OPEN ACCESS
Journal of Innovative Optical Health Sciences
Vol. 17, No. 2 (2024) 2350015 (12 pages)
© The Author(s)
DOI: 10.1142/S1793545823500153

World Scientific
www.worldscientific.com

Deep neural network based on multi-level wavelet and attention for structured illumination microscopy

Yanwei Zhang*,†, Song Lang*,†, Xuan Cao‡, Hanqing Zheng† and Yan Gong*,†,§
*School of Biomedical Engineering (Suzhou)
Division of Life Sciences and Medicine
University of Science and Technology of China
Hefei, Anhui 230026, P. R. China

†Suzhou Institute of Biomedical Engineering and Technology
Chinese Academy of Sciences
Suzhou, Jiangsu 215163, P. R. China

‡School of Physical Science and Technology
Suzhou University of Science and Technology
Suzhou, Jiangsu 215009, P. R. China
§gongy@sibet.ac.cn

Received 11 April 2023
Revised 9 May 2023
Accepted 17 May 2023
Published 28 June 2023

Structured illumination microscopy (SIM) is a popular and powerful super-resolution (SR) technique in biomedical research. However, the conventional reconstruction algorithm for SIM heavily relies on the accurate prior knowledge of illumination patterns and signal-to-noise ratio (SNR) of raw images. To obtain high-quality SR images, several raw images need to be captured under high fluorescence level, which further restricts SIM's temporal resolution and its applications. Deep learning (DL) is a data-driven technology that has been used to expand the limits of optical microscopy. In this study, we propose a deep neural network based on multi-level wavelet and attention mechanism (MWAM) for SIM. Our results show that the MWAM network can extract high-frequency information contained in SIM raw images and accurately integrate it into the output image, resulting in superior SR images compared to those generated using wide-field images as input data. We also demonstrate that the number of SIM raw images can be reduced to three, with one image in each illumination orientation, to achieve the optimal tradeoff between temporal and spatial resolution. Furthermore, our MWAM network exhibits superior reconstruction ability on low-SNR images compared to conventional SIM algorithms. We have also analyzed the adaptability of this network on other biological samples and successfully applied the pretrained model to other SIM systems.

§Corresponding author.

This is an Open Access article. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

Keywords: Super-resolution reconstruction; multi-level wavelet packet transform; residual channel attention; selective kernel attention.

1. Introduction

Structured illumination microscopy (SIM) is a powerful super-resolution (SR) technique that can enhance spatial resolution by up to two times compared to wide-field (WF) microscopic imaging systems.[1–3] SIM offers numerous advantages, including high spatial and temporal resolution, low phototoxicity, universal fluorescent dyes, and three-dimensional imaging. As a result, it has become a popular choice for biomedical research, especially for observing live cells.[4–7] In practice, SIM requires a series of raw images to be captured under different illumination patterns, such as sinuous illumination with varying angles and phases. These images are then reconstructed using an algorithm to produce a super-resolution (SR) image. However, like other SR microscopic imaging technologies, such as stimulated emission depletion microscopy (STED) and stochastic optical reconstruction microscopy (STORM), SIM also must involve a trade-off between spatial and temporal resolution.[8,9]

The conventional reconstruction algorithm for SIM involves a reverse engineering process based on SIM imaging theory. The accuracy of the prior knowledge, such as the period, angle, phase, and modulation of illumination pattern, and the point spread function (PSF) of the imaging system, are critical for stable reconstruction. In addition, studies have shown that the signal-to-noise ratio (SNR) of raw images also directly affects the reconstruction quality. The reconstructed SR image can be degraded by various forms of artifacts if the *a priori* knowledge is skewed and the fluorescent level is low.[10–13] In recent years, novel algorithms for SIM have been proposed to improve reconstruction quality. Hessian SIM is a deconvolution algorithm that uses the continuity of biological structures and the randomness of noise as prior knowledge to guide image reconstruction based on Hessian matrices. This algorithm requires raw images at a relatively higher acquisition frequency and is suitable for long-term observation of living cells.[14] HiFi SIM is a high-fidelity SIM reconstruction algorithm that engineers the effective PSF into an ideal form, effectively reducing commonly seen artifacts without loss of fine structures and improving axial sectioning for samples with strong background.[15] The sparse deconvolution algorithm improves the resolution of SR fluorescence microscopy by taking advantage of *a priori* knowledge about the sparsity and continuity of biological structures.[16] Although these algorithms have improved SR performance to a great degree, they inevitably depend on the parameters of the imaging system and SNR of the raw image.

Deep learning (DL) is powerful tool that can perform complex tasks by utilizing multilayered artificial neural networks trained on a large amount of tagged data.[17] Since 2017, DL has been gradually applied to optical microscopes to expand the limits of imaging systems. For example, DL methods can predict fluorescent or histological staining information from label-free images,[18,19] enhance lateral and axial resolution, temporal resolution, SNR, and optical section.[20–23] Since 2020, a series of DL methods have been proposed to improve the performance of SIM. These methods learn a mapping from a set of low-resolution raw images to a high-resolution SIM image through deep neural networks. Despite the significant differences between wide-field microscopic images and structured illuminated images, these methods still perform reasonably well in SIM. Under the prerequisite of guaranteed SR quality, the temporal resolution of SIM can be enhanced by reducing the number of raw images or exposure time.[24–30] Notably, the SIM raw images contain high-frequency information that cannot be obtained by microscopy, but this information is modulated to low frequency by structured illumination. However, most DL algorithms applied to SIM do not give enough consideration to the differences between WF and SIM raw images. Additionally, the adaptability of DL algorithms to different samples and SIM systems has not been fully discussed.

The quality of ground truth (GT) images directly determine the best effect that the DL algorithm can achieve, as is an end-to-end algorithm. In this paper, the original SIM-SR images in the BioSR dataset[24] are processed to improve the SNR and contrast, and the improved SIM-SR images are used

as the GT images for network training. Additionally, a network based on multi-level wavelet packet transform (WPT) and attention mechanism is designed, namely multi-level wavelet and attention mechanism (MWAM) network. The network decomposes the input raw images into different spectral channels through WPT,[31] processes these channels using residual channel attention block (RCAB),[32] and selective kernel attention block (SKAB),[33] and finally outputs the SR images through inverse wavelet transform (IWT), multi-channel fusion, and up-sampling. Through a series of training and validation experiments, the quality of SR reconstruction using SIM raw images as input is significantly superior to the SR quality using WF images as input. This proves that the MWAM network can effectively extract the high-frequency information contained in SIM raw images and make an important contribution to the output SR image.

