PhotoniX

RESEARCH Open Access

Ensemble deep learning-enabled single-shot composite structured illumination microscopy (eDL-cSIM)

Jiaming Qian[1,2,3†], Chunyao Wang[1,2,3†], Hongjun Wu[1,2,3*], Qian Chen[3*] and Chao Zuo[1,2,3*]

†Jiaming Qian and Chunyao Wang contributed equally to this work.

*Correspondence: wuhongjun2022@njust.edu.cn; chenqian@njust.edu.cn; zuochao@njust.edu.cn

[1] Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, No. 200 Xiaolingwei Street, 210094 Nanjing, Jiangsu Province, China
[2] Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, 210094 Nanjing, Jiangsu Province, China
[3] Jiangsu Key Laboratory of Visual Sensing & Intelligent Perception, No. 200 Xiaolingwei Street, 210094 Nanjing, Jiangsu Province, China

Abstract

Structured illumination microscopy (SIM) has emerged as a powerful super-resolution technique for studying protein dynamics in live cells thanks to its wide-field imaging mode and high photon efficiency. However, conventional SIM requires at least nine raw images to achieve super-resolution reconstruction, which limits its imaging speed and increases susceptibility to rapid sample dynamics. Moreover, the reliance of SIM on illumination parameters and algorithmic post-processing renders it vulnerable to reconstruction artifacts, especially at low signal-to-noise ratios. In this work, we propose a single-shot composite structured illumination microscopy method using ensemble deep learning (eDL-cSIM). Without modifying the original SIM setup, eDL-cSIM employs only one composite structured illumination pattern generated by 6-beam interferometry. The resultant composite-coded raw image, which contains multiplexed high-frequency spectral information beyond the diffraction limit, is further processed using ensemble deep learning to predict a high-quality, artifact-free super-resolved image. Experimental results demonstrate that eDL-cSIM integrates the advantages of various state-of-the-art neural networks, enabling robust super-resolution image predictions across different specimen types or structures of interest, and outperforms classical physics-driven methods in terms of imaging speed, reconstruction quality and environmental robustness, while avoiding intricate and specialized algorithmic procedures. These collective advantages make eDL-cSIM a promising tool for fast and robust live-cell super-resolution microscopy with significantly reduced phototoxicity and photobleaching.

Keywords: Structured illumination microscopy, Super-resolution, Deep learning

Introduction

Over the past few decades, fluorescence super-resolution imaging techniques that break the Abbe diffraction limit, such as stochastic optical reconstruction microscopy (STORM) [1], photoactivated localization microscopy (PALM) [2], structured illumination microscopy (SIM) [3], and stimulated emission depletion microscopy (STED) [4], have become important research tools in the field of life sciences, enabling human beings to observe fine subcellular structures at the nanoscale or even the single-molecule level [5]. Among various super-resolution techniques, SIM, a wide-field method

capable of doubling the lateral resolution, has attracted widespread attention in dynamic super-resolution imaging of live cells thanks to its advantages of full-field imaging, high photon efficiency, and compatibility with regular fluorescent dyes [6–13].

Typically, SIM utilizes grating-structured illuminations to modulate the high-frequency information of the sample beyond the diffraction limit into the support domain within the cutoff frequency of the system, thereby achieving super-resolution [14, 15]. In order to recover the high-frequency information, at least three phase-shifting images at three distinct orientations ($\geq$ 9 frames) need to be acquired for demodulating and expanding the lateral full-field Fourier-domain sub-apertures [16, 17]. Despite the faster imaging speed compared to STED, STORM, etc., the multi-frame nature still causes SIM to suffer from decreased temporal resolution and increased photodamage, which exacerbates its susceptibility to specimen motion and subcellular organelle dynamics, making it less conducive to rapid, long-term live-cell imaging [18, 19]. In addition, high-quality SIM super-resolution reconstruction usually requires precision knowledge of the experimental illumination parameters, the estimation of which is computationally cumbersome and time-consuming, and minor parameter errors may lead to severe reconstruction artifacts [20–25]. At low signal-to-noise ratios (SNRs), noise will make the SIM reconstructions more vulnerable to artifacts that affect quantization and fidelity [26, 27]. In order to maximize imaging speed and minimize photodamage for rapid live-cell imaging, researchers tend to compress the photon dose required for reconstruction by shortening the exposure time and decreasing the excitation power, but at the cost of dramatically deteriorating SNRs. Although numerous improved reconstruction approaches, e.g., total variation-based SIM (TV-SIM) [28], Hessian-SIM [26], high-fidelity SIM (Hifi-SIM) [29], Sparse-SIM [27], flexible SIM (FlexSIM) [30], principal component analysis-based SIM (PCA-SIM) [31], have been proposed to mitigate noise-induced artifacts, they are not always effective for artifacts arising from all possible factors and somewhat increase the algorithm complexity, with more parameter settings affecting the generalizability of SIM. Furthermore, these methods cannot fundamentally address the motion susceptibility issue caused by the multi-frame nature of SIM, as they still need to acquire a series of illumination patterns for each reconstructed high-resolution SIM image. Reducing the number of raw patterns required for reconstruction is a more effective and direct scheme to improve photon efficiency [32, 33]. However, the reconstruction quality of existing frame-reduced algorithms is inferior to that of regular 9-frame reconstructions in practical imaging experiments due to their reliance on strict optical setups or complex noise models [34–37].

Recently, deep learning has brought about significant breakthroughs in SIM, indicating the great potential to revolutionize its future development [32, 38–47]. It has been proven that the application of deep learning can reduce the number of raw images while retrieving super-resolution information, thereby increasing the speed of SIM [32, 38] and even enabling single-shot SIM reconstruction [46, 48, 49]. However, since a single image in conventional illumination mode contains super-resolved information at only one orientation, the quality of network prediction using it as input is compromised [48, 49]. To modulate isotropic super-resolution information in a single illumination, some researchers leveraged multiple side-branching galvanometer sets to direct three shunted beams and generate interferometric lattice illumination, on the basis of which

deep learning-enabled single-frame SIM was realized [46]. However, both imaging stability and resolution are compromised due to the increased complexity of the imaging equipment (particularly the need for additional reflective mirror sets driven by piezo stages to introduce phase shifts for generating conventional illumination patterns) and the lower frequency of the illumination pattern generated by three-beam interference compared to that in conventional illumination mode (see Supplementary Note S1 for detailed analysis).

In this work, we propose a single-shot composite structured illumination microscopy method enabled by ensemble deep learning (eDL-cSIM). eDL-cSIM adopts a composite structured illumination instead of the conventional means, which encodes the lateral full-field high-frequency spectrum information with only a single modulation to break through the diffraction limit, increasing the imaging speed by 9 times without complicating the optical setup. Furthermore, different from conventional reliance on a single deep neural network (DNN) with limited learning attributes, we introduce ensemble learning [50–52] into SIM reconstruction to achieve superior performance over several state-of-the-art DNNs through an adaptive ensemble model based on the Transformer architecture across different specimen types or structures of interest. Experiments demonstrate that eDL-cSIM achieves faster imaging speed, higher reconstruction quality and better noise immunity compared to physics-driven methods at only one-ninth of the conventional photon dose, making it a promising imaging tool for fast, long-duration observation of live subcellular structures.

Principle of eDL-cSIM

Composite structured illumination for single-shot isotropic lateral spectrum modulation

In the classical SIM optical setup, the illumination field is generated by the interference of ± 1-order diffraction light from a collimated laser irradiated on a spatial light modulator (SLM) displaying the dense grating pattern. Since a single illumination can only modulate the sample spectrum distributed in the direction orthogonal to the fringe orientation, at least three illumination fields with orientation angles rotated by 120° from each other are required to implement isotropic lateral super-resolution. Figures 1a and b illustrate the well-designed mask used to filter the ± 1-order diffraction light, in which diffraction light of various orientations passes through the mask at different time points, respectively. In order to maximize the modulation efficiency, the illuminations in the temporal domain can be compressed into the spatial domain, i.e., generating diffraction light of different orientations at the same time point by 6-beam interference instead of the regular two-beam interference [Fig. 1c and d]. Accordingly, the classical grating structured illumination is replaced with a composite structured illumination, which superimposes standard grating patterns of three orientations, thus realizing single-shot full-field spectrum modulation without modifying the optical setup, as shown in Fig. 1e and f. Figure 1g illustrates the customized SIM setup based on a commercial inverted fluorescence microscope (IX73, Olympus, Japan) using the above-mentioned composite structured illumination strategy. In the illumination path, three laser beams of different wavelengths (405 *nm*, 488 *nm* and 561 *nm*) were coupled through a series of dichroic mirrors and subsequently expanded and collimated. The composite excitation pattern was then generated by a ferroelectric liquid crystal SLM (QXGA- 3DM, Fourth

