM 2.0 has strongly pushed the spatiotemporal limit of current SRM / 3D fluorescence microscopy and could be a powerful and accessible tool that helps a broad range of biology research to demystify the worlds inside the cell. The paradigm shift it shows could be also beneficial to propelling other microscopy techniques towards deeper, faster and clearer imaging.

References

1. Valm, A.M. et al. Applying systems-level spectral imaging and analysis to reveal the organelle interactome. *Nature* **546**, 162-167 (2017).
2. Sahl, S.J., Hell, S.W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat Rev Mol Cell Biol* **18**, 685-701 (2017).
3. Choquet, D., Sainlos, M. & Sibarita, J.B. Advanced imaging and labelling methods to decipher brain cell organization and function. *Nat Rev Neurosci* **22**, 237-255 (2021).
4. Schermelleh, L. et al. Super-resolution microscopy demystified. *Nat Cell Biol* **21**, 72-84 (2019).
5. Schermelleh, L., Carlton, P. M., Haase, S., Shao, L., Winoto, L., Kner, P., ... & Sedat, J. W. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320.5881**, 1332-1336 (2008).
6. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813 (2008).
7. Shao, L., Kner, P., Rego, E.H. & Gustafsson, M.G. Super-resolution 3D microscopy of live whole cells using structured illumination. *Nat Methods* **8**, 1044-1046 (2011).
8. York, A.G. et al. Instant super-resolution imaging in live cells and embryos via analog image processing. *Nat Methods* **10**, 1122-1126 (2013).
9. Chen, B.C. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).
10. Li, D. et al. ADVANCED IMAGING. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).
11. Wu, Y. et al. Multiview confocal super-resolution microscopy. *Nature* **600**, 279-284 (2021).
12. Li, X. et al. Three-dimensional structured illumination microscopy with enhanced axial resolution.

Nature Biotechnology **41**, 1307-1319 (2023).
13. Huff, J. The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. *Nature Methods* **12**, i-ii (2015).
14. Gao, L. et al. Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens. *Cell* **151**, 1370-1385 (2012).
15. Broxton, M. et al. Wave optics theory and 3-D deconvolution for the light field microscope. *Optics express* **21**, 25418-25439 (2013).
16. Cohen, N. et al. Enhancing the performance of the light field microscope using wavefront coding. **22**, 24817-24839 (2014).
17. Prevedel, R. et al. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature methods* **11**, 727-730 (2014).
18. Guo, C., Liu, W., Hua, X., Li, H. & Jia, S. Fourier light-field microscopy. *Opt Express* **27**, 25573-25594 (2019).
19. Wu, J. et al. Iterative tomography with digital adaptive optics permits hour-long intravital observation of 3D subcellular dynamics at millisecond scale. *Cell* **184**, 3318-3332 e3317 (2021).
20. Zhang, Z. et al. Imaging volumetric dynamics at high speed in mouse and zebrafish brain with confocal light field microscopy. *Nat Biotechnol* **39**, 74-83 (2021).
21. Hua, X., Liu, W. & Jia, S. High-resolution Fourier light-field microscopy for volumetric multi-color live-cell imaging. *Optica* **8**, 614-620 (2021).
22. Lu, Z. et al. Virtual-scanning light-field microscopy for robust snapshot high-resolution volumetric imaging. *Nat Methods* (2023).
23. Han, K. et al. 3D super-resolution live-cell imaging with radial symmetry and Fourier light-field microscopy. *Biomedical Optics Express* **13** (2022).
24. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods* **15**, 1090-1097 (2018).
25. Qiao, C. et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nat Biotechnol* (2022).
26. Wang, Z. et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nat Methods* **18**, 551-556 (2021).
27. Zhu, L., Yi, C. & Fei, P. A practical guide to deep-learning light-field microscopy for 3D imaging of biological dynamics. *STAR Protoc* **4**, 102078 (2023).
28. Zhu, T. et al. High-speed large-scale 4D activities mapping of moving C. elegans by deep-learning-enabled light-field microscopy on a chip. *Sensors and Actuators B: Chemical* **348** (2021).
29. Lempitsky, V., Vedaldi, A. & Ulyanov, D. in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 9446-9454 (IEEE, 2018).
30. Guo, Y. et al. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 5407-5416 (2020).
31. Sapoznik, E. et al. A versatile oblique plane microscope for large-scale and high-resolution imaging of subcellular dynamics. *Elife* **9** (2020).
32. Wong, Y.C., Ysselstein, D. & Krainc, D. Mitochondria-lysosome contacts regulate mitochondrial fission via RAB7 GTP hydrolysis. *Nature* **554**, 382-386 (2018).
33. Kleele, T. et al. Distinct fission signatures predict mitochondrial degradation or biogenesis. *Nature* **593**, 435-439 (2021).

