Received: 4 November 2020 | Revised: 3 February 2021 | Accepted: 5 February 2021

DOI: 10.1002/jbio.202000444

FULL ARTICLE

JOURNAL OF BIOPHOTONICS

Double-flow convolutional neural network for rapid large field of view Fourier ptychographic reconstruction

Minglu Sun[1,2] | **Lina Shao**[3] | **Youqiang Zhu**[1,2] | **Yuxi Zhang**[1,2] | **Shaoxin Wang**[1] | **Yukun Wang**[1] | **Zhihui Diao**[1] | **Dayu Li**[1*] | **Quanquan Mu**[1,2*] | **Li Xuan**[1,2]

[1]State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China

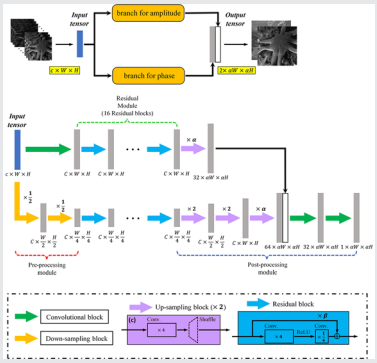
[2]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing, China

[3]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Science, Changchun, China

***Correspondence**
Dayu Li and Quanquan Mu, State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 130033 Changchun, China.
Email: lidayu@ciomp.ac.cn (D. L.) and
Email: muquanquan@ciomp.ac.cn (Q. M.)

Funding information
CAS Interdisciplinary Innovation Team; Changchun Science and Technology Project, Grant/Award Number: 18SS003; National Natural Science Foundation of China, Grant/Award Numbers: 11704378, 12004381, 31901073; Science and Technology Development Plan Project of Jilin Province, Grant/Award Number: 20180201043GX; Science Foundation of State Key Laboratory of Applied Optics

Abstract



Fourier ptychographic microscopy is a promising imaging technique which can circumvent the space-bandwidth product of the system and achieve a reconstruction result with wide field-of-view (FOV), high-resolution and quantitative phase information. However, traditional iterative-based methods typically require multiple times to get convergence, and due to the wave vector deviation in different areas, the millimeter-level full-FOV cannot be well reconstructed once and typically required to be separated into several portions with sufficient overlaps and reconstructed separately, which makes traditional methods suffer from long reconstruction time for a large-FOV (of the order of minutes) and limits the application in real-time large-FOV monitoring of live sample in vitro. Here we propose a novel deep-learning based method called DFNN which can be used in place of traditional iterative-based methods to increase the quality of single large-FOV reconstruction and reducing the processing time from 167.5 to 0.1125 second. In addition, we demonstrate that by training based on the simulation dataset with high-entropy property (Opt. Express 28, 24 152 [2020]), DFNN could has fine generalizability and little dependence on the morphological features of samples. The superior robustness of DFNN against noise is also demonstrated in both simulation and experiment. Furthermore, our model shows more robustness against the wave vector deviation. Therefore, we could achieve better results at the edge areas of a single large-FOV reconstruction. Our method demonstrates a promising way to perform real-time single large-FOV reconstructions and

Abbreviations: Adam, adaptive moment estimation; AP, alternative projection; CTF, coherent transfer function; DCNN, deep convolutional neural networks; DFNN, double flow neural network; DL, deep learning; FPM, Fourier ptychographic microscopy; MAE, mean absolute error; NMSE, normalized mean square error; PV, peak-to-velly value; ReLU, rectified linear unit activation; SBP, space-bandwidth product; SSIM, structural similarity.

Dayu Li and Quanquan Mu are considered as joint corresponding author.

provides further possibilities for real-time large-FOV monitoring of live samples with sub-cellular resolution.

KEYWORDS

convolutional neural networks, deep learning, Fourier ptychographic, optical microscopic imaging systems, rapid large-FOV reconstructions

1 | INTRODUCTION

High-resolution wide-field imaging is essential for various applications in different fields, such as digital pathology, biological and bio-medical research, which requires large space-band product (SBP) to provide computational and statistical analysis for thousands of cells across a wide field-of view (FOV). However, due to the limitation of the system SBP, almost all conventional microscopes suffer from a trade-off between the spatial resolution and the expansion of the FOV [1]. To solve this problem, a novel coherent imaging technique called Fourier ptychographic microscopy (FPM) was developed in 2013 [1–3] which can obtain higher reconstruction resolution while maintaining the size of the FOV, therefore circumvents the limitation of the system SBP. In addition, due to the phase retrieval technique employed, the quantitative phase information could be acquired without direct phase measurement, which is of great importance in the case of imaging phase object that cannot be well sensed by traditional microscopy.

Since the advent of the original FPM, various modifications have been implemented to further improve the performance. Several methods for suppressing the negative influence of the background noise have been proposed [4–8]. By embedding the pupil recovery procedure, the impacts of the optical aberrations could be eliminated [9, 10]. In addition, many new optimization methods have been developed to improve the robustness of FPM [11–16]. In order to perform real-time live sample reconstruction, the temporal resolution of data collection should be improved, various strategies such as multiplexed illumination [17–19] and self-learning [20] have been applied, the data acquisition time of single frame could be less than 1 s [18]. Furthermore, in order to achieve real-time FPM full-FOV monitoring of live samples we need to reduce the full-FOV reconstruction time for single-frame to below the data acquisition time. However, for the currently used iterative-based methods, multiple iterations are usually needed to get convergence. Moreover, as discussed in this paper, due to the wave vector deviation at the edge FOV, traditional iterative methods could not obtain a fine result when directly reconstruct a large FOV. Commonly, it needs to separate the full-FOV into several small portions and reconstruct them separately, then stitch them together to output a full-FOV reconstruction results, and typically, a sufficient overlap rate in each direction between adjacent portions is required to ensure the quality of the fusion result [1, 21], which introduces extra calculations and makes the time consumption of reconstruction much longer than the time consumption of data acquisition. Therefore, it becomes necessary to find a rapid FPM reconstruction algorithm for real-time monitoring of live samples with full-FOV and high-quality.

In recent years, with the rapid development of the deep learning (DL) technique, algorithms based on deep convolutional neural network (DCNN) have been proposed to solve many image processing problems, such as image de-noising [22], single image super-resolution [23–25] and phase retrieval [26, 29]. Since the purpose of FPM is to synthesize a high-resolution complex field from multiple low-resolution images, several algorithms have employed DCNN to solve the FPM problem [27–32] which greatly improve the reconstruction speed. However, in Reference [27] they need traditional method to generate a preliminary result as the input of network, and then the network is trained to optimize the input instead of using the low-resolution images to reconstruct, and in Reference [28] the performance of the network is mainly demonstrated by simulation and lack the description of the generalization to the actual experimental dataset. Moreover, most of the DL-based methods mentioned above require retraining or transfer learning for new actual sample distributions [29–32]. Although using the technique of transfer learning can make the network quickly adapt to another sample with less re-training times and smaller size of the re-training dataset [32], they still need other algorithms to generate high-resolution ground truth of this new sample, which is not practical [31, 32]. The main reason of this generalization problem is that a dataset of a certain morphological information with low-entropy property [33] is used during the training process which will certainly speed up the training progress [31, 32], but also make the network rely too much on the certain morphological feature to reconstruct, thereby reducing the generalization for different samples.

In this paper, we propose a novel network based on DCNN called DFNN to solve the FPM problem, we employ a large number of simulated pictures with different morphological features and high Shannon entropy to generate training dataset which has been discussed to improve the generalization of a network [33]. After the first training process, we can directly use it to perform reconstructions on experimental dataset and obtain fine results without a secondary training, which indicates that our model has fine generalizability and little data dependence on the morphological features, thereby the practicality of the algorithm could be further improved. Noted that the experimental dataset should be captured with the same system parameters. Besides, the reconstruction results on both simulation and actual experimental dataset indicate that the network has stronger robustness to imaging noise. In the comparison of experimental results, we perform single high-throughput reconstructions (large FOV) and compare the results with traditional iterative-based method. The comparison shows that our model is less sensitive to the deviation of the wave vector which is essential to traditional methods, leading to better reconstruction results at the edge areas of the FOV. Moreover, due to the end-to-end structure and graphic processing units (GPU) acceleration technology, the time consumption of DFNN for single large-FOV reconstruction could be reduced by nearly 1500 times which greatly improves the possibility of FPM in real-time full-FOV monitoring of live samples.

This paper is structured as follows: In section 2, we briefly introduce the overall structure and the physical imaging process of FPM, we also introduce the reconstruction process of commonly used alternative projection (AP) method [1, 4, 5]. In section 3, we describe the structure of DFNN and the training strategy, and the feasibility of the network is verified through simulation. Meanwhile, the robustness of DFNN at different noise levels is also compared with AP method. In section 4, in order to illustrate the generalizability of DFNN, we do not retrain or fine-tune the network on the experimental dataset. First, we evaluate the resolution enhancement performance with USAF dataset and compare the results with traditional AP method. Then we use biological samples to perform single large-FOV reconstructions and by comparing the results with AP method, we demonstrate that our method has less sensitive to wave vector deviations and could obtain fine generalization property and better results with higher contrast and more details. Finally, section 5 concludes the paper with summaries and discussions.

2 | THE PRINCIPLE OF FOURIER PTYCHOGRAPHIC MICROSCOPY

In the traditional FPM system, an LED matrix is utilized to provide incident light with different angles. During the imaging process, a thin sample, which could be represented by its complex transmission function $o(\boldsymbol{r})$, is placed at the front focal plane of the objective lens, where $\boldsymbol{r} = (x, y)$ represents the 2-dimensional (2D) coordinates in the spatial domain.