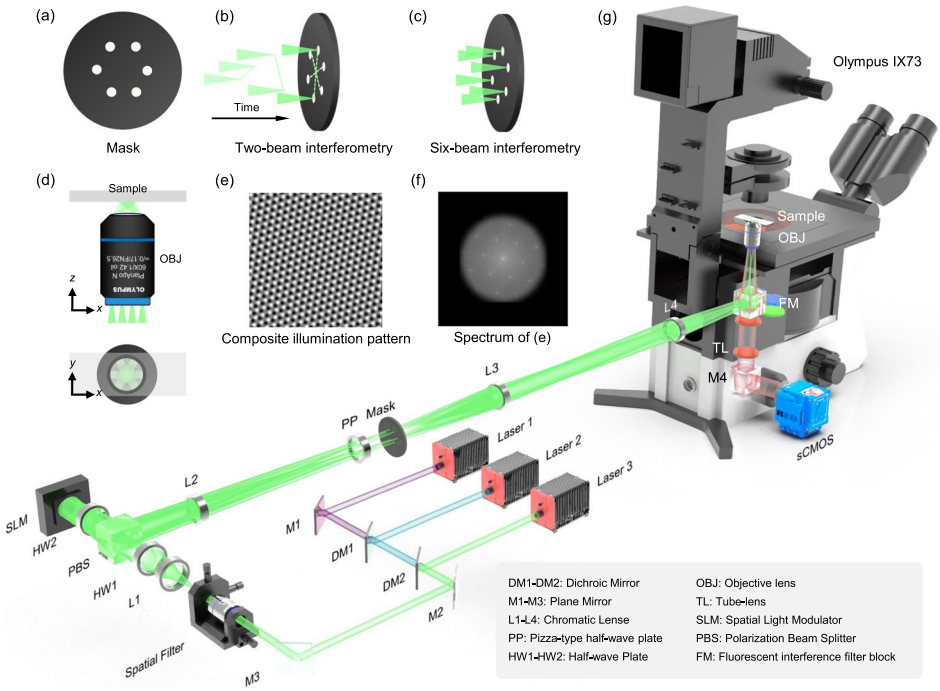


Fig. 1 Composite structured illumination for single-shot isotropic lateral spectrum modulation. **a** The well-designed mask used to filter the ± 1 diffraction light. **b** The classical structured illumination based on two-beam interference, which requires the use of at least three illuminations to modulate the full-field spectrum information. **c** The composite structured illumination based on 6-beam interference, which modulates the full-field spectrum information using only one illumination. **d** 6-beam interference on the sample surface. **e** The composite illumination pattern that superimposes standard grating patterns of three orientations. **f** The spectrum image of (**e**). **g** The customized SIM setup based on a commercial inverted fluorescence microscope using composite structured illumination

Dimension Displays, UK) with a high frame rate. In the imaging path, a high numerical aperture (NA) objective and an sCMOS camera (PCO Edge 5.5, PCO, Germany) with 60% peak quantum efficiency were used to detect the emitted fluorescence. The display of SLM was triggered by the sCMOS exposure to ensure timing synchronization. More information on the customized SIM setup is provided in Methods and Supplementary Figs. S1-S3.

High-quality super-resolution reconstruction based on ensemble learning

Although we have achieved single-shot isotropic spectrum modulation, the inverse problem of super-resolution reconstruction is severely ill-posed due to the fact that, in addition to the desired six spectrum orders carrying the highest illumination frequencies, there are other spectrum components generated by the interference of diffracted light from different orientations, as shown in Supplementary Fig. S2. Unlike conventional reconstruction methods that reformulate a well-posed regression task at the expense of increasing the number of raw images (see Supplementary Note S2 for more details), we apply deep learning to bypass the obstacle of solving the ill-posed inverse problem and directly establish the pseudo-inverse mapping relation between the composite-coded raw image and the high-quality super-resolution image (see Supplementary Note S3 for

more details). More specifically, an adaptive ensemble neural network with the Transformer architecture as the main module is designed to integrate the features extracted by three state-of-the-art base convolutional neural networks (CNNs), thus achieving better reconstruction performance than any individual base CNN. Moreover, this combination leverages the advantages of CNN and the Transformer architecture, allowing the model to accommodate both global information and local features. Figure 2 illustrates the overall network architecture of eDL-cSIM.

First, we employ U-Net [53, 54], which combines a simple and efficient network structure with excellent performance, as Base Network 1. U-Net can learn multiscale features and map them at the pixel level through hierarchical cross-connections in the spatial domain, making it applicable to various biomedical visual tasks [55–61]. In addition, the encoder-decoder structure ensures that shallow feature information can be well preserved. Based on the above characteristics, we utilize U-Net to learn diffraction-limited wide-field information from a composite-coded raw illumination image, and set its architecture to five groups of encoders and decoders with residual connections established, as shown in Fig. 2a1. Next, two additional base networks are applied to recover the high-frequency information of the sample

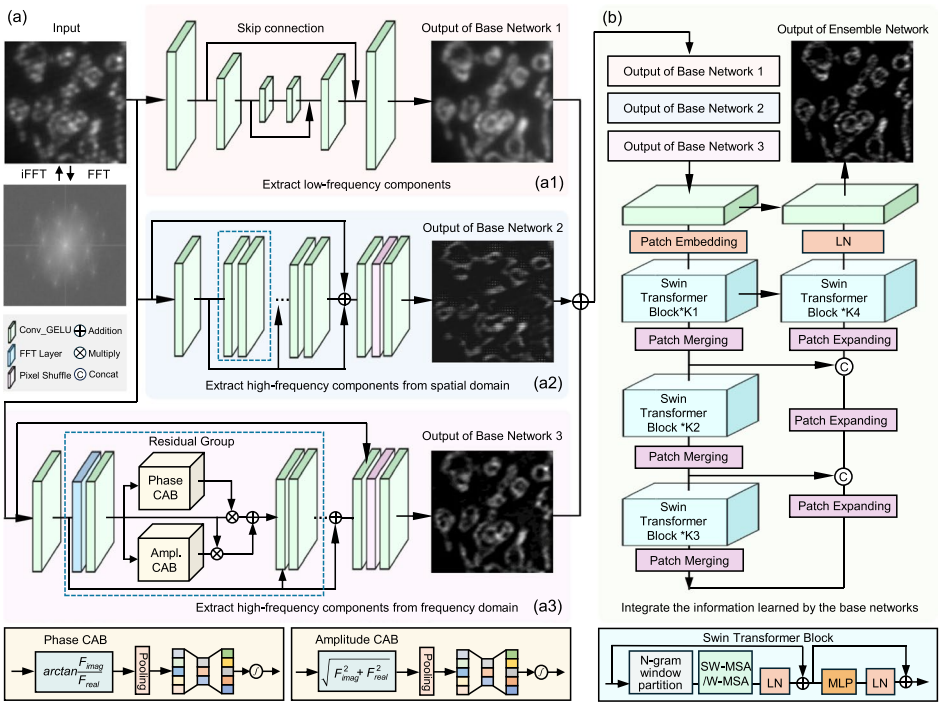


Fig. 2 The overall network architecture of eDL-cSIM. **a** Architectures of the base networks. Base Network 1 is a U-Net (a1), which is used to learn the wide-field information from a composite-coded raw input. Base Network 2 is a recursive residual network (a2), which is used to establish the mapping relationship between a single-frame input and the super-resolution image from the real-space domain. Base Network 3 is a DNN with a Fourier channel attention mechanism of amplitude and phase dual branches (a3), which is used to recover the high-frequency information from the Fourier domain. **b** Architecture of the ensemble network, which is used to integrate the output information of the three base networks, thus achieving better reconstruction performance. F_{imag} and F_{real} represent the imaginary and real parts of the input spectrum respectively, 'CAB' means 'Channel Attention Block', 'SW-MSA' means 'Shifted Window Multihead Self Attention', 'W-MSA' means 'Window Multihead Self Attention', 'LN' means 'LayerNorm', and 'MLP' is 'Multilayer Perceptron'

from the real-space domain and the Fourier domain, respectively. For Base Network 2, we adopt a recursive residual network [62] to establish the mapping relationship between the composite-coded input and the super-resolution image [Fig. 2a2]. The recursive residual network consists of residual blocks primarily composed of convolutional layers, including local residual links and global recursive residual links. Local residual links enable information retention and gradient alleviation, while global recursive residual links connect each residual block directly to the input without increasing network complexity, thereby predicting high-frequency information from the composite-coded input more effectively in the real-space domain. Furthermore, considering that the high-frequency components to be recovered are more easily distinguishable in the Fourier domain, we construct Base Network 3 through a Fourier channel attention mechanism with amplitude and phase dual branches [63] [Fig. 2a3]. Such network architecture can leverage the spectrum differences among various features to learn precise representations of the high-frequency information in biological structures, offering more reliable guarantees for high-quality super-resolution reconstruction. More implementation details on the base networks are provided in Supplementary Notes S4.1-S4.3 and Figs. S4-S6.

