Supplementary Materials for

Video-level and high-fidelity super-resolution SIM reconstruction enabled by deep learning

Hanchu Ye,[a] Zitong Ye,[a] Yunbo Chen,[a] Jinfeng Zhang,[a] Xu Liu,[a] Cuifang Kuang,[a,b,c,*] Youhua Chen,[a,c,*] Wenjie Liu[a,b,*]
[a]State Key Laboratory of Extreme Photonics and Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou, 310027, China
[b]Zhejiang Lab, Hangzhou, 311121, China
[c]Ningbo Innovation Center, Zhejiang University, Ningbo 315100, China

*Address all correspondence to Wenjie Liu, wenjieliu@zju.edu.cn; Youhua Chen, chenyh21012@zju.edu.cn; Cuifang Kuang, cfkuang@zju.edu.cn

S1. Channel pruning for VDL-SIM network

Network pruning is a technique to reduce model size and computational burden by removing unnecessary parameters from a neural network[1]. These unnecessary parameters have limited contribution to the performance of the model. Channel pruning[2] is a simple way of pruning the network, which focuses on the channels in the convolutional layer of the neural network. The goal of pruning is to reduce the size of the model, thereby reducing the computational cost, increasing the speed of inference and making it more suitable for deployment in resource-limited environments[3-4]. However, channel pruning needs to be carefully balanced to ensure that the speedup is not accompanied by a deterioration of the performance of the task.

After the initial construction of the VDL-SIM network, we adopt the way of channel pruning to further improve the reconstruction speed of the network.

Under the FOV of 512 pixel×512 pixel, the reconstruction speeds of the models with 64, 32 and 16 channels are 5, 15 and 43 frame/s, respectively. The resolution of the corresponding reconstructed images are calculated by decorrelation analysis method[5], in which the model performance is better for 64 and 32 channel sizes, which are 149 nm and 159 nm [Fig. S1(b-c)], respectively. However, the performance of the 16-channel size model is impaired with a resolution of only 187 nm (Fig. S1(d)). For the two close microtubules in the zoomed-in boxes, we plot their intensity distribution profiles [Fig. S1(a)]. It shows that the distinguishing ability is similar for 64 and 32 channel sizes, whereas the reconstruction ability of 16 channel size is obviously deteriorated. Therefore, considering both reconstruction speed and quality, the network with a channel size of 32 is the best choice.

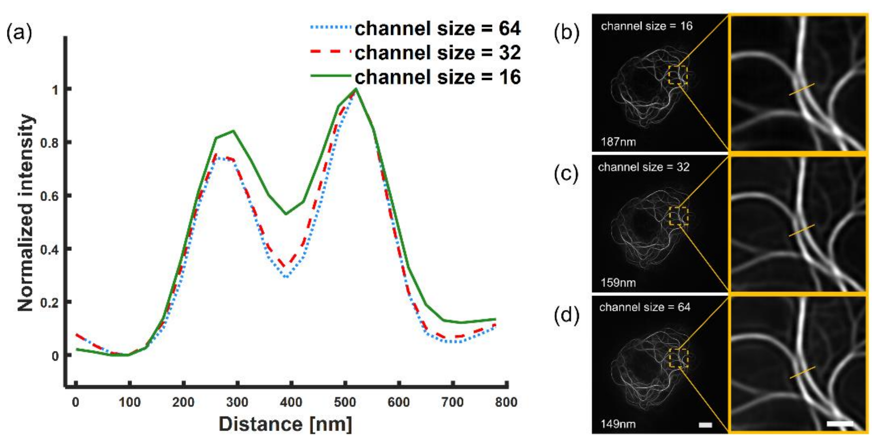


Fig. S1 Comparison of VDL-SIM reconstructed images for different channel sizes. (a) The intensity distribution profiles for two close microtubules in (b-c) enlarged boxes, at channel size of 64, 32, and 16. (b) The reconstructed image of VDL-SIM network with channel size 16 and its enlarged view. (c) The reconstructed image of VDL-SIM network with channel size 32 and its enlarged view. (d) The reconstructed image of VDL-SIM network with channel size 64 and its enlarged view. Scale bars: 3.28 μm (left image) and 0.75 μm (right boxed magnified images).

