Article

Deep Multi-Feature Transfer Network for Fourier Ptychographic Microscopy Imaging Reconstruction [†]

Xiaoli Wang ^{1,2}, Yan Piao ^{1,*}, Jinyang Yu ², Jie Li ^{2,*}, Haixin Sun ², Yuanshang Jin ², Limin Liu ² and Tingfa Xu ³

¹ Information and Communication Engineering, Electronics Information Engineering College, Changchun University of Science and Technology, Changchun 130022, China; wangxl@ccu.edu.cn

² Electrical and Electronic Teaching Center, Electronics Information Engineering College, Changchun University, Changchun 130022, China; 190401067@mails.ccu.edu.cn (J.Y.); haixin_s@hotmail.com (H.S.); 200401080@mails.ccu.edu.cn (Y.J.); 200401097@mails.ccu.edu.cn (L.L.)

³ School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; ciom_xtf1@bit.edu.cn

* Correspondence: piaoyan@cust.edu.cn (Y.P.); lij69@ccu.edu.cn (J.L.)

[†] This paper is an extended version of Jinyang Yu, Jie Li, Xiaoli Wang, Jizhou Zhang, Limin Liu and Yuanshang Jin. Microscopy image reconstruction method based on convolution network feature fusion. In Proceedings of the 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 23–26 September 2021.

Abstract: Fourier ptychographic microscopy (FPM) is a potential imaging technique, which is used to achieve wide field-of-view (FOV), high-resolution and quantitative phase information. The LED array is used to irradiate the samples from different angles to obtain the corresponding low-resolution intensity images. However, the performance of reconstruction still suffers from noise and image data redundancy, which needs to be considered. In this paper, we present a novel Fourier ptychographic microscopy imaging reconstruction method based on a deep multi-feature transfer network, which can achieve good anti-noise performance and realize high-resolution reconstruction with reduced image data. First, in this paper, the image features are deeply extracted through transfer learning ResNet50, Xception and DenseNet121 networks, and utilize the complementarity of deep multiple features and adopt cascaded feature fusion strategy for channel merging to improve the quality of image reconstruction; then the pre-upsampling is used to reconstruct the network to improve the texture details of the high-resolution reconstructed image. We validate the performance of the reported method via both simulation and experiment. The model has good robustness to noise and blurred images. Better reconstruction results are obtained under the conditions of short time and low resolution. We hope that the end-to-end mapping method of neural network can provide a neural-network perspective to solve the FPM reconstruction.

Citation: Wang, X.; Piao, Y.; Yu, J.; Li, J.; Sun, H.; Jin, Y.; Liu, L.; Xu, T. Deep Multi-Feature Transfer Network for Fourier Ptychographic Microscopy Imaging Reconstruction. *Sensors* **2022**, *22*, 1237. <https://doi.org/10.3390/s22031237>

Academic Editor: Alessandro Artusi

Received: 10 December 2021

Accepted: 2 February 2022

Published: 6 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Fourier ptychographic microscopy; transfer learning; neural network; deep multi-feature

1. Introduction

The development of microscopic imaging technology has opened the door for human beings to explore the microscopic world, microscope has become an indispensable tool in the field of life sciences, and every progress in microscopic imaging technology has promoted the development of various fields. Although the microscopic imaging technology has made great progress in recent years, the imaging mechanism of its technology has not fundamentally changed. It is still based on the imaging principle of traditional lens, that is, the imaging mode of what is seen and what is obtained. Although the traditional imaging mechanism is simple and easy to implement, it still faces many bottleneck problems. It is difficult to solve the challenges brought by the new application requirements [1,2]. Fourier ptychographic microscopy is an imaging technology with large field of view, high resolution and quantitative phase calculation developed in recent years [3–5]. This technology uses an optical microscope and LED lighting array [6], which combines phase retrieval algorithm to

recover the high-resolution intensity and phase of the sample [7,8]. Fourier ptychographic microscopy is based on the intensity information acquired in the spatial domain and a fixed mapping relationship in the frequency domain to iterate alternately. In the spatial domain, the amplitude is constrained by the low-resolution image, and in the frequency domain, the spectrum support domain is constrained by the CTF (Coherent Transfer Function). It overcomes the problems brought by the traditional lens imaging mechanism and solves the contradiction between resolution and large field of view from the perspective of calculation. Moreover, the phase retrieval mechanism can not only reconstruct the intensity image, but also generate the quantitative phase image of the sample at the same time.