Assuming that the LED matrix is far enough away from the sample and the reconstruction area is small enough so that the incident light could be approximated as an oblique plane wave. When the m^{th} LED is activated, the spectrum of the sample will be shifted accordingly. The amount of shift is equal to the wave vector of the incident light $\boldsymbol{u}_m$, which could be formulated as:

$$\boldsymbol{u}_m = \left(\frac{\sin\theta_x^{(m)}}{\lambda}, \frac{\sin\theta_y^{(m)}}{\lambda} \right), \tag{1}$$

where $\theta_x^{(m)}$ and $\theta_y^{(m)}$ represent the illumination angle of the m^{th} LED, which is determined by the relative position of the LED to the reconstructed area. λ is the illumination wavelength. Therefore, the spectrum of the sample at the Fourier plane could be expressed as:

$$\mathcal{F}\{o(\boldsymbol{r})\exp(i2\pi\boldsymbol{u}_m\boldsymbol{r})\} = O(\boldsymbol{u}-\boldsymbol{u}_m), \tag{2}$$

where $\boldsymbol{u} = (f_x, f_y)$ represents the 2D coordinates in the frequency domain, F represents the Fourier transform and $O(\boldsymbol{u} - \boldsymbol{u}_m)$ refers to the spectrum of the sample which is shifted to be centered around $\boldsymbol{u}_m$. Due to the limitation of NA, the spectrum is low-filtered by the system pupil function $P(\boldsymbol{u})$. Therefore, according to Equation (2), the intensity image recorded by the sensor corresponding to the m^{th} LED could be formulated as:

$$I_m(\boldsymbol{r}) = \left| \mathcal{F}^{-1}\{O(\boldsymbol{u}-\boldsymbol{u}_m)\cdot P(\boldsymbol{u})\} \right|^2, \tag{3}$$

where F^{-1} represents the inverse Fourier transform. When the LEDs are sequentially activated, we can obtain a series of low-resolution images containing information from different sub-regions of the spectrum.

The main principle of FPM algorithm is to synthesize the estimated spectrum of the sample $O_e(\boldsymbol{u})$ using the captured low-resolution intensity images, then get the estimated high-resolution complex field $o_e(\boldsymbol{r}) = F^{-1}\{O_e(\boldsymbol{u})\}$. However, since only the intensity information is captured, a phase retrieval algorithm needs to be employed to obtain the phase information. At present, the most commonly used algorithm called AP method which alternatively

constrains the reconstruction process in the spatial and frequency domain. First, AP method need an initial guess of the reconstructed spectrum $O_e(\boldsymbol{u})$, which is the Fourier transform of an initial complex field in which the amplitude is usually the up-sampled low-resolution image under the normal incidence condition and the phase is set to zero. Therefore, we can get the estimated low-resolution complex field $E_{e,m}(\boldsymbol{r})$ according to the m^{th} LED.

$$E_{e,m}(\boldsymbol{r}) = F^{-1}\{O_e(\boldsymbol{u}-\boldsymbol{u}_m)\cdot P(\boldsymbol{u})\} \tag{4}$$

In which the pupil function $P(\boldsymbol{u})$ is commonly considered as the coherent transfer function (CTF) of the system and can be expressed as:

$$P\left(f_x,f_y\right) = \text{CTF} = \begin{cases} 1, & \text{if } \left(f_x^2+f_y^2\right) < \left(\frac{NA}{\lambda}\right)^2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Second, the amplitude of $E_{e,m}(\boldsymbol{r})$ is replaced by the actual measured intensity $\bar{I}_m(\boldsymbol{r})$ and the phase remained still to enforce the spatial domain restrain. Then the corresponding sub-region of the spectrum is updated by the Fourier transform of the new complex field:

$$O_e(\boldsymbol{u}-\boldsymbol{u}_m)\cdot P(\boldsymbol{u}) = \mathcal{F}\left\{\frac{\sqrt{\bar{I}_m(\boldsymbol{r})}}{|E_{e,m}(\boldsymbol{r})|}E_{e,m}(\boldsymbol{r})\right\}. \tag{6}$$

All sub-regions of the spectrum need to be updated during one iteration, and AP method usually need multiple iterations to converge to the final result with a wider pass-band. Moreover, to ensure the accuracy of the wave vector $\boldsymbol{u}_m$, a large-FOV is usually needed to be separated into multiple portions and reconstructed separately with the correct wave vectors [1, 2], then the multiple results are stitched together using image fusion technique. In addition, to ensure the fusion quality, sufficient overlap between adjacent portions is needed [1, 21, 32], therefore, the reconstruction speed of traditional AP methods for a large-FOV is usually sacrificed, leading to a low reconstruction temporal resolution. In this paper, we are committed to achieving a higher reconstruction speed for large-FOV by utilizing deep learning technique while obtaining good generalizability and fine reconstruction results.

3 | METHOD

3.1 | The structure of DFNN

We take a series of low-resolution images captured by the camera as input to the network. Since we expect the network to output high-resolution complex amplitudes, therefore the output needs to be a 2-channels tensor consisting of the high-resolution amplitude and phase information. In order to avoid the crosstalk, the reconstruction network is divided into two branches, that is, one for the amplitude and the other for the phase. Both branches are identical in structure and contain multiple residual connection blocks [34]. The general structure of one branch with dimension information is shown in Figure 1A.

Each branch of the network has two data flows and each flow contains a pre-processing module, a residual module and a post-processing module. In the pre-processing module, the channel number of the input tensor need to be extracted into C by passing through a convolutional layer, moreover, in order to increase the receptive field of the network and improve the ability to extract information of different scales [35], the tensor in one flow is going to down-sampled by ×4. Instead of applying the pooling layer, we utilize two convolutional layers with stride as 2 to carry out the ×4 down-sampling operation. In this way we could implement the down-sampling and convolution operations simultaneously.

The pre-processing module is followed by the residual module which is an essential part of the network. The residual module is composed of 16 residual blocks, inspired by the structure proposed in Reference [32], we build the residual block by using two convolutional layers and a rectified linear unit activation (ReLU) layer, as shown in Figure 1B. The first convolutional layer expands the channel number of the input tensor by a factor of 4, then after the tensor is modulated by the ReLU layer, the number of channels is restored into C by the second convolutional layer. Finally, the tensor is added up with the input through a mapping operation which is modulated by a parameter β and output as the result of the current block. By utilizing the residual structure which connects the output and input of the block, the performance and the convergence speed of the network could be significantly improved [34]. In addition, compared with the traditional residual structure, increasing the number of channels before the activation layer could further improve the performance in image reconstruction [36, 37].

The output of residual module is then passed through a post-processing module, in which a series of operations are performed, including up-sampling, concatenation and the output of the result. During the up-sampling operation, the tensors of two data flows are up-sampled into the desired size. The super-resolution factor is set to α, therefore the up-sampling ratios for two flows are $\times\alpha$ and $\times 4\alpha$ respectively. The up-sampled block is composed of a convolutional layer and a pixel shuffle operation, as shown in Figure 1C, an ×2 up-sampled block consists of a convolutional layer which expands the channels by a

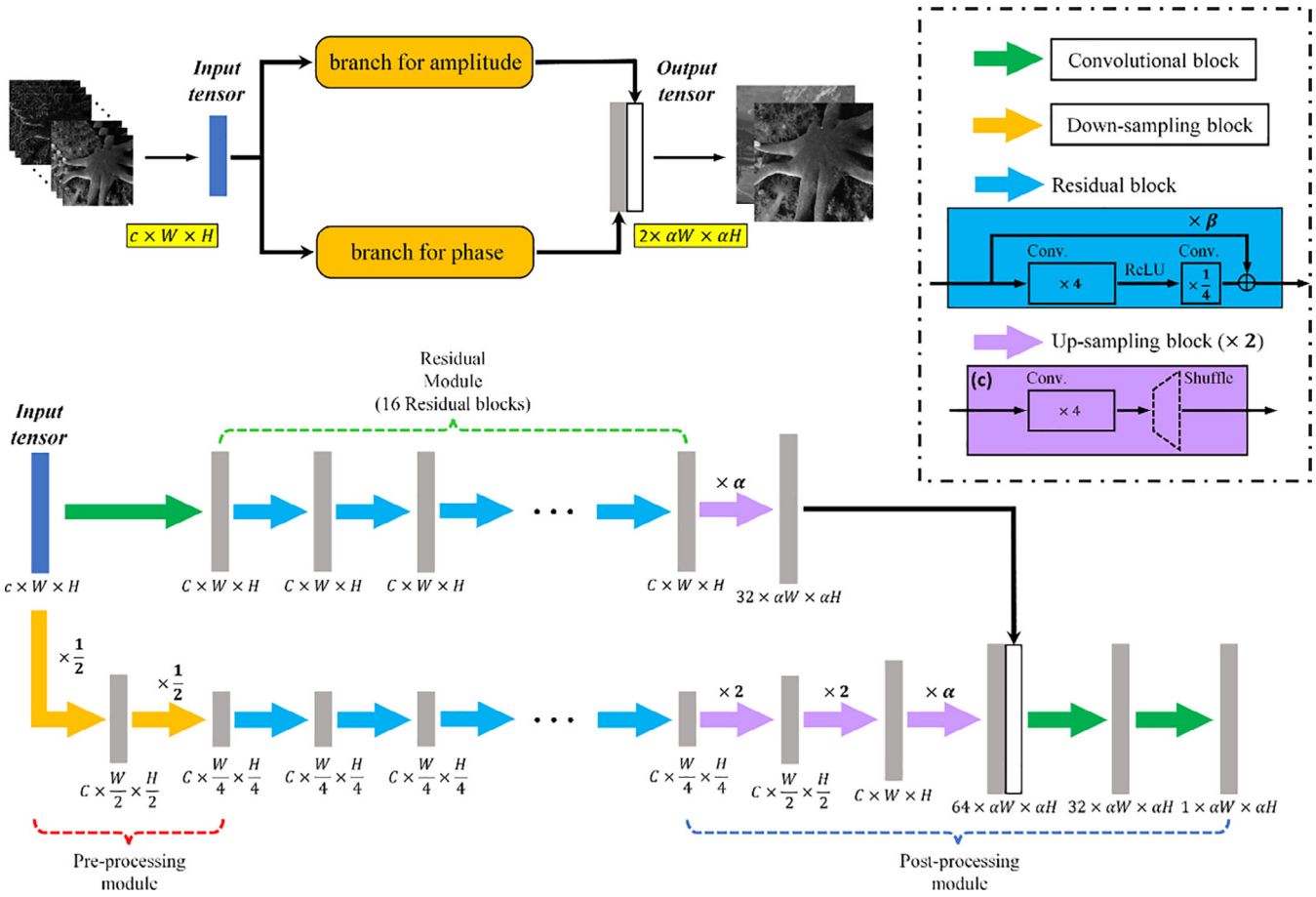


FIGURE 1 Diagram of the proposed DFNN model. A, The general architecture of one branch. B, The structure of a residual block in the residual module. C, The structure of the ×2 up-sampling block

factor of 4 and a pixel shuffle layer which up-samples the size ×2 by merging the information in channels and restores the channel number [38]. After the up-sampling module, the output tensors of two data flows are concatenated together and pass through the last two convolutional blocks of this branch, which perform the final operation to output a 1-channel tensor (see Figure 11 in Section 6 for more details about the functional blocks and parameter setting we used).