S2. Training datasets

The training dataset used in this work is based on the open-source BioSR. Specifically, we selected relatively complex structure of F-actin in BioSR as the training structure, which allows the network to learn and understand complex patterns better and thus to have higher generalizability to other structures. In our experiments, we focus on the training datasets of three biological structures: the ER [Fig. S2(b)], the microtubules [Fig. S2(a)], and the F-actin [Fig. S2(c)], which represent the increasing complexity of the structures. (Fig. S2) shows the reconstructed outputs of our network for microtubules after training based on different structures. Comparing Figs. S2(d-f), it is obvious that the reconstructed super-resolution image of the network trained out based on the ER will have similarity to the ER structure with serious distortion [Fig. S2(e)]. In contrast, the network trained based on the microtubules structure reconstructs with more detail [Fig. S2(d)], but the learning of complex structures is still not as accurate as the F-actin structure [Fig.S2(f)]. Fig. S2(h) shows line profiles of neighboring microtubules in the enlarged region of the images. The output of the F-actin trained VDL-SIM contains two neighboring microtubules with the distance between the peaks of the profiles greater than the gap between the ER and microtubules in the same cropped region.

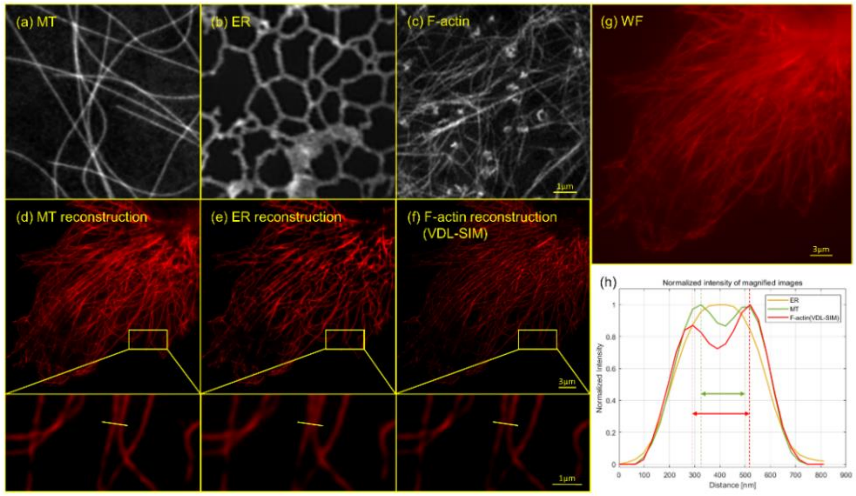


Fig. S2 Comparison of reconstructed images of the training dataset for three different biological structures. (a) (b) (c) The GT images of the training dataset for microtubules (MT), ER and F-actin, respectively. (d) (e) (f) The reconstructed microtubules images and their enlarged images after training based on the biological structures microtubules (MT), ER and F-actin, respectively. (g) Shows the wide-field image common to (d) (e) (f) microtubules. (h) The line profile of neighboring microtubules in the magnified images. Scale bars,1μm for the GT images of the training dataset. scale bars,3μm for the reconstructed images. and boxed magnified images, Scale bars:1 μm.

The F-actin structure was selected to perform data augmentation on the training dataset. We selected 50 different regions of interest, each with nine SNR levels, and randomly rotated the images to extend the datasets. Datasets of 40 regions are used for training, and datasets of the remaining 10 regions are used for validation. To train the network model, we use a supervised learning approach. The widefield images are treated as the network inputs. The paired reference images for the network are the traditional SIM reconstructed image after background removal by the rolling ball algorithm. The information is more concentrated after the background suppression, which can help the network to better understand the structural features while reducing the demand of computational resources.