Methods

LFM Add-on device

An add-on device was designed to provide easy light-field imaging of live cells on a commercial inverted fluorescent microscope (Olympus IX73), thus enabling wider applications for biology researchers without optics background. The device only requires the assembling of off-the-shelf lenses into a customized mounting base without any complex alignment of the optical path. To precisely position MLA at the native image plane of the microscope, a customized lens tube 1 was designed to fix the MLA (RPC Photonics, MLA-S100-f28) at the Flange focal distance away from the camera port. Then a pair of relay lens (Thorlabs, TTL100-A) was integrated into another customized lens tube 2 to 1:1 relay the back focal plane of MLA (light-field image plane) onto the camera sensor. A high-precision zoom housing (SM1ZM, Thorlabs) was used to interconnect the 2 lens tubes and allow fine focusing of the light-field image plane (Supplementary Fig. 6a, b). The complete add-on device containing the two lens tubes and the zoom housing was then mounted between the microscope's camera port and the sCMOS camera (Photometrics Prime BSI Express), as shown in Supplementary Fig. 6c, converting the ordinary inverted microscope into an advanced light-field microscope.

VCD 2.0 network

Previous deep-learning-based light-field reconstruction network enabled resolution enhancement with training the light-field projections and high-resolution ground truths in an end-to-end manner. This one-step network produces hallucination or artifacts when solving hard inverse problem (ref multi stage super resolution). Similar with the physical process of light-field image generation, our VCD 2.0 network disentangles this inverse problem into four sub-problems: denoising, de-aliasing, 3D reconstruction. The three sub-problems are divided and conquered by jointly optimizing task-specific networks with weighted objective function.

The architectures of the denoising sub-network are based on channel-attention mechanism[37, 38] to reweight the extracted feature from input views with different SNR. This denoise module can be formulated as

$$x_1 = Fusion\left(Attention^{(2)}\left(Conv_{dilated}(x)\right) + Attention^{(1)}\left(Conv_{dilated}(x)\right)\right)$$

where $Conv_{dilated}$ means dilated convolution that enlarge receptive field. $Fusion$ is feature element-wise adding. In order to suppress the noise fluctuation, L1 loss is used as loss function during denoising stage:

$$Loss_{denoising} = MAE(x_1, y_1)$$

Due to the angular and spatial information encoded in LFs, directly applying traditional 2D super resolution network like ResNet[39], RCAN[40] on extracted views induces the under-exploration of the rich angular features. For LF de-aliasing, we adopt two types of dilated convolution operations on 4D light-field images to separate 2D spatial information and 2D angular information[41]. These two convolution and series of activation function and feature fusion layers form Spatial-Angular (SA) block. According to sequential connection of four SABlocks and Upsampling convolution

, the missing spatial information caused by sparse sampling rate of microlens array can be recovered:

$$x_2 = Upsampling(Fusion(Conv_{dilated}(x_1) + \underbrace{SABlock(\ldots SABlock(x_1))}_{n\times SABlock}))$$

where $Fusion$ is channel concatenate operation. Considering the potential parallax information across different views, epipolar constraint ($EPI_{consistency}$) on de-aliasing prediction is proposed to keep geometry relationship among various views. The objective function of de-aliasing can be formulated as:

$$Loss_{de-aliasing} = MSE(x_2, y_2) + EPI_{consistency}(x_2, y_2)$$

where MSE means mean squared error between network output and de-aliased LF.