2. Materials and Methods

2.1. *Image preprocessing*

The public BioSR SIM dataset[24] is utilized for the training and testing of the proposed MWAM network. It comprises four distinct biological structures, namely clathrin-coated pits (CCPs), endoplasmic reticulum (ER), microtubules (MTs), and F-actin. The original SR images in the dataset were reconstructed from raw images with ultrahigh SNR using conventional SIM algorithm. Although these SR images have no apparent artifacts, they are still subject to system noise, defocus, and other factors, resulting in structural discontinuity, low contrast, and background blurring.

To enhance the quality of the SIM-SR images, the BM3D algorithm is employed to denoise and improve the structure continuity and completeness.[34] Although the BM3D algorithm can preserve high-frequency details a great extent, it still results in minor loss of spatial resolution. Subsequently, the sparse deconvolution algorithm is applied to enhance the resolution of the denoised SIM-SR images.[16] By adjusting the algorithm's parameters, such as sparse iteration, deconvolution times, sparsity, and background level, the improved SIM-SR images with best quality are obtained. The preprocessed results of the SIM-SR images are presented in Fig. 1, with the images arranged in columns representing WF images, original SIM-SR images, denoised SIM-SR images, improved SIM-SR images and zoom-in images, respectively. As shown in Fig. 1, the improved SIM-SR images exhibit lower background noise, higher contrast, and no loss of resolution.

2.2. *Network architecture*

In order to obtain the low-frequency and high-frequency information from SIM raw images and achieve SR reconstruction, we designed a neural network based on multi-level WPT and attention mechanism. This network is called the MWAM network. Initially, the input raw images are decomposed into multiple subband images using multi-level WPT, which are then concatenated in the channel dimension. These subband images are then fed into multiple branches. Each branch mainly consists of the RCAB, IWT, and SKAB. The feature maps extracted from each branch are scaled and summed up. Finally, the MWAM network outputs the SR image after up-sampling and output convolution layers. The architecture of the proposed MWAM network is illustrated in Fig. 2, which comprises three-level WPT and corresponding three branches. The subsequent sections will provide a detailed explanation of the MWAM network's design.

Wavelet packet transform is a technique used to analyze and process signals in a multi-resolution framework. The 2D WPT employs four filters, namely f_{LL}, f_{LH}, f_{HL}, and f_{HH} to convolve with an image. The convolution results are then down-sampled to obtain four subband images, i.e., LL, LH, HL, and HH. The subband images can describe the image information at different scales and orientations.[35] Due to the biorthogonal property of WPT filters, the original image can be accurately reconstructed by the IWT, as shown in Fig. 3. In multi-level wavelet packet transform, each subband image is further decomposed into four subband images. WPT is invertible, ensuring that all information is retained by the down-sampling scheme. Furthermore, WPT can extract information in different frequencies, which is specifically suitable for extracting feature maps from SIM raw images.

The residual channel attention (RCA) unit combines the channel attention mechanism with residual structure. The residual structure enables low-frequency information to be directly transmitted through skip connections, and then allows the

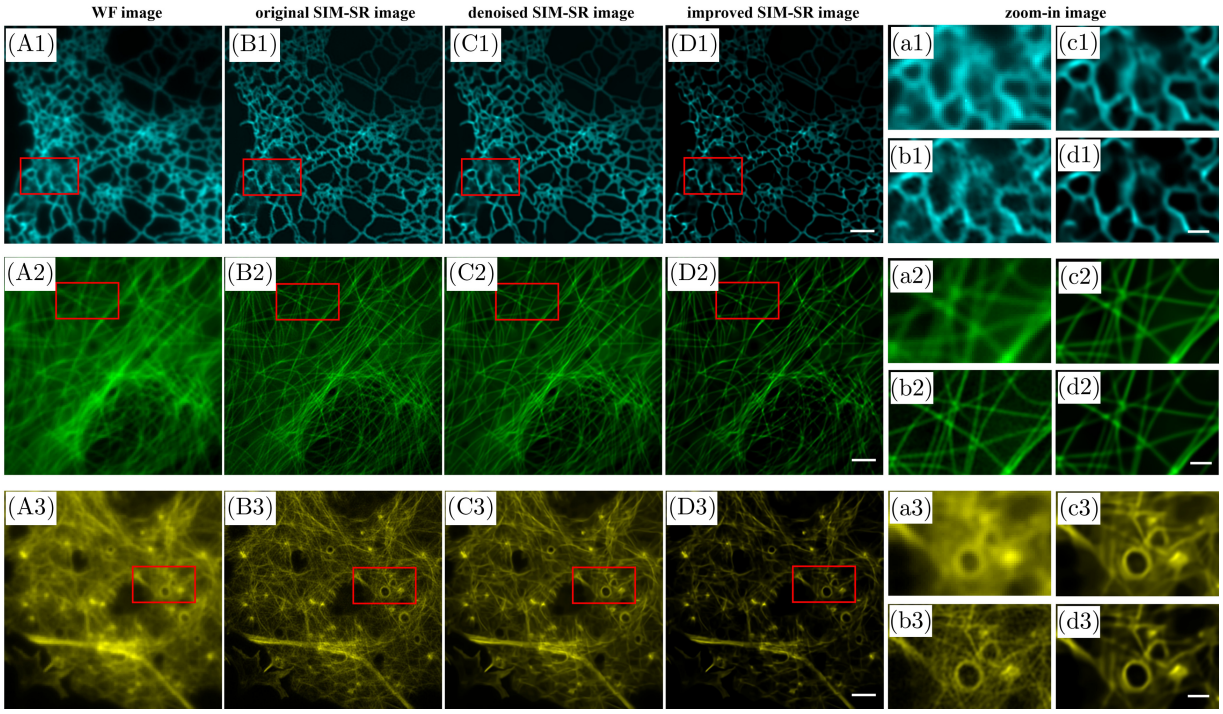

Fig. 1. The preprocessed results of SIM-SR images. The WF images for reference (first column), the original SIM-SR images obtained by conventional reconstruction algorithm (second column), the denoised SIM-SR images (third column), the improved SIM-SR images (fourth column). The zoom-in images of red box are shown in the fifth column. The images arranged in rows represent ER sample, MTs sample and F-actin sample respectively. Scale Bar: 3 μm (images in first to fourth column) and 1 μm (zoom-in images).