To adaptively integrate and model long-range dependencies from information extracted by the base networks, we employ Vision Transformer (VIT) as the primary module to construct the ensemble network, as illustrated in Fig. 2b. By dividing the input into multiple sub-windows and computing the multi-head attention mechanism, VIT can process different aspects of input sequences in parallel to provide multiple representation subspaces for attention layers, thus enabling the network to efficiently capture global information and long-distance dependencies in images and improve the accuracy of feature extraction. To fully utilize the interactive information between different windows, we adopt the N-Gram context method to partition the windows and compute the sliding window attention mechanism with Swin Transformer, which treats adjacent local windows as a set of uni-grams [64]. Consequently, the image embedding operation is redefined, i.e., adopting a group convolution with a stride of M and groups of $C/2$ to embed the input into a $C/2 \times Wh \times Ww$ dimension (where C represents the number of channels, M is the window size, and Wh and Ww represent the width and height of the window). In each N-Gram unit, there are M^2 pixels. The sliding window attention mechanism is implemented using the sliding operation in convolution, with padding applied to the image edges, thus achieving bidirectional window-self-attention (WSA). The WSA weights are shared across all directions. After computing the forward and backward WSA, the bidirectional WSA results are concatenated and a 1×1 convolution is used to generate the N-gram context. In a single Transformer block, residual connections are employed, and patch merging is performed before the next stage. Between two stages, max-pooling is used to halve the resolution of the feature map and expand the number of channels, aiding the network in learning meaningful representations. Finally, a 3×3 convolution is applied to learn the representation within each channel space to output the desired image. Supplementary Note S4.4 anf Figs. S7 and S8 provide more implementation details about the ensemble network.

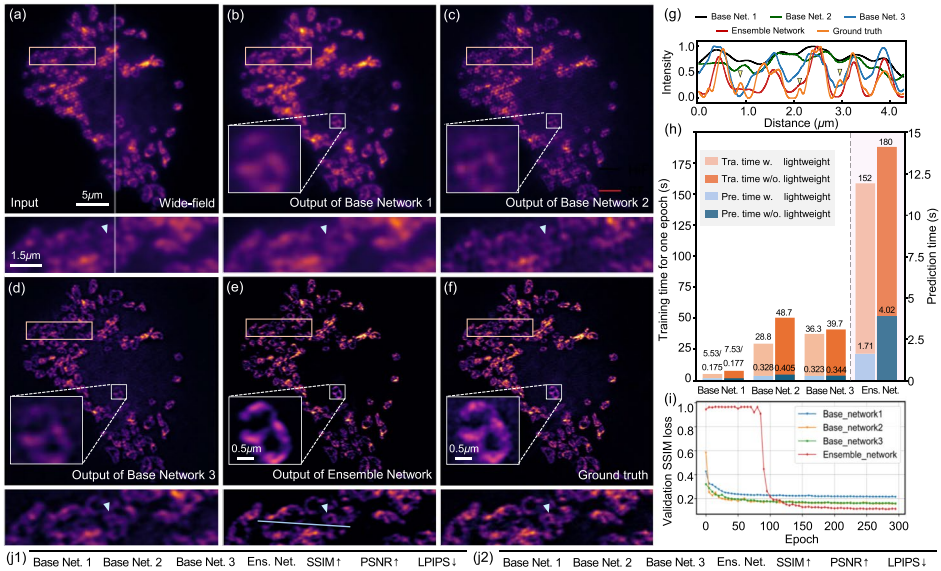
Network lightweighting to reduce the training and predicting time

Since we utilize multiple base networks and an ensemble network, the training and predicting time will increase dramatically compared to using only a single DNN. To address this issue, network lightweighting is performed to achieve more efficient data training and prediction. First, to reduce the memory consumption of the base networks in the graphics processing unit (GPU), partial convolution (PConv) is used as the basic convolutional block to replace the convolutional layers in the base networks [65]. Compared to the conventional convolution operation, PConv features advantages of high computational efficiency and fewer model parameters, while ensuring effective extraction of spatial features. More technical details about PConv are provided in Supplementary Note S5 and Fig. S9. For the ensemble network, unlike the symmetric encoder-decoder structure of Swin-Unet [66], we merge the outputs of each stage in the encoder through patch expansion, utilizing PixelShuffle [67] and concatenation to skip-connect information from each stage of the encoder, as can be seen in Fig. 2b. This operation concentrates multi-scale information delivery to the decoder and significantly enhances network efficiency.

Results

Experiments to demonstrate superior super-resolution reconstruction performance over individual base networks through ensemble learning

To test the performance of ensemble learning, we used a composite-coded raw image of cellular mitochondria that did not appear in the training set as input to the base networks and the ensemble network, respectively. The test object was a fixed sample of bovine pulmonary artery endothelial (BPAE) cells with DAPI-labeled nucleus, Alexa Fluor^{TM} 568-labeled actin and MitoTracker^{TM} Green FM-labeled mitochondria. The prediction results of different neural networks are illustrated in Fig. 3, from which it can be seen that Base Network 1 removes the honeycomb-shaped illumination field in the input and achieves the recovery of the low-frequency component of the sample, i.e., the wide-field image [Fig. 3a and b, and Supplementary Fig. S11]. Distinctly, Base Networks 2 and 3 obtain super-resolution images from the real-space domain and the Fourier domain, respectively [Fig. 3c and d]. However, although DNNs focusing on high-frequency information allow for resolution improvements, the imaging quality of a single neural network remains to be enhanced compared to the ground truth (obtained by PCA-SIM [31] from nine conventional illumination patterns) since the composite-coded input overlaps numerous spectrum components, making the super-resolution reconstruction task highly ill-posed. In contrast, after integrating the features extracted by different base networks, a high-fidelity super-resolution image can be derived, with a quality almost comparable to that of the ground truth [Fig. 3e and f]. Furthermore, we find that for regions with low SNRs, such as where the cell nucleus is located, the results reconstructed by the physics-driven method suffer from artifacts interfering with the fidelity due to the influence of the out-of-focus background, whereas eDL-cSIM achieves artifact-free super-resolution reconstruction, as shown by the magnified mitochondrial details from the white-boxed regions in Fig. 3e and f. Further analysis of the fluorescence intensities along the blue line from Fig. 3e in different results indicates that eDL-cSIM



(j1)

Base Net. 1	Base Net. 2	Base Net. 3	Ens. Net.	SSIM↑	PSNR↑	LPIPS↓
✓				0.49	15.02	0.39
	✓			0.82	19.71	0.19
		✓		0.84	22.12	0.17
			✓	0.77	20.03	0.23

(j2)

Base Net. 1	Base Net. 2	Base Net. 3	Ens. Net.	SSIM↑	PSNR↑	LPIPS↓
✓	✓		✓	0.82	20.23	0.17
	✓	✓	✓	0.90	22.67	0.12
✓		✓	✓	0.88	22.30	0.14
✓	✓	✓	✓	**0.95**	**25.54**	**0.08**

Fig. 3 Prediction results of the base networks and the ensemble network for mitochondria of a fixed BPAE cell samples with DAPI-labeled nucleus, Alexa Fluor^{TM} 568-labeled actin and MitoTracker^{TM} Green FM-labeled mitochondria. **a** Comparison of the composite-coded raw image and the wide-field image. The raw SIM images (with the resolution of 512 × 512) were captured through a 100× objective (UPlanSApo 100×/1.40 Oil, Olympus, Japan). **b**-**d** Prediction results of Base Networks 1–3 using the composite-coded raw image as input, where Base Network 1 outputs the wide-field image, and Base Networks 2 and 3 output the super-resolution image from the real-space domain and the Fourier domain, separately. **e** Prediction result of the ensemble network that integrates features focused on by different base networks. **f** Ground truth obtained from nine raw images in conventional illumination mode by PCA-SIM [31]. **g** Intensity profiles of the results of different networks along the light blue line in **e** (normalized to maximum). **h** Training and prediction time for one epoch of different base networks and the ensemble network before and after lightweighting. **i** Loss convergence curves for different base networks and the ensemble network. **j** The ablation study to evaluate the contribution of the sub-networks to the final result. 'Net.' means 'Network', 'w' and 'w/o' mean 'with' and 'without', and 'Tra.' and 'Pre.' mean 'Training' and 'Prediction'. Experiments were repeated ten times independently with similar results. Colored arrows point to regions where result differences are distinct. Scale bars: 5 μm (full-field-of-view images in **a**-**e**); 1.5 μm (magnified results for the rectangles in **a**-**e**); 0.5 μm (magnified results for the squares in **e** and **f**)

implements the resolution improvement comparable to the ground truth without producing the artifacts pointed out by the yellowish arrows [Fig. 3g].

Regarding the time consumption of network training, the total training time for one epoch of multiple networks is about 223.63 *s*, which is ∼ 19.32% less than that (275.93 *s*) without network lightweighting, and only 23.68% more than that (180.00 *s*) of using a single neural network (the ensemble network) [Fig. 3h]. As shown in Fig. 3i, the loss curves of Base Networks 1, 2, 3 and the ensemble network converged after about 70, 150, 160 and 240 epochs, respectively. For prediction, the efficiency improvement is more significant, with time consumption before and after network lightweighting being 4.95 *s* and 2.54 *s*, respectively, the latter even 63.08% less than that (4.02 *s*) of using a single network [Fig. 3h].