Since the FPM was proposed, various methods have been proposed to improve the reconstruction performance and reduce the computational complexity of FPM. In terms of system setting, pupil aberration correction [9,10], LED misalignment correction [11,12] and more efficient LED module design [13,14] are proposed to improve reconstruction quality. In terms of reconstruction methods, multiplexing strategy is introduced [15,16] to accelerate the image acquisition process, and optimization theories such as Wirtinger-flow, adaptive step-size strategy, and convex relaxation [17] are proposed to improve reconstruction speed and robustness. Recently, with the rapid development of convolutional neural networks, people have introduced a convolutional neural network (CNN) into the FPM reconstruction process and can learn the mapping from low resolution intensity input to high resolution complex output to solve the problem of FPM reconstruction. Jiang et al. [18] modeled FPM through convolutional neural network and reconstructed FPM through back propagation. Zhang et al. [19] introduced Zernike aberration recovery and total variation constraints based on neural network to ensure the quality of FPM reconstruction and at the same time to correct aberration. The above work is still based on iterative algorithms and does not give full play to the advantages of deep learning. Deep learning finds the linear mapping between input and output by building Deep Neural Networks (DNN) [20,21], and obtains the weight information of Deep Neural Networks in the case of mass data training to solve the reconstruction quality problem. In recent years, DNN-based methods have achieved good results in image reconstruction [22–25] and phase recovery [26,27], and some researchers have proposed the FPM reconstruction method using DNN [28–30], which performs well in high-resolution image reconstruction. One of the authors of this paper proposed a fmccnn algorithm at the 2021 International Conference on electronic information engineering and computer science. This algorithm is a Fourier superposition micro reconstruction method based on multi convolution network feature fusion, which realizes image reconstruction from low resolution to high resolution. Additionally, it constantly adjusts the super parameters of the network model to make it have better reconstruction effect and robustness to noisy images [31].

In this paper, we focus on solving the FPM reconstruction quality problem by deep learning. Using ResNet50 [32], Xception [33] and Densenet121 [34] transfer learning networks to extract image features. Cascaded feature fusion strategy and pre-upsampling reconstruction network are used to build deep multi-feature transfer network (DMFTN) for Fourier ptychographic microscopy imaging model to solve the problem of FPM reconstruction. DMFTN fully extracts the features at each level in the input original low-resolution image, preserves and amplifies the information features, which makes the information transmission more complete and efficient. DMFTN uses the complementarity of the extracted features of different migratory learning networks to increase the redundancy of the extracted feature information and better reconstructs the high-resolution image, and adopts the pre-upsampling reconstruction. DMFTN is capable of reconstructing high-resolution complex amplitudes of many types of samples. In order to train the reconstruction network, the acquired FPM intensity images are synthesized into two-channel complex amplitudes such as network model inputs, and the reconstructed outputs of network model outputs are also two-channel complex amplitudes. The 400 animal tissue images were acquired and used to build a simulation dataset through the FPM imaging model. The simulation

data set is divided into training data set and test data set to train and test the proposed network model.

2. Related Works

The FPM consists of two main components, the imaging model and the reconstruction model. The imaging model is an objective description of the imaging acquisition process, representing the process of plane waves passing through the microscope and imaged on the sensor. The reconstruction model is the reconstruction of the high-resolution complex amplitude process using the low-resolution image acquired by the sensor combined with the phase recovery algorithm.

2.1. Imaging Model of Fourier Ptychographic Microscopic

FPM is a computational method that combines the original low-resolution data into high-resolution and wide-field images. The imaging process is shown in Figure 1.

Figure 1. Schematic diagram of FPM imaging model.

In the process of imaging, the objective lens and cylinder lens in the system carry out two Fourier transforms of the light waves of the object in turn. First of all, the sample function is expressed by the light emitted by LED. When the light emitted by the sample is illuminated on the sample, the light wave is equivalent to a monochromatic plane wave that is oblique incidence. The sample function interacts with the tilted plane waves $e^{ik_{xi}x}$, $e^{ik_{yi}y}$ to produce the radiation field $e(r) = o(x, y)e^{ik_{xi}x}e^{ik_{yi}y}$, where k_{xi} and k_{yi} are the wavenumber in the x and y directions, respectively. For the case where the n th LED is lit, the wave vector of the incident light can be expressed as:

$$k_n = \left(\frac{\sin \alpha_{xn}}{\lambda}, \frac{\sin \alpha_{yn}}{\lambda} \right) \quad (n = 1, 2, \dots, N_{LED}) \quad (1)$$

where $(\alpha_{xn}, \alpha_{yn})$ represents the incident angle of the illuminating light wave, λ represents the wavelength of the light, and n represents the number of LED. When $n = 1$, it indicates positive incidence. Assuming that the intensity of the incident light is 1 and the initial phase is 0, then the incident light can be expressed as $\exp(jk_n r)$ when the incident wave vector is k_n . After passing through the sample, the plane wave emits light $e(r) = o(x, y)e^{ik_{xi}x}e^{ik_{yi}y}$. The objective lens transforms the field by Fourier transform, so it is expressed as:

$$F\{e(r)\} = F\{o(x, y)e^{ik_{xi}x}e^{ik_{yi}y}\} = O(k - k_n) \quad (2)$$

This is equivalent to moving the original spectral center of the sample to the k_n position.