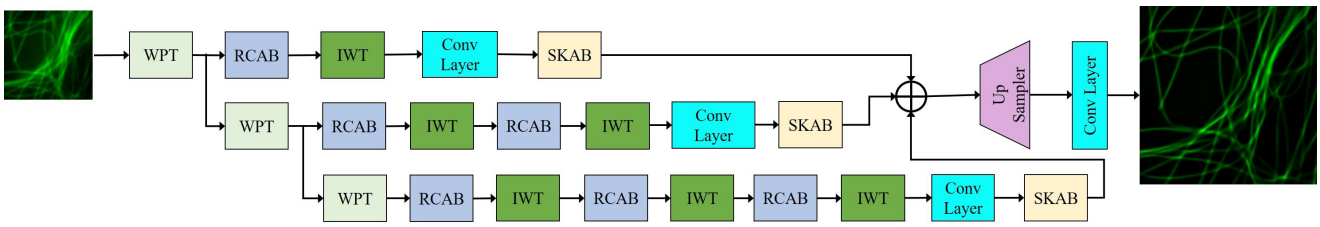

Fig. 2. Architecture of the proposed network MWAM network comprising three branches.

network to focus more on high-frequency information learning. Additionally, by considering the interdependence between channels and readjusting the feature maps, the network's discriminative learning ability can be enhanced. It is important to note that the input data to RCA unit are subband images obtained from WPT. The architecture of the RCA unit is illustrated in Fig. 4(a). The RCAB is a block comprising several RCA units.

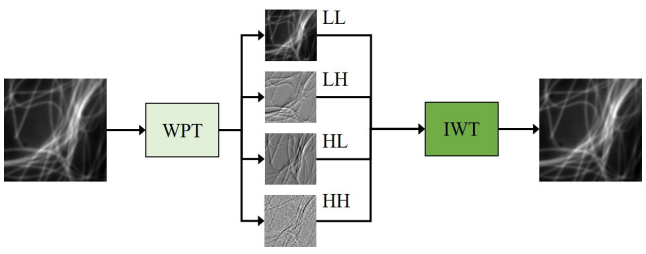

Fig. 3. Architectures of WPT and IWT.

The selective kernel attention (SKA) unit enhances the representation power of deep convolutional neural networks by fusing information from different receptive fields. The SKA unit sets up several branches with different convolution kernel sizes. Typically, a large kernel size can provide more semantic information, while a small kernel size can bring local texture information. Next, attention vectors of different kernels and channels are

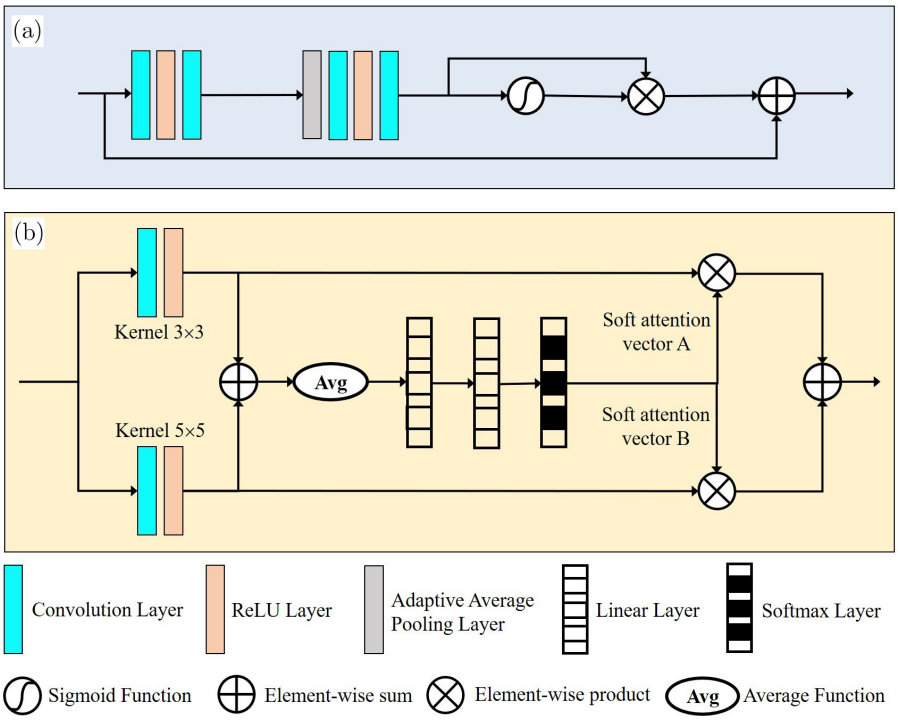

Fig. 4. (a) Architecture of RCA unit; (b) Architecture of selective kernel attention unit.

calculated through a series of layers and used to refine the feature maps from each branch. Finally, the adaptive adjusted feature maps are output. The architecture of the SKA unit with two different kernel size branches is illustrated in Fig. 4(b). The SKAB is a block comprising several SKA units.

2.3. *Dataset*

Cropping and flipping transformations were applied to the raw images and the improved SIM-SR images, resulting in the generation of 4400 pairs of raw images (128×128 pixels) and GT-SIM (256×256 pixels) images. Out of these, 4240 pairs were used for training, while 160 pairs were used for testing, as listed in Table 1.

Table 1. Dataset for training and testing.

Sample	Image number for training	Image number for testing
CCPs	586	40
ER	1612	40
MTs	859	40
F-actin	1183	40
All	4240	160

Six types of input data are designed, namely one raw image (denoted by raw-1-input), three raw images in one illumination angle (denoted by raw-3-input), three raw images in three illumination angles (denoted by raw-3angle-input), nine raw images (denoted by raw-9-input), wide-field image averaged by three raw images (denoted by WF-3-input), and wide-field image averaged by nine raw images (denoted by WF-9-input). These input data have differences at number of channels, temporal resolution, SNR, and level of high-frequency information, as listed in Table 2. The temporal resolution only considers the exposure time and ignores other time factors such as switching illumination pattern. The exposure time of a single raw image is set to ΔT and the SNR of a single raw image is set to α. Since WF-3-input and WF-9-input images are averaged by multiple raw images, their SNR is 3α and 9α, respectively. Compared with the WF images, the SIM raw images contain high-frequency information, which is modulated by structured illumination. Both raw-1-input and raw-3-input images theoretically contain the same high-frequency information in one orientation. Similarly, both raw-3angle-input and raw-9-input images contain the same high-frequency information in three orientations. However, the conventional

Table 2. Six input data types.