In order to further validate that the reconstruction performance improvement is mainly contributed by ensemble learning, Supplementary Fig. S12 provides a

comparison of the super-resolution prediction results using the single ensemble network and using the multi-network ensemble strategy, indicating that ensemble learning is more conducive to superior reconstruction quality. In addition, we conducted an ablation study to evaluate the contribution of the sub-networks to the final result. As shown in Fig. 3j, the learning performance of each base network alone as well as removing any base network in the ensemble learning strategy was tested using three metrics: structural similarity coefficient (SSIM), peak signal-to-noise ratio (PSNR), and learned perceptual image patch similarity (LPIPS). The results indicate that Base Networks 2 and 3 focusing on high-frequency information contribute significantly more than that centered on low-frequency information, but the lack of any sub-network affects the final reconstruction performance, demonstrating that the ensemble learning strategy can achieve superior reconstruction quality than any individual sub-network.

Comparative experiments to demonstrate the high reconstruction quality, high generalization and well environment robustness of eDL-cSIM

To further validate the advancement of eDL-cSIM, we compared it with two other state-of-the-art DNNs for SIM reconstruction, i.e., Amplitude-Phase Channel Attention Network (APCAN) [63] and Fourier Enhanced and Shifting Tiered Network (FESTN) [46], where FESTN is a deep learning method for single-frame SIM. Both networks used for comparison were trained with the composite-code illumination image and the corresponding super-resolution image as input and output. The prediction results of the BPAE cells obtained by these three methods are shown in Fig. 4. Since the severe spectrum overlapping of the input image poses great challenges to the performance of deep learning, the reconstruction quality of FESTN is relatively compromised, with the predicted results of both mitochondria and actin displaying no significant super-resolution effects [Fig. 4b3 and c3]. Although APCAN obtains resolution-enhanced reconstructions of mitochondria, its fidelity is affected by the spectrum overlapping and its results for actin remain compromised [Fig. 4b4 and c4]. In contrast, eDL-cSIM recovers higher-quality super-resolution images, especially in the results of actin, where fine structures that other DNNs cannot distinguish are well resolved [Fig. 4b5 and c5]. Figure 4d provides the SSIM between the results obtained by different methods and the ground truth, indicating that eDL-cSIM outperforms APCAN and FESTN in terms of reconstruction fidelity. We display the fluorescence intensity profiles of the results obtained by different methods and calculate the full width at half maximum (FWHM) for selected fine structures, as shown in Fig. 4e. It can be seen that eDL-cSIM achieves a resolution comparable to the ground truth, with FWHM of the selected mitochondrial structure being 115.62 *nm* and 116.24 *nm*, and FWHM of the actin structure being 117.27 *nm* and 114.56 *nm*, respectively. Supplementary Figs. S13-S15 present more comparative experimental results on subcellular organelles of BPAE cells, further verifying the advantages of eDL-cSIM in reconstruction quality and fidelity. We then tested the structures not covered in the training set using APCAN, FESTN, and eDL-cSIM. The super-resolution results of the different networks on the microtubules of CV- 1 in Origin Simian- 7 (COS-7) cells and the autofluorescent ascaris sample are illustrated in Fig. 5, from which it can be seen that eDL-cSIM still yields the best reconstruction quality and fidelity among the three methods. Supplementary Figs. S16-S18 provide more comparative experimental

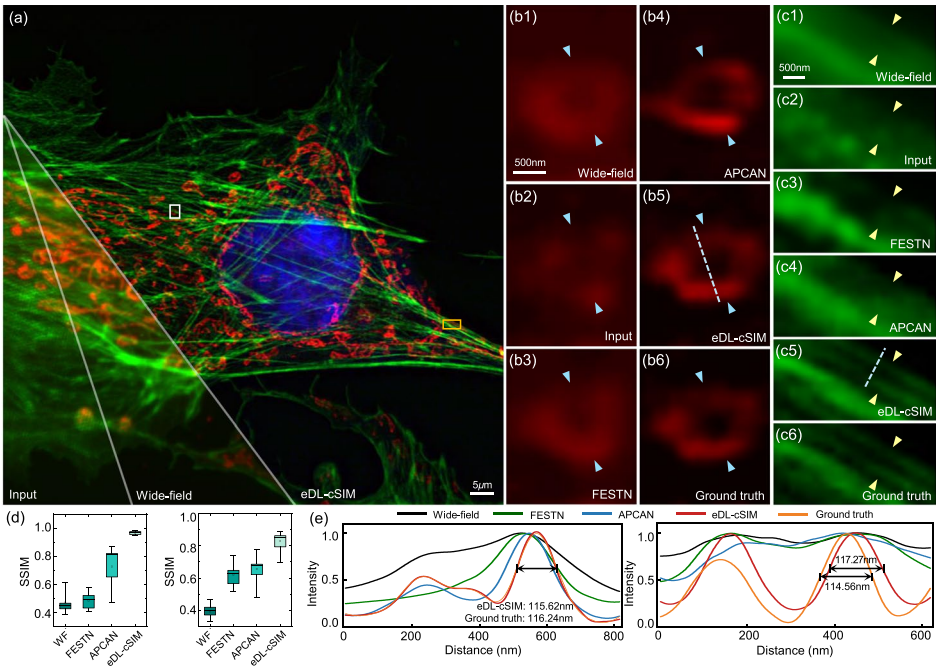


Fig. 4 Comparison of the multi-color super-resolution experimental results on a fixed BPAE cell sample. **a** Comparison of the composite illumination input, wide-field image and the super-resolution image obtained by eDL-cSIM. The raw SIM images with the resolution of 1024 × 1024 were captured through a 100× objective (UPlanXApo 100×/1.45 Oil, Olympus, Japan). For easy distinguishing, we show the mitochondria, actin and nucleus of the BPAE cell in red, green and blue, respectively. **b** Magnified wide-field image, composite illumination image and super-resolution images of mitochondria from the white-boxed regions in (**a**) obtained by different methods (FESTN [46], APCAN [63], eDL-cSIM and PCA-SIM [31]). **c** Magnified wide-field image, composite illumination image and super-resolution images of actin from the yellow boxed regions in (**a**) obtained by different methods. **d** SSIM between super-resolution results obtained by different methods and ground truth for the regions in (**b**) and (**c**). **e** Intensity profiles of the results obtained by different methods along the light blue line in (**b**) and (**c**) (normalized to maximum). 'WF' means 'Wide-field'. Experiments were repeated ten times independently with similar results. Colored arrows point to regions where reconstruction differences are distinct. Scale bars: 5 μm (**a**); 500 *nm* (**b**, **c**)

results for different samples, such as actin and microtubules in COS- 7 cells (Supplementary Fig. S16), microtubules and nucleus in buffalo green monkey kidney cells (BSC-1) (Supplementary Fig. S17), a stratified squamous epithelium sample [Supplementary Figs. S18(a) and S18(b)], a cross section of dicot stem [Supplementary Fig. S18(c)], a human blood smear [Supplementary Fig. S18(d)], microspore mother cells in diplotene of lily [Supplementary Fig. S18(e)] and a stem cross-section of small-leaved linden [Supplementary Fig. S18(f)]. All these experiments demonstrate that eDL-cSIM can achieve higher-quality super-resolution predictions than other state-of-the-art DNNs from only a single-frame composite-coded image, while possessing well generalization.

Next, we tested the performance of eDL-cSIM in complex imaging environments. We gradually reduced the excitation power from the rated value (55 mW) to 25% of the original (with a corresponding reduction in the average photon count of the raw images from 252 to 60) and sequentially captured nine raw images of BPAE mitochondria in regular illumination mode and one image in composite illumination mode. PCA-SIM and eDL-cSIM were applied to obtain super-resolution images under regular and composite illumination modes, respectively. As shown in Supplementary Fig. S19, the results