For 3D reconstruction model, we extended original VCD to handle super-resolution reconstruction. According to adding MultiRes blocks[42] and modifying activation function, the capability of VCD network is enhanced achieving lower fitting errors when solving SR reconstruction:

$$x_3 = Unet_{MultiRes}(Conv(SubpixelConv(x_2)))$$

Correspondingly, for 3D reconstruction task, we adopt MSE loss and lateral gradients loss to preserve the structure information of ground truth.

$$Loss_{Reconstruction} = MSE(x_3, y_3) + Gradient_{xy}(x_3, y_3)$$

These three networks are optimized synchronously based on designed weighted objective function:

$$Loss = \lambda_1 Loss_{denoising} + \lambda_2 Loss_{de-aliasing} + \lambda_3 Loss_{reconstruction}$$

where λ_1, λ_2, λ_3 are weight coefficients for different network and set as 0.1, 0.2, 0.8, respectively.

For VCD 2.0 network training, all models were trained on NVIDIA GeForce RTX 3090 or 4090 with Python version 3.8 and Tensorflow version 1.15.0. The training process of a VCD 2.0 model was composed of two stages. First, pretrain 3 sub-networks, including denoising network, de-aliasing network and 3D reconstruction network. Specifically, denoising network was trained with a learning rate of 1×10^{-4} and a training epoch of 51. For de-aliasing network and 3D reconstruction network, they used the same learning rate (5×10^{-4}) but different training epoch (51 and 151, respectively). The pretraining process lasted for 16 hours. Secondly, train VCD 2.0 network with these pretrained model under the designed weighted loss. During this decomposed-progressive optimization, we adopted step-based learning rate schedules with a decay factor of 0.5 and a decay step of 25 to enhance model performance. The initial learning rate was set to 1×10^{-4} and the total training epoch was 101. The training time of second phase was roughly 2 hours. For more details about VCD 2.0 network implementations, see Supplementary Note 1,3 and our open-source code.

Fine-tuning strategy for reconstruction of unseen samples

Unlike previous supervised learning strategy[26] only solving the inverse problem for specific samples similar to the training data, VCD 2.0 adopted a fine-tuning strategy to overcome the hallucination and abnormal structure features when facing unseen data for trained model. We developed an adaptive-tuning strategy to reconstruct HiFi 3D signals from LF captures of unseen sample via the fine-tuning model, in which only an *in-situ* 2D WF image captured together with the LF imaging was required instead of the supervision of 3D label data from another scanning microscope. During the fine-tuning phase, we adopted an alternative optimization approach to re-update parameters of the trained VCD 2.0 model. The paired synthetic training data (e.g. the outer membrane of

mitochondria) containing 3D stack were used to provide the volumetric priors to ensure the fidelity of 3D prediction of original model while the experimental LF captures (termed LF_{unseen}) and corresponding WF images (termed WF_{unseen}) act as prior knowledge of new domain (e.g. the outer membrane of lysosome) to prevent data bias brought by trained model. For synthetic data training, the designed weighted loss of original VCD 2.0 was adopted:

$$Loss_{syn} = \lambda_1 Loss_{denoising} + \lambda_2 Loss_{de-aliasing} + \lambda_3 Loss_{reconstruction}$$

For the new domain transfer, reprojection loss was to provide constraints on the different structure of unseen samples:

$$Loss_{projection} = MSE(reprojection, WF_{unseen})$$

where $reprojection$ was computed from the sum and maximum projection of network 3D prediction of LF_{unseen}. To ensure the accuracy of this loss computation, the two detection modalities (LF and WF) were pre-aligned with the aid of fluorescent beads to yield paired LF images and registered WF images. The paired LF-WF dataset in fine-tuning phase contained ~3 cells. In this work, $Loss_{syn}$ and $Loss_{projection}$ were calculated every two consecutive iterations N and N+1, respectively. The training time of fine-tuning phase was typically 70~100 folds less than the consumption of building a new VCD 2.0 model, with a total mini-batch iteration of ~30,000 finished 10-15 minutes.