Input data type	Number of input channel	Temporal resolution	SNR of single input image	Level of high-frequency information
Raw-1-input	1	ΔT	α	+
Raw-3-input	3	$3\Delta T$	α	+
Raw-3angle-input	3	$3\Delta T$	α	+++
Raw-9-input	9	$9\Delta T$	α	+++
WF-3-input	1	$3\Delta T$	3α	/
WF-9-input	1	$9\Delta T$	9α	/

reconstruction algorithm has to separate and extract high-frequency information from at least three raw images with different illumination phases.

2.4. *Network hyperparameters and training*

In this paper, the configuration parameters of the MWAM network are set as follows: The number of RCA units in RCAB is 4, the number of SKA units in SKAB is 5, and the output channel number of the convolution layer before SKAB is 64. A MWAM network model is learned for each type of input data. The MWAM network was trained by optimizing the Charbonnier loss[36] and the optimizer is Adam algorithm.[37] The batch size is set to 16. The initial learning rate is set to 0.0002, which is then decayed by half every 50 epochs for a total of 300 epochs. All experiments are conducted in the Python 3.9 environment running on a PC with 11th Gen Intel(R) Core (TM) i7-11700 @ 2.50 GHz CPU and a Nvidia GeForce RTX3080 Ti GPU. The learning algorithm converges very fast and it takes about 27 h to train a MWAM network model.

3. Results

3.1. *Reconstructed results with different input data*

The train loss and test loss for six types of input data are shown in Fig. 5. The train loss has been consistently decreasing in the 300 epochs, while the test loss has remained almost stable after approximately 50 epochs. The average PSNR and SSIM are used to evaluate the quality of reconstructed images, as listed in Table 3. The SR images reconstructed by different trained models are shown in

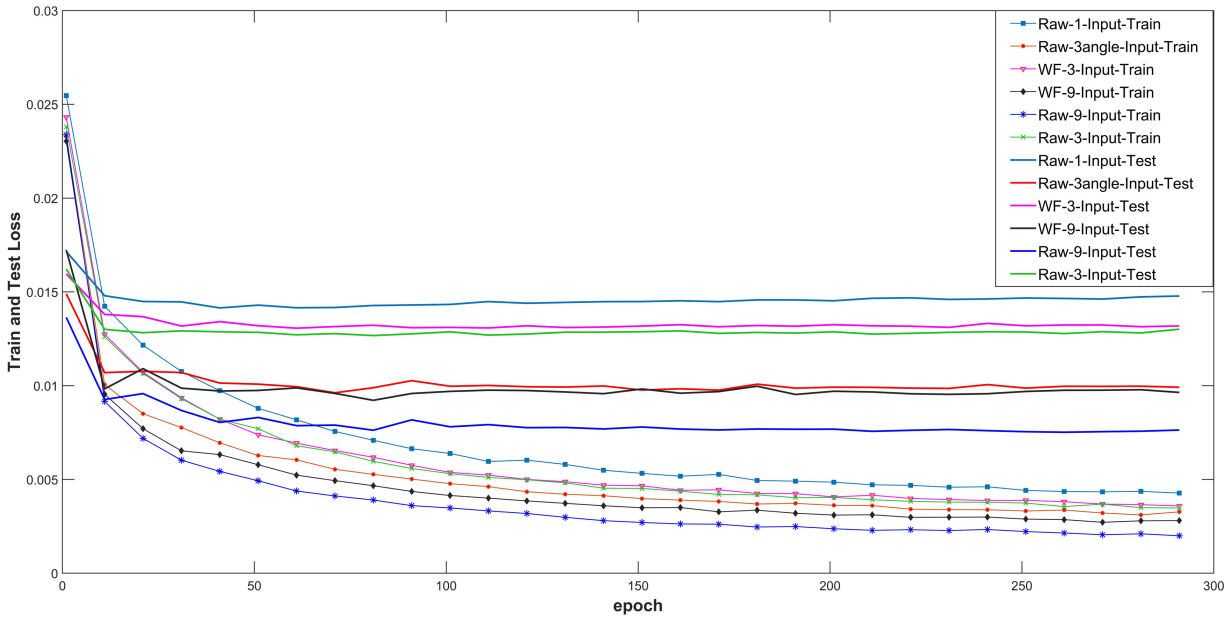

Fig. 5. Train loss and test loss for six types of input data.

Table 3. Average PSNR and SSIM results for six types of input data.

Input data type	PSNR (dB)	SSIM
Raw-1-input	31.58	0.9039
Raw-3-input	32.90	0.9251
Raw-3angle-input	35.17	0.9498
Raw-9-input	38.41	0.9684
WF-3-input	32.56	0.9208
WF-9-input	35.83	0.9543

Fig. 6. For samples with relatively simple structures, such as ER, the output SR images from six different models have almost no difference, and can achieve comparable quality and resolution with GT-SIM images, as shown in the first row of Fig. 6. However, the models with raw-1-input and WF-3-input perform poorly on samples with complex structures, such as CCPs, MTs, and F-actin. These two models cannot accurately reconstruct some fine structures, such as two very close MTs, as shown in the fourth row of Fig. 6. By comparing the models with raw-3angle-input and WF-3-input, which have the same temporal resolution, we can find that the former produces significantly superior reconstruction quality. The same conclusion applies to the models with raw-9-input and WF-9-input. The MWAM network proposed in this paper can extract high-frequency information from SIM raw images, and obtain SR images with better quality than those obtained using WF images as input data. When the input data consist of nine SIM raw images, the trained model achieves optimal reconstruction quality, and the average PSNR and SSIM values of output SR images are 38.41 dB and 0.9684 dB, respectively. In general, the model with raw-3angle-input data achieves the best balance between reconstruction quality and temporal resolution.