S3. The effect of the rolling ball algorithm on VDL-SIM

The rolling ball algorithm is a commonly used image processing algorithm for background estimation and subtraction. It is based on the assumption of smoothness of the background in the image and approximates the background by fitting a rolling sphere. The basic idea of the algorithm is to scroll a sphere from the top to the bottom of the image, and the radius of the sphere is adjusted according to the changing gray value of the image pixels. When the sphere intersects with the background part of the image, the surface of the sphere does not overlap with the foreground part of the image, so the gray intensity inside the sphere can be considered as an approximation of the background. The equation of the rolling ball algorithm can be described as

$$B(x,y) = \min\left\{ I(x+a, y+b) - r^2 \right\} \tag{S1}$$

where $B(x,y)$ is the pixel value of the background estimated image, $I(x,y)$ is the pixel value of the raw image, r is the radius of the sphere, a and b are the offsets of the center of the sphere with respect to the pixel (x,y). The equation indicates that for a given pixel, the background value can be estimated by calculating the corresponding minimum value inside the sphere. Then, the background of the whole image is calculated by scrolling the sphere from the top to the bottom of the image.

To implement the algorithm, the radius and the center of the sphere need to be adjusted to accommodate different background levels. The reconstructed results of VDL-SIM with and without rolling ball processing are shown in Fig. S3(a-c), the background of the reconstructed image is suppressed and the image contrast is improved after processing. As shown in Fig. S3(d), the valley of the curve for r=5 almost coincides with r=10, which indicates that the radius of 5 is sufficient for background suppression. The results thus show that the rolling ball algorithm can reduce the influence of the background on the VDL-SIM reconstruction.

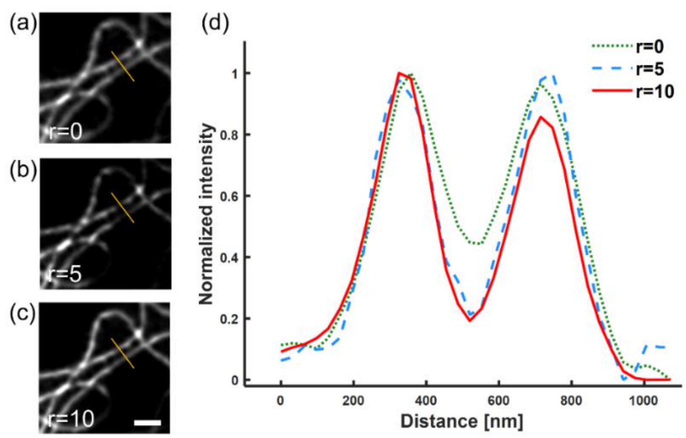


Fig. S3 The influence of the rolling ball radius on VDL-SIM reconstruction. (a) Super-resolution image reconstructed by VDL-SIM without rolling ball operation on the training images. (b) Super-resolution image reconstructed by VDL-SIM with rolling ball operation on the training images (radius size of 5). (c) Super-resolution image reconstructed by VDL-SIM with rolling ball operation on the training images (radius size of 10). (d) The intensity distribution profiles alone the yellow lines in (a-c). Scale bars: 1 μm.

S4. Extremely low SNR imaging

When imaging biological living specimens, there are many application situations that require lower light intensities and exposure times to minimize damage to the organisms. The reason for this requirement is to maintain the life activities and to avoid cell damage, cell death, or other irreversible changes in morphology and structure. In this regard, low SNR imaging conditions are necessary to help maintain the physiology of living specimens. In the following, we compare the imaging results of VDL-SIM with the conventional algorithms HiFi and IM SIM at extremely low SNR.

Imaging results with low SNR greater than 10 have been compared in detail in section 3.3. Here we compare the imaging conditions with extremely low SNR below 10. In this condition the traditional algorithms are no longer able to estimate the parameters and lose the ability to image. VDL-SIM, in contrast, still has the ability to discriminate biological structures and provides a useful tool for SIM imaging at extremely low SNR.

