Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy

Chang Qiao[1,2,3,4,8], Yunmin Zeng[1,8], Quan Meng[5,6,8], Xingye Chen[1,2,3,4,8], Haoyu Chen[5,6], Tao Jiang[5,6], Rongfei Wei[5], Jiabao Guo[5,6], Wenfeng Fu[5,6], Huaide Lu[5,6], Di Li[5], Yuwang Wang[7], Hui Qiao[1,2,3,4], Jiamin Wu[1,2,3,4], Dong Li[5,6,*], Qionghai Dai[1,2,3,4,*]

[1] Department of Automation, Tsinghua University, Beijing, 100084, China

[2] Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing, 100084, China

[3] Beijing Key Laboratory of Multi-dimension & Multi-scale Computational Photography, Tsinghua University, Beijing, 100084, China

[4] Beijing Laboratory of Brain and Cognitive Intelligence, Beijing Municipal Education Commission, Beijing, 100010, China

[5] National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China

[6] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

[7] Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, 100084, China

[8] These authors contributed equally

*Correspondence: lidong@ibp.ac.cn (D.L.), qhdai@tsinghua.edu.cn (Q.D.)

Abstract

Computational super-resolution (SR) methods, including conventional analytical algorithms and deep learning models, have substantially improved optical microscopy. Among them, supervised deep neural networks have demonstrated outstanding SR performance, however, demanding abundant high-quality training data, which are laborious and even impractical to acquire due to the high dynamics of living cells. Here, we develop zero-shot deconvolution networks (ZS-DeconvNet) that instantly enhance the resolution of microscope images by more than 1.5-fold over the diffraction limit with 10-fold lower fluorescence than ordinary SR imaging conditions in an unsupervised manner without the need for either ground truths or additional data acquisition. We demonstrate the versatile applicability of ZS-DeconvNet on multiple imaging modalities, including total internal reflection fluorescence microscopy, three-dimensional (3D) wide-field microscopy, confocal microscopy, lattice light-sheet microscopy, and multimodal structured illumination microscopy (SIM), which enables multi-color, long-term, super-resolution 2D/3D imaging of subcellular bioprocesses from mitotic single cells to multicellular embryos of mouse and C. elegans.

Introduction

Optical fluorescence microscopy (FM) is an essential tool for biological research. The recent developments of super-resolution (SR) techniques provide unprecedented resolvability to visualize the fine dynamic structures of diverse bioprocesses[1]. However, the gain in spatial resolution via any SR method comes with trade-offs in other imaging metrics, e.g., duration or speed, which are equally important for dissecting bioprocesses[1, 2]. Recently, computational SR methods have gained considerable attention for their ability to instantly enhance the image resolution in silico[3], enabling significant upgradation of existing FM systems and extension of their applicable range.

In general, existing computational SR methods can be classified into two categories: analytical model-based methods, e.g., deconvolution algorithms[4-6], and deep learning-based methods, e.g., SR neural networks[7-12]. The former category often employs analytical models prescribing certain assumptions about the specimen and image properties, e.g., sparsity[5] and local symmetry[13], to improve the image resolution with multiple tunable parameters. Parameter tuning is experience-dependent and time-consuming, and the outputs of analytical models greatly depend on the parameter sets[5, 13-15]. Moreover, in practical experiments, handcrafted models with certain assumptions cannot address the full statistical complexity of microscope imaging, thus lacking robustness and prone to generating artifacts, especially under low signal-to-noise ratio (SNR) conditions[9]. On the other hand, deep learning-based SR (DLSR) methods have achieved stunning success in learning the end-to-end image transformation relationship according to large amounts of exemplary data without the need for an explicit analytical model[7-12]. Of note, the data-driven inversion scheme via deep learning can approximate not only the pseudoinverse function of the image degradation process but also the stochastic characteristics of the SR solutions. Nevertheless, the training of DLSR models requires acquiring large amounts of paired low-resolution (LR) input images and high-quality ground truth (GT) SR images, which are extremely laborious and sometimes even impractical due to the rapid dynamics or the low fluorescence SNR in biology specimens[3, 8, 16]. In addition, the performance of DLSR methods strongly depends on the quality and quantity of training data[16]. These factors significantly hinder the wide application of DLSR methods in daily imaging experiments despite their compelling SR performance compared to analytical model based methods[3, 16].

Here, we present a zero-shot deconvolution deep neural network (ZS-DeconvNet) framework that is able to train a DLSR network in an unsupervised manner using as few as only one single

planar image or volumetric image stack of low-resolution and low-SNR. As such, compared to state-of-the-art DLSR methods[7-12, 17-21], the ZS-DeconvNet can adapt to diverse bioimaging circumstances, where the bioprocesses are too dynamic, too light-sensitive to acquire the ground-truth SR images, or the image acquisition process is affected by unknown and nonideal factors. We characterized that ZS-DeconvNet can improve the resolution by more than 1.5-fold with high fidelity and quantifiability, even when trained on a single low SNR input image and without the need for image-specific parameter-tuning[5, 13]. We demonstrated that the properly trained ZS-DeconvNet could infer the high-resolution (HR) image on millisecond timescale, achieving high throughput long-term SR 2D/3D imaging of multiple organelle interactions, cytoskeletal and organellar dynamics during the light sensitive processes of migration and mitosis, and subcellular structures and dynamics in developing C. elegans and mouse embryos.

Results

Development and characterization of ZS-DeconvNet

The concept of ZS-DeconvNet is based on the optical imaging forward model informed unsupervised inverse problem solver:

$$\arg\min_{\theta} \left\| \boldsymbol{y} - \left(H \times f_{\theta}(\boldsymbol{y})\right)_{\downarrow} \right\|_2^2 \quad (1)$$

where $\boldsymbol{y}$ denotes the noisy LR image, H denotes the points spread function (PSF) matrix, and f_{θ} represents a deep neural network (DNN) with trainable parameters θ. $\times$ and $(\cdot)_{\downarrow}$ indicate matrix multiplication and downsampling, respectively. If the DNN is trained directly via the above objective function, both the desired biology structures and unwanted noise from the acquisition process, e.g., shot noise, could be enhanced simultaneously, which will substantially degrade the quality of output images at low SNR conditions[22] (Supplementary Fig. 1a). To improve the noise robustness of ZS-DeconvNet while maintaining its unsupervised characteristic, we adopted an image recorrupting scheme[23] that generates two noise-independent recorrupted images from the original image using a mixed Gaussian-Poisson noise model, which are then used as inputs and GTs in the network training (Methods). Furthermore, we introduce the Hessian regularization term, which has been demonstrated to be useful for mitigating reconstruction artifacts in microscopy images[24, 25], to regulate the network convergence (Supplementary Fig. 1b-e). Taken together, the overall objective function of ZS-DeconvNet can be formulated as:

$$\arg\min_{\theta} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(\boldsymbol{y_i} - D^{-1}\boldsymbol{z}, \left(H \times f_\theta(\boldsymbol{y_i} + D^T\boldsymbol{z})\right)_{\downarrow}\right) + \lambda\mathcal{R}_{Hessian}\left(f_\theta(\boldsymbol{y_i} + D^T\boldsymbol{z})\right) \quad (2)$$

where N is the total number of images to be processed, D is an invertible noise control matrix that can be calculated according to the signal and noise levels (Methods), and z is a random noise map that is sampled from a standard normal distribution. We refer to the first part of the objective function as the degradation term, which accounts for the inference fidelity, and the second part as the regularization term, rationalizing the SR outputs.

After defining the objective function, we devised a dual-stage DNN architecture composed of two sequentially connected U-nets[26] for different subtasks (Fig. 1a, b and Supplementary Fig. 2a). The first stage serves as a denoiser to generate noise-free images according to the denoising loss (Methods), and the second stage enhances the image resolution according to the unsupervised deconvolution loss described above. We empirically found that the dual-stage architecture and the physical model-regulated loss function stabilize the training procedures and endow interpretability for the overall network model.

To characterize and evaluate ZS-DeconvNet, we first simulated the microscopy images of punctate and tubular structures contaminated by Gaussian-Poisson noise at escalating signal levels from 5 to 25 average photon counts, which allowed us to systematically test how the recorrupting hyperparameter settings at different imaging conditions influence the final outputs. We found that the optimal hyperparameters are independent of the image contents and signal levels (Supplementary Figs. 3-5), thus enabling parameter tuning-free applications of ZS-DeconvNet onto various biological specimens and imaging configurations. Next, we compared the performance of the ZS-DeconvNet models trained with the data augmented by recorrupting a single noisy image with analytical deconvolution algorithms or the models trained with numbers of simulated or independently acquired images. To do so, we employed the total internal reflective fluorescence (TIRF) illumination mode of our home-built multimodal structured illumination