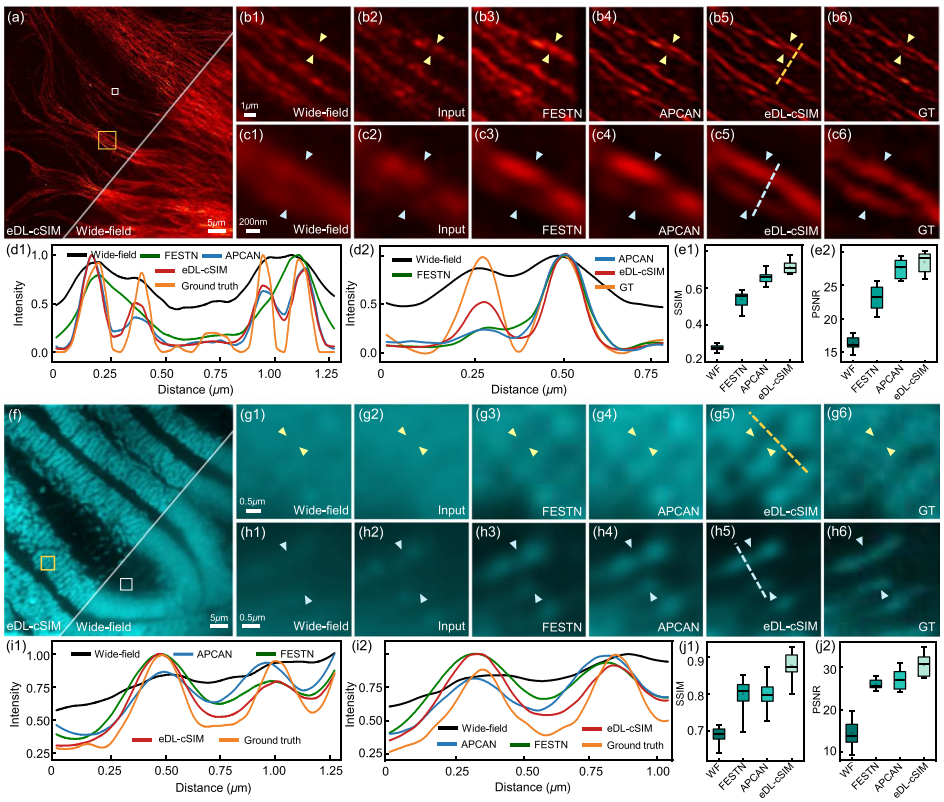


Fig. 5 Comparison of the super-resolution experimental results on COS- 7 microtubules labeled by BODIPY^R FL goat anti-mouse IgG and an autofluorescent ascaris sample. **a** Comparison of the wide-field image and the super-resolution image of COS- 7 microtubules obtained by eDL-cSIM. The raw SIM images with the resolution of 1024 × 1024 were captured through a 100× objective (UPlanXApo 100×/1.45 Oil, Olympus, Japan). **b** Magnified wide-field image, composite illumination image and super-resolution images from the yellow-boxed regions in (**a**) obtained by different methods (FESTN [46], APCAN [63], eDL-cSIM and PCA-SIM [31]). **c** Magnified wide-field image, composite illumination image and super-resolution images from the white-boxed regions in (**a**) obtained by different methods. **d** Intensity profiles of the results obtained by different methods along the yellow and blue lines in (**b**) and (**c**) (normalized to maximum). **e** SSIM and PSNR between super-resolution results obtained by different methods and ground truth shown in (**a**)-(**c**). **f** Comparison of the wide-field image and the super-resolution image of the ascaris sample obtained by eDL-cSIM. The raw SIM images with the resolution of 1024 × 1024 were captured through the same 100×, 1.45 NA objective. **g** Magnified wide-field image, composite illumination image and super-resolution images from the yellow-boxed regions in (**f**) obtained by different methods. **h** Magnified wide-field image, composite illumination image and super-resolution images from the white-boxed regions in (**f**) obtained by different methods. **i** Intensity profiles of the results obtained by different methods along the yellow and blue lines in (**g**) and (**h**) (normalized to maximum). **j** SSIM and PSNR between super-resolution results obtained by different methods and ground truth shown in (**f**)-(**h**). 'GT' means 'Ground truth'. Experiments were repeated ten times independently with similar results. Colored arrows point to regions where reconstruction differences are distinct. Scale bars: 5 μm (**a**, **f**); 1 μm (**b**); 200 *nm* (**c**); 0.5 μm (**g**, **h**)

of PCA-SIM appear increasingly serious reconstruction artifacts as SNR decreases, with the SSIM between the results in different noise environments and that at high SNR reducing from 0.77 to 0.47, whereas eDL-cSIM always maintains stable reconstruction quality, with the SSIM being above 0.85. Experimental results in Fig. 3e and f, and Supplementary Fig. S18 can also demonstrate that eDL-cSIM features excellent noise robustness, in which the results of PCA-SIM are interfered by ambient noise or out-of-focus backgrounds, leading to reconstruction artifacts, while eDL-SIM suppresses these artifacts well and obtains higher quality super-resolution images. In addition to

environmental noise, out-of-focus is also a key factor affecting the reconstruction quality. We sampled the same region of BPAE mitochondria at intervals over a long period of time without interfering with the imaging setup. Due to perturbations in the surrounding environment, the focal plane changed over time, making the image blurred and compromising the reconstruction quality of PCA-SIM (Supplementary Fig. S20). In contrast, eDL-cSIM is relatively less sensitive to out-of-focus, maintaining its SSIM above 0.75 in the case of defocusing where the resultant SSIM of PCA-SIM has dropped to 0.51. The excellent performance of eDL-cSIM under out-of-focus conditions can also be proven in the experiments on thick samples as shown in Supplementary Fig. S17. The above performance enhancement of eDL-cSIM at low SNRs is actually achieved by bypassing the spectrum separation and reorganization in conventional reconstruction algorithms (which are prone to artifacts due to the carrying of various errors into the high-frequency regions [29]) to directly establish a pseudo-mapping between low-quality inputs and high-resolution outputs. In addition, unlike the conventional step-by-step procedure prone to error accumulation, deep learning is more conducive to achieving better global performance by synthesizing various spatio-temporal information [68]. The single-shot property and high reconstruction quality also result in less photodamage with eDL-cSIM. To verify this aspect, we compare the reconstruction performance of PCA-SIM and eDL-cSIM under the same illumination dose. As shown in Supplementary Fig. S21, when the total exposure time of capturing 9 images in conventional illumination mode is the same as that in composite illumination mode, the SNR of the conventionally illuminated image suffers from a significant degradation to the extent that substantial artifacts occur in its reconstructed image. However, eDL-cSIM maintains high reconstruction quality, whereas the conventional 9-frame reconstruction method requires at least nine times the illumination dose to achieve the same quality. All these experimental results demonstrate that eDL-SIM is capable of robustly obtaining high-quality, artifact-free super-resolution results from only a single-frame composite-coded input in complex, low SNR environments.

Dynamic super-resolution imaging of live cells based on eDL-cSIM

The lower phototoxicity, photobleaching and faster imaging speed resulting from the single-shot feature, as well as the high reconstruction quality and well noise robustness enabled by ensemble learning make eDL-cSIM highly suitable for super-resolution observation of live cells. To demonstrate the potential for live-cell imaging, we imaged the mitochondria of live COS- 7 cells labeled with MitoTracker^{TM} Green FM using eDL-cSIM and reconstructed some interesting mitochondrial dynamic events. As can be seen in Fig. 6a and b, a clustered annular mitochondrion (pointed to by the yellow arrow) tried to fuse with a long mitochondrion (pointed to by the blue arrow) below, which was deformed at the point of fusion due to being dragged (24.09 $s \sim 33.96$ s). Meanwhile, the annular mitochondrion also emitted a mitochondrial tubule to fuse with another long mitochondrion composed of multiple small ones at the lower left (pointed to by the green arrow). In almost the same time interval (24.09 $s \sim 30.69$ s), mitochondrial fission occurred on another annular mitochondrion (pointed to by the green and yellow arrows) in the region below, but one of the fissioned sites rapidly fused with the two mitochondria on the right (pointed to by the green arrow), as shown in Fig. 6. Figure 6d-f

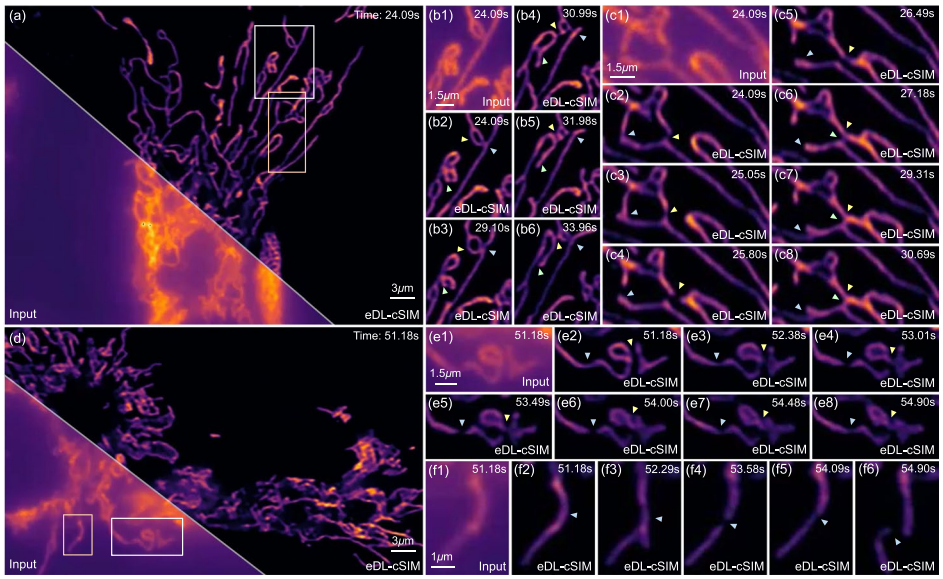


Fig. 6 Dynamic super-resolution images of mitochondria of live COS- 7 cells labeled by MitoTracker^{TM} Green FM at different time points. **a**, **d** Comparison of input composite illumination images and super-resolution images obtained by eDL-cSIM. The raw images with the resolution of 512 × 512 were captured through a 60× objective (UPlanXApo 60×/1.42 Oil, Olympus, Japan). **b**, **e** Magnified super-resolution images from the white-boxed regions in (**a**) and (**d**) at different time points. **c**, **f** Magnified super-resolution images from the yellow-boxed regions in (**a**) and (**d**) at different time points. Colored arrows point to regions where dynamics are distinct. Scale bars: 3 μm (**a**, **d**); 1.5 μm (**b**, **c**, **e**); 1 μm (**f**)

illustrate similar mitochondrial dynamic events, where the mitochondrial fission and fusion occurred in the regions indicated by the blue and yellow arrows, respectively. Supplementary Movie S1 provides the complete super-resolution results of the COS- 7 mitochondria, as well as the input composite-coded images for comparison. These mitochondria constantly underwent a dynamic balance of fusion and division, the dysregulation of which is associated with a range of human diseases, including neurodegenerative diseases, diabetes, tumors and so on [69]. Therefore, the study of these mitochondrial dynamic events mentioned above is of positive significance for the diagnosis, prognosis, and development of individualized therapeutic means for related diseases. In addition, as seen in Supplementary Movie S1, although the region of interest was constantly changed and the focus was frequently adjusted, eDL-cSIM always provides reliable and robust reconstruction, verifying that it can be applied to high-quality, long-duration dynamic super-resolution imaging of live cells in complex environments.