Training data processing pipeline

A data processing pipeline was developed to generate all training data (including Noisy LF, Clean LF, De-aliased LF and 3D SR) from the 3D SR data. In this paper, 3D SR data were acquired from AiryScan microscopy by imaging fixed cells. Through applying axial-to-lateral self-supervised isotropic learning to Airy scan results[24, 43], the airyscan image achieved an isotropic resolution of 120 nm. A SAS LFP pipeline was designed to create De-aliased LF images to guide the LF de-aliasing task. The 3D SR volumes were shifted by a distance less than the size of the microlens along the horizontal (both x and y respectively) and then were projected by convolving with 5D PSF of LFM to yield multiple light field projections that included more spatial information through multiple sampling. The shifted times depended on the under-sampling rate of the microlens, which is 5x5 times with a step of 3 pixels for the setup of Nnum = 15 in this paper. The SAS LF projections were then realigned to generate the De-aliased LF images according to the arrangement of the light. Then, Clean LF images (the center ones with shift = 0 of the De-aliased LF images) were used to guide the denoising task. To generate Noisy LF with accurate SNR matched with the experimental LF, the Clean LF images was normalized to the same range of experimental LFs and various noises in the range of the highest and lowest noise in the experimental LFs during long-term observation were added to the Clean LFs to generate Noisy LFs. The SNR was calculated by: $\mathrm{SNR} = \frac{\mu}{\sigma} = \frac{S - I_{background}}{\sqrt{S - I_{background} + \sigma^2_{background}}}$, where S is the average signal intensity value in the image, $I_{background}$ and $\sigma_{background}$ are the mean and the standard deviation of the background, respectively. Finally, we obtained De-aliased LF images, Clean LF images, Noisy LF images, which were all conveniently generated from the same 3D SR images, for the VCD 2.0 network training.

Network comprehensive performance pyramid (NCPP).

The NCPP visualized the model ability by quantifying the resolution and fidelity of VCD 2.0 and VCD inference results during model optimization. To clarify the fitting ability of these two models on label data, such resolution and fidelity metrics were derived from the cut-off frequency difference (ΔKc) and the structural similarity discrepancy (ΔSSIM=1-SSIM) between the 3D reconstruction of the network and corresponding GT , where Kc is computed by decorrelation analysis[44]. Both ΔKc and ΔSSIM range between 0 and 1, where values close to 0 indicate a favorable combination of resolution and fidelity, while values close to 1 signify a low-quality image reconstruction. In our study, 196 LF patches (Size: 360×360) of mitochondrial outer membrane data were reconstructed by VCD and VCD 2.0 during whole model training process. Specifically, we chose five different epochs (10, 100, 200, 210, and 300) to track this comprehensive performance of two models: For VCD 2.0, 10~100 epochs denoted the decomposed optimization on each sub-network while 200~300 epochs represented the progressive optimization process with multi-stage data (Noisy LFs, Clean LFs, De-aliased LFs and 3D SR); For VCD, 10~300 denoted the optimization process under SR stacks and corresponding Noisy LFs. By calculating ΔKc and ΔSSIM of network inferences under the epochs, the performance scattering plots in network optimization process were produced (Fig. 1). Besides, the convex hull computation of these scattering plots was used to generate the boundary line of statistics distribution and evaluate the performance deviation.