We concatenate several low-resolution images as the input of DFNN, which could be expressed as a tensor with a dimension of $c \times W \times H$, where the W and H indicate the width and height of the low-resolution image respectively, c represents the channels of the tensor which is also the number of low-resolution images used, noted that, for better performance, the channel number C inside the network should be larger than the channels of input tensor. Therefore, after the transition of the input tensor through two branches and a concatenate operation, we obtain a tensor with a dimension of $2 \times \alpha W \times \alpha H$ which contains the reconstructed high-resolution amplitude and phase information.

3.2 | Loss function

Instead of simply applying the L_1 loss function to constrain the training process [31], we divide the loss function into two parts representing the losses in the spatial and frequency domains. The loss function of the spatial domain incorporates the L_1-norm term and the structural similarity index (SSIM) term, which could be written as the following form:

$$\text{loss}_{\text{spatial}} = \eta_1 L_1\left(\tau_{\text{GT}}, \tau_{\text{output}}\right) + \eta_2 \text{SSIM}\left(\tau_{\text{GT}}, \tau_{\text{output}}\right), \quad (7)$$

where τ_{GT} and τ_{output} represent the ground truth tensor and the reconstruction result of the network respectively. The L_1-norm term is formulated as:

$$L_1\left(\tau_{\text{GT}}, \tau_{\text{output}}\right) = \frac{1}{N_{\text{pixels}}}\left(\left|\tau_{\text{GT}} - \tau_{\text{output}}\right|\right), \quad (8)$$

where N_{pixels} denotes the total number of pixels in each tensor. This function describes the absolute difference

between each pixel of the reconstruction result and the corresponding ground truth. Moreover, in order to make the DFNN model pay attention to the structural information of the image, the SSIM term is added as a part of the loss function [39], which is defined as:

$$\mathrm{SSIM}\left(\tau_{\mathrm{GT}},\tau_{\mathrm{output}}\right)=\frac{\left(2\mu_{\mathrm{GT}}\mu_{\mathrm{output}}+C_1\right)\left(2\sigma_{\mathrm{GT,output}}+C_2\right)}{\left(\mu_{\mathrm{GT}}^2+\mu_{\mathrm{output}}^2+C_1\right)\left(\sigma_{\mathrm{GT}}^2+\sigma_{\mathrm{output}}^2+C_2\right)}, \tag{9}$$

where μ_{GT} and μ_{output} denote the averages of τ_{GT} and τ_{output}, respectively, $\sigma_{\mathrm{GT,output}}$ represents the covariance of τ_{GT} and τ_{output}, σ_{GT}^2 and $\sigma_{\mathrm{output}}^2$ are the variances of τ_{GT} and τ_{output}, respectively, C_1 and C_2 are the constant values that stabilize the division when the averages are close to zero which are set to 1×10^{-4} and 9×10^{-4}, respectively.

In the frequency domain, the higher frequency information of an image is indicated farther from the center, which makes it easier to be identified than that in the spatial domain, Therefore, we add a spectral domain loss to the loss function which could be formulated as:

$$\mathrm{loss}_{\mathrm{freq}}=\eta_3 L_1\left(F_t\{\tau_{\mathrm{GT}}\},F_t\{\tau_{\mathrm{output}}\}\right), \tag{10}$$

where F_t denotes the Fourier transform of a 2-channels tensor, which results in another 2-channels tensor consisting of the real and imaginary parts of the spectrum. According to Equations (5) and (8), the loss function of DFNN is defined as:

$$\begin{aligned}\mathrm{loss}=&\eta_1 L_1\left(\tau_{\mathrm{GT}},\tau_{\mathrm{output}}\right)+\eta_2\mathrm{SSIM}\left(\tau_{\mathrm{GT}},\tau_{\mathrm{output}}\right)\\&+\eta_3 L_1\left(\mathcal{F}\{\tau_{\mathrm{GT}}\},\mathcal{F}\{\tau_{\mathrm{output}}\}\right),\end{aligned} \tag{11}$$

where (η_1, η_2, η_3) are the hyper-parameters which indicate the relative weight of each component.

3.3 | Training and testing on simulation

In order to quantitatively evaluate the reconstruction performance of DFNN, we utilize the DIV2K dataset [40] to create a simulation dataset based on the FPM imaging principle. All the 800 high-resolution images in the DIV2K dataset are first reshaped into 512×512, and then we randomly employ these images as amplitude and phase to obtain 400 high-resolution complex amplitudes. During the simulation, the NA of the objective lens is set to 0.2, the wavelength is set to 0.514 μm, the distance between the sample and the 7×7 LED matrix is 67.5 mm, the gap between adjacent LEDs is 4 mm and the magnification of the system is set to ×8.15. The synthesized NA could reach to 0.44, which leads to a 2-fold improvement in the resolution ($\alpha=2$). Due to the memory limitation, we consider the ×8 up-sampling operation shown in Figure 1A as three connected ×2 up-sampling blocks and the channel number C inside the network is set to 64.

During the process of acquiring low-resolution images, the pupil function is set to the CTF. In order to make the network learn the strict FPM reconstruction process, the CTFs shifted corresponding to all simulated 'LEDs' are merged together to build a synthesized low-pass filter, and the original high-resolution spectrum is passed through this synthesized filter to obtain the ground truth complex field.

Through the simulation, 400 patches of low-resolution image tensors with the size of $49\times256\times256$ and the corresponding high-resolution ground truth with the size of $2\times512\times512$ are achieved. Then the simulation dataset is randomly separated into 300 and 100 patches to act as the training dataset and testing dataset respectively. Moreover, in order to improve the robustness of DFNN to image noise, Gaussian distribution noise with zero mean and standard deviation of 2×10^{-3} is added on the low-resolution image tensor. DFNN is trained on the training dataset with 200 epochs and the batch-size is set to 2 due to the limitation of memory. Especially, we set the hyper-parameters (η_1, η_2, η_3) of the loss function to (0.3, −0.3 0.2) and the adaptive moment estimation (Adam) as the optimizer [41] with the initial learning rate being 1×10^{-3} and exponential decay rate for the first and second moment estimates being 0.9 and 0.999 ($\beta1=0.9$, $\beta2=0.999$). During the training process, we also employ the *L2-penalty* weight decay rate of 1×10^{-3} and weight normalization method [42] as the weight regularization of the network, the learning rate is multiplied by 0.2 for every 40 epochs. The entire network is built with Python based on the PyTorch library and runs on a PC with an Intel Core i7-8700HQ processor and NVIDIA RTX TITIAN graphic card.

Once DFNN is trained, a rapid reconstruction process could be performed by simply inputting the low-resolution images. We use the test dataset to quantitatively evaluate the effectiveness of the network, Figure 2 exhibits the boxplots of mean absolute error (MAE) and SSIM between the output and the ground truth.

The results are normalized into 0-1 before calculate the MAE and SSIM. It can be seen from Figure 2 that MAE of the amplitude and phase reconstructed by DFNN could be less than 0.010 and 0.015, respectively, and SSIM of the reconstructed amplitude and phase could exceed 0.990 and 0.985, respectively. Each patch of this

simulated dataset has completely different morphological features from others, the only thing in common is the process of getting low-resolution images through the FPM forward imaging model as mentioned in section 2. Therefore, by training DFNN on this dataset, it could pay less attention to the morphological features of the dataset itself, and focus on the reverse process to get high-resolution results. On the other hand, for FPM reconstruction, the input–output ratio of the algorithm is usually much larger than 1, DFNN only needs to learn how to extract the detailed information contained in low-resolution images [43, 44] and "merge" them together, instead of relying on learning the morphological features of a large amount of dataset to know how to "add" detailed information like single image super-resolution technology. Therefore, DFNN could show less dependence on the morphological features, and for the testing dataset with different morphological features, DFNN still shows good reconstruction ability and fine generalization property (Figure 2).

Moreover, in order to verify the robustness of DFNN, we build a new test dataset by contaminating the low-resolution images with different levels of Gaussian distribution noise and compare the reconstruction results of the network with traditional AP method. The MAE and SSIM of the two methods at different levels of noise are shown in Figure 3, in which the standard deviation of Gaussian noise is set from 8×10^{-4} to 8×10^{-3}. In order to ensure the convergence, the AP algorithm is iterated for 40 times.

From Figure 3, we can see that the reconstruction quality of AP is rapidly degraded with the increase of the noise level and DFNN can obtain a better result in terms of MAE and SSIM with only slightly decline as the noise level increases.

For better illustration, Figure 4 exhibits the reconstructed amplitudes, phases and spectra of these two methods at the maximum noise level (8×10^{-3}) along with the ground truth. Inserts two small highlighted sub-regions of amplitude and phase and their error maps with the ground truth. It can be seen that, in the case of a high-level noise, the iterative-based AP method would converge to the result suffering from serious artifacts (Figure 4A), However, the DFNN model could still obtain a clearer result with more details and fewer errors (Figure 4B), which indicates that the proposed DFNN has a much stronger robustness and better performance under noise.

In order to quantitatively evaluate the performance of these two methods from the reconstructed spectrum, we calculate the normalized mean square error (NMSE) between the L_1-norms of the reconstructed spectra and the ground truth spectrum as shown in Table 1 [9, 10]. We can see that DFNN presents the best result in NMSE, which means the reconstructed spectrum is more similar to the ground truth.