Discussion and conclusion

We have presented a single-shot composite structured illumination microscopy approach enabled by ensemble learning (eDL-cSIM). eDL-cSIM achieves single-shot laterally isotropic spectrum modulation via an efficient six-beam interferometric composite structured illumination strategy without modifying the SIM optical setup. Ensemble learning is utilized to solve the ill-posed inverse problem of super-resolution reconstruction by integrating the base networks focusing on features of different frequency components in different domains, rather than conventional deep learning approaches relying

only on individual neural networks. Experimental results demonstrate that eDL-cSIM enables higher-quality super-resolution image reconstruction in low SNR environments compared to state-of-the-art data- and physics-driven methods at 9 times the imaging speed of conventional illumination modes. Faster imaging speed, less photodamage, higher reconstruction quality and better noise immunity make eDL-cSIM an efficient and robust imaging tool to study the dynamics of nanoscale fine structures in living cells.

We also demonstrate that eDL-cSIM exhibits well generalization to various samples [Figs. 5, 6, S16-S18] and imaging environments [Figs. S19-S21] that are not covered by the training data. This is primarily attributed to the prior knowledge learned from the training data, including the abstract representation of the physical characteristic of extrapolating high-frequency signals (which is independent of specific samples and further enhanced by the ensemble learning strategy) [50–52] and the generalized high-dimensional features statistically extracted from the sample structural information [39, 70]. To practice eDL-cSIM on other SIM microscopes or samples of completely diverse types, re-preparing the dataset according to our proposed framework is recommended, since, for supervised learning, the dataset is a crucial factor that affects the actual performance of the model. Notably, the proposed composite illumination strategy requires no modifications to the original SIM setup, allowing dataset preparation to be easily generalized to other SIM microscopes. Further employing transfer learning [71] or data augmentation [72], which is implemented by fine-tuning an already trained model or increasing the diversity of the training set, can somewhat reduce the reliance on new datasets. This will be a primary focus for future work.

Regarding the 6-beam composite illumination modulation, in addition to the spectral components contributing to the highest resolution enhancement, other components with smaller wavelet vectors impose severe spectral aliasing, making the inverse reconstruction problem even more ill-posed. In fact, these spectral components can assist in compensating for the missing-cone effect in the three-dimensional (3D) optical transfer function (OTF) of the system, thus improving the axial resolution and optical sectioning capability. In the future, we will consider utilizing the super-resolution images demodulated from the composite illumination images as the ground-truth data of the training set for single-shot super-resolution and optical sectioning.

Despite the impressive results achieved by eDL-cSIM, caution should still be exercised for the essentially ill-posed inverse problem of SIM reconstruction, where network prediction is theoretically impossible to obtain the full ground-truth details. To further ensure the fidelity and unambiguity of the prediction results, future efforts can focus on enhancing network architecture, e.g., embedding physical models in the neural network to improve the network performance [47]. In addition, although we have improved the time consumption for network training and prediction by network lightweighting, especially the prediction time is reduced by about 1.95 times, which is even 63.08% lower than that using a single network, the network runtime still needs to be improved for long-time continuous observation of live cells. In the future, we will further lighten the neural networks in terms of both volume and speed while maintaining the accuracy, and optimize the structure and number of base networks to avoid computational redundancy due to performance conflicts, based on which the local deployment of pre-trained

models will be considered in order to achieve real-time, robust deep-learning-based single-shot SIM super-resolution imaging.

Methods

SIM setup

Our customized SIM setup is shown in Supplementary Fig. S1. We couple three laser beams (Laser 1: OBIS LX405, Coherent, USA; Laser 2: OBIS LX561, Coherent, USA; and Laser 3: Sapphire 488LP- 200, Coherent, USA) with different wavelengths through a series of dichroic mirrors (DM1: ZT561 dcrb, Chroma, USA; and DM2: ZT488 dcrb, Chroma, USA) and plane mirrors (M1, M2: OMM1-A1, JCOPTiX, China), and then expand and collimate them through a spatial filter and an achromatic lens (L1: LSB08-A 150 mm, Thorlabs, USA). The laser beam is then modulated to p-polarized light by a half-wave plate (HW1: GCL- 0604, Daheng Optics, China), passes through a polarizing beam splitter (PBS: PBS251, Thorlabs, USA) used to reduce laser loss, and is projected onto a ferroelectric liquid crystal spatial light modulator (SLM: QXGA-3DM, Fourth Dimension Displays, UK) displaying the standard or composite grating patterns. The generated diffracted light is modulated by another half-wave plate (HW2: GCL- 0604, Daheng Optics, China) into the s-polarized state and reflected by the PBS and an achromatic lens (L2: LSB08-A 250 mm, Thorlabs, USA). A well-designed mask blocks the zero-order beam at the focusing plane of the diffracted light, allowing only the ± 1-order beams from three directions to pass through. Typically, a six-zoned half-wave plate (AHWP25-VIS-A- 6P-M, LBTEK, China) is employed to control the polarization directions of the diffracted beams parallel to the directions of the resulting interference fringes to maximize the depth of illumination modulation [73]. These beams are then focused onto the rear focal plane of the objective lens by a pair of lenses (L3: LSB08-A 200 mm, Thorlabs, USA; L4: LSB08-A 175 mm, Thorlabs, USA), generating the interference patterns on the sample surface. The emitted fluorescence light is collected by the same objective lens, passes through the fluorescent interference filter block, and is finally captured by an sCMOS camera with a quantum efficiency of 60% (PCO Edge 5.5, PCO, Germany). In contrast to regular two-beam interferometry, we use an illumination strategy of six-beam interferometry, which does not require any modification of the optical setup, and can realize the modulation of the isotropic spectrum from only one illumination by simply replacing the image displayed in SLM with a composite pattern superimposed on the gratings in three orientations. Supplementary Fig. S2 illustrates the spectrum modulation in the case of six-beam interference. It can be seen that the interference of the six amplitude vectors in three directions will generate 19 spatial frequency components, among which the black components are generated by the interference of diffracted light with a direction difference of 180 degrees, the blue components are generated by the interference of diffracted light with a direction difference of 120 degrees, and the green components are generated by the interference of diffracted light with a direction difference of 60 degrees.

Multi-head attention mechanism calculation

The multi-head attention mechanism within a sub-window can be defined as:

$$Attention(Q, K, V) = softmax\frac{QK^T}{\sqrt{d_k}}V \tag{1}$$

$$head_i = Attention\left(xW_i^Q, xW_i^K, xW_i^V\right) \tag{2}$$

$$MultiHead(head_{i(i=1...h)}) = Concat(head_1, ...head_h)W^O \tag{3}$$

where *Attention*, *head* and *MultiHead* represent the operation of calculating the self-attention mechanism, single-head attention mechanism and multi-head attention mechanism respectively, the matrices Q, K, and V are derived by linearly transforming the feature matrix x through the weight matrices W_Q, W_K, and W_V (Q=xW^Q, K=xW^K, V=xW^V), d_k denotes the dimension of Q, K, and V, *softmax* is the normalized exponential function, W^O denotes the weight matrix for linear transformation of the multi-head attention mechanism, and *Concat* represents the operation of concatenating $head_{i(i=1...h)}$ in d_k dimensions.

Activation function and loss function

In eDL-cSIM, both the base networks and the ensemble network use the Gaussian Error Linear Unit (GELU) [74] as the global activation function to introduce nonlinear characteristics, which can be described as follows:

$$GELU(x) = 0.5x\left\{1 + Tanh\left[\sqrt{\frac{2}{\pi}}\left(x + 0.447515x^3\right)\right]\right\} \tag{4}$$

where *Tanh* represents the hyperbolic tangent function.