Assessment of the resolution, fidelity and SBP of VCD 2.0 network

We used the structural similarity (SSIM) function in Matlab to assess the fidelity of our VCD 2.0 reconstructions and other LFM techniques including LFD, VS-LFD, VCD utilizing the AiryScan data as reference in both the xy and yz planes. We applied decorrelation analysis to quantify the resolution of VCD 2.0 results and other 3D microscopy implementations. The analysis was conducted using MATLAB. The axial resolution was measured by the sectorial resolution mode of the decorrelation analysis[44], in which only the resolution in a narrow sectorial region along z direction was calculated. The space-bandwidth product (SBP) was calculated by $SBP=FOV/(0.5\delta)^3$ with δ as the system's resolution, and the factor 0.5 stemming from the Nyquist-Shannon sampling theorem.

Cell culture and fluorescence labeling

U2OS cells were grown in culture medium containing McCoy's 5A medium (Thermo Fisher Scientific) supplemented with 1% antibiotic-antimycotic (Thermo Fisher Scientific) and 10% fetal bovine serum (Thermo Fisher Scientific) at 37°C with 5% CO_2 in a humidified incubator.

For labeling lysosomes in fixed U2OS cells, cells were first transfected with EGFP-Rab7A using Lipofectamine 2000 according to the standard protocol and cultured at 37 °C with 5% CO_2 for an additional 24 h. Before imaging, the cells were fixed with 2% glutaraldehyde for 20 min.

For labeling tubulin in fixed U2OS cells, cells were seeded onto coverslips at 37°C with 5% CO_2 for 12 h. Before fixation, the cells were washed with phosphate buffered saline (PBS, Thermo Fisher Scientific) at 37°C and then treated with fixing buffer (containing 3% paraformaldehyde (Electron Microscopy Sciences), 0.1% glutaraldehyde (Electron Microscopy Sciences), 0.2% Triton X-100 (Sigma-Aldrich)) for 15 min, then incubated with 0.2% Triton X-100 for 15 min and blocked with blocking buffer (3% bovine serum albumin (Sigma-Aldrich) and 0.05% Triton X-100 (Sigma-Aldrich)) for 20 min at room temperature. After that, cells were incubated with an anti-alpha tubulin antibody (Abcam, 1:500 dilution) overnight at 4°C. Subsequently, the primary antibody was

removed and the cells were washed twice with PBS. Next, the cells were incubated with a secondary antibody (Abcam, labeled with Alexa Fluor 488, 1:400 dilution) for another 2 h at room temperature. The antibody was then removed and the cells were washed three times with PBS.

For labeling outer membranes of mitochondria and lysosomes in live U2OS cells, cells were first transfected with Tomm20-EGFP and Rab7A-mCherry2 using Lipofectamine 2000 according to the standard protocol and cultured at 37°C with 5% CO_2 for an additional 24 h. Before imaging, remove old media and add fresh media.

For labeling mitochondrial matrix and chromosomes in live U2OS cells, cells were first transfected with Cox4-EGFP and H2B-mCherry2 using Lipofectamine 2000 according to the standard protocol and cultured at 37°C with 5% CO_2 for an additional 8 h. After 6-8 hours of transfection, the cells were digested with 0.25% trypsin, seeded on cell culture dishes (20 mm diameter), and incubated for 36 hours at 37°C in 5% CO_2.

During the imaging process, cells were cultivated in phenol red free McCoy's 5A medium (customized, Boster Biological Technology) within the confocal dishes. To ensure a stable environment, the cell in confocal dishes were cultured in the live cell microscope incubation system (TOKAIHIT) to maintain a consistent temperature of 37°C and a 5% CO_2 atmosphere.