4 | PERFORMANCE ON EXPERIMENT

4.1 | Testing the performance on resolution improvement

As discussed in section 3, DFNN can improve the resolution of the input image, therefore, in order to

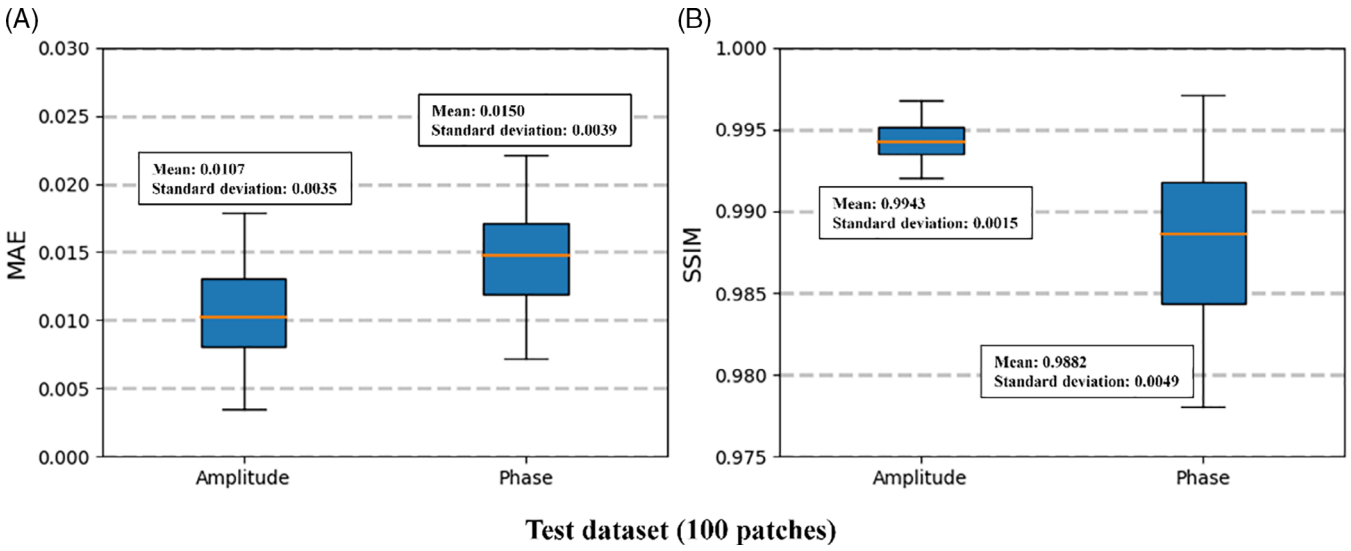


FIGURE 2 The quantitative evaluation of the reconstruction results using the test dataset (100 patches), mean absolute error (MAE) and structual similarity (SSIM) are calculated after normalizing the results into 0-1. A, The boxplots of MAE of the reconstructed amplitude and phase. B, The boxplots of SSIM of the reconstructed amplitude and phase

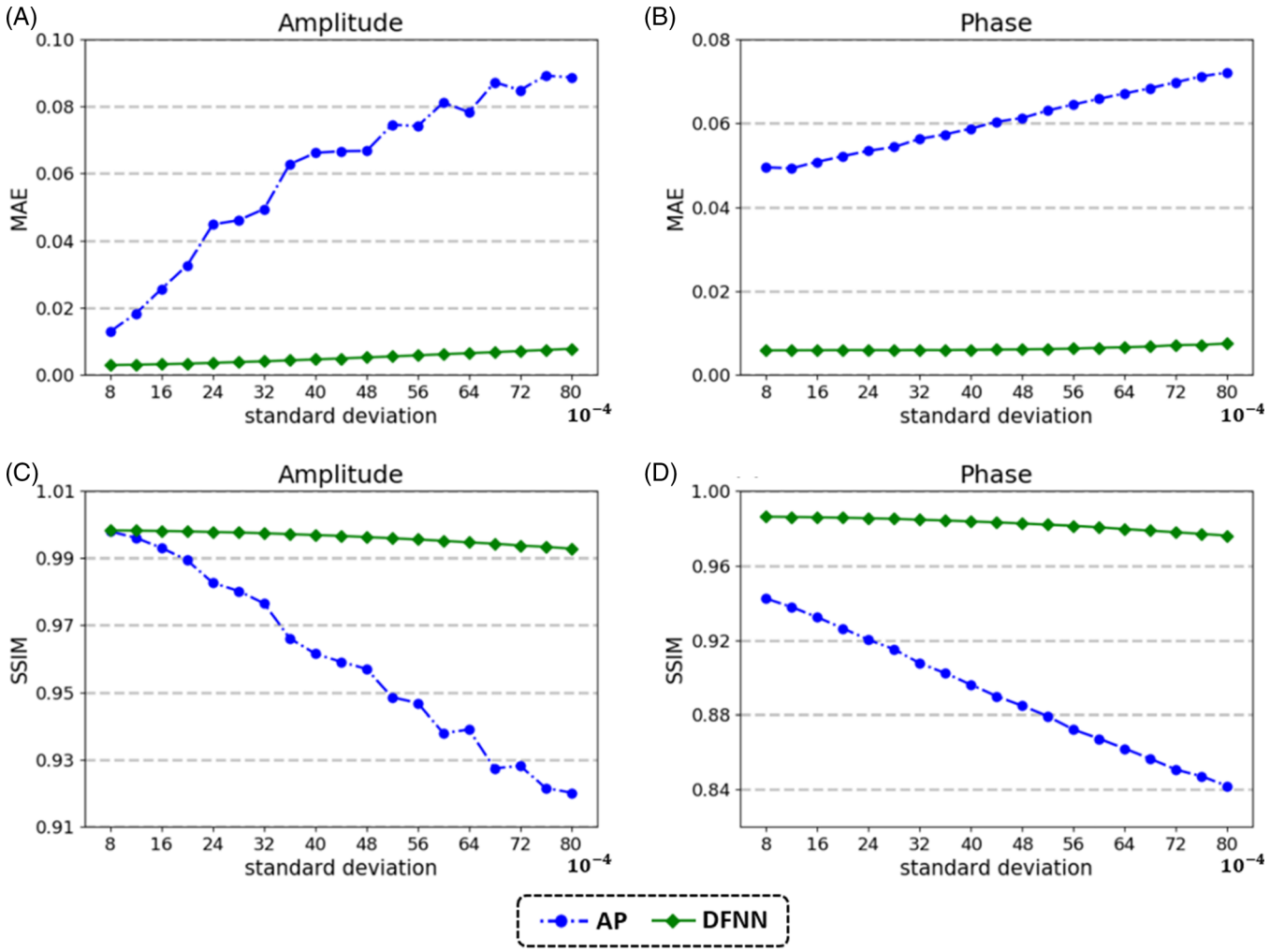


FIGURE 3 Mean absolute error (MAE) and structual similarity (SSIM) of the reconstruction results obtained by alternative projection (AP) and DFNN at different levels of noise, MAE and SSIM are calculated after normalizing the results into 0-1. A,B, MAE of the reconstructed amplitudes and phases. C,D, SSIM of the reconstructed amplitudes and phases

quantitatively illustrate the performance of DFNN in improving the resolution on actual measured dataset, in this section, we first test the performance of our network on an open-sourced USAF dataset [4]. The NA (NA_{obj}) of the employed object lens is 0.1, the magnification of the system is ×4, the pixel size of the camera is 6.5 μm, the distance between the sample and a 11 × 11 LED matrix is 87.5 mm, the distance between adjacent LEDs is 5 mm, therefore, the synthetic NA (NA_{syn}) of the system is up to 0.37 in both X and Y direction which results in a ×4 resolution improvement.

It is worth noting that since the experiment parameters of this dataset are different from those used in our previous simulation, our proposed network need to be retrained according to the new system parameters. In order to demonstrate that DFNN trained based on simulation could have fine generalizability to experimental dataset, the training dataset is obtained from the same simulation dataset (DIV2K) as in section 3 and according to the new parameters, we get 300 patches of low-resolution image tensors with the size of 121 × 128 × 128 (the same Gaussian distribution noise is added as before) and the corresponding high-resolution ground truth with the size of 2 × 512 × 512. Besides, we add an extra ×2 upsampling operation at the beginning of the network, therefore, the overall resolution improvement could reach to ×4 and the channel number C is set to 128 for better performance. Then we use this new simulation dataset to train our network with a similar training process as before, after the training is complete, we directly test our network with the USAF dataset and compare the reconstruction results with AP method which iterates for 15 times to ensure the convergence. The comparison is shown in Figure 5.

From Figure 5D we can see that at the normal incidence, the resolution limit of the system is group 7,

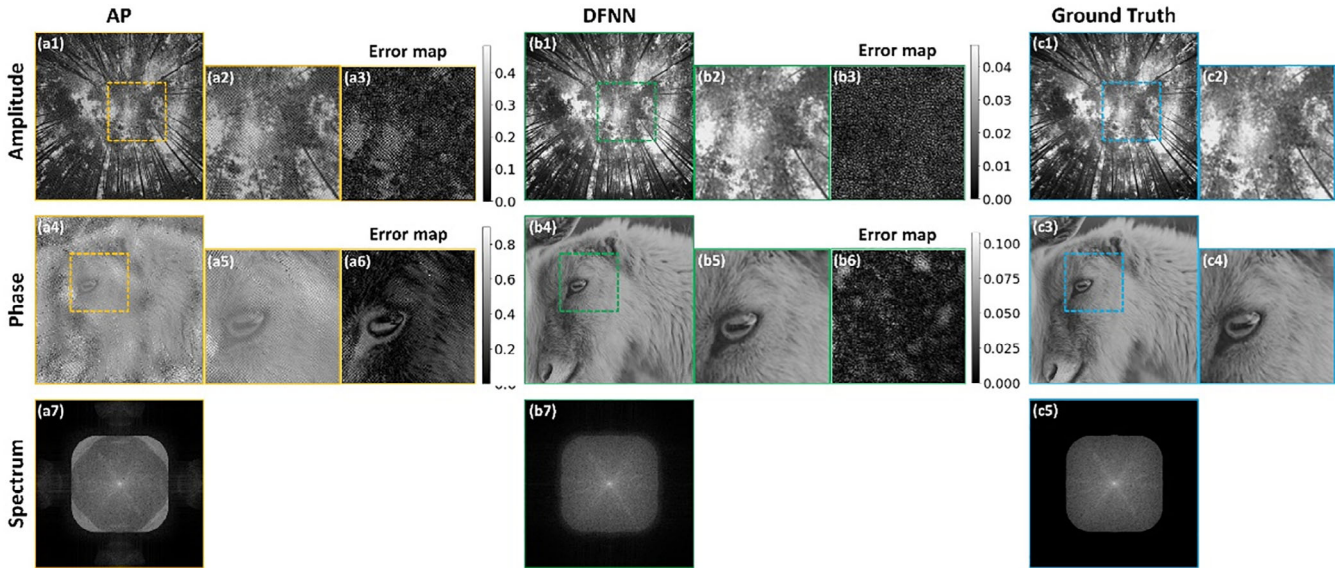


FIGURE 4 The reconstruction results of two methods at the maximum noise level (8×10^{-3}). Inserts the detailed features and error maps with the ground truth

TABLE 1 The comparison of reconstructed spectra

Methods	AP	DFNN	Ground Truth
NMSE[a] ($\times e^{-3}$)	37.1	1.9	0.0

Abbreviations: AP, alternative projection; NMSE, normalized mean square error.