Regarding the loss function, since the tasks vary between networks, we set different loss functions for the sub-network stage and the ensemble stage to optimize the network parameters. The loss function of the base networks is defined as a combination of mean square error (MSE) loss and SSIM loss:

$$Loss_{base}(Y, \hat{Y}) = MSE(Y, \hat{Y}) + \alpha[1 - SSIM(Y, \hat{Y})] \tag{5}$$

where Y and $\hat{Y}$ represent the network output and the ground truth respectively, and α is used to adjust the weight between MSE loss and SSIM loss. For Base Network 1, which is tasked with extracting low-frequency components, α is set to 0.05 to enhance the ability of averaging local details and ensure pixel-level accuracy while balancing dynamic range. For Base Networks 2 and 3, α is set to 0.5 to enhance the structural similarity. The loss function of the ensemble network is defined as the combination of SSIM loss and perception loss (PLoss) [75], where PLoss compares $\hat{Y}$ with Y by extracting different depth feature maps using a pre-trained CNN, and ensures that both content and structural information are considered. In this case, we use a pre-trained VGG16 network and select features from the 4 th, 9 th, and 16 th layers [76]. Then PLoss is described as:

$$Ploss = \sum_i \frac{1}{C_i H_i W_i} \left\| vgg_i(Y) - vgg_i(\hat{Y}) \right\|_2^2 \tag{6}$$

where i represents the layer number, and $C_i H_i W_i$ denotes the size of the feature map at the i-th layer. The loss function of the ensemble network with PLoss considered is denoted as:

$$Loss_{ensemble}(Y, \hat{Y}) = [1 - SSIM(Y, \hat{Y})] + \beta Ploss(Y, \hat{Y}) \tag{7}$$

where the weight β is empirically set to 0.2.

Dataset collection and network training

Through the customized SIM setup, we collected 200 sets of data on mitochondria and actin in BPAE cells. Each dataset contains nine raw images captured in regular illumination mode and one image in composite illumination mode. Note that all mentioned neural networks were trained based on this dataset and all used test data are not included in the training set. To ensure the quality of the subsequent reconstruction, sufficient laser power and exposure time were provided to acquire a high imaging SNR, while the regions of interest where thin samples with relatively clear backgrounds were located were captured to avoid the emergence of out-of-focus artifacts. We utilized PCA-SIM [31], which enables high-precision illumination parameter estimation to ensure high-quality reconstruction, to obtain the super-resolution image from nine regular illumination patterns (more technical details about PCA-SIM are provided in Supplementary Note S2). For Base Network 1, the composite illumination image and the wide-field image derived from multi-frame averaging of nine images of conventional illumination are used as an input-output (ground truth) pair. For Base Networks 2 and 3, the outputs are changed to the reconstructed super-resolution images. For the ensemble network, the inputs and outputs are the predictions of the three base networks and the reconstructed super-resolution images, respectively. The resolutions of the raw image, the wide-field image and the super-resolution image are 512×512, 512×512 and 1024×1024, respectively. By random cropping, a total of 12, 800 sets of input images with 128×128 resolution and ground truth with 256×256 resolution (or input images and ground truth both with 256×256 resolution) were obtained, and random horizontal/vertical flipping and random scaling were used to further enhance the dataset. Supplementary Fig. S10 presents a set of raw acquisition data and the input-output pairs of different neural networks. When training the networks, the ratio of the training set to the validation set is 8 : 2. We used the Adam optimizer [77] (with learning rate α= 0.003, and attenuation rates $\beta 1$ =0.9, $\beta 2$ =0.999) to update the network parameters and a dynamic learning rate adjustment strategy to make more refined adjustments to the model weights. Specifically, the initial learning rate was set to 0.05. From 0 to 60 epochs, the learning rate decayed by 0.5 times every 15 epochs. From 61 to 150 epochs, the learning rate decayed by 0.5 times every 30 epochs. From 150 to 300 epochs, if the validation loss did not decrease for 20 consecutive epochs, the learning rate decayed by 0.5 times. The constructed neural networks are computed based on the PyTorch platform (version 2.0.1,

using Python 3.10.0) on a workstation with Intel Core i7 - 13700 KF CPU, 32GB RAM and NVIDIA GeForce RTX4090.

Sample preparation

The live COS- 7 cells were incubated in H-DMEM containing 10% fetal bovine serum and 1% penicillin-streptomycin under a humidified environment of 5% CO2 at 37 °C. COS- 7 cells were seeded at a density of 5×10^8 cells in the confocal dishes with three parallel samples, which were cultured overnight. MitoTracker^{TM} Green FM (Beyotime, China) were diluted to 200 nM in H-DMEM, which should be pre-warmed to 37 °C. They were washed with 0.5 mL Phosphate Buffered Saline (PBS) for 3 min. The prepared solution was incubated for 10 min at a dose of 0.5 mL/well under dark conditions. To remove interference from free dyes, all samples were washed with PBS 3 times.

Statistical analysis

Except for Fig. 6, all the experiment results showed the representative data from 10 representative experiments. The SSIM, PSNR and LPIPS shown in Figs. 4 and 5, and Supplementary Figs. S11-S13, S15 and S21 were presented as box plots (center line, average; limits, 75% and 25%; whiskers, maximum and minimum) in graphs. The intensity profiles in Figs. 3, 4 and 5, and Supplementary Figs. S11-S13 and S15-S17 are interpolated by linear interpolation in MATLAB.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43074-025-00171-w.

Additional file 1.

Additional file 2.

Acknowledgements

This work was supported by National Natural Science Foundation of China (62405136, 62275125, 62275121, 12204239, 62175109), China Postdoctoral Science Foundation (BX20240486, 2024M754141), Youth Foundation of Jiangsu Province (BK20241466, BK20220946), Jiangsu Funding Program for Excellent Postdoctoral Talent (2024ZB671), Fundamental Research Funds for the Central Universities (30922010313), Fundamental Research Funds for the Central Universities (2023102001), and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (JSGP202201, JSGPCXZNGZ202402).

Authors' contributions

Z.C. and J.Q. conceived the project. Z.C. and Q.C. supervised the research. J.Q. and C.W. programmed the deep learning algorithm and analyzed the data. J.Q. and C.W. built the SIM system. J.Q., H.W. and C.W. performed the experiments. C.Z. supervised the experiments. All authors discussed the results and contributed to writing the paper.

Data availability

The customized Python codes of eDL-SIM, as well as the corresponding datasets are publicly available at https://doi.org/10.6084/m9.figshare.26894011.v2.

Declarations

Competing interests

The authors declare no competing interests.

Received: 3 February 2025 Revised: 18 March 2025 Accepted: 18 April 2025
Published online: 07 May 2025

References

1. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods. 2006;3(10):793–6.
2. Hess ST, Girirajan TP, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys J. 2006;91(11):4258–72.
3. Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc. 2000;198(2):82–7.
4. Klar TA, Jakobs S, Dyba M, Egner A, Hell SW. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. Proc Natl Acad Sci. 2000;97(15):8206–10.
5. Zhao K, Xu X, Ren W, Jin D, Xi P. Two-photon MINFLUX with doubled localization precision. eLight. 2022;2(1):5.
6. Li D, Shao L, Chen BC, Zhang X, Zhang M, Moses B, et al. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. Science. 2015;349(6251):aab3500.
7. Shao L, Kner P, Rego EH, Gustafsson MG. Super-resolution 3D microscopy of live whole cells using structured illumination. Nat Methods. 2011;8(12):1044–6.
8. Zhanghao K, Chen X, Liu W, Li M, Liu Y, Wang Y, et al. Super-resolution imaging of fluorescent dipoles via polarized structured illumination microscopy. Nat Commun. 2019;10(1):4694.
9. Johnson KA, Hagen GM. Artifact-free whole-slide imaging with structured illumination microscopy and Bayesian image reconstruction. Gigascience. 2020;9(4):giaa035.
10. Wang Z, Zhao T, Hao H, Cai Y, Feng K, Yun X, et al. High-speed image reconstruction for optically sectioned, super-resolution structured illumination microscopy. Adv Photon. 2022;4(2):026003.
11. Huang G, Liu Y, Wang D, Zhu Y, Wen S, Ruan J, et al. Upconversion nanoparticles for super-resolution quantification of single small extracellular vesicles. eLight. 2022;2(1):20.
12. Chen X, Zhong S, Hou Y, Cao R, Wang W, Li D, et al. Superresolution structured illumination microscopy reconstruction algorithms: a review. Light Sci Appl. 2023;12(1):172.
13. Lai-Tim Y, Mugnier LM, Orieux F, Baena-Gallé R, Paques M, Meimon S. Jointly super-resolved and optically sectioned Bayesian reconstruction method for structured illumination microscopy. Opt Express. 2019;27(23):33251–67.
14. Müller M, Mönkemöller V, Hennig S, Hübner W, Huser T. Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ. Nat Commun. 2016;7(1):10980.
15. Saxena M, Eluru G, Gorthi SS. Structured illumination microscopy. Adv Opt Photon. 2015;7(2):241–75.
16. Wang Z, Zhao T, Cai Y, Zhang J, Hao H, Liang Y, et al. Rapid, artifact-reduced, image reconstruction for super-resolution structured illumination microscopy. Innovation. 2023;4(3):100425.
17. Descloux A, Müller M, Navikas V, Markwirth A, Van den Eynde R, Lukes T, et al. High-speed multiplane structured illumination microscopy of living cells using an image-splitting prism. Nanophotonics. 2020;9(1):143–8.
18. Lyu J, Qian J, Xu K, Huang Y, Zuo C. Motion-resistant structured illumination microscopy based on principal component analysis. Opt Lett. 2023;48(1):175–8.
19. Turcotte R, Liang Y, Tanimoto M, Zhang Q, Li Z, Koyama M, et al. Dynamic super-resolution structured illumination imaging in the living brain. Proc Natl Acad Sci. 2019;116(19):9586–91.
20. Gustafsson MG, Shao L, Carlton PM, Wang CR, Golubovskaya IN, Cande WZ, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys J. 2008;94(12):4957–70.
21. Wicker K, Mandula O, Best G, Fiolka R, Heintzmann R. Phase optimisation for structured illumination microscopy. Opt Express. 2013;21(2):2032–49.
22. Wicker K. Non-iterative determination of pattern phase in structured illumination microscopy using auto-correlations in Fourier space. Opt Express. 2013;21(21):24692–701.
23. Dan D, Lei M, Yao B, Wang W, Winterhalder M, Zumbusch A, et al. DMD-based LED-illumination super-resolution and optical sectioning microscopy. Sci Rep. 2013;3(1):1116.
24. Ball G, Demmerle J, Kaufmann R, Davis I, Dobbie IM, Schermelleh L. SIMcheck: a toolbox for successful super-resolution structured illumination microscopy. Sci Rep. 2015;5(1):15915.
25. Qian J, Xu K, Feng S, Liu Y, Ma H, Chen Q, et al. Fast-Structured Illumination Microscopy Based on Dichotomy-Correlation Parameter Estimation (dCOR-SIM). ACS Photonics. 2024;11(5):1887–92.
26. Huang X, Fan J, Li L, Liu H, Wu R, Wu Y, et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. Nat Biotechnol. 2018;36(5):451–9.
27. Zhao W, Zhao S, Li L, Huang X, Xing S, Zhang Y, et al. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. Nat Biotechnol. 2022;40(4):606–17.
28. Chu K, McMillan PJ, Smith ZJ, Yin J, Atkins J, Goodwin P, et al. Image reconstruction for structured-illumination microscopy with low signal level. Opt Express. 2014;22(7):8687–702.
29. Wen G, Li S, Wang L, Chen X, Sun Z, Liang Y, et al. High-fidelity structured illumination microscopy by point-spread-function engineering. Light Sci Appl. 2021;10(1):70.
30. Soubies E, Nogueron A, Pelletier F, Mangeat T, Leterrier C, Unser M, et al. Surpassing light inhomogeneities in structured-illumination microscopy with FlexSIM. J Microsc. 2024;296(1):94–106.
31. Qian J, Cao Y, Bi Y, Wu H, Liu Y, Chen Q, et al. Structured illumination microscopy based on principal component analysis. eLight. 2023;3(1):4.
32. Jin L, Liu B, Zhao F, Hahn S, Dong B, Song R, et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. Nat Commun. 2020;11(1):1934.
33. Qian J, Cao Y, Xu K, Bi Y, Xia W, Chen Q, et al. Robust frame-reduced structured illumination microscopy with accelerated correlation-enabled parameter estimation. Appl Phys Lett. 2022;121(15):153701.
34. Ströhl F, Kaminski CF. Speed limits of structured illumination microscopy. Opt Lett. 2017;42(13):2511–4.
35. Orieux F, Sepulveda E, Loriette V, Dubertret B, Olivo-Marin JC. Bayesian estimation for optimized structured illumination microscopy. IEEE Trans Image Process. 2011;21(2):601–14.
36. Lal A, Shan C, Zhao K, Liu W, Huang X, Zong W, et al. A frequency domain SIM reconstruction algorithm using reduced number of images. IEEE Trans Image Process. 2018;27(9):4555–70.