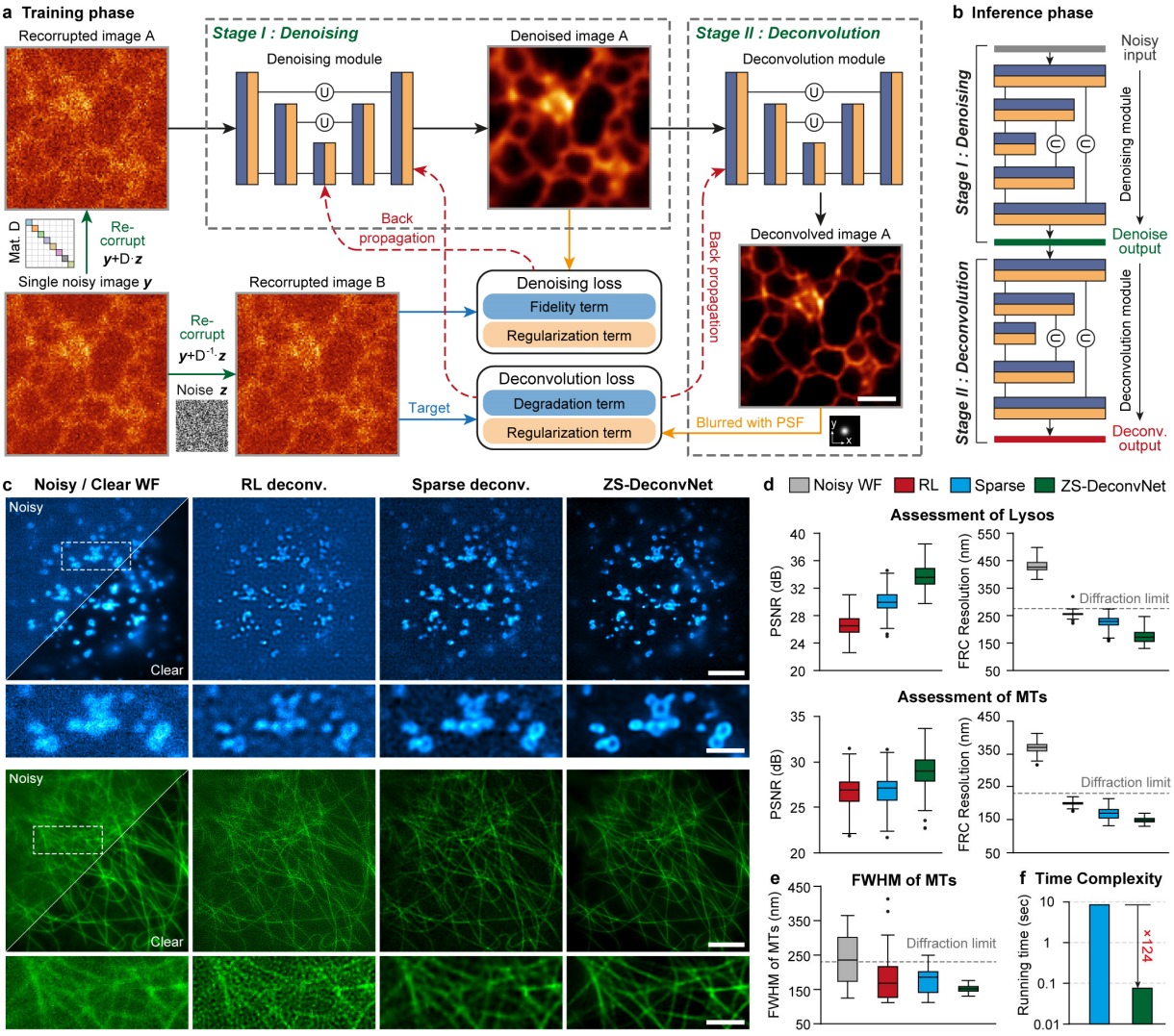


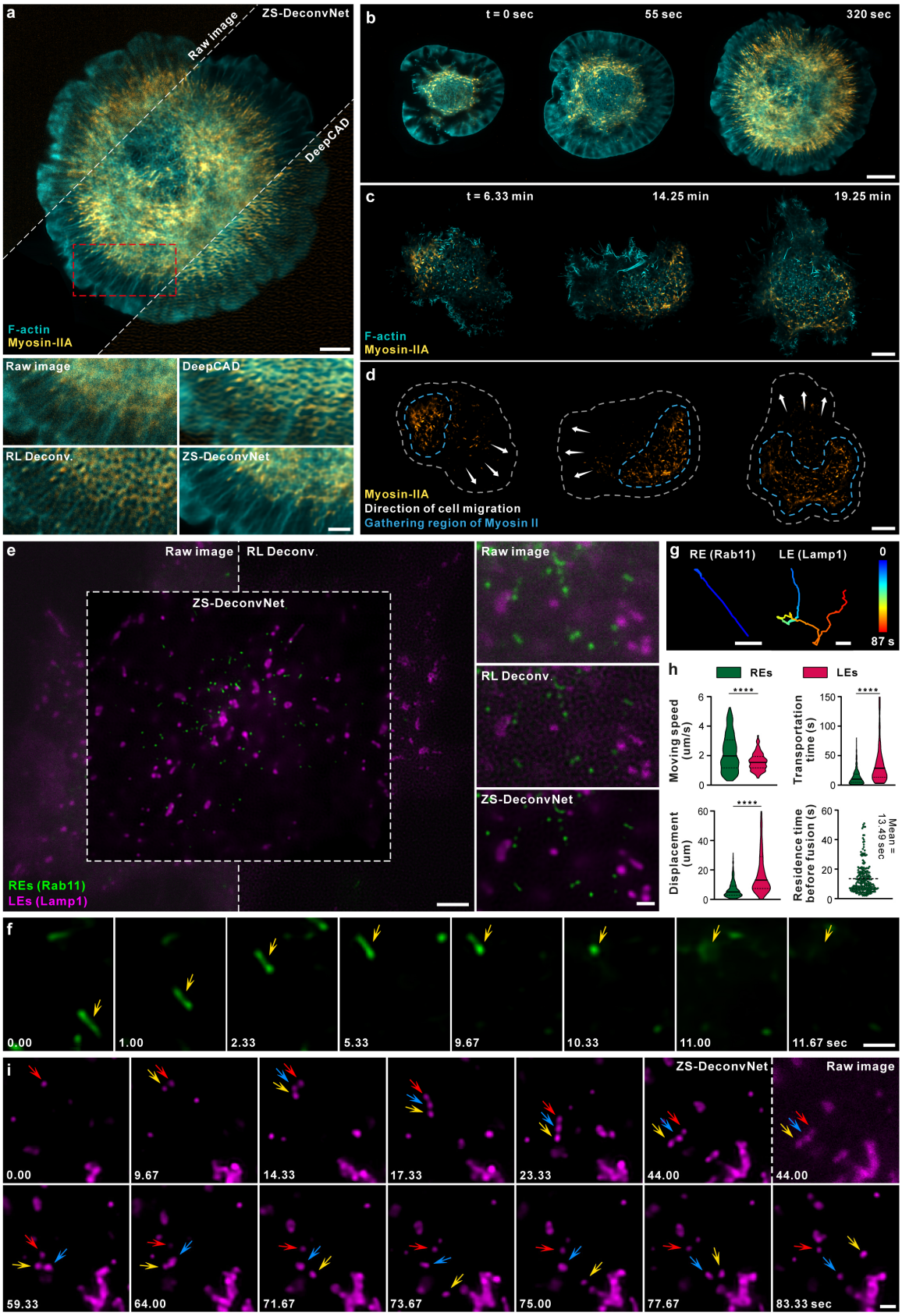
Fig. 1 | Zero-shot deconvolution networks. a, The dual-stage architecture of ZS-DeconvNet and the schematic of its training phase. **b**, The schematic of the inference phase of ZS-DeconvNet. **c**, Representative SR images of Lyso and MTs reconstructed by RL deconvolution (second column), sparse deconvolution (third column) and ZS-DeconvNet (fourth column). The clear WF images are displayed for reference. **d**, Statistical comparisons of RL deconvolution, sparse deconvolution and ZS-DeconvNet in terms of PSNR and FRC resolution (n=100). **e**, Full width at half maximum (FWHM) comparisons of clear WF images and processed images via RL deconvolution, sparse deconvolution and ZS-DeconvNet. The theoretical diffraction limit is labelled with the gray dashed line for reference. **f**, Running time comparison between GPU-based sparse deconvolution and ZS-DeconvNet (average from 25 testing images of 1024×1024 pixels). Center line, medians; limits, 75% and 25%; whiskers, the larger value between the largest data point and the 75th percentiles plus 1.5× the interquartile range (IQR), and the smaller value between the smallest data point and the 25th percentiles minus 1.5× the IQR; outliers, data points larger than the upper whisker or smaller than the lower whisker. Scale bar, 1.5 μm (a), 5 μm (c), 2 μm (zoom-in regions in c).

microscopy (Multi-SIM)[8, 27] (Methods) to acquire ~20 sets of images at low- and high-SNR for each subcellular structure of lysosomes (Lyso) and microtubules (MTs). We found that although

the ZS-DeconvNet was trained with the augmented data from a single input image, the perceptual quality and quantified metrics of its output images were comparable with the images from the model trained with larger amounts of data (Supplementary Fig. 6). Moreover, the peak signal-to-noise ratio (PSNR) and resolution of ZS-DeconvNet images were substantially better than those generated by analytical algorithms, such as the classic Richardson-Lucy (RL) and the latest developed sparse deconvolution[5] (Fig. 1c-e). In addition, the throughput rate of a well-trained ZS-DeconvNet is >100-fold higher than that of the sparse deconvolution algorithm (Fig. 1f). These characterizations demonstrate that ZS-DeconvNet is able to generate high-quality DLSR images of 1.5-fold resolution improvement relative to the diffraction limit while using the least training data, which holds great potential to upgrade the imaging performance of diverse microscope systems, and extend their applicability into a wide variety of bioprocesses that are challenging for conventional methods.

Long-term observation of bioprocesses sensitive to phototoxicity

Cell adhesion and migration are essential in morphogenetic processes and contribute to many diseases[28]. Visualizing cytoskeletal dynamics at high resolution during the adhesion/migration process is critical for elucidating the underlying mechanism. However, due to severe photosensitivity, the whole processes of cell adhesion and migration are typically recorded at low framerates, i.e., several seconds per frame, and low light intensities[9, 29]. Under these imaging conditions, either RL deconvolution or temporal continuity-based self-supervised learning[30] (Methods) fails to recover and sharpen the intricate structure of F-actin and myosin-II (Fig. 2a, Supplementary Fig. 7, and Supplementary Video 1). In contrast, the ZS-DeconvNet model effectively improves both the SNR and resolution of the two-color time-lapse recordings of cell spreading processes after dropping a cell coexpressing mEmerald-Lifeact and mCherry-myosin-IIA onto a coverslip (Fig. 2b and Supplementary Video 2). Intriguingly, we observed that in certain substances cells crawled around the contact site to explore the neighborhood before spreading and adhering (Fig. 2c and Supplementary Video 3). The cell crawling was preceded by the polarized accumulation of myosin-II at the cell rear, leading to cell migration in the opposite direction driven by posterior myosin-II contractility. Moreover, the migration direction could be swiftly changed in response to the dynamic redistribution of myosin-II within the cell (Fig. 2d). These results demonstrate that the kinetics of cell adhesion and migration can be faithfully recorded by