Live-cell imaging

Light-field imaging of the lysosome outer membrane in live U2OS cells was implemented using our add-on light-field device with a × 60/1.3NA objective (Olympus UPlanSApo60XS2) at a volumetric imaging rate up to 333 Hz. This high-speed light-field imaging of live cells continued for 2 minutes with less than 50 % photobleaching, yielding 40000 SR reconstructions using VCD-LFM 2.0. As a result, the fine deformation of lysosome outer membranes occurred in milliseconds could be observed by our VCD-LFM 2.0 in three dimensions. To compare the photon bleaching rate with other 3D microscopy implementations, we also imaged the lysosome outer membrane by Airyscan and LSFM microscope. The laser intensity and exposure time was carefully adjusted to ensure that the image SNRs by these microscopes were similar. To maintain sufficient SNR, Airyscan and LSFM required 4 minutes and 4s for imaging a whole cell, respectively. The bleaching rates were then calculated using Matlab. Multiple areas were cropped and a threshold was used to discriminate the signal areas and background areas automatically. The photon bleaching rate over time was finally calculated using the following equation: $\mathrm{I(t)} = I_{signal}(t) - I_{background}(t))/MAX(I_{signal}(t) - I_{background}(t)$, where I_{signal} and $I_{background}$ are the mean intensity value of the multiple regions of signal and background images, respectively.

Quantitative analysis of the interaction between mitochondria and lysosome

Interaction of mitochondrial and lysosomes was identified when the distance between a certain lysosome and the constriction site of a certain mitochondrion was measured smaller than 240 nm (close proximity) in at least three consecutive frames. We four-dimensionally captured such interaction between mitochondria and lysosomes in 17 live cells, where 41 ROIs that met the Lyso-Mito contact criteria in 2D MIP mode were identified, with 24 ROIs containing lysosome-mediated mitochondrial fission and other 17 ROIs containing mitochondrial fusion. Then, the real distance between lysosomes and mitochondria in these ROI was measured in true 3D mode using the commercial Imaris software. As a result, 10 false contacts were found in 2D results. We further calculated the fission and fusion velocity using self-written MATLAB code that included the

following steps:
(i) Obtaining the coordination of the mitochondrial endpoints. For each frame of the time-series data: first, use the 'imbinarize()' function in Matlab to binarize the image. Then, employ the 'bwskel()' function to skeletonize the binary image. Finally, use the 'endpoints' method in the 'bwmorph3()' function to mark the endpoints of the skeletonized image and record the endpoint coordinates of all mitochondria.
(ii) Tracking the endpoints and identify the trajectories related to mitochondrial fission and fusion events. The code uses Matlab's 'assignDetectionsToTracks' function to track the sequence of the recorded endpoint coordinates along the time axis for the cropped ROIs, providing the motion trajectory of all endpoint along the time axis. Then, two longest motion trajectories were filtered as the trajectories related to the fission or fusion events while the discrete trajectories were removed. As long as the discontinuities occurred between frames in tracking, the 'interp1()' function is applied to complete the trajectories.
(iii) Velocity Acquisition. For each motion trajectory, the code calculates the distance between endpoints in all adjacent frames. Since frame interpolation has been performed, the instantaneous velocity were calculated through dividing the distance by the time interval. The calculated sequence of instantaneous velocity along the time corresponding to each motion trajectory were then averaged to obtain the overall fission or fusion velocity of this ROI.

Experimental setup for long-term live-cell imaging

To observe the mitochondrial dynamics and using chromosomes to identify the cell stages, we used a 488/561 nm dual-band optical filter block (Chroma, 59904) in the microscope. Given that the cell division is highly sensitive to the phototoxicity, the intensity of LED (CoolLED, PE-800) were adjusted to 1% in 470 nm (30 μw) and 1% in 550 nm (30 μw). Since the tracking of high-speed fission and fusion event requires at least 10-seconds imaging rates and the entire cell cycle needs at least 48 hours observation, we designed a non-uniform imaging strategy for this long-term experiment. During this 30-min mid-term observation window, the imaging pipeline included the following steps: (i) Implementing a VCD 2.0-based focusing strategy to correct sample drift along z axis. Then the 3D distribution of the mitochondria was quickly reconstructed by VCD 2.0 network and compared with the original distribution to calculate the z-drift distance, which would be further corrected by moving the objective using piezo scanner (COREMORROW, P73.Z). This process typically takes merely 1 second. (ii) Imaging the mitochondria dynamics at fixed depth of interest for 30 minutes, with an exposure time of 500 ms and interval time of 10 seconds. Meanwhile, identifying the related cell stage through the imaging of chromosomes in cell nuclei with an exposure time of 500 ms and interval time of 100 seconds.