Deep Fourier channel attention network (DFCAN) is a novel deep neural network developed by Chang Qiao *et al.* in 2021.[24] The test data are MTs images with and without background fluorescence, the SR images reconstructed by MWAM and DFCAN models are shown in Fig. 7. The models utilized in the first to fifth columns of Fig. 7 are DFCAN-SIM model, DFCAN-SISR model, MWAM-Raw-9-input model, MWAM-WF-9-input

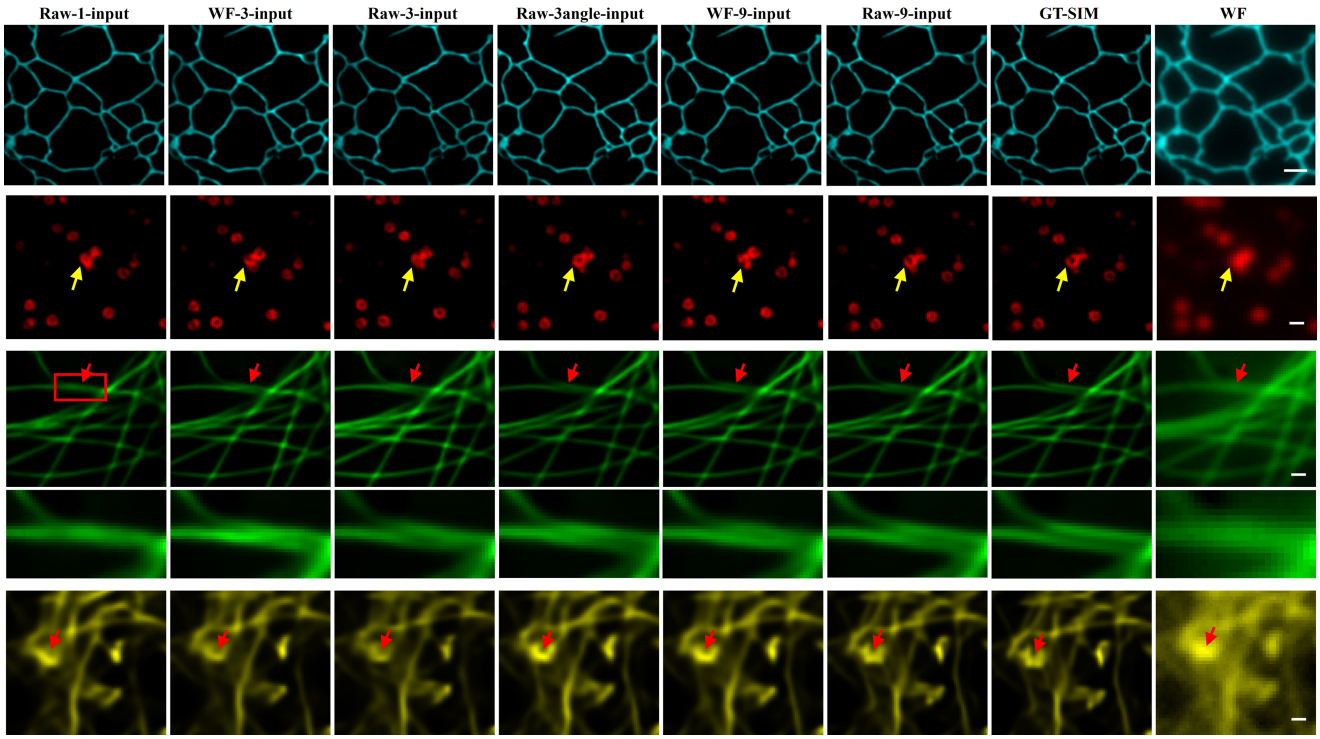

Fig. 6. Comparison of the SR images reconstructed by different models. The images in first to sixth columns are SR images reconstructed by the model with Raw-1-input data, WF-3-input data, Raw-3-input data, Raw-3angle-input data, WF-9-input data and Raw-9-input data, respectively. The GT-SIM images (seventh column) and wide-field images (eighth column) are shown for reference. Scale bars: 1.5 μm.