[a]NMSE is calculated over the center 256 × 256 pixels area which is enough overlapped.

element 4 (2.76 μm), therefore the maximum resolving limit of the reconstructed results in both X and Y direction should be $2.76/(\mathrm{NA_{syn}}/\mathrm{NA_{obj}}) = 0.745\ \mu m$, which is slightly lower than the line width of group 9 element 3 (0.775 μm).

As shown in Figure 5D, DFNN can successfully reconstruct the fringe of group 9, element 2–3 to the theoretical limit with fine contrast, therefore, it can be concluded that DFNN can surely improve the resolution to the desired level which also demonstrates the resolution improvement ability of the proposed network. Furthermore, since the morphological information of the USAF target is completely different from the simulated training dataset, this reconstruction effect also, to a certain extent, verifies the generalizability of DFNN as mentioned in section 3.

Due to the sufficient overlap rate of the dataset in the frequency domain, AP method could obtain fine result with good contrast at the theoretical limit (Figure 5A2 and Figure 5D) [1, 2]. However, the reconstructed result obtained by AP method suffers from serious background noise (Figure 5A). In contrast, as shown in Figure 5B2 and Figure 5D, since the DCNN structure employed in DFNN could capture the invariants while filtering out other random fluctuations [45, 46], DFNN can achieve a better result with minimal background noise and slightly higher contrast which further verifies the robustness of DFNN against noise.

Meanwhile, as shown in Table 2. due to the end-to-end structure and GPU acceleration technology employed, DFNN can achieve a reconstruction speed about 50× faster than iterative-based AP method.

4.2 | Testing the performance on biological samples for large-FOV reconstruction

In this section, we utilize the open-sourced experimental dataset [17, 18] (stained Human Bone Osteosarcoma Epithelial U2OS sample captured on a Nikon TE300 inverted microscope) to test the generalizability and reconstruction performance of DFNN in terms of both amplitude and phase. The system parameters are the same as those in section 3, therefore, we could directly apply the DFNN trained based on simulation to reconstruct the U2OS sample. In order to better illustrate the advantage of DFNN in reconstructing large FOV result, we choose the region with the size of 1800 × 1800 as input, which is equivalent to 1.44 mm × 1.44 mm at the object plane. Meanwhile, we compare the result with traditional AP method in terms of both the reconstruction quality and speed. As for the iteration times of AP method, we employ the adaptive step strategy [4] to make the algorithm automatically judge whether the result is

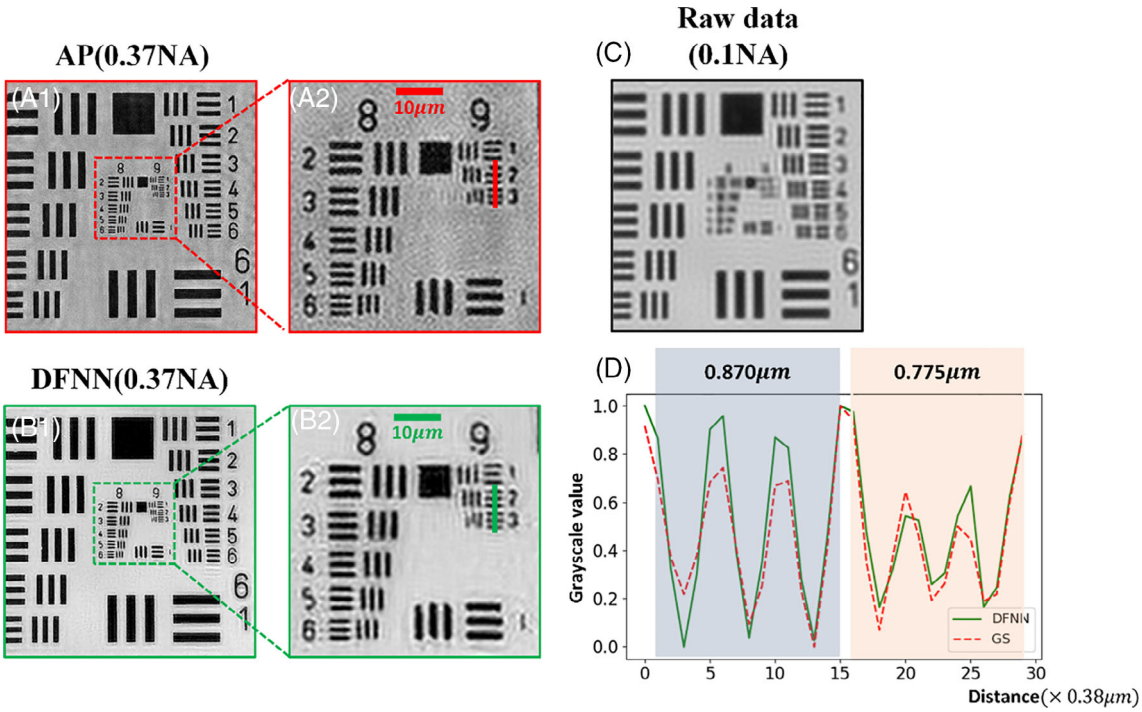


FIGURE 5 The comparison of reconstructed results between different methods using USAF dataset. A,B, The reconstructed results by alternative projection (AP) and DFNN respectively. C, the low-resolution image captured under the illumination of the center LED. D, the fringe contrast of group 9, element 2–3 obtained by AP and DFNN respectively

TABLE 2 The comparison of reconstruction time for USAF sample

Methods	Iteration times	Time
AP[a]	15	2.01 s
DFNN[b]	0	35.75 ms

Abbreviation: AP, alternative projection.

[a]The code of AP is open-sourced and provided in Ref. 1 and implemented with MATLAB.

[b]DFNN is implemented based on Python and tested on the same PC as AP.

converged and jump out of the iterative process. The pupil function used in AP method is set to CTF. The comparison is shown in Figure 6.

In Figure 6, in order to better illustrate the reconstruction performance of these two algorithms, we randomly zoom in and demonstrate seven different sub-regions scattered from the center to the edge FOV. Figure 6A illustrates the raw image captured under the normal incidence. It can be seen from the comparison of Figure 6B-E that the main advantage of DFNN over traditional AP method for this dataset lies in the phase reconstruction. All 7 sub-regions of the phase reconstructed by DFNN show better contrast, lower fluctuation and much clearer details than those by AP method.

Moreover, in Figure 7, we show the contrast curve of the plasmodesmata feature indicated in the sub-region of Figure 6C,E, DFNN can successfully reconstruct this detailed information with higher contrast than traditional AP method. Furthermore, in Figure 7B,C, we demonstrate the background fluctuation at the upper right corner of the phase results shown in Figure 6C,E, from the 3D surface graph, it can be seen that the peak-to-valley value (PV) of the background obtained by AP method is about 0.27; however, the PV value of the background by DFNN is only about 0.08.

As for the reconstruction speed of two methods, we show the time comparison in Table 3. Since the adaptive strategy used in AP method, it will automatically stop after 12 iterations. At the same time, due to the large reconstruction image size, the advantage of the DL-based DFNN in reconstruction speed is fully revealed, the time consumption is only 0.1125s which is about 1.5×10^3 faster than AP method.

It is worth noting that such phenomenon of large background fluctuation appeared in the reconstructed phase of AP method (Figure 7B) could be explained by the effect of the inaccurate wave vector [47, 48]. Since the system uses a point light source (LED) for illumination, the incident light received by the sample is strictly a spherical wave, therefore, the wave vector corresponding to the edge FOV is different from the center. For the

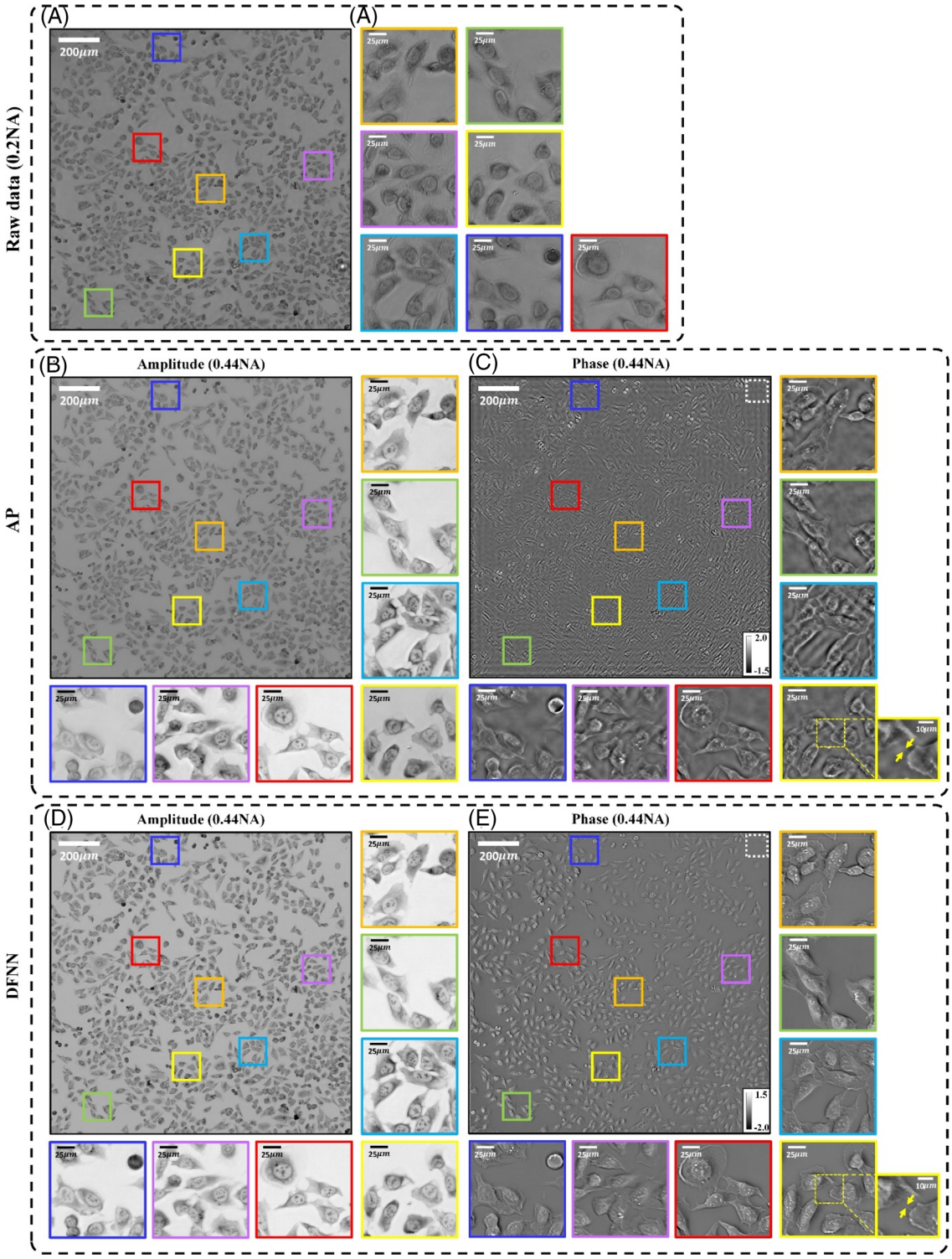


FIGURE 6 The comparison of the large field-of-view (FOV) reconstruction results using the open-sourced dataset (U2OS). A, the large-FOV raw image captured under the normal incidence and the seven randomly selected enlarged sub-regions. B-E, The reconstructed complex amplitudes of alternative projection (AP) and DFNN. Inserted with 7 randomly selected enlarged sub-regions