37. Dong S, Liao J, Guo K, Bian L, Suo J, Zheng G. Resolution doubling with a reduced number of image acquisitions. Biomed Opt Express. 2015;6(8):2946–52.
38. Ling C, Zhang C, Wang M, Meng F, Du L, Yuan X. Fast structured illumination microscopy via deep learning. Photon Res. 2020;8(8):1350–9.
39. Wang H, Rivenson Y, Jin Y, Wei Z, Gao R, Günaydın H, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nat Methods. 2019;16(1):103–10.
40. Shah ZH, Müller M, Wang TC, Scheidig PM, Schneider A, Schüttpelz M, et al. Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. Photon Res. 2021;9(5):B168–81.
41. Christensen CN, Ward EN, Lu M, Lio P, Kaminski CF. ML-SIM: universal reconstruction of structured illumination microscopy images using transfer learning. Biomed Opt Express. 2021;12(5):2720–33.
42. Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods. 2021;18(2):194–202.
43. Qiao C, Li D, Liu Y, Zhang S, Liu K, Liu C, et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. Nat Biotechnol. 2023;41(3):367–77.
44. Qiao C, Zeng Y, Meng Q, Chen X, Chen H, Jiang T, et al. Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy. Nat Commun. 2024;15(1):4180.
45. Chen X, Qiao C, Jiang T, Liu J, Meng Q, Zeng Y, et al. Self-supervised denoising for multimodal structured illumination microscopy enables long-term super-resolution live-cell imaging. PhotoniX. 2024;5(1):4.
46. Wu H, Li Y, Sun Y, Yin L, Sun W, Ye Z, et al. Single-frame structured illumination microscopy for fast live-cell imaging. APL Photon. 2024;9(3):036102.
47. Saurabh A, Brown PT, Bryan JS 4th, et al. Approaching maximum resolution in structured illumination microscopy via accurate noise modeling. npj Imaging. 2025;3(5).
48. Zhang, Qinnan, et al. Deep learning-based single-shot structured illumination microscopy. Opt Lasers Eng. 2022;155:107066.
49. Cheng X, Li J, Dai Q, Fu Z, Yang J. Fast and lightweight network for single frame structured illumination microscopy super-resolution. IEEE Trans Instrum Meas. 2022;71:1–11.
50. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. Eng Appl Artif Intell. 2022;115:105151.
51. Rahman MSS, Li J, Mengu D, Rivenson Y, Ozcan A. Ensemble learning of diffractive optical networks. Light Sci Appl. 2021;10(1):14.
52. Feng S, Xiao Y, Yin W, Hu Y, Li Y, Zuo C, et al. Fringe-pattern analysis with ensemble deep learning. Adv Photon Nexus. 2023;2(3):036010.
53. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Switzerland: Springer; 2015. pp. 234–41.
54. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: Institute of Electrical and Electronics Engineers (IEEE); 2016. pp. 770–8.
55. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. 2021. arXiv preprint arXiv:2102.04306. https://arxiv.org/abs/2102.04306.
56. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Switzerland: Springer; 2016. pp. 424–32.
57. Li Z, Sun J, Fan Y, Jin Y, Shen Q, Trusiak M, et al. Deep learning assisted variational Hilbert quantitative phase imaging. Opto-Electron Sci. 2023;2:220023.
58. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. 2018. arXiv preprint arXiv:1804.03999. https://arxiv.org/abs/1804.03999.
59. Wang Z, Zhu L, Zhang H, Li G, Yi C, Li Y, et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. Nat Methods. 2021;18(5):551–6.
60. Li X, Hu X, Chen X, Fan J, Zhao Z, Wu J, et al. Spatial redundancy transformer for self-supervised fluorescence image denoising. Nat Comput Sci. 2023;3(12):1067–80.
61. Chen R, Tang X, Zhao Y, Shen Z, Zhang M, Shen Y, et al. Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. Nat Commun. 2023;14(1):2854.
62. Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 3147–55.
63. Fan J, Tang X, Bi X, Miao Y, Li W, Xiao B, et al. High-Fidelity Reconstruction of Structured Illumination Microscopy by an Amplitude-Phase Channel Attention Network With Multitemporal Information. IEEE Trans Instrum Meas. 2023;72:1–13.
64. Choi H, Lee J, Yang J. N-gram in swin transformers for efficient lightweight image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: Institute of Electrical and Electronics Engineers (IEEE); 2023. pp. 2071–81.
65. Chen J, Kao Sh, He H, Zhuo W, Wen S, Lee CH, et al. Run, don't walk: chasing higher FLOPS for faster neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: Institute of Electrical and Electronics Engineers (IEEE); 2023. pp. 12021–31.
66. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision. Switzerland: Springer; 2022. pp. 205–18.
67. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: Institute of Electrical and Electronics Engineers (IEEE); 2016. pp. 1874–83.

68. Zuo C, Qian J, Feng S, Yin W, Li Y, Fan P, et al. Deep learning in optical metrology: a review. Light Sci Appl. 2022;11(1):1–54.
69. Li YJ, Cao YL, Feng JX, Qi Y, Meng S, Yang JF, et al. Structural insights of human mitofusin-2 into mitochondrial fusion and CMT2A onset. Nat Commun. 2019;10(1):4914.
70. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
71. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data. 2016;3:1–40.
72. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):1–48.
73. Sun Y, Zhu H, Yin L, Wu H, Cai M, Sun W, et al. Fluorescence interference structured illumination microscopy for 3D morphology imaging with high axial resolution. Adv Photon. 2023;5(5):056007.
74. Hendrycks D, Gimpel K. Gaussian error linear units (gelus). 2016. arXiv preprint arXiv:1606.08415. https://arxiv.org/abs/1606.08415.
75. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: Institute of Electrical and Electronics Engineers (IEEE); 2018. pp. 586–95.
76. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. arXiv preprint arXiv:1409.1556. https://arxiv.org/abs/1409.1556.
77. Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980. https://arxiv.org/abs/1412.6980.

Publisher's Note