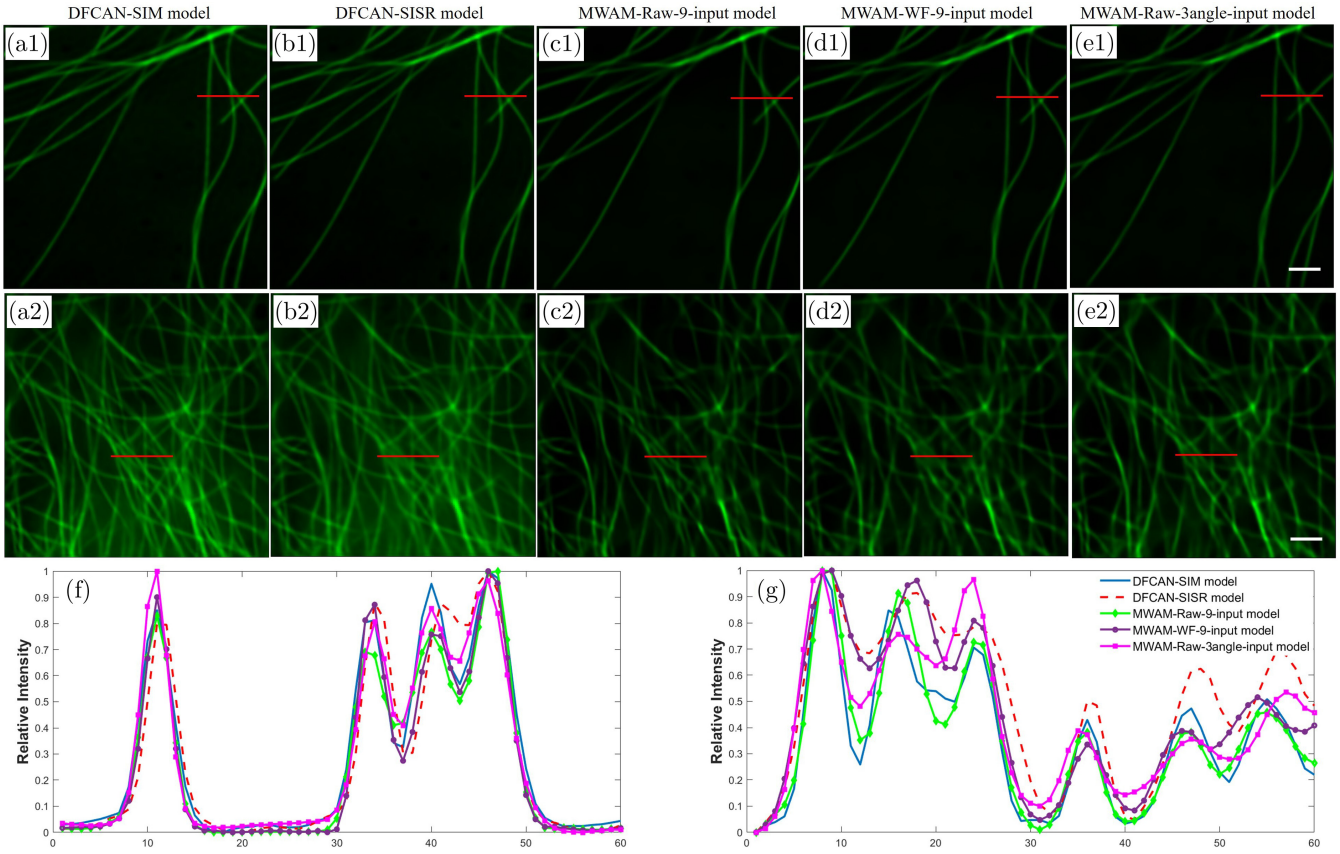

Fig. 7. Comparison of the SR images reconstructed by DFCAN and MWAM models. The images in first to fifth columns are SR images reconstructed by the DFCAN-SIM model, DFCAN-SISR model, MWAM-Raw-9-input model, MWAM-WF-9-input model, and MWAM-Raw-3angle-intput model, respectively. (f) and (g) show the intensity distributions along the red line in the first and second row images. Scale bars: 1.0 μm.

model, and MWAM-Raw-3angle-input model, respectively. The input images of DFCAN-SIM model and MWAM-Raw-9 input model are nine raw images, while the input images of the DFCAN-SISR model and MWAM-WF-9-input model are WF image averaged by nine raw images. Figures 7(f) and 7(g) show the intensity distributions along the red line in the first and second row images in Fig. 7, respectively. We can see that, the MWAM models produce comparable reconstruction performance to the DFCAN models. Due to the optimization of SIM-SR images in this paper, the reconstructed SR images by MWAM models contain weaker background fluorescence. In addition, the MWAM network is uniquely capable of producing exceptional results even when only three raw images are used as input, thus enabling a better tradeoff between temporal and spatial resolution.

3.2. *Reconstructed results under low fluorescence level*

In certain biological experiments and specific samples, it is crucial to strictly control the illumination density and exposure time to prevent damage or bleaching of the sample. However, this can often result in obtaining raw images with poor SNR. To address this issue, we utilized two other datasets with fluorescence levels 0.6 and 0.3, respectively to train the MWAM network. The fluorescence level of raw images with optimal SNR is set to 1.0 for reference. The input data consist of three raw images captured at different illumination angles. The SR images reconstructed by the MWAM model and conventional SIM algorithm are presented in Fig. 8. The results demonstrate that the MWAM network can reconstruct SR images from low SNR raw

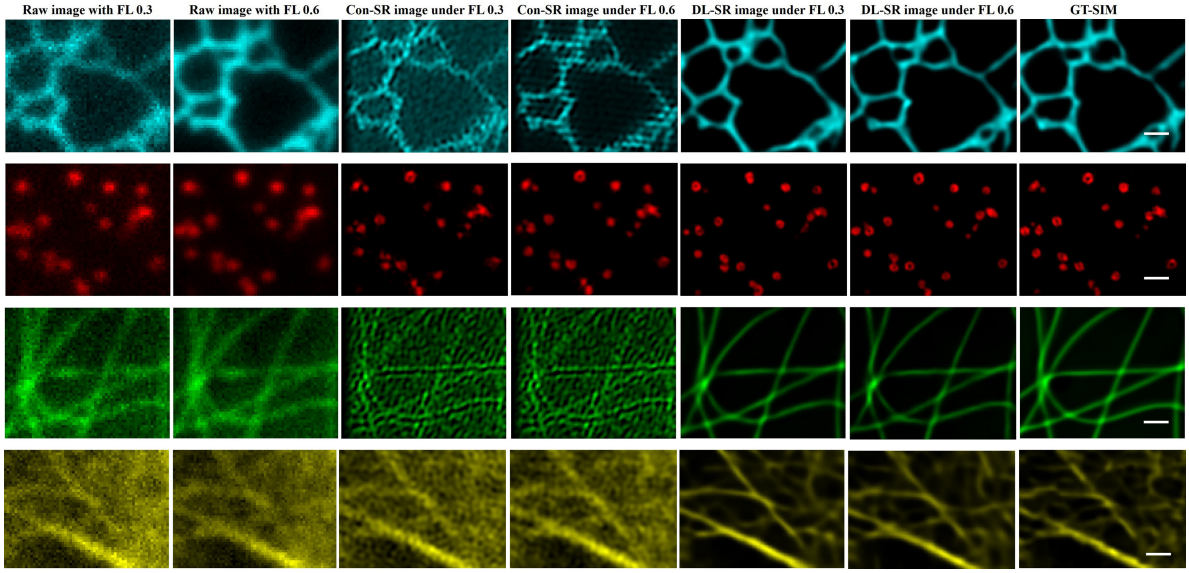

Fig. 8. Reconstructed SR images under low-fluorescence level. Raw images with fluorescence level 0.3 (first column), raw images with fluorescence level 0.6 (second column), SR images reconstructed by the conventional SIM algorithm under fluorescence level 0.3 (third column), SR images reconstructed by the conventional SIM algorithm under fluorescence level 0.6 (fourth column), SR images reconstructed by MWAM network under the fluorescence level 0.3 (fifth column), SR images reconstructed by MWAM network under the fluorescence level 0.6 (sixth column), GT-SIM images (seventh column). Scale bars: 1 μm.

images with fewer errors and artifacts, particularly for relatively simple structures, as depicted in the fifth and sixth column of Fig. 8. Average PSNR and SSIM values of SR images output by models are 32.24 dB/0.8976 (fluorescence level 0.6) and 30.58 dB/0.8618 (fluorescence level 0.3), respectively. In contrast, the conventional SIM algorithm struggles to obtain accurate illumination parameters and eliminate high-frequency noise under low fluorescence level, leading to severely degraded SR images, as depicted in the third and fourth column of Fig. 8.