U2OS dataset used above, the wave vector angle from the center LED to the edge FOV is about 1.75 degree different from the center FOV which leads to a total deviation of 182 pixels in the amount of movement on the spectrum (91 pixels in the X direction, 91 pixels in the Y direction). Since the large FOV is reconstructed at the same time, the wave vector corresponding to the center FOV is used, therefore, the inaccuracy of the wave vector at the edge FOV will iteratively contaminate the reconstruction result until the AP algorithm converges [47, 48]. As

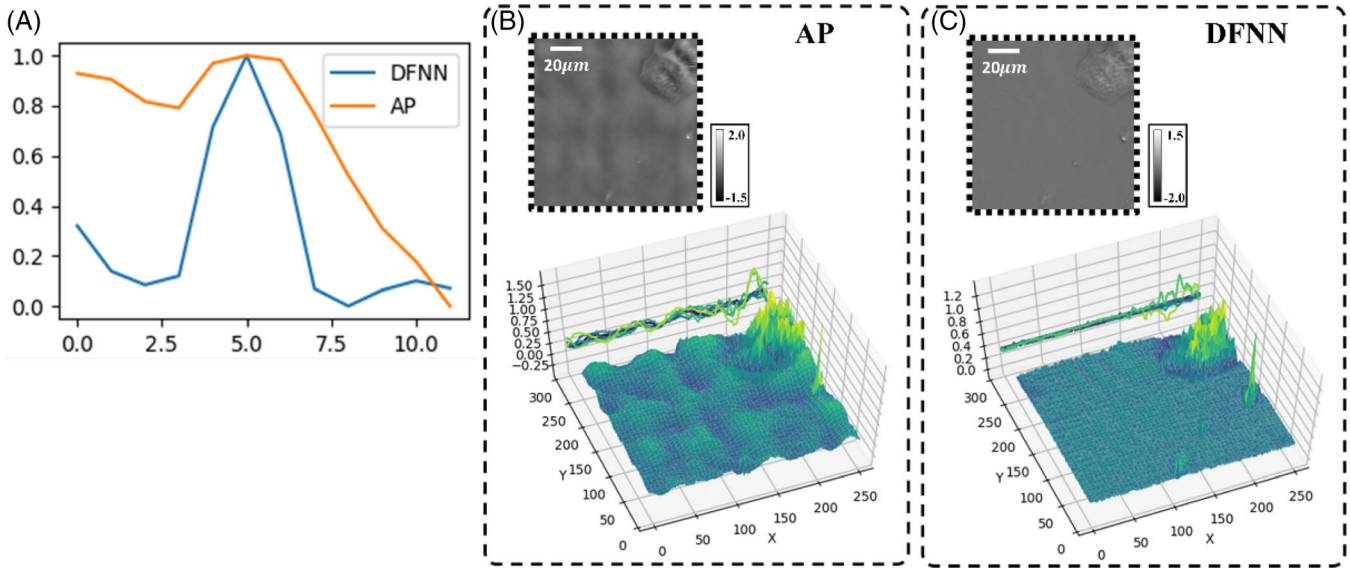


FIGURE 7 A, The contrast curve of the phase results across the plasmodesmata information which is drawn after normalize into 0 ~ 1. B, C, The enlarged sub-regions of the upper-right corner of the phase results along with the 3D surface graphs

TABLE 3 The comparison of reconstruction time in single large-FOV reconstruction

Methods	Iteration times	Time(s)
AP	12	167.5
DFNN	0	0.1125

shown in Figure 8, the comparison between the reconstructed sub-regions at the edge FOV using the wrong wave vector and the correct one. The sub-region is the same as shown in Figure 7B,C, and Figure 8A is the copy of Figure 7B, in which the wave vector is inaccurate for this sub-region. Figure 8B is the reconstruction result using the correct wave vector, it can be seen that the background fluctuation is effectively reduced.

In general, to ensure the accuracy of the wave vector, traditional AP method can only reconstruct a small FOV at once, and then stitch multiple small FOV results together using image fusion technology such as the alpha-blending stitching method [1]. However, the uneven brightness of multiple reconstruction results will more or less affect the stitching result, therefore, sufficient overlap of adjacent small portions is required [1, 21, 32] to ensure the fusion quality which further introduces additional calculations and makes the method less practical. As for the DL-based DFNN method, since there is no iterative process and the DCNN structure used captures the invariants while filtering out other random fluctuations [45, 46], the wave vector error has little influence on the reconstruction results leading to a smoother background and higher contrast as shown in Figure 7. Meanwhile, the substantial increase in reconstruction speed also makes it possible to perform real-time single full-FOV reconstruction.

To further test the reconstruction performance and the generalization property of DFNN on different samples, we use the open-sourced unstained vitro-Hela dataset to test the reconstruction performance. Noted that this sample could be treated as a phase object which contains significant intensity differences from the previous one [32]. The system parameters are the same as before, however, here we employ 11×11 images to increase the synthesis NA to 0.58 which leads to the upsample factor of 3 ($\alpha = 3$). And the channel number C is set to 128. The training dataset is still obtained based on the simulation dataset (DIV2K) and the training strategy is similar as before. We enlarge two sub-regions located at the center and edge of the FOV and compare the results with those obtained by AP method. To ensure the accuracy of wave vectors, we use AP method to reconstruct the small sub-regions separately with the corresponding wave vectors. The comparison is shown in Figure 9.

The reconstructed result shown in Figure 9A has the size of 0.77 $mm \times 0.77$ mm at the object plane due to the CUDA memory limitation. Figure 9C1,C2 represent the contrast of two reconstructed details indicated in Figure 9A3,A4 and Figure 9B3,B4, it can be visually seen that the results obtained by DFNN show higher contrast.

Moreover, such generalization property over experimental dataset could also be explained by the Shannon entropy of the training dataset, as mentioned in [33], a trained DCNN can have a better generalization

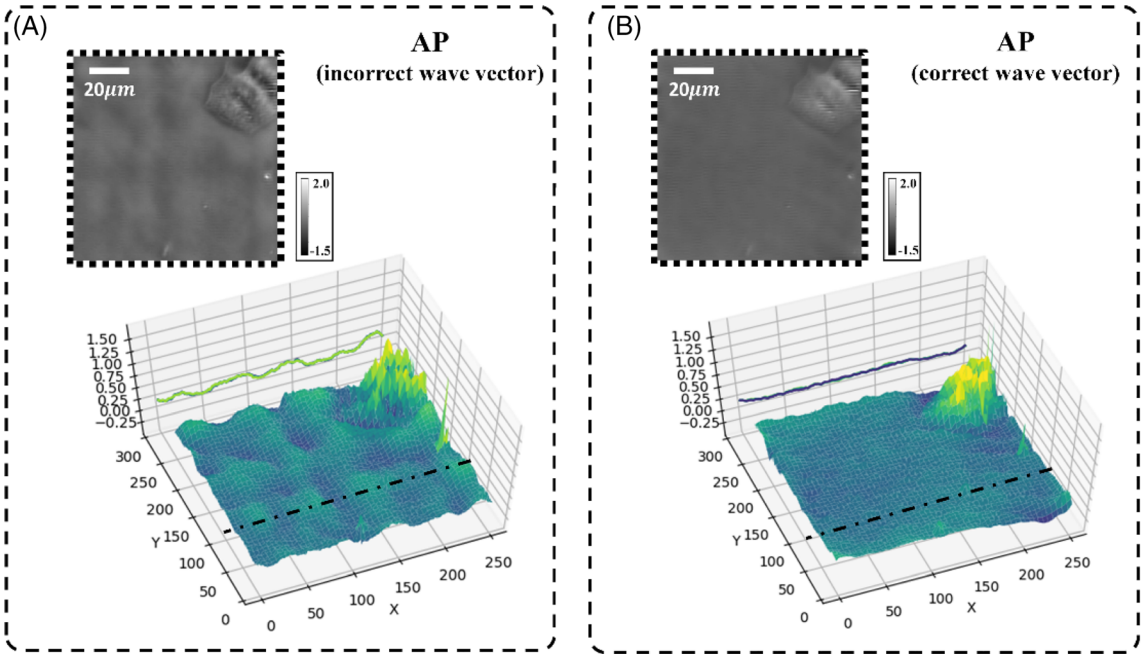


FIGURE 8 The comparison of the edge sub-region reconstructed by alternative projection (AP) method with the wrong wave vector and the correct one. A, The reconstructed phase of the sub-region with the wrong wave vector which is the same as shown in Figure 7B. B, The reconstructed phase of the sub-region with the correct wave vector in which the background fluctuation is smaller than the other one

performance when the training dataset is more general and high-entropy. Compare with the strategy of using the open-sourced dataset (DIV2K) to generate the training dataset as mentioned in section 3, other networks for similar reconstruction task usually employ the experimental dataset to train the networks and the high-resolution ground-truth images are typically generated by other iterative-based algorithms such as AP method [31, 32]. Here, we compute the Shannon entropy of the images in DIV2K and the reconstructed results of the experimental dataset (U2OS) obtained by AP method and show their histograms in Figure 10.