Tracking of the fates and morphology changes of mitochondria

We applied Mitometer[45] to track the mitochondria dynamics reconstructed by VCD 2.0. Then the confident tracks and the corresponding morphology parameters can be exported as a mat file. To obtain the morphology changes of mitochondria throughout entire cycles, we first extracted the morphology parameter including major axis length, minor axis length, Volume, Solidity from the mat file in each timepoint. Then we mapped these parameters along the time and observed the changes throughout entire cycles. To further track the fates of each mitochondrion, we used the self-written Matlab code to extract the fission and fusion events in the mat file of the confident tracks.

Then the mitochondria that exhibited the first fission were extracted with the fission sites being identified. We compared the volumes of two divided mitochondria in the next frame after fission and the smaller one will be defined as the small daughter. If the volume of the smaller one is less than 25% of the total length of two divided mitochondria, the fission will be defined as a peripheral fission, otherwise it is a midzone fission. Then the number of fission and fusion events occurred in these daughters were extracted and the daughters that undergo no fission or fusion were defined as no event. As a result, the event rates of these peripheral or midzone daughters were calculated and mapped.

Data availability. The datasets generated and analyzed in this study are available from the corresponding authors upon request.

Code availability. The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

References

34. Mishra, P. & Chan, D.C. Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Biol* **15**, 634-646 (2014).
35. Chen, A. et al. in Proceedings of the IEEE/CVF International Conference on Computer Vision 14124-14133 (2021).
36. Mildenhall, B. et al. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**, 99-106 (2021).
37. Zhang, Y. et al. in Proceedings of the European conference on computer vision (ECCV) 286-301 (2018).
38. Mo, Y., Wang, Y., Xiao, C., Yang, J. & An, W. Dense dual-attention network for light field image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology* **32**, 4431-4443 (2021).
39. He, K., Zhang, X., Ren, S. & Sun, J. in Proceedings of the IEEE conference on computer vision and pattern recognition 770-778 (2016).
40. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat Methods* **18**, 194-202 (2021).
41. Wang, Y. et al. in Computer Vision – ECCV 2020. (eds. A. Vedaldi, H. Bischof, T. Brox & J.-M. Frahm) 290-308 (Springer International Publishing, Cham; 2020).
42. Ibtehaz, N. & Rahman, M.S. MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks* **121**, 74-87 (2020).
43. Zhao, F. et al. Deep-learning super-resolution light-sheet add-on microscopy (Deep-SLAM) for easy isotropic volumetric imaging of large biological specimens. *Biomedical Optics Express* **11** (2020).
44. Descloux, A., Grußmayer, K.S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods* **16**, 918-924 (2019).
45. Lefebvre, A.E.Y.T., Ma, D., Kessenbrock, K., Lawson, D.A. & Digman, M.A. Automated segmentation and tracking of mitochondria in live-cell time-lapse images. *Nature Methods* **18**, 1091-1102 (2021).

Acknowledgements

We are grateful to X. Duan for providing us the fluorescent cell samples. This work was supported by the funding from National Key Research and Development Program of China (2022YFC3401102). National Natural Science Foundation of China (T2225014, 21927802, 62375095).

Author contributions

P.F., L.Z. and C.Y. conceived the idea. P.F. and D.L. oversaw the project. L.Z., and J.S. developed the optical setups and acquired the experimental images. L.Z., J.S. and C.Y. developed the programs. L.Z. and J.S. processed the images. L.C., Y.Z., Y.H., C.Z., J.W., G.N., J.T., Y.Z., L.Z., C.Y., J.S., M.H. and D.L. analyzed the data. L.Z., D.L, and P.F discussed and wrote the paper.

Competing interests

The authors declare no conflicts of interest.