3.3. *The adaptability of MWAM network*

To investigate the adaptability of the MWAM network to various biological structures, we trained the model on four different datasets, namely CCPs, ER, MT, and F-action, respectively, and evaluated performance on other three datasets that are not included in the training set. The average PSNR and SSIM values of the validation results are presented in Table 4. Although it is preferable to train the network for each type of biological structure, we observed that the model trained on complex structures, such as F-actin and MTs, exhibited good generalization performance on other structures, such as and CCPs.

On the other hand, we obtained raw images using the 2D SIM mode of Nikon NSIM system, which had a different system configuration, such as numerical aperture (NA) of objective, magnification, PSF. The tested sample is bovine pulmonary artery endothelial (BPAE) cells stained with a combination of fluorescent dyes, including mitochondria

Table 4. Average PSNR/SSIM results under different training datasets and validation datasets.

		Dataset for validation			
Input data type	Dataset for training	CCPs	ER	MTs	F-actin
Raw-3angle-input	CCPs	35.79/0.9791	24.38/0.6886	27.37/0.5821	27.72/0.5267
	ER	31.47/0.8841	33.53/0.9526	30.55/0.8653	31.47/0.8534
	MTs	31.28/0.9366	30.59/0.8989	34.95/0.9380	35.84/0.9226
	F-actin	31.88/0.9435	30.89/0.9145	34.59/0.9292	36.67/0.9331

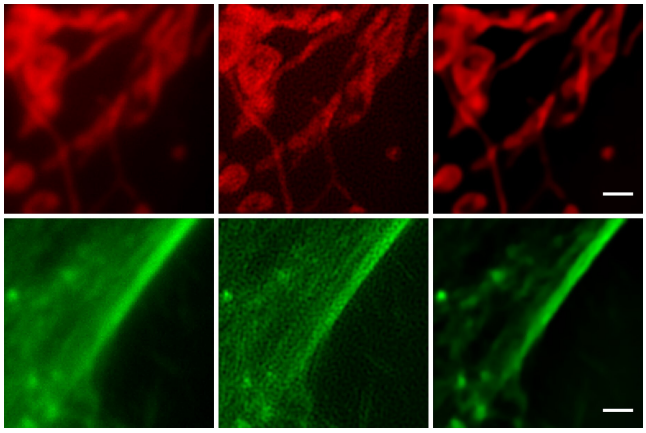

Fig. 9. The adaptability of the MWAM model to Nikon NSIM system. WF images (first column), SR images reconstructed by Nikon NSIM system (second column), and SR images output by pretrained MWAM model (third column). Scale bar: 2 μm.

labeled with red-fluorescent MitoTracker® Red CMXRos and F-actin labeled with green-fluorescent Alexa Fluor® 488 phalloidin. The WF images and SR images reconstructed by Nikon NSIM are presented in the first and second column of Fig. 9. To compare the performance, we fed these raw images into the pretrained MWAM model on the BioSR dataset, and the output results are shown in the third column of Fig. 9. It can be seen that the SR images reconstructed by the network model exhibit lower background noise and fewer artifacts. Therefore, the proposed MWAM network can be effectively applied to the 2D SIM mode of Nikon NSIM system. However, the network model cannot be applied to images obtained using the 3D SIM mode, due to the presence of clear fringe structures in the raw images.

4. Discussion

In this paper, we demonstrate that the deep neural network MWAM is highly effective for super-resolution reconstruction of SIM. The pretrained MWAM model can output high-fidelity SR image using just three raw images and can achieve remarkable reconstruction results when the fluorescence level is as low as 0.3. These research results have surpassed the limitations of conventional reconstruction algorithms, which are essential in improving the temporal resolution and minimizing the phototoxicity.

The application of DL to optical microscopes confronts the issue of adaptability. In this paper, we have shown that the MWAM model trained on complex structures exhibited good generalization performance on other simple structures. However, it should be noted that if the structures of the sample are completely 'unfamiliar' to the model, it may lead to reconstruction errors. Training on datasets with sufficiently complex structures can effectively improve the applicability of network model. Complex structures can be obtained through artificial synthesis or multicolor colocalization experiments. Another solution is to obtain high-quality small datasets of new biological samples under high fluorescence level, and use transfer learning methods to train a new model suitable for this sample. These two aspects of work are our main research directions in the future.

5. Conclusions

SIM is a widely used and powerful super-resolution technique in biomedical research that uses structured patterns to illuminate the sample, resulting in up to two times higher spatial resolution compared to traditional microscopy. However, as a computational imaging method, the quality of reconstructed SR images heavily relies on accurate prior knowledge of illumination patterns and SNR of raw images. This limits the temporal resolution SIM, as multiple raw images with high SNR need to be captured. DL is an emerging technology that has brought significant innovation to many fields, including optical microscopy. However, it is important to note that SIM raw images differ significantly from wide-field images, as SIM images contain high-frequency information modulated by structured illumination, which is crucial for reconstructing SR images.

In this study, we propose a deep neural network based on MWAM for SIM. The raw images are decomposed into sub band images through multi-level WPT, then several branches including RCAB, IWT, and SKAB are used to extract multi-level feature maps. Finally, the network outputs SR images through multi-channel fusion and up-sampling. The study results show that the MWAM network can extract high-frequency information contained in SIM raw images and output superior SR images compared to those generated using WF images as input data. To achieve the optimal tradeoff between temporal and spatial resolution, the number of SIM raw images can be reduced to three. Furthermore, the proposed MWAM network

outperforms conventional algorithms under low-fluorescence level. We have also demonstrated that the model trained on complex structures exhibited good generalization performance on other simple structures. Finally, the pretrained MWAM model based on BioSR dataset is successfully applied to the 2D SIM data measured by Nikon NSIM systems.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62005307 and 61975228).

Conflicts of Interest

The authors have no conflicts of interest relevant to this article.