167

Fig. 2 | Long-term SR imaging of rapid and photo-sensitive bioprocesses via ZS-DeconvNet. a, Representative SR images reconstructed by ZS-DeconvNet of F-actin cytoskeleton and myosin-II in a COS-7 cell co-expressing mEmerald-lifeact and mCherry-myosin-IIA. Comparisons of raw noisy TIRF image and images processed by RL deconvolution, DeepCAD-based deconvolution and ZS-DeconvNet are displayed. **b**, Two-color time-lapse SR images enhanced via ZS-DeconvNet showing the coordinated dynamics of F-actin (cyan) and myosin-II (yellow) over the whole spreading process after placing a COS-7 cell onto a coverslip (Supplementary Video 2). **c**, **d**, Two-color time-lapse SR images enhanced via ZS-DeconvNet of F-actin and myosin-II in a crawling COS-7 cell showing that myosin-II preferentially concentrates to the rear of the cell (outlined by yellow dashed lines in d), opposite to crawling direction (indicated by the white arrows in d) (Supplementary Video 3). **e**, Representative SR image generated via ZS-DeconvNet of recycling endosomes (REs, green) and late endosomes (LEs, magenta) in a gene-edited SUM-159 cell endogenously expressing EGFP-Rab11 and mCherry-Lamp1 (Supplementary Video 4). **f**, Time-lapse images illustrate the directional movement of a RE in rod-like shape, and the subsequent fusion with plasma membrane. **g**, Typical trajectories of RE (top) and LE (bottom) movements showing the rapid directional motility of RE, and the bidirectional nature of LE. **h**, Comparisons of the speed, displacement, and travelling time between Lyso/LEs and REs, and quantification of the residence time of REs near their exocytosis sites before fusing with plasma membrane. A small number of data points exceeding transportation time of 150s or displacement of 60 μm were not displayed for better presentation of the distributions. We provided all source data in Extended Data files. **i**, Time-lapse images illustrate three LEs tether each other and co-migrate for certain distance before splitting into individual LEs. Scale bar, 5 μm (a, c, and d), 2 μm (zoom-in regions in a), 8 μm (b), 3 μm (e), 0.5 μm (zoom-in region in e), 1 μm (g, f, and i).

ZS-DeconvNet-assisted imaging without perturbing this lengthy and vulnerable process.

Visualizing the rapid dynamics of the endolysosomal system

The endolysosomal system includes diverse types of vesicles that function in a highly dynamic, yet well-organized manner. Although live-cell fluorescence imaging has remarkably improved our understanding of the endolysosomal system, most studies had to overexpress the proteins of interest to record their rapid dynamics[27], which often resulted in artifact morphologies or behaviors. With ZS-DeconvNet, we were able to image the knock-in SUM-159 cell line endogenously expressing EGFP-Rab11 and mCherry-Lamp1 for 1,500 frames at ~150 nm resolution and 3 frames per second (fps) in two colors (Fig. 2e and Supplementary Video 4), thereby allowing us to visualize and track the rapid motion of recycling-endosomes (REs) and lysosomes or late endosomes (LEs) on a substantially finer spatiotemporal scale and longer observation window than previously achieved[31]. As exemplified in Fig. 2f and g, we found that the majority of REs (n = 505 tracks) experienced a directional movement, with a total displacement of 6.7 ± 5.4 μm at a high speed of 2.2 ± 1.2 μm/s (instantaneous speed exceeding 5.3 μm/s), with a rare intermediate pause, then stopped at specific sites for a period of 13.5 ± 10.3 sec (Fig. 2h) before fusing with the plasma membrane. This observation suggests that REs might be efficiently transported over long ranges

to regions near the plasma membrane to facilitate subsequent exocytosis. Unexpectedly, ZS-DeconvNet captured multiple fission events of the Rab11-positive REs, in which both separated REs underwent exocytosis sequentially (Supplementary Fig. 8a) or one RE moved away (Supplementary Fig. 8b). This observation indicates that the highly specialized Rab11-positive REs may be subject to further cargo sorting right before exocytosis.

In contrast, the movements of LEs were typically discontinuous and proceeded in a bidirectional stop-and-go manner at a relatively slow speed of 1.6 ± 0.6 μm/s (n = 230 tracks) (Fig. 2h, i). Although the transportation of LEs seemed inefficient, the LEs often persisted for a long period of 91.8 s with a total displacement as long as 23.6 μm (averaged from n = 230 tracks) (Fig. 2h). Interestingly, we noticed that two or more LEs sometimes tended to tether each other in a kiss-and-stay fashion and migrate for a certain distance before splitting into individual LEs again (Fig. 2i and Supplementary Fig. 8c), which might facilitate the directional movement of LEs without sufficient motor-protein-adaptors for long-range transportation. These complex dynamics of LEs suggest that their positioning and mobility are delicately regulated by multiple factors, such as MT-based motors and membrane contacts.

3D ZS-DeconvNet for lattice light-sheet microscopy

Volumetric live-cell imaging conveys more biological information than 2D observations; however, it is subject to much severer phototoxicity, photobleaching and out-of-focus fluorescence contamination. To extend the superior capability of ZS-DeconvNet to volumetric SR imaging, we upgraded the backbone of the dual-stage network architecture into a 3D RCAN, which has been demonstrated to be suitable for volumetric image restoration[9, 32] (Fig. 3a, b and Supplementary Fig. 2b). Next, we integrated our previously proposed spatially interleaved self-supervised (SiS) learning scheme[9] with the physical model-informed self-supervised inverse problem solver to construct the 3D ZS-DeconvNet. The 3D ZS-DeconvNet with SiS scheme follows a simpler data augmentation procedure (Methods), while achieving comparative or even better performance than the recorruption-based strategy (Supplementary Fig. 9).

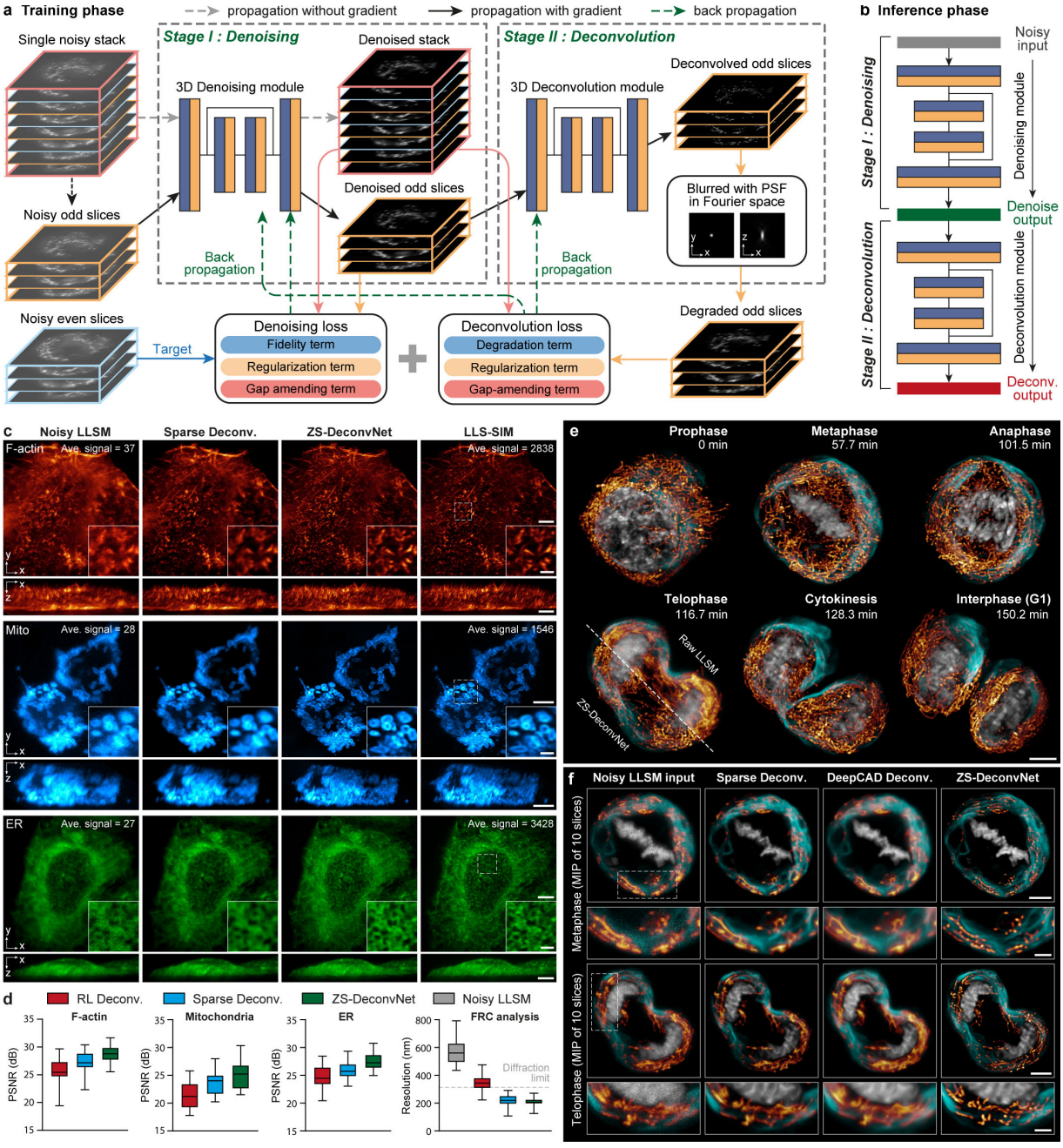


Fig. 3 | Characterizations and demonstrations of 3D ZS-DeconvNet. a, The network architecture of 3D ZS-DeconvNet and the schematic of its training phase. **b**, The schematic of the inference phase of 3D ZS-DeconvNet. **c**, Representative maximum intensity projection (MIP) SR images (MIP) of F-actin, Mito outer membrane, and ER reconstructed by sparse deconvolution (second column), 3D ZS-DeconvNet (third column), and LLS-SIM (fourth column). Average sCMOS counts of the highest 1% pixels for raw images before processed are labelled on the top right corner. **d**, Statistical comparisons of RL deconvolution, sparse deconvolution and ZS-DeconvNet in terms of PSNR and resolution on different specimens (n=40). The resolution was measured by Fourier ring correlation analysis[33] with F-actin image stacks. Center line, medians; limits, 75% and 25%; whiskers, maximum and minimum. **e**, Time-lapse three-color 3D rendering images reconstructed via 3D ZS-DeconvNet of ER, H2B, and Mito, showing

their transformations in morphology and distribution as well as interaction dynamics during mitosis (Supplementary Video 5). **f**, Representative three-color images obtained with conventional LLSM (first column), sparse deconvolution (second column), DeepCAD based deconvolution (third column) (Methods), and 3D ZS-DeconvNet (fourth column). The comparisons are performed on two typical timepoints of the time-lapse data shown in e. Scale bar, 5 μm (c, e, and f), 1.5 μm (zoom-in regions of c), 2 μm (zoom-in regions of f).