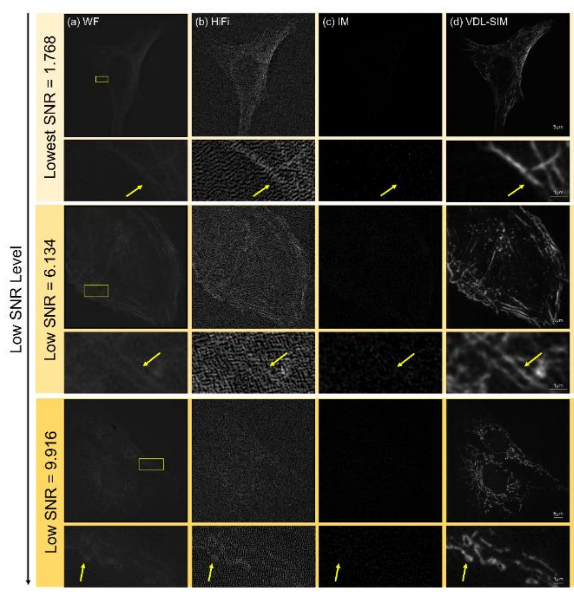
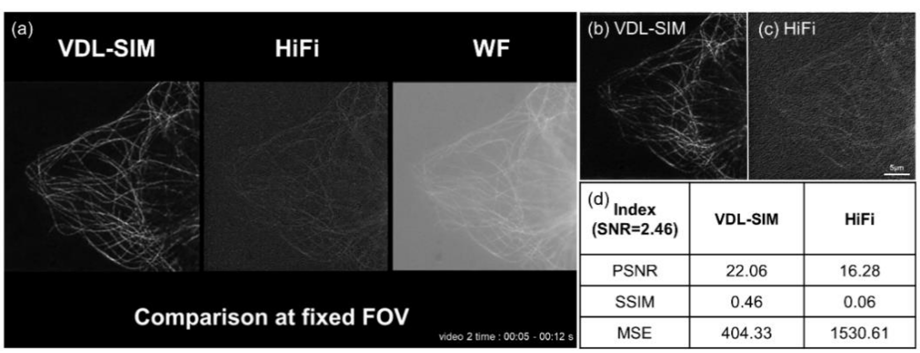


Fig. S4 Comparison of extremely low SNR reconstructed images. (a) Reconstructed wide-field (WF) images. (b) Conventional HiFi SIM algorithm reconstructed images. (c) Conventional IM SIM algorithm reconstructed images. (d) Reconstructed images by VDL-SIM algorithm. Larger images, scale bar: 5 μm. Enlarged images, scale bar: 1 μm. Three extremely low SNR levels with shallow to deep corresponding to increased SNR, from low to high are 1.768, 6.134 and 9.916.

Fig. S4 we show three SNR levels below 10. At SNR=1.768 and 6.734, the HiFi algorithm can only shadow a little bit of biological structures in the noise artifacts [Fig. S4(b)]. With the improvement of SNR, the noise artifacts are reduced at SNR=9.916, but still without reconstruction ability. The imaging results of the IM algorithm at extremely low SNR are even more unsatisfactory [Fig. S4(c)]. However, VDL-SIM can be used as a complementary technique for this application scenario, giving the observer a reference of the biological structure [Fig. S4(d)]. Benefit from deep learning for large-scale data training, feature learning and abstract representation. These make it possible to better adapt to imaging conditions with extremely low SNR and provide new solutions when parameters are difficult to estimate. Nevertheless, for specific tasks, it is still necessary to choose the right method and tuning approach depending on the characteristics of the problem and the accuracy requirements.

Fig. S5 VDL-SIM video frame performance evaluation



(d) Index (SNR=2.46)	VDL-SIM	HiFi
PSNR	22.06	16.28
SSIM	0.46	0.06
MSE	404.33	1530.61

Fig. S5 VDL-SIM video frame performance evaluation. (a) Screenshot of the video content of the fixed imaging region from 00:05-00:12s in Video 2. (b) Microtubules imaging results of VDL-SIM. (c) Microtubules imaging results of the conventional HiFi SIM algorithm. (d) Performance comparison table of VDL-SIM and HiFi reconstructed images. The imaging condition, SNR is 2.46. Exposure time is set to 15 ms.

References

1. S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both Weights and Connections for Efficient Neural Networks," https://doi.org/10.48550/arXiv:1506.02626v3 (2015).
2. P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning Convolutional Neural Networks for Resource Efficient Inference," https://doi.org/10.48550/arXiv:1611.06440v2 (2017).
3. Y. He, et al., "AMC: AutoML for Model Compression and Acceleration on Mobile Devices," https://doi.org/10.48550/arXiv:1802.03494v4 (2019).
4. B. Jacob, et al., "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference," https://doi.org/10.48550/arXiv:1712.05877v1(2017).
5. A. Descloux, K. S. Grußmayer, and A. Radenovic, "Addendum: Parameter-free image resolution estimation based on decorrelation analysis," *Nat. Methods* 17(10), 1061-1063 (2020).