References

1. M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *J. Micros.* **198**(2), 82–87 (2000).
2. M. G. Gustafsson, L. Shao, P. M. Carlton, C. R. Wang, I. N. Golubovskaya, W. Z. Cande, J. W. Sedat, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," *Biophys. J.* **94**(12), 4957–4970 (2008).
3. K. Samanta, J. Joseph, "An overview of structured illumination microscopy: recent advances and perspectives," *J. Opt.* **23**(12), 123002 (2021).
4. Y. Guo, D. Li, S. Zhang, Y. Yang, J. J. Liu, X. Wang, D. Li, "Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales," *Cell* **175**(5), 1430–1442 (2018).
5. V. Navikas, S. M. Leitao, K. S. Grussmayer, A. Descloux, B. Drake, K. Yserentant, G. E. Fantner, "Correlative 3D microscopy of single cells using super-resolution and scanning ion-conductance microscopy," *Nat. Commun.* **12**(1), 4565 (2021).
6. C. Qiao, D. Li, Y. Liu, S. Zhang, K. Liu, C. Liu, D. Li, "Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes," *Nat. Biotechnol.* **41**, 367–377 (2023).
7. L. Xu, Y. Zhang, S. Lang, Y. Gong, "Exploration of deformation of F-Actin during macropinocytosis by confocal microscopy and 3D-structured illumination microscopy," *Photonics* **9**(7), 461 (2022).
8. S. W. Hell, J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.* **19**(11), 780–782 (1994).
9. M. Bates, B. Huang, G. T. Dempsey, X. Zhuang, "Multicolor super-resolution imaging with photo-switchable fluorescent probes," *Science* **317**(5845), 1749–1753 (2007).
10. G. Ball, J. Demmerle, R. Kaufmann, I. Davis, I. M. Dobbie, L. Schermelleh, "SIMcheck: a toolbox for successful super-resolution structured illumination microscopy," *Sci. Rep.* **5**(1), 15915 (2015).
11. M. Müller, V. Mönkemöller, S. Hennig, W. Hübner, T. Huser, "Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ," *Nat. Commun.* **7**(1), 10980 (2016).
12. Y. Zhang, S. Lang, H. Wang, J. Liao, Y. Gong, "Super-resolution algorithm based on Richardson–Lucy deconvolution for three-dimensional structured illumination microscopy," *J. Opt. Soc. Am. A* **36**(2), 173–178 (2019).
13. J. Demmerle, C. Innocent, A. J. North, G. Ball, M. Müller, E. Miron, L. Schermelleh, "Strategic and practical guidelines for successful structured illumination microscopy," *Nat. Protoc.* **12**(5), 988–1010 (2017).
14. X. Huang, J. Fan, L. Li, H. Liu, R. Wu, Y. Wu, L. Chen, "Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy," *Nat. Biotechnol.* **36**(5), 451–459 (2018).
15. G. Wen, S. Li, L. Wang, X. Chen, Z. Sun, Y. Liang, H. Li, "High-fidelity structured illumination microscopy by point-spread-function engineering," *Light-Sci. Appl.* **10**(1), 70 (2021).
16. W. Zhao, S. Zhao, L. Li, X. Huang, S. Xing, Y. Zhang, L. Chen, "Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy," *Nat. Biotechnol.* **40**(4), 606–617 (2022).
17. Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
18. Y. Rivenson, K. de Haan, W. D. Wallace, A. Ozcan, "Emerging advances to transform histopathology using virtual staining," *BME Front.* **2020**, 1–11 (2020).
19. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nat. Meth.* **15**(11), 917–920 (2018).
20. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).

21. V. Mannam, Y. Zhang, X. Yuan, S. Howard, Deep learning-based super-resolution fluorescence microscopy on small datasets, *In Single Molecule Spectroscopy and Superresolution Imaging XIV*, Vol. 11650, pp. 60–68, Society of Photo-Optical Instrumentation Engineers, Bellingham, Washington, America (2021).
22. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, E. W. Myers, "Content-aware image restoration: Pushing the limits of fluorescence microscopy," *Nat. Meth.* **15**(12), 1090–1097 (2018).
23. Y. Zhao, M. Zhang, W. Zhang, Y. Zhou, L. Chen, Q. Liu, Y. H. Zhang, "Isotropic super-resolution light-sheet microscopy of dynamic intracellular structures at subsecond timescales," *Nat. Meth.* **19**(3), 359–369 (2022).
24. C. Qiao, D. Li, Y. Guo, C. Liu, T. Jiang, Q. Dai, D. Li, "Evaluation and development of deep neural networks for image super-resolution in optical microscopy," *Nat. Meth.* **18**(2), 194–202 (2021).
25. L. Jin, B. Liu, F. Zhao, S. Hahn, B. Dong, R. Song, K. M. Hahn, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," *Nat. Commun.* **11**(1), 1934 (2020).
26. C. Ling, C. Zhang, M. Wang, F. Meng, L. Du, X. Yuan, "Fast structured illumination microscopy via deep learning," *Photon. Res.* **8**(8), 1350–1359 (2020).
27. Q. Zhang, J. Chen, J. Li, E. Bo, H. Jiang, X. Lu, J. Tian, "Deep learning-based single-shot structured illumination microscopy," *Opt. Laser. Eng.* **155**, 107066 (2022).
28. Z. H. Shah, M. Müller, T. C. Wang, P. M. Scheidig, A. Schneider, M. Schüttpelz, W. Schenck, "Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images," *Photon. Res.* **9**(5), B168–B181 (2021).
29. C. N. Christensen, E. N. Ward, M. Lu, P. Lio, C. F. Kaminski, "ML-SIM: Universal reconstruction of structured illumination microscopy images using transfer learning," *Biomed. Opt. Exp.* **12**(5), 2720–2733 (2021).
30. X. Cheng, J. Li, Q. Dai, Z. Fu, J. Yang, "Fast and lightweight network for single frame structured illumination microscopy super-resolution," *IEEE Trans. Instrum. Meas.* **71**, 1–11 (2022).
31. P. Liu, H. Zhang, K. Zhang, L. Lin, W. Zuo, Multi-level wavelet-CNN for image restoration, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 773–782, IEEE, Piscataway, NJ, America (2018).
32. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, *Proc. Eur. Conf. Computer Vision (ECCV)*, pp. 286–301, Springer, Berlin, Germany (2018).
33. X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 510–519 (2019).
34. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).
35. I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory* **36**(5), 961–1005 (1990).
36. W. S. Lai, J. B. Huang, N. Ahuja, M. H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE Trans. Pattern Anal.* **41**(11), 2599–2613 (2018).
37. D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," *Comput. Sci.* **1412**, 80 (2014).