We demonstrate that even when trained with a single noisy image stack, 3D ZS-DeconvNet significantly improves the SNR and provides a laterally isotropic spatial resolution enhancement by over 1.5-fold (Supplementary Fig. 10). We systematically assessed the 3D ZS-DeconvNet model with datasets of three different biological specimens acquired via our home-built lattice light-sheet structured illumination microscopy[34] (LLS-SIM) (Methods). We found that 3D ZS-DeconvNet successfully reconstructed the elaborate filaments of F-actin, the hollowing structure of the mitochondrial (Mito) outer membrane, and the intricate networks of the endoplasmic reticulum (ER) with high fidelity and resolution comparable to LLS-SIM images acquired under high-SNR conditions (Fig. 3c). The quantifications of PSNR and resolution illustrate that the 3D ZS-DeconvNet model substantially outperforms conventional analytical model-based approaches in diverse biological specimens (Fig. 3d).

Long-term volumetric super-resolution imaging enabled by 3D ZS-DeconvNet

Volumetric observation of cell division at high spatiotemporal resolution is of vital importance for exploring mitosis-related biological mechanisms, such as the mechanism that allocates the numerous distinct organelles in the cytoplasm into each daughter cell[35, 36]. Due to the extreme light sensitivity and vulnerability of mitotic cells, previous volumetric SR imaging of this process has relied on the low-light LLS-SIM system and supervised learning-based SR reconstruction[9]. However, collecting high-quality training data is extremely laborious and sometimes impractical because the morphology and distribution of organelles usually undergo dramatic changes during mitosis[35]. Here, we demonstrate that the self-supervised 3D ZS-DeconvNet model can be generally applied to superresolve the fine subcellular structures of the ER, Mito, and chromosomes from noisy LLSM volumes without the need for additional training data, thus enabling fast and long-term volumetric SR observation of multiple organelles for 1,000 timepoints at 10 sec intervals in a mitotic HeLa cell (Fig. 3e and Supplementary Video 5). Moreover, the unsupervised property of ZS-DeconvNet allows us to integrate the transfer adaptation learning strategy[37] to fully exploit the structural content in each noisy volume, which yielded the best 3D SR performance (Methods). In

contrast, the conventional prior-dependent deconvolution algorithm[5] and temporally interleaved self-supervised learning[9, 30, 38] method both failed to restore the high-frequency details of the specimens because of the low SNR condition and weak temporal consistency between adjacent timepoints (Fig. 3f and Methods). Furthermore, according to the low invasiveness provided by 3D ZS-DeconvNet, a group of mitotic HeLa cells labeled with H2B-mCherry and HeLa-mEmerald-SC35 were imaged in a large field of view (FOV) of 100×50×25 μm^3 for more than 300 timepoints, thereby recording the entire disassembly and reassembly processes of nuclear speckles at a high spatiotemporal resolution (Supplementary Fig. 11 and Supplementary Video 6). In brief, 3D ZS-DeconvNet allows biologists to easily explore various light-sensitive bioprocesses with low invasiveness at substantially higher spatiotemporal resolution without the need for any additional datasets or optical setup modifications.

ZS-DeconvNet for confocal and wide-field microscopy

The ZS-DeconvNet relies on the randomness of noises and the low-pass filter characteristic of optical microscopes, which are common for various types of microscopy modalities. On this basis, we expect that ZS-DeconvNet can be generally applied to all microscopy, e.g., the most commonly used confocal microscopy and wide-field (WF) microscopy. To investigate the performance of 3D ZS-DeconvNet on confocal data, we employed our home-built confocal microscope to acquire a four-color volume of the mouse early embryo immunostained for the microtubule, chromosomes, actin, and apical domain (Methods), which play key roles in the first cell fate decision and are critical for embryo development[39-41]. We then trained 3D ZS-DeconvNet models on this single noisy volume and processed the original data with the trained models. As shown in Fig. 4a, b, 3D ZS-DeconvNet significantly enhances the SNR, contrast, and resolution of the confocal data volume and resolves the fine structures of microtubule bridges and actin rings (Fig. 4c, d, Supplementary Fig. 12, and Supplementary Video 7). These results indicate that ZS-DeconvNet enables a higher spatial resolution at a lower photon budget for confocal microscopy in imaging specimens on large scale, e.g., mouse early embryos, which is critical to research on cell polarity[41] intracellular transport and blastocyst formation[40].

We next imaged Caenorhabditis elegans embryos marked apical junctions, cell membranes and lysosomes using the 3D WF mode of our Multi-SIM system (Methods). To ensure that C. elegans embryo development was not disturbed, we acquired raw image stacks at relatively low light

excitation in intervals of 30 seconds for more than 200 timepoints. However, under such conditions, the WF images are heavily contaminated by both out-of-focus background and noise (Fig. 4e, f). Even in this challenging situation, 3D ZS-DeconvNet images presented considerable suppression upon noise and background while enhancing the spatial resolution of the subcellular details (Fig. 4e, g and Supplementary Video 8), thus allowing us to investigate the elaborate process of embryonic development, e.g., hypodermal cell fusion[42] (Fig. 4h), even via a simple WF microscope.

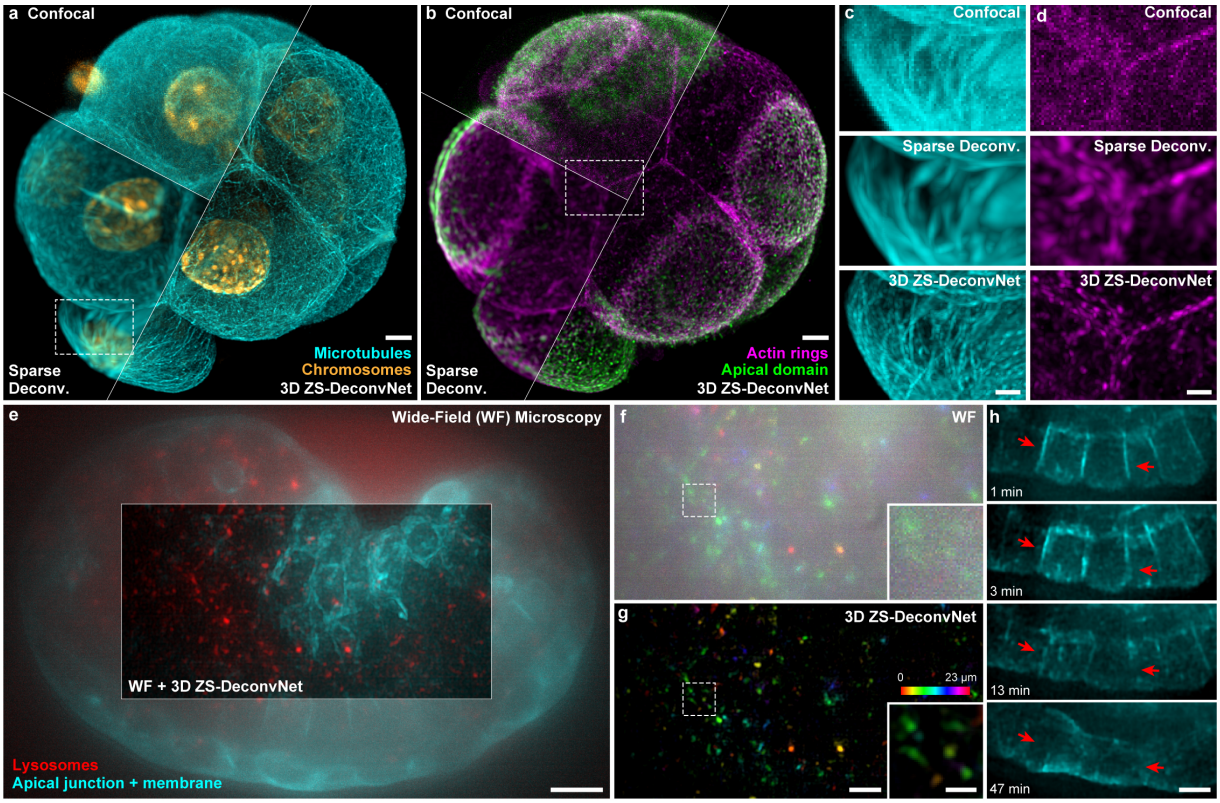


Fig. 4 | Generalization of ZS-DeconvNet to multiple imaging modalities. a, **b**, Representative confocal (top left), sparse deconvolution (bottom left), and 3D ZS-DeconvNet enhanced (right) images of an early mouse embryo immunostained for microtubule (cyan), chromosomes (orange), actin rings (magenta), and apical domain (green). **c**, **d**, Magnified regions of microtubule bridges (c) and actin rings (d) labelled with while dashed boxes in (a) and (b) acquired via confocal microscopy, sparse deconvolution, and 3D ZS-DeconvNet. **e**, Representative WF (center region) and 3D ZS-DeconvNet enhanced (surrounding region) images of a C. elegans embryo with apical junction, cell membrane (cyan) and lysosomes (red) labelled. **f**, **g**, Lysosome channel of the central region in (c) color-coded for distance from the substrate. Both WF (d) and 3D ZS-DeconvNet processed images are shown for comparison. **h**, Time-lapse 3D ZS-DeconvNet enhanced images showing the process of hypodermal cell fusion (red arrows) during the development of a C. elegans embryo. Scale bar, 5 μm (a, b, e), 2μm (c, d), 3 μm (g, h), 1 μm (zoom-in region of g). Gamma value, 0.7 for cytomembrane and lysosomes in the C. elegans embryo.