Noted that to ensure the ground-truth images in two strategies have the same size (512 × 512), the large format raw image of U2OS (2560 × 2160) is first divided into 300 portions with the size of 256 × 256 (125-pixel and 136-pixel overlap in the vertical and horizontal directions) and then reconstructed into 512 × 512 by AP method using the correct wave vectors, the reconstructed results are divided into amplitude and phase and computed separately.

It can be seen that most images in DIV2K dataset have their entropy lie in 7.0 to 8.0 (mean 7.382, standard deviation 0.423), which is larger than the reconstructed results of U2OS by AP method (mean 5.787, SD 0.478 for amplitude; mean 5.352, SD 0.535 for phase). Therefore, due to the "higher-entropy" property of the simulated dataset we use, DFNN could show better generalizability than the networks only trained based on experimental dataset and obtain fine reconstruction results of various samples with complete different morphological features.

5 | CONCLUSION AND DISCUSSION

In this paper, we have proposed and demonstrated a DL-based method called DFNN for rapid FPM reconstruction. By separating the network data flow into two branches, DFNN can simultaneously achieve the high-resolution amplitude and phase information without crosstalk. Each branch consists of two data flows with multiple residual blocks which utilize the structure from Reference 32 for better performance. In order to quantitatively evaluate the feasibility and the reconstruction quality of DFNN, we carry out a simulation using the DIV2K dataset. In addition, by adding different levels of noise, we demonstrate that the proposed DFNN model has stronger robustness than traditional iterative-based AP method. The superior of robustness is also illustrated through the experiment on actual USAF dataset.

We further discuss the importance of the accuracy of the wave vector to traditional AP method, in order to obtain fine large-FOV reconstruction results, traditional AP method needs to separate the full-FOV image into several portions and reconstruct them separately using the accurate wave vectors, and then stitch the results together using image fusion technology such as alpha-

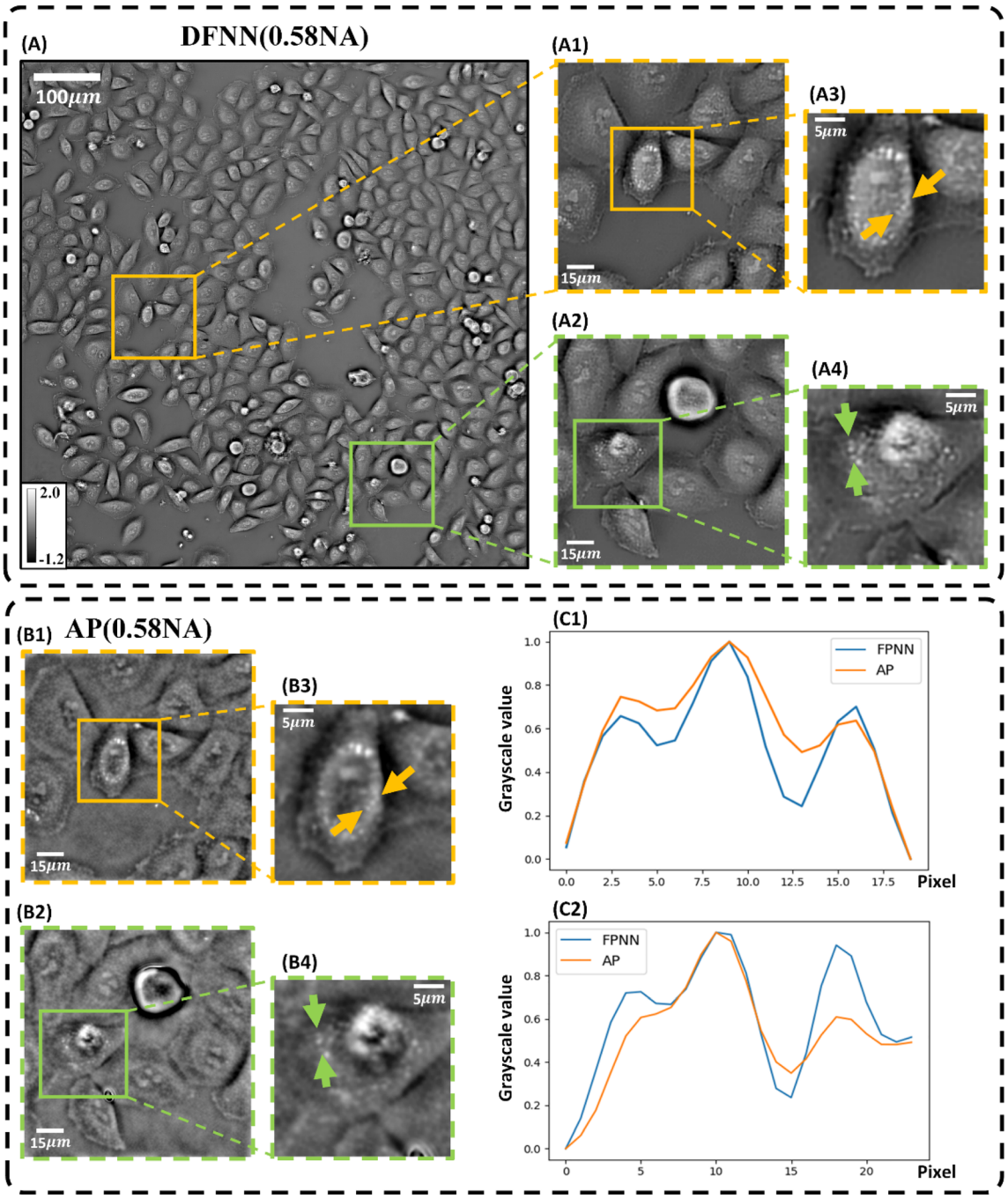


FIGURE 9 The comparison of the center and edge sub-regions reconstructed by DFNN and alternative projection (AP) methods. A, The large field-of-view (FOV) reconstruction result obtained by DFNN with the size of 2880 × 2880 corresponding to 0.77 mm × 0.77 mm at the object plane. A1-A4, The enlarged two sub-regions located at the center and edge FOV. B1-B4, the reconstruction results of the two sub-regions using AP method which will automatically stop the iteration after 12, respectively. C1, the contrast curve of the detail structures indicated in, A3 and B3 which is drawn after normalized into 0 ~ 1. C2, the contrast curve of the detail structures indicated in, A4 and B4 which is drawn after normalized into 0 ~ 1

blending strategy [1], furthermore, in order to ensure the quality of the fusion results, sufficient overlap between adjacent portions is required. As a result, extra calculation will be introduced leading to a lower reconstruction speed. In addition, we demonstrate the performance of DFNN and traditional AP method in large FOV reconstruction. Due to the end-to-end structure, no iterative process is needed for DFNN method. Moreover, since the model could capture the invariants while filtering out other random fluctuations [45, 46], the sensitivity to the wave vector deviations could be greatly reduced. Therefore, DFNN could obtain better reconstruction quality

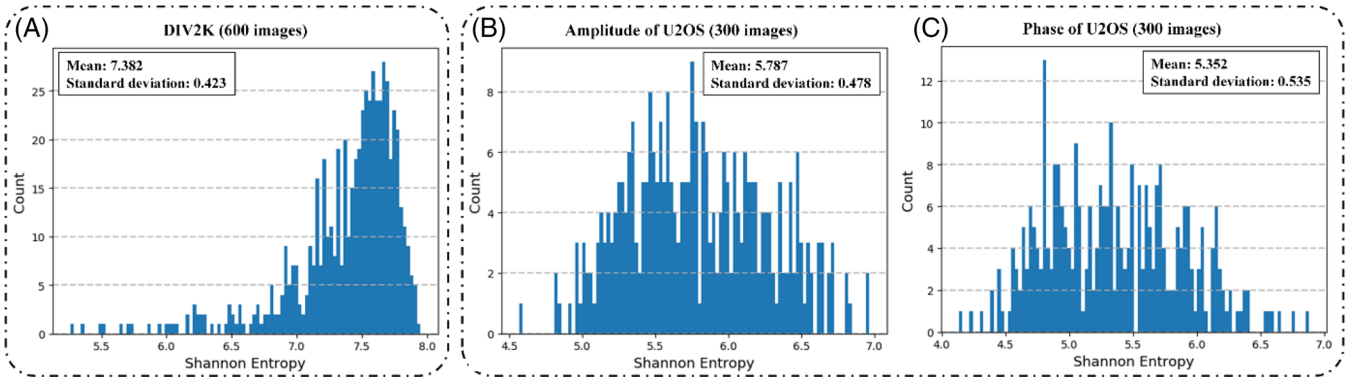


FIGURE 10 The entropy histogram of DIV2K and the reconstructed results of U2OS by AP method. A, The entropy histogram of DIV2K computed based on 100 bins and 600 images. B1, B2, The entropy histograms of the reconstructed amplitude and phase of U2OS dataset by alternative projection (AP) method and computed based on 100 bins and 300 images separately