ZS denoising and resolution enhancement in multimodal SIM images

Among the various forms of SR microscopy, structured illumination microscopy (SIM) is often recognized as a balanced option for SR live-cell imaging because it needs less than ten raw modulated images to provide a twofold improvement in spatial resolution[1, 2]. Nevertheless, conventional SIM has two critical limitations: first, further resolution enhancement requires considerably more raw data, i.e., at least 25 raw images are needed for nonlinear SIM to obtain a sub-80 nm resolution[43, 44]; second, the postreconstruction of SIM images generally requires raw images with a high SNR to eliminate noise-induced reconstructed artifacts, thus impairing fast, low-light, and long-term live-cell imaging. Recent studies have explored supervised learning approaches by either denoising SIM images[9, 45] or reconstructing SR SIM images directly from noisy raw images[8, 20] to achieve low-light SIM reconstruction; however, these methods require abundant training data and do not further enhance the resolution. In light of the superb denoising and SR capability of ZS-DeconvNet, we integrated the zero-shot learning scheme with the conventional SIM reconstruction algorithm. We designed the ZS-DeconvNet enhanced SIM (ZS-DeconvNet-SIM) model to simultaneously denoise and sharpen SR SIM images in an unsupervised manner (Fig. 5a, and Methods). Resorting to the remarkable improvement in both SNR and resolution provided by ZS-DeconvNet-SIM (Supplementary Fig. 13a-d), the hollow structure of clathrin-coated pits (CCPs) in a SUM-159 cell and the densely interlaced cytoskeletons in a COS-7 cell, which are indistinguishable in WF and conventional SIM images, were clearly resolved (Fig. 5b, c).

Furthermore, we integrated 3D ZS-DeconvNet with LLS-SIM to develop the 3D ZS-DeconvNet LLS-SIM modality. By incorporating the anisotropic PSF of conventional LLS-SIM[34] into the training process, 3D ZS-DeconvNet LLS-SIM not only prominently enhanced the contrast and resolution in all three dimensions but also provided an isotropic lateral resolution of approximately 150 nm (Fig. 5d, e, and Supplementary Fig. 13e-g). These successful applications of ZS-DeconvNet to multimodal SIM systems demonstrate its capability to further extend the spatiotemporal resolution bandwidth of existing SR techniques.

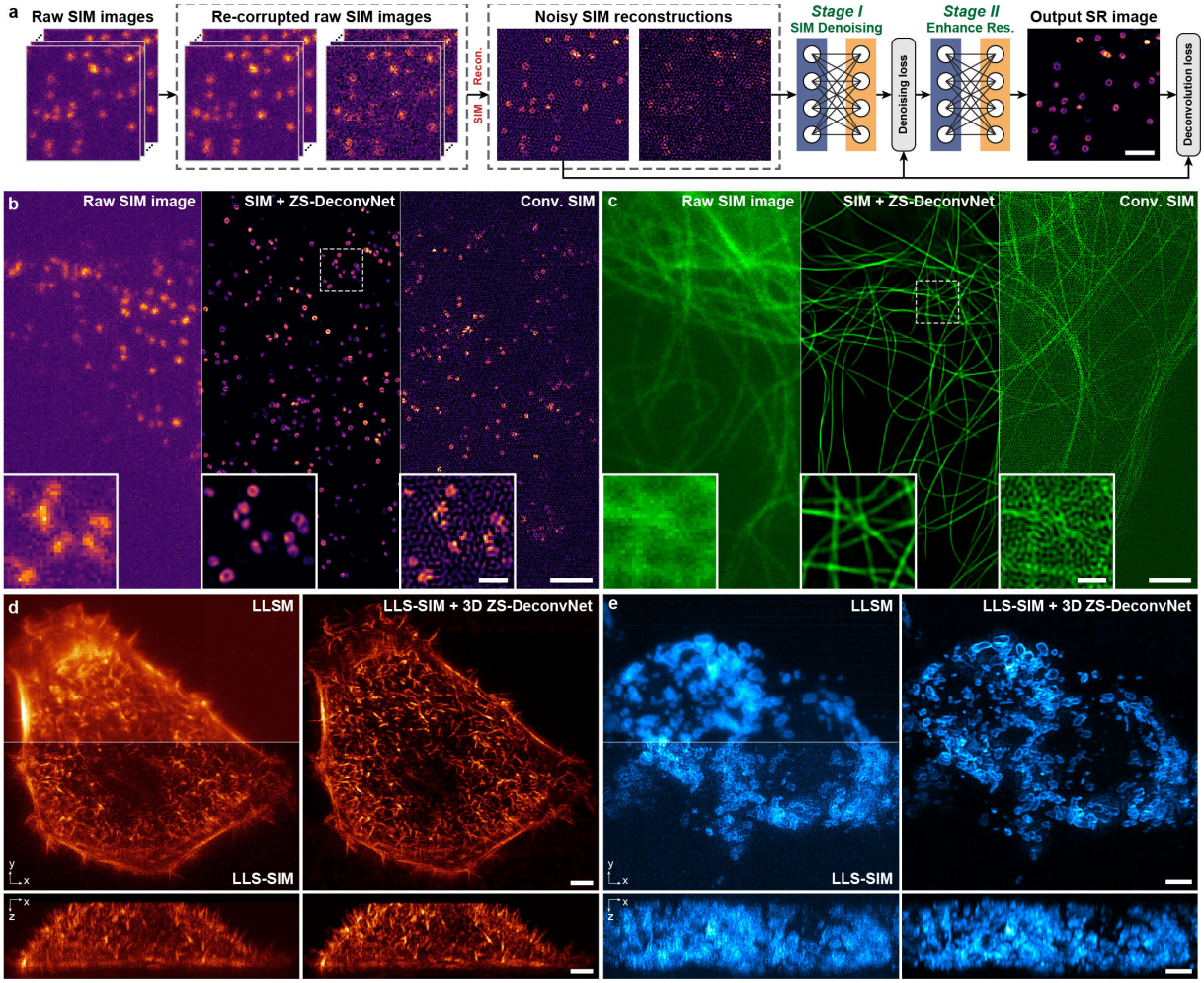


Fig. 5 | Zero-shot denoising and resolution enhancement in multimodal SIM data. a, Schematic of the training procedure of ZS-DeconvNet for SIM. **b**, Progression of SNR and resolution improvement across the CCPs in a SUM-159 cell, from raw SIM images (left), conventional SIM image (right), and ZS-DeconvNet enhanced SIM image (middle). **c**, Progression of SNR and resolution improvement across the microtubules in a COS-7 cell, from raw SIM images (left), conventional SIM image (right), and ZS-DeconvNet enhanced SIM image (middle). **d**, Representative maximum intensity projection (MIP) images of F-actin in a HeLa cell obtained via LLSM, LLS-SIM, and LLS-SIM enhanced by 3D ZS-DeconvNet across three dimensions. **e**, Representative MIP images of mitochondrial outer membrane labelled with TOMM20 in a 293T cell obtained via LLSM, LLS-SIM, and LLS-SIM enhanced by 3D ZS-DeconvNet across three dimensions. Scale bar, 1 μm (a), 2 μm (b, c), 0.5 μm (zoom-in regions in b, c), 3 μm (d, e).

Discussion

The ultimate goal of live imaging is to collect the most spatiotemporal information about bioprocesses with the least invasiveness to biological specimens. However, the mutual restrictions between imaging speed, duration, resolution, and SNR in fluorescence microscopy together result in the spatiotemporal bandwidth limitation[46], which limits the synergistic improvement in all these

aspects. For instance, to obtain higher spatial resolution, conventional SR techniques have to rely on repetitive acquisitions or additional excitation[1], which aggravates phototoxicity and photobleaching, impeding fast, long-term observations of bioprocesses. To address the spatiotemporal bandwidth limitations in microscopy, in this manuscript, we proposed the versatile framework ZS-DeconvNet, which can be incorporated with any optical fluorescence microscope to instantly enhance image SNR and resolution without compromising other imaging properties. We emphasize that the application of ZS-DeconvNet does not require image-specific parameter-tuning thus avoiding errors caused by irrational parameter selections (Supplementary Fig. 14) and that ZS-DeconvNet can be well trained with only one slice or stack of raw images (Supplementary Figs. 6, 10) without using assumptions of structural sparsity[5] and spatiotemporal continuity[25, 30, 38]. The qualitative and quantitative evaluations on both simulated and experimental data show that our methods substantially enhance the image quality and resolution by more than 1.5-fold with high fidelity, even under low-light conditions, thereby permitting fast, long-term, super-resolution observations of multiple subcellular dynamics.

The proposed ZS-DeconvNet method has wide functionality for various types of imaging modalities, from scanning-based microscopy, e.g., confocal microscopy and two-photon microscopy, to wide-field detection-based microscopy, e.g., TIRF, 3D WF microscopy, LLSM, and multimodal SIM. We demonstrate its capabilities with more than 10 distinct fixed- or live-specimens imaged via six different microscopy setups, including planar and volumetric imaging of multiple organelles in single cells, observations of subcellular dynamics and interactions during cell mitosis, and multi-color 3D imaging of early mouse embryos and C. elegans embryos. Moreover, we demonstrate that our ZS-DeconvNet can conveniently be implemented on any existing imaging system to notably improve the SNR, contrast, and resolution (Figs. 4, 5). The functionality and convenience of ZS-DeconvNet demonstrate its great potential in upgrading the performance of existing optical microscopy.

Several improvements and extensions of ZS-DeconvNet can be envisioned. First, a recently published work[12] that devised a network architecture incorporating the image formation process to accelerate deconvolution suggests that the backbone architecture of ZS-DeconvNet can be further improved in terms of a lightweight structure and effectiveness, thereby obtaining better SR performance with less computation time. Second, although we presented only the applications of ZS-DeconvNet on SIM, it can be reasonably speculated that other optics-based SR techniques,

392 such as photoactivated localization microscopy[47], stimulated emission depletion microscopy[48], and
393 image scanning microscopy[49], can be improved by integrating ZS-DeconvNet into their image
394 processing pipelines. Third, due to the lack of generalization, users need to train a specialized
395 model for each type of specimen to achieve the best performance. Incorporating domain
396 adaptation[37] or domain generalization[50] techniques with our methods may effectively alleviate the
397 burden of applying trained models into unseen domains. Finally, we used a spatially invariant PSF
398 for the well-calibrated imaging systems in this work. With spatially varying PSF, the functionality
399 of ZS-DeconvNet can be further extended to various image processing tasks, such as phase space
400 light-field reconstruction and digital adaptive optics[51].

401

References

1. Schermelleh, L. et al. Super-resolution microscopy demystified. *Nature Cell Biology* **21**, 72-84 (2019).
2. Wu, Y. & Shroff, H. Faster, sharper, and deeper: structured illumination microscopy for biological imaging. *Nature Methods* **15**, 1011-1019 (2018).
3. Belthangady, C. & Royer, L.A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature Methods* **16**, 1215-1225 (2019).
4. Sage, D. et al. DeconvolutionLab2: An open-source software for deconvolution microscopy. *Methods* **115**, 28-41 (2017).
5. Zhao, W. et al. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. *Nature Biotechnology* **40**, 606-617 (2021).
6. Guo, M. et al. Rapid image deconvolution and multiview fusion for optical microscopy. *Nature Biotechnology* **38**, 1337-1346 (2020).
7. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103-110 (2019).
8. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* **18**, 194-202 (2021).
9. Qiao, C. et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nature Biotechnology* (2022).
10. Yanny, K., Monakhova, K., Shuai, R.W. & Waller, L. Deep learning for fast spatially varying deconvolution. *Optica* **9**, 96-99 (2022).
11. Zhao, Y. et al. Isotropic super-resolution light-sheet microscopy of dynamic intracellular structures at subsecond timescales. *Nature Methods* **19**, 359-369 (2022).
12. Li, Y. et al. Incorporating the image formation process into deep learning improves network performance. *Nature Methods* (2022).
13. Gustafsson, N. et al. Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications* **7**, 12471 (2016).
14. Richardson, W.H. Bayesian-based iterative method of image restoration. *JoSA* **62**, 55-59 (1972).
15. Lucy, L.B. An iterative technique for the rectification of observed distributions. *The astronomical journal* **79**, 745 (1974).
16. Laine, R.F., Arganda-Carreras, I., Henriques, R. & Jacquemet, G. Avoiding a replication crisis in deep-learning-based bioimage analysis. *Nature Methods* **18**, 1136-1144 (2021).
17. Park, H. et al. Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy. *Nature Communications* **13**, 3297 (2022).
18. Qiao, C. et al. 3D Structured Illumination Microscopy via Channel Attention Generative Adversarial Network. *IEEE Journal of Selected Topics in Quantum Electronics* **27**, 1-11 (2021).

19. Fang, L. et al. Deep learning-based point-scanning super-resolution imaging. *Nature Methods* **18**, 406-416 (2021).
20. Jin, L. et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nature Communications* **11**, 1934 (2020).
21. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* **36**, 460-468 (2018).
22. He, Y. et al. Self-supervised deep-learning two-photon microscopy. *Photonics Research* **11**, 1-11 (2023).
23. Pang, T., Zheng, H., Quan, Y. & Ji, H. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2043-2052 (2021).
24. Lefkimmiatis, S., Bourquard, A. & Unser, M. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing* **21**, 983-995 (2011).
25. Huang, X. et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nature Biotechnology* **36**, 451-459 (2018).
26. Ronneberger, O., Fischer, P. & Brox, T. in International Conference on Medical image computing and computer-assisted intervention 234-241 (Springer, 2015).
27. Guo, Y. et al. Visualizing Intracellular Organelle and Cytoskeletal Interactions at Nanoscale Resolution on Millisecond Timescales. *Cell* **175**, 1430-1442 e1417 (2018).
28. Parsons, J.T., Horwitz, A.R. & Schwartz, M.A. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature reviews Molecular cell biology* **11**, 633-643 (2010).
29. Burnette, D.T. et al. A role for actin arcs in the leading-edge advance of migrating cells. *Nature cell biology* **13**, 371-382 (2011).
30. Li, X. et al. Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nature Biotechnology*, 1-11 (2022).
31. Guo, M. et al. Single-shot super-resolution total internal reflection fluorescence microscopy. *Nature Methods* **15**, 425-428 (2018).
32. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nature Methods* **18**, 678-687 (2021).
33. Nieuwenhuizen, R.P. et al. Measuring image resolution in optical nanoscopy. *Nature methods* **10**, 557-562 (2013).
34. Chen, B.C. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).
35. Carlton, J.G., Jones, H. & Eggert, U.S. Membrane and organelle dynamics during cell division. *Nature Reviews Molecular Cell Biology* **21**, 151-166 (2020).
36. Moore, A.S. et al. Actin cables and comet tails organize mitochondrial networks in mitosis. *Nature* **591**, 659-664 (2021).

37. Zhang, L. & Gao, X. Transfer adaptation learning: A decade survey. *IEEE Transactions on Neural Networks and Learning Systems* (2022).

38. Lecoq, J. et al. Removing independent noise in systems neuroscience data using DeepInterpolation. *Nature Methods* **18**, 1401-1408 (2021).

39. Zenker, J. et al. A microtubule-organizing center directing intracellular transport in the early mouse embryo. *Science* **357**, 925-928 (2017).

40. Zenker, J. et al. Expanding actin rings zipper the mouse embryo for blastocyst formation. *Cell* **173**, 776-791. e717 (2018).

41. Zhu, M. et al. Developmental clock and mechanism of de novo polarization of the mouse embryo. *Science* **370**, eabd2703 (2020).

42. Mohler, W.A., Simske, J.S., Williams-Masson, E.M., Hardin, J.D. & White, J.G. Dynamics and ultrastructure of developmental cell fusions in the Caenorhabditis elegans hypodermis. *Current biology* **8**, 1087-1091 (1998).

43. Gustafsson, M.G. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences* **102**, 13081-13086 (2005).

44. Li, D. et al. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).

45. Shah, Z.H. et al. Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. *Photonics Research* **9**, B168-B181 (2021).

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

47. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *science* **313**, 1642-1645 (2006).

48. Klar, T.A., Jakobs, S., Dyba, M., Egner, A. & Hell, S.W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences* **97**, 8206-8210 (2000).

49. Muller, C.B. & Enderlein, J. Image scanning microscopy. *Physical Review Letters* **104**, 198101 (2010).

50. Wang, J. et al. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering* (2022).

51. 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 (2021).

52. Castello, M. et al. A robust and versatile platform for image scanning microscopy enabling super-resolution FLIM. *Nature methods* **16**, 175-178 (2019).

Methods

Optical systems

Multi-SIM system

The Multi-SIM system was built based on an invented fluorescence microscope (Ti2E, Nikon). Three laser beams of 488 nm (Genesis-MX-SLM, Coherent), 560 nm (2RU-VFL-P-500-560, MPB Communications), and 640 nm (LBX-640-500, Oxxius), were combined collinearly, and then passed through an acousto-optic tunable filter (AOTF, AOTFnC-400.650, AA Quanta Tech), which serves to select the desired laser wavelength and control its power and exposure time. Afterwards the selected laser light was expanded and sent into an illumination modulator, which is composed of a ferroelectric spatial light modulator (SLM, QXGA-3DM, Forth Dimension Display), a polarization beam splitter, and an achromatic half-wave plate. Different illumination modes were generated by adjusting the patterns displayed on the SLM, e.g., grating patterns of 3-phase × 3-orientation at 1.41 NA for TIRF-SIM or 1.35 NA for GI-SIM. Next, the modulated light was passed through a polarization rotator consisting of a liquid crystal cell (Meadowlark, LRC-200) and a quarter-wave plate, which rotated the linear polarization to maintain the s-polarization necessary, thus maximizing the pattern contrast for all pattern orientations. The diffraction orders, except for ±1 orders for TIRF/GI-SIM, were filtered out by a spatial mask, and then relayed onto the back focal plane of the objectives (1.49 NA, Nikon). The raw SIM images excited by different illumination patterns were sequentially collected by the same objective, then separated by a dichroic beam splitter (Chroma, ZT405/488/560/647tpc), finally captured with a sCMOS camera (Hamamatsu, Orca Flash 4.0 v3). Besides TIRF/GI-SIM modes used in this work, the Multi-SIM system integrated diverse SIM modalities including nonlinear-SIM and 3D-SIM into a single setup, which has been commercially available from NanoInsights Inc. (nanoinsights-tech.com).

LLS-SIM system

The home-built LLS-SIM system was developed from the original design[34]. Similar to the laser combinator and pattern modulator used in our Multi-SIM system, three lasers of 488 nm, 560 nm, and 640 nm (MPB Communications) were selected and controlled by an AOTF, and then modulated by the lattice patterns displayed on the SLM. The excitation light was then filtered by an annular mask equivalent to 0.5 outer NA and 0.375 inner NA for the excitation objective

(Special Optics). Subsequently, the filtered excitation light passed through a pair of galvo mirrors (*x*- and *z*-galvo) (Cambridge Technology, 6210H). In LLS-SIM mode, the lattice patterns of 3-phase were sequentially displayed on the SLM and synchronized with the programmed "ON" time of AOTF, and then scanned by the sample piezo in a step size of 0.39 μm, which equals to a z-interval of 0.2 μm, to acquire the volumetric LLSM images. In LLSM mode, a fixed lattice pattern was quickly dithered by *x*-galvo, and then scanned by the sample piezo. In particular, we used the triangle wave when reversing the scanning direction of the piezo stage to minimize the flyback time to an extreme. Live cell specimens were held in a customized microscope incubator (OKO lab, H301-LLSM-SS316) to maintain the physiology condition of 37℃ and 5% CO2 during imaging. The emission light was collected by the detection objective (Nikon, CFI Apo LWD 25XW, 1.1NA) and captured by a sCMOS camera (Hamamatsu, Orca Fusion).

Confocal system

The home-built confocal microscopy was developed as a modification of the image-scanning microscopy system[52] based on a commercial invented fluorescence microscope (Ti2E, Nikon). Four laser beams of 405 nm, 488 nm, 561 nm, and 640 nm (BDL-405-SMN, BDL-488-SMN, BDS-561-SMY-FBE, and BDL-640-SMN, Becker & Hickel) were collinearly combined and then expended by 6.25 times. After being reflected by a multi-band dichroic mirror (Di03-R405/488/561/635, Semrock), the lasers were passed through two galvanometer scanners (8315k, CT Cambridge Technology) and then directed toward the objective (CFI SR HP Plan Apo Lambda S 100XC/1.35NA, Sil, Nikon) via a scan lens and a tube lens. The emission fluorescence was collected by the same objective, descanned, and passed through the multi-band dichroic mirror and then separated into the green channel and then red channel by a dichroic beam splitter (FF573-DI01, Semrock). The green-channel signals (filtered by FF02-447/60, FF03-525/50, Semrock) were collected by a single photon counting module (SPCM-AQRH-44, Excelitas) and finally counted by a digital counter (BNC-2121, National Instruments). The red-channel signals (filtered by FF01-609/57, FF01-679/41, Semrock) were collected by a fiber bundle and then captured by a multi-channel photomultiplier tube (PML-16-GASP) and quantified by a single photon counter (SPC-164-PCI, Becker & Hickel). The pinhole was kept open during image acquisition and the overall magnification factor was 333× for the green channel and 666× for the red channel. The data acquisition/visualization/processing was operated by a home-developed software based on LabView (National Instruments) and the software also controlled all microscope devices during

the image acquisition, such as the galvanometer scanners, the axial piezo stage, and the laser power by sending analog signals via a field-programmable-gate-array card (NI PXIe-7868R, National Instruments).