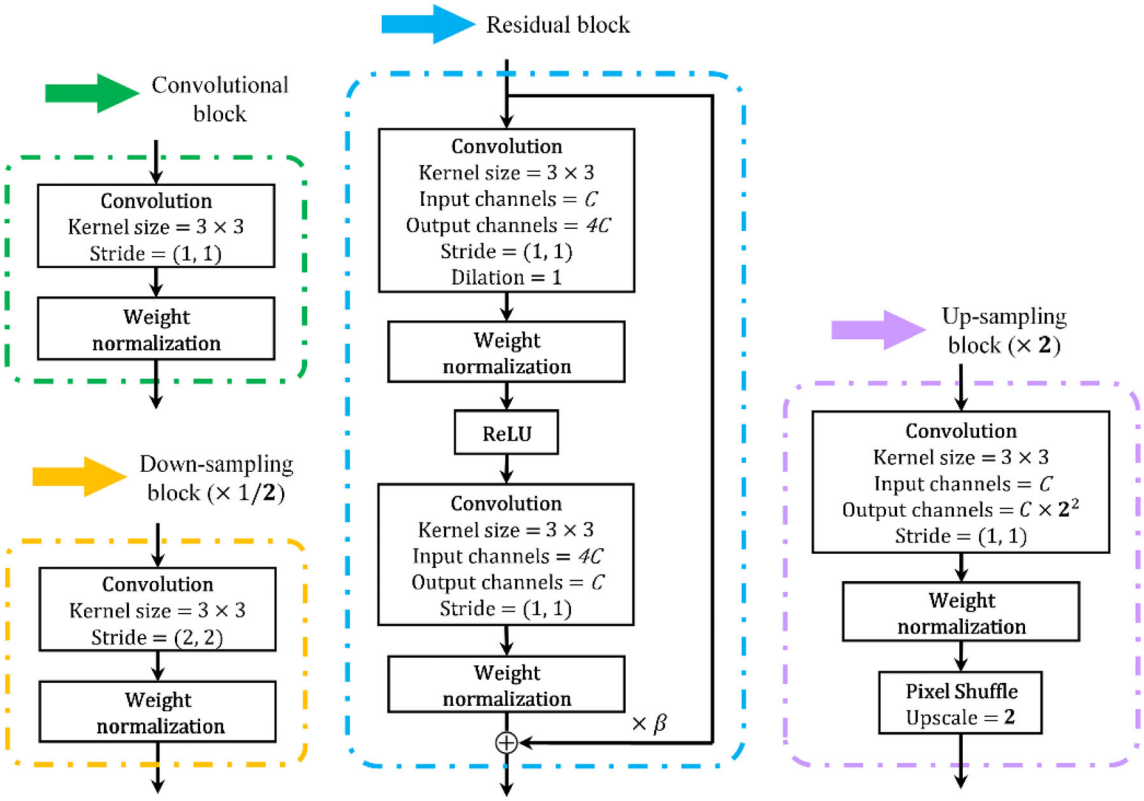


FIGURE 11 The detailed information of each function block in DFNN

than AP method especially at the edge FOVs. Meanwhile, high-throughput reconstruction could better reveal the advantage of the DL-based DFNN in reconstruction speed, which results in about 1500 times faster. In section 4, we show that the Shannon entropy of our training dataset is higher than the dataset generated from the experimental images which is typically used in other networks with similar task [31, 32]. And according to [33], by training based on a "high-entropy" dataset, the generalizability of the network could be improved, which is further be verified by the testing on experimental samples in sections 3 and 4.

It is worth noting that the temporal resolution of FPM technology does not mainly rely on the reconstruction speed but on the raw-data acquisition speed, however, in order to achieve real-time full-FOV FPM monitoring, the time consumption of full-FOV reconstruction for single frame should be less than the time consumption of data acquisition. In recent years, since the introduction of multi-LEDs illumination strategy, the

single frame acquisition speed of FPM could be above 1 Hz [18]. Therefore, it becomes vital to reduce the time consumption of full-FOV reconstruction to less than 1 s while ensuring the quality, which makes the DL-based reconstruction method become promising. Moreover, we could combine the deep learning strategy with image acquisition, and through training, we can reasonably reduce the data acquisition time while maintaining the reconstruction quality, which will be our future work.

6 | MORE DETAILS OF DFNN

In section 3.1, we introduce the general architecture of one flow in DFNN (Figure 1). Here, in Figure 11, we provide the detailed lay-wise information of each function block in Figure 1, along with the parameter setting. It can be seen that each convolution layer is followed by a weight normalization operation which could accelerate the training convergence of DFNN [42]. In the residual block, we employ the strategy mentioned in [36] which expands the channel number in the first convolution layer while restores the channel number in the second convolution layer. Noted that in order to increase the receptive field [35] of the block without changing the size of the tensor, we use dilated convolution in the first convolution layer (Dilation = 1). In the side-branch of the residual block, there is a parameter β which could modulate the retention of the input tensor in the output result and is set to 1.0. It is worth noting that in the last $\times\alpha$ upsampling block shown in Figure 1A, the output channels in the convolution layer should be $32 \times \alpha^2$ and the upscale in the pixel-shuffle layer is α.

ACKNOWLEDGMENTS

We acknowledge CAS Interdisciplinary Innovation Team for supporting this work; We sincerely acknowledge the open-sourced dataset of USAF provided by Zuo et al [4] and the open-sourced dataset of U2OS and unstained Hela cells provided by Tian et al [17]. We also acknowledge the DIV2K dataset provided by NTIRE challenges.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Github at https://github.com/MLWise112358/DFNN_for_FPM_reconstruction.

ORCID

Minglu Sun https://orcid.org/0000-0001-8952-9995
Dayu Li https://orcid.org/0000-0002-6535-0285

REFERENCES

[1] G. Zheng, R. Horstmeyer, C. Yang, *Nat. Photonics* **2013**, *7*(9), 739.
[2] G. Zheng, *IEEE Photonics J.* **2014**, *6*, 1.
[3] K. Guo, S. Dong, G. Zheng, *IEEE J. Sel. Top. Quantum Electron.* **2016**, *22*(4), 77.
[4] C. Zuo, J. Sun, Q. Chen, *Opt. Express* **2016**, *24*(18), 20724.
[5] Y. Fan, J. Sun, Q. Chen, M. Wang, C. Zuo, *Opt. Commun.* **2017**, *404*, 23.
[6] Y. Zhang, P. Song, Q. Dai, *Opt. Express* **2017**, *25*(9), 168.
[7] L. Hou, H. Wang, M. Sticker, L. Stoppe, J. Wang, M. Xu, *Appl. Optics* **2018**, *57*(7), 1575.
[8] S. Chen, T. Xu, J. Zhang, X. Wang, Y. Zhang, *IEEE Photonics J.* **2019**, *11*(1), 1.
[9] M. Sun, X. Chen, Y. Zhu, D. Li, Q. Mu, L. Xuan, *Opt. Express* **2019**, *27*(17), 24161.
[10] X. Ou, G. Zheng, C. Yang, *Opt. Express* **2014**, *22*(5), 4960.
[11] Y. Zhang, W. Jiang, Q. Dai, *Opt. Express* **2015**, *23*(26), 33822.
[12] L. H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, L. Waller, *Opt. Express* **2015**, *23*(26), 33214.
[13] A. Zhou, W. Wang, N. Chen, E. Y. Lam, B. Lee, G. Situ, *Opt. Express* **2018**, *26*(18), 23661.
[14] L. Bian, J. Suo, G. Zheng, K. Guo, F. Chen, Q. Dai, *Opt. Express* **2015**, *23*(4), 4856.
[15] R. Horstmeyer, R. Y. Chen, X. Ou, B. Ames, J. A. Tropp, C. Yang, *New J. Phys.* **2014**, *17*(5), 053044.
[16] A. Pan, Y. Zhang, T. Zhao, Z. Wang, D. Dan, M. Lei, B. Yao, *J. Biomed. Opt.* **2017**, *22*(9), 1.
[17] L. Tian, X. Li, K. Ramchandran, L. Waller, *Biomed. Opt. Express* **2014**, *5*(7), 2376.
[18] L. Tian, Z. Liu, L.-H. Yeh, M. Chen, J. Zhong, L. Waller, *Optica* **2015**, *2*, 904.
[19] S. Dong, R. Shiradkar, P. Nanda, G. Zheng, *Biomed. Opt. Express* **2014**, *5*, 1757.
[20] Y. Zhang, W. Jiang, L. Tian, L. Waller, Q. Dai, *Opt. Express* **2015**, *23*(14), 18471.
[21] J. Sun, C. Zuo, L. Zhang, Q. Chen, *Sci. Rep.* **2017**, *7*(1), 1187.
[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, *IEEE Trans. Image Process.* **2017**, *26*(7), 3142.
[23] C. Dong, C. C. Loy, K. He, X. Tang, *IEEE Trans. Pattern Anal. Machine Intell.* **2015**, *38*(2), 295.
[24] J. Kim, J. K. Lee, and K. M. Lee, Accurate Image Super-Resolution Using Very Deep Convolutional Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1646–1654 (**2016**).
[25] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 105–114 (**2017**).
[26] Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, A. Ozcan, *Light. Sci. Appl.* **2017**, *1*, 1.
[27] J. Zhang, T. Xu, Z. Shen, Y. Qiao, Y. Zhang, *Opt. Express* **2019**, *27*(6), 8612.

[28] F. Shamshad, F. Abbas, A. Ahmed, Deep ptych: Subsampled Fourier ptychography using generative priors, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (**2018**).
[29] L. Boominathan, M. Mainiparambil, H. Gupta, R. Baburaj, K. Mitra, *arXiv* **2018**, *1805*, 3593.
[30] A. Kappeler, S. Ghosh, J. Holloway, O. Cossairt, and A. Katsaggelos, "Ptychnet: CNN based fourier ptychography, IEEE International Conference on Image Processing (ICIP), 1712–1716 (**2017**).
[31] Y. F. Cheng, M. Strachan, Z. Weiss, M. Deb, D. Carone, V. Ganapati, *arXiv* **2018**, *1810*, 3481.
[32] T. Nguyen, Y. Xue, Y. Li, L. Tian, G. Nehmetallah, *Opt. Express* **2018**, *26*, 26470.
[33] M. Deng, S. Li, Z. Zhang, I. Kang, N. X. Fang, G. Barbastathis, *Opt. Express* **2020**, *28*, 24152.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778 (**2016**).
[35] W. Luo, Y. Li, R. Urtasun, R. Zemel, *arXiv* **2017**, *1701*, 4128.
[36] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, T. Huang, *arXiv* **2018**, *1808*, 8718.
[37] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2160–7516 (**2017**).
[38] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert and Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1874–1883 (**2016**).
[39] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, *IEEE Trans. Med. Imaging* **2004**, *13*(4), 600.
[40] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1122–1131 (**2017**).
[41] D. P. Kingma, J. Ba, *arXiv* **2014**, *1412*, 6980.
[42] T. Salimans, D. P. Kingma, *arXiv* **2016**, *1602*, 7868.
[43] M. Kellman, E. Bostan, M. Chen, L. Waller, *arXiv* **2019**, *1904*, 4175.
[44] L. Tian, L. Waller, *Opt. Express* **2015**, *23*, 11394.
[45] Y. Bengio, A. Courville, P. Vincent, *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*(8), 1798.
[46] M. D. Zeiler, R. Fergus, *Visualizing and Understanding Convolutional Networks, European Conference on Computer Vision*, Springer, Switzerland **2014**, p. 818.
[47] R. Eckert, Z. F. Phillips, L. Waller, *Appl. Optics* **2018**, *57*(19), 5434.
[48] A. Zhuo, W. Wang, N. Chen, E. Y. Lam, B. Lee, G. Situ, *arXiv* **2018**, *1803*, 0395.

How to cite this article: Sun M, Shao L, Zhu Y, et al. Double-flow convolutional neural network for rapid large field of view Fourier ptychographic reconstruction. *J. Biophotonics*. 2021;e202000444. https://doi.org/10.1002/jbio.202000444