Architectures and objective functions of ZS-DeconvNet

ZS-DeconvNet adopts a dual-stage architecture, which factorizes low-SNR super-resolution task into two sequential subdivisions of denoising and deconvolution, and each stage is responsible for one subtask, respectively. The dual-stage design is helpful to regulating the training procedures and eliminating the noise-induced artifacts in the final outputs[11]. For 2D images, a simplified U-net models[26] with four down- and up-sampling modules are used as the backbone of each stage. The overall network architecture of ZS-DeconvNet we used for 2D image SR in this work is shown in Supplementary Fig. 2a. In the training phase, we designed a combined loss function consisting of a denoising term and a deconvolution term, which respectively corresponds to the denoising stage and the deconvolution stage:

$$\mathcal{L}(\hat{\boldsymbol{y}}, \tilde{\boldsymbol{y}}) = \mu \mathcal{L}_{den}(\hat{\boldsymbol{y}}, \tilde{\boldsymbol{y}}) + (1-\mu)\mathcal{L}_{dec}(\hat{\boldsymbol{y}}, \tilde{\boldsymbol{y}}) \quad (3)$$

where $(\hat{\boldsymbol{y}}, \tilde{\boldsymbol{y}})$ indicates the recorrupted image pair (see next section for the details of image recorruption), and μ is a scalar weighting factor to balance the two terms, which we empirically set as 0.5 in our experiments.

For 3D ZS-DeconvNet, we deploy 3D RCAN as the backbone model for the two stages, each of which includes two residual groups consisting of four channel attention blocks. The overall architecture is illustrated in Supplementary Fig. 2b. During training procedures, the 3D ZS-DeconvNet is optimized iteratively following a similar combined loss function to its 2D versions, nevertheless, with two major modifications in detail: first, the image pairs used for training were generated by axial sampling rather than via recorruption, resulting in a totally parameter-free data augmentation strategy; second, the gap amending regularization (GAR)[10] was implemented in both denoising term and deconvolution term to correct the inconsistency between the inputs and targets which are originally interleaved in the same noisy image stack.

It is noteworthy that since the theoretical basis of ZS-DeconvNet is model-agnostic, both U-net and RCAN are not the only applicative backbone models but the widely adopted and efficient ones. Equipping ZS-DeconvNet with other state-of-the-art network architectures, e.g., DFCAN[8] and RLN[12], may further improve its denoising and SR capability.

Implementation details of ZS-DeconvNet

The image pairs $(\hat{y}, \tilde{y})$ used for training 2D ZS-DeconvNet models were generated following a modified scheme from the original recorrupted to recorrupted strategy[23] under the assumption of mixed Poisson-Gaussian noise distributions, where three hyperparameters β_1, β_2, α needed to be pre-characterized. The recorruption procedure from a single noisy image y can be represented in matrix form as:

$$\hat{\boldsymbol{y}} = \boldsymbol{y} + D^T \boldsymbol{z} \tag{4}$$

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - D^{-1} \boldsymbol{z} \tag{5}$$

where $D = \alpha I$ is an invertible matrix defined as a magnified unit matrix by a factor of α, which controls the overall magnitude of added noises, and z is a random noise map sampled from a Gaussian distribution with zero means:

$$\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}) \tag{6}$$

$$\sigma^2 = \beta_1 H(\boldsymbol{y} - \boldsymbol{b}) + \beta_2 \tag{7}$$

where β_1 is the Poissonian factor affecting the variance of the signal-dependent shot noise, and β_2 is the Gaussian factor representing the variance of additive Gaussian noises. $\boldsymbol{b}$ is the background, approximately regarded as a fixed value related to the camera, by subtracting which we extracted fluorescence signals from the sample. $H(\cdot)$ is a linear low-pass filter used to preliminarily smooth the image and reduce the noise, and we adopted an averaging filter with a size of 5 pixels in our experiments.

The theoretically optimal value of both β_1 and α is 1, while β_2 is dependent to the camera and can be estimated from the sample-free region of the image itself or pre-calibrated following standard protocols[53]. Evaluations on simulated data has shown that the best denoising and SR performance are achieved at the theoretically optimal values of these hyperparameters regardless of the structure and SNR of the testing images (Supplementary Figs. 3, 4). Lower signal levels may lead to heavier artifacts in the output images of ZS-DeconvNet, but will not change the optimal value of hyperparameters, while our method still outperforms other unsupervised deconvolution approaches (Supplementary Fig. 5).

For ZS-DeconvNet implementations of 2D-SIM, i.e., TIRF/GI-SIM (Fig. 5b, c), every set of raw SIM images were first augmented into two sets of recorrupted raw images through Eq. 4 and 8, and reconstructed into a pair of SR SIM images via the conventional SIM reconstruction algorithm. The generated SIM image pairs were then used for self-supervised training in a similar

manner to 2D WF images. For 3D modalities such as 3D-SIM and LLS-SIM (Fig. 5d, e), post-reconstructed volumetric SIM data instead of the raw images were axially sampled into two SIM stacks respectively containing odd and even slices, which were used in subsequent training procedures of 3D ZS-DeconvNet models with loss functions. The schematic workflow of ZS-DeconvNet-SIM is shown in Fig. 5a.

In this work, ZS-DeconvNet models were trained on a PC with an Intel Core i7-11700 processor and an RTX 3090 graphic processing card (NVIDIA) under the software environment of TensorFlow 2.5.0 and python 3.9.7. Training was typically conducted with the Adam optimizer and an initial learning rate of 0.5×10^{-4}, which would decay with a factor of 0.8 every 10,000 iterations. Training batch size was 4 for 2D images and 3 for 3D stacks. The entire training process usually required 50,000 iterations for 2D images and 10,000 iterations for 3D stacks. Elapsed time of training 50,000 iterations for 2D images of 128×128 pixels and 10,000 iterations for 3D stacks of 64×64×13 voxels was ~45 minutes and ~4 hours, respectively. To eliminate the edge artifacts induced by deconvolution, we typically padded 2 blank slices at the top and bottom of 3D stacks and a margin of 8 pixels for each xy-slice in both training and inference processes (Supplementary Fig. 15a). Particularly, when processing the time-lapsing data of cell mitosis (Fig. 3e, f), the unsupervised property of ZS-DeconvNet enabled a transfer adaptation learning strategy[37] in which we first trained a general model for each biological structure with data of the entire process and then finetuned the pre-trained model for each timepoint with a small number of training steps (typically 50 iterations taking ~1 min) to fully exploit the structural information of the raw data and obtain the optimal SR performance.

Data post-processing and SR image evaluation

For imaging modalities employing wide-filed detection such as LLSM, the fixed pattern noise (FPN) which are induced by the nonuniformity in the pixel sensitivity of the camera cannot be removed by noise2noise-based schemes[54]. In our implementation of ZS-DeconvNet, the FPN would be enhanced in the deconvolution stage and became nonnegligible especially at imaging conditions of extremely low SNR. For sCMOS sensors, which are the most common in fluorescence microscopy, the fixed pattern usually presents a regular appearance of horizontal or vertical stripes attributed to the column amplifier. To this end, we simply applied an apodization mask in Fourier domain to suppress the stripy artifacts while preserving other frequency

components of the samples (Supplementary Fig. 15b). It is noted that the fixed pattern noise can also be fundamentally removed by pre-calibration for the acquired raw images before sent into the network model following the well-established procedures[53, 55, 56].

Other computational SR approaches compared in this work, i.e., the sparse deconvolution[5], DeepCAD-based deconvolution[30], and SRRF[13] are implemented following the instructions in the original papers. Specifically, we tried our best to select the optimal hyperparameters for sparse deconvolution and SRRF to obtain a reconstructed image with the least artifacts and the highest resolution. And the DeepCAD-based deconvolution (Figs. 2a and 3f) was carried out by integrating the temporally sampling scheme into our ZS-DeconvNet framework, that was, using images temporally sampled from the time-lapsing data for training our dual-stage network models.

To quantitatively evaluate the SR performance of 2D ZS-DeconvNet and other computational SR approaches with only diffraction limited references, we calculated PSNR between clear WF targets and SR images degraded with the PSF by following three steps: (1) Convolving the SR image with the corresponding PSF and down-sampling the convolved image y to the size of GT; (2) Normalizing the GT image x in the range of [0, 1] and then applying a linear transformation[8, 46] to the convolved image y to match its dynamic range with x:

$$\tilde{\boldsymbol{y}} = a\boldsymbol{y} + b \tag{8}$$

$$(a, b) = \underset{(\theta_1,\theta_2)\in\mathbb{R}^2}{argmin}\left(\|\theta_1\boldsymbol{y} + \theta_2 - \boldsymbol{x}\|_2^2\right) \tag{9}$$

(3) Calculating the PSNR between the normalized GT image x and linearly transformed image $\tilde{y}$.

For PSNR evaluation of 3D ZS-DeconvNet (Fig. 3d), we directly leveraged the LLS-SIM images as the reference in that both of LLS-SIM and our 3D ZS-DeconvNet provided a resolution improvement by ~1.5-fold theoretically. The overall calculation process is similar to the 2D cases, except that the SR stacks were not convolved and the PSNR was only calculated within the feature-only regions with a threshold of 0.02 to avoid obtaining an abnormally high value of PSNR.

Biological sample preparation

Cell culture, transfection, and staining

Cos7 and 293T cells were cultured in DMEM (Gibco), supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin in 37℃ with 5% CO_2. For live cell imaging, the coverslips were pre-coated with 50μg ml^{-1} of collagen and cells were seeded onto coverslips with about 70%

density before transfection. After 12 h, cells were transfected with plasmids using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol. Cells were imaged for 12-24 hours after transfection. HeLa stable cell lines marked endoplasmic reticulum protein and Cos7 cells marked Lifeact were constructed by lentivirus and retrovirus packaging methods respectively, and also cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin in 37℃ with 5% CO_2. Cells were plated about 80% density 12-24 h before imaging. All live cell imaging were performed in a stage top incubator (Okolab) to maintain condition at 37℃ with 5% CO_2. Where indicated, the cells transfected with Halo Tag plasmids were labelled with JF549 ligand following the published protocol [57], and the cells were imaged immediately afterward. The plasmid constructs used in this study include Lifeact-mEmerald, clathrin-mEmerald, 3xmEmerald-Ensconsin, Lamp1-Halo, 2xmEmerald-Tomm20, Myosin2-Halo, KDEL-mCherry and Halo-Calnexin.

Genome edited cell lines

SUM159 cells were genome edited sequentially to incorporate EGFP to the N-terminus of Rab11A and then Halo to the C-terminus of Lamp1 using the CRISPR/Cas9 approach[58, 59]. The single-guide RNA (sgRNA) targeting sequences are 5'-TCGCTCCTCGGCCGCGCAAT-3' for *RAB11A* and 5'-CTATCTAGCCTGGTGCACGC-3' for *LAMP1*.

SUM159 were transfected with the EGFP-Rab11A donor plasmid, the plasmid coding for the spCas9 and the free PCR product containing the sgRNA targeting sequence using Lipofectamin 3000 (Invitrogen) according to the manufacturer's instruction. The cells expressing EGFP were enriched by fluorescence-activated cell sorting (FACS) (FACSAria II, BD Biosciences), and further subjected to single cell sorting to 96-well plates. The monoclonal cells with successful EGFP incorporation were identified by PCR screening using GoTaq Polymerase (Promega). The clonal SUM159 cells expressing EGFP-Rab11A$^{+/+}$ were subjected to the second round of genome editing to incorporate Lamp1-Halo in the genome as described above. The transfected cells were stained by the Janelia Fluor 646 HaloTag Ligands (Promega) and then enriched by FACS. The monoclonal SUM159 cells expressing both EGFP-Rab11A$^{+/+}$ and Lamp1-Halo$^{+/+}$ were confirmed by PCR and western blot analysis.

SUM159 cells were genome edited to incorporate EGFP to the C-terminus of clathrin light chain A (clathrin-EGFP) using the TALEN-based approach as described[60]. The clathrin-EGFP expressing cells were enriched by two sequential bulk sorting.

HeLa cells lines were genome edited to incorporate mEmerald into the N- terminus of human genomic SC35 using CRISPR-Cas9 gene editing system. The sgRNA targeting sequence is 5'-CGAGCAGCACTCCTAATGAT-3' for SC35, and the sgRNA was ligated into pX330A-1×2 (Addgene, 58766). To construct donor vector, mEmerald flanked with about 1800bp homology arms complementary to the 3' end of human genomic SC35 locus were ligated to pEASY-blunt (Transgene, CB101). HeLa cells were transfected with sgRNA vectors and donor vectors at a 3:1 ratio. After 48 hours, cells positive expressed mEmerald were sorted using FACS (FACSAria III, BD Biosciences). After one week, H2B-mCherry lentivirus were infected sorted cells and then separated monoclonal cell into 96-well. About two weeks, cell clones were identified by PCR and western blot, homozygous cells were used for the study.

Lentivirus packaging and stable cell line

For lentivirus packaging, the lentiviral transfer vector DNA, together with psPAX2 packaging and pMD2.G envelope plasmid DNA were co-transfected to HEK293T cells using Lipofectamine 3000 (Invitrogen) following the manufacturer's protocol. After 2 days, supernatant containing viral particles was harvested and filtered with a 0.45-mm filter (Millipore), 200 µl of filtered viral supernatant was added to the cells. Forty-eight hours after transduction, the positive cells were enriched by flow cytometer (FACSAria III, BD Biosciences) and then plated one cell per well into 96-well plates, Monoclonal cells were used for our experiments. Specifically, Calnexin-halo for HeLa used in Fig. 3; Lifeact-mEmerald for Cos7 used in Figs. 3 and 5; calnexin-mEmerald, Mito-dsRed and Halo-H2B for HeLa cells used in Fig. 3; H2B-mCherry for HeLa-mEmerald-SC35 used in Supplementary Fig. 11.

C. elegans embryo preparation

C. elegans strains were cultured at 20 °C on nematode growth medium (NGM) plates seeded with OP50 following standard protocols[61]. TV52712*[wyEx51119[dlg-1p::GFP::PLCdPH]*; *jcIs1[ajm-1::GFP+UNC-29(+)+rol-6(su1006)]*; *qxIs257 [ced-1p::nuc-1::mCherry + unc-76(+)]]* was used in this study. The plasmid *dlg-1p::GFP::PLCdPH* was constructed following the Clontech In-Fusion PCR Cloning System[62] and microinjected to *jcIs1;qxIs257*. Extrachromosomal array

wyEx51119 marked epidermal cell membrane. *jcIs1* marked the apical junctional domain of C. elegans[62]. *qxIs257* marked lysosomes in epidermal cells[63].

About 50 L4 stage transgenic worms were put onto NGM plates with freshly OP50 48 to 60 hours before experiments. Transgenic eggs were collected under the dissecting fluorescent microscope (Olympus MVX10), and mounted on 3% agarose pads. Lima bean to 2-fold stage embryos were then imaged using the 3D WF mode of our Multi-SIM system.

Mouse embryo preparation

Mice used in this study were of C57BL/6J background. All animal experiments were approved by the Animal Care and Use Committees (IACUC) of the Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. Pre-implantation embryos were isolated from 5-6-week-old females, superovulated by intraperitoneal injection of 5 international units (IU) of pregnant mares' serum gonadotropin (PMSG; LEE BIOSOLUTIONS) and 5 IU human chorionic gonadotropin (hCG; Millipore) 48 h later, and mated with male mice. Zygotes were recovered at E0.5 in M2 medium (Millipore) and cultured in KSOM medium (Millipore) in CO2 incubator (Thermo Scientific) at 37℃ with 5% CO2 until the late 8-cell stage.

For immunofluorescence, embryos were fixed with 4% paraformaldehyde in PBS for 30 min at room temperature (RT) and washed with PBS three times. Embryos were then permeabilized in 0.5% TritonX-100 (Sigma) in PBS for 20 min at RT, washed in PBS three times, blocked in 1% bovine serum albumin in PBS for 1 h at RT and incubated with anti-pERM antibody (Abcam, ab76247), anti-alpha-tubulin-FITC (Sigma, F2168-.2ML) and Phalloidin-Rhodamine (Molecular Probes, R415) overnight at 4℃. Then, embryos were washed in PBS three times, incubated with secondary antibodies (Life technologies) for 1 h at RT, stained with Hoescht 33342 (Thermo) for 15 min at RT, washed in PBS three times and imaged by the home-built confocal microscope.

3D image visualization

The axially color-coded images of lysosomes shown in Fig. 4f, g were generated with Fiji. The 3D rendering images of mitosis cell and mouse embryos shown in Figs. 3e, f were visualized and generated by using of the commercial software Amira.

Methods only references

53. Castello, M. et al. A robust and versatile platform for image scanning microscopy enabling super-resolution FLIM. *Nature Methods* **16**, 175-178 (2019).

54. Liu, S. et al. sCMOS noise-correction algorithm for microscopy images. *Nature Methods* **14**, 760-761 (2017).

55. Lehtinen, J. et al. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189* (2018).

56. Mandracchia, B. et al. Fast and accurate sCMOS noise correction for fluorescence microscopy. *Nature Communications* **11**, 1-12 (2020).

57. Diekmann, R. et al. Photon-free (s)CMOS camera characterization for artifact reduction in high- and super-resolution microscopy. *Nature Communications* **13**, 3362 (2022).

58. Grimm, J.B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods* **12**, 244-250 (2015).

59. He, K. et al. Dynamics of phosphoinositide conversion in clathrin-mediated endocytic traffic. *Nature* **552**, 410-414 (2017).

60. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281-2308 (2013).

61. Sanjana, N.E. et al. A transcription activator-like effector toolbox for genome engineering. *Nature Protocols* **7**, 171-192 (2012).

62. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).

63. Köppen, M. et al. Cooperative regulation of AJM-1 controls junctional integrity in Caenorhabditis elegans epithelia. *Nature Cell Biology* **3**, 983-991 (2001).

64. Li, Y. et al. The lysosomal membrane protein SCAV-3 maintains lysosome integrity and adult longevity. *Journal of Cell Biology* **215**, 167-185 (2016).

Acknowledgements

The authors thank T. Kirchhausen for the donor plasmids used for genome editing and help in generating the genome-edited cell lines, and thank the Caenorhabditis Genetics Center and Prof. Xiaochen Wang for C. elegans strains. This work was supported by grants from the National Natural Science Foundation of China (31827802, 32125024, 31970659, 32271513, 62071271, 62088102, and 62222508); the Ministry of Science and Technology (2021YFA1300303 and 2020AA0105500); the Youth Innovation Promotion Association of Chinese Academy of Sciences (2020094); the Chinese Academy of Sciences (ZDBS-LY-SM004 and XDA16021401); the Collaborative Research Fund of the Chinese Institute for Brain Research, Beijing (2021-NKX-XM-03); China Nation Postdoctoral Program for Innovative Talents (BX2021159); China

Postdoctoral Science Foundation (2022M721842); the Tencent Foundation through the XPLORER PRIZE; the Shuimu Tsinghua Scholar Program; and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2020094).

Author contributions

Q.D. and Dong Li supervised the research. Q.D., Dong Li, and C.Q. conceived and initiated this project. C.Q. designed the detailed implementations under the instruction of Q.D. and Dong Li. Y.Z, C.Q., and X.C developed the python code, performed simulations, and processed relevant imaging data. T.J., R.W, C.Q, H.L., Di Li, and J.G. prepared samples and performed imaging experiments. C.Q., Y.Z., X.C., and Q.M. analyzed the data with conceptual advice from Q.D., Dong Li, J.W, and H.Q. C.Q., Y.Z, and Q.M. composed the figures and videos under the supervision of Q.D. and Dong Li. Q.D., Dong Li, and C.Q. wrote the manuscript, with input from all authors. All authors discussed the results and commented on the manuscript.