The spectrum is filtered by the coherent transfer function $H(k)$ of the objective lens, and the components of the spectrum that can be accepted by the optical system after filtering are:

$$G_n(k) = O(k - k_n)H(k) \quad (3)$$

The obtained spectrum is filtered by pupil aperture low-pass filter, and the cutoff frequency is $NA \cdot 2\pi/\lambda$, where NA is the numerical aperture and λ is the optical wavelength. Through coordinate transformation, Equation (3) can be written as follows:

$$G_n(k + k_n) = O(k)H(k + k_n) \quad (4)$$

Finally, the tube lens carries out the second Fourier transform of the filtered spectrum $O(x, y)$, and the sensor captures the final intensity image $I(x, y)$.

The outgoing light wave is imaged on the image plane and is received by the image sensor and converted into a digital signal.

$$I_{nc}(x, y) = |g_n(r)|^2 = \left| F^{-1}\{G_n(k + k_n)\} \right|^2 = \left| F^{-1}\{O(k)H(k + k_n)\} \right|^2 \quad (5)$$

where $g_n(r)$ is the complex amplitude of reaching the image plane. As the image sensor only responds to the intensity of light, the final image is $I_{nc}(r)$, and the subscript c represents the actual value collected.

2.2. Reconstruction Model of Fourier Ptychographic Microscopy

The reconstruction process of Fourier ptychographic microscopy combines the concepts of phase recovery and synthetic aperture. Phase recovery is first used to solve the problem of phase loss in electronic diffraction. In 1971, Gerchberg and Saxton first proposed their phase iterative recovery algorithm, so their phase iterative recovery algorithm is also known as the GS algorithm. It uses the Fourier transform relationship between the spatial domain and the frequency domain, and adds constraints in the spatial domain and the frequency domain to iterative alternately, so as to continuously approximate and converge to the real complex amplitude of the object. FPM is committed to iteratively reduce the difference between the low-resolution amplitude corresponding to the guessed object function and the actual low-resolution amplitude, which can be seen as a nonconvex optimization problem.

The concrete steps of the reconstruction process are:

1. Initial guess

At the beginning of the algorithm, it is necessary to initially guess the high-resolution object function and the coherent transfer function. In the case of sufficient data redundancy, FPM can converge no matter what kind of guess is used, such as complete guess, random guess, and the initial guess has no effect on the final result. In order to improve the convergence speed and reduce the number of iterations, the low-resolution image under normal incidence is often used as the initial strength guess of the object, and the initial phase guess of the object is generally initialized to zero. CTF carries on the initial guess according to the ideal CTF without aberration, as shown;

$$O_0^0(k) = F\left\{\sqrt{B(I_{1c}(r))} \exp(jk \cdot \varphi_0)\right\}H(k) \quad (6)$$

where $O_0^0(k)$ denotes the initial value of the sample spectrum. B denotes the bilinear interpolation of the image for amplification. φ_0 denotes the initialized phase, which is generally taken as 0.

2. Low resolution light field in computational imaging face

Under the illumination of LED at the corresponding angle, the light wave field emitted by the object is Fourier transformed and then passes through the low pass of the optical system. The light wave field after the low pass reaches the imaging surface through an inverse Fourier transform. The corresponding low-resolution estimated light field is

$$g_n^i(r) = F^{-1}\{G_n^i(k + k_n)\} = F^{-1}\{O_{n-1}^i(k)H(k + k_n)\} \quad (7)$$

where i represents the number of current iterations ($i = 1, 2, \dots, i_{\max}$). n represents the different incidence angles ($n = 1, 2, \dots, N_{\text{LED}}$). $O_{n-1}^i(k)$ denotes the sample spectrum after the i^{th} round update of $(n-1)^{\text{th}}$ sub-spectrum, and $O_0^i(k)$ represents the initial value of the i^{th} initial value of the spectrum updated in the first round, and $g_n^i(r)$ denotes the complex amplitude estimate corresponding to the n^{th} sub-spectrum in the i^{th} update round.

3. Update the amplitude of the light field

Keeping the phase information of the low resolution estimated light field unchanged, the corresponding low resolution intensity image is used to update its amplitude information to obtain the updated low resolution estimated light field:

$$\overline{g_n^i} = \sqrt{I_{\text{nc}}(r)} \frac{g_n^i(r)}{|g_n^i(r)|} \quad (8)$$

where $\overline{g_n^i}$ denotes the updated sample complex amplitude for the n^{th} illumination, and $I_{\text{nc}}(r)$ is the actual acquired low-resolution intensity image corresponding to the n^{th} illumination.

4. Spectrum of update function

The updated low-resolution estimated optical field is transformed into the frequency domain by Fourier transform $\overline{G_n^i}(k + k_n) = F\{\overline{g_n^i}\}$, and the object function spectrum in the subaperture is updated. Other regions remain unchanged:

$$O_n^i(k) = O_{n-1}^i(k)[1 - H(k + k_n)] + \overline{G_n^i}(k + k_n)H(k + k_n) \quad (9)$$

where $O_n^i(k)$ denotes the sample spectrum for the n^{th} sub-spectrum updated in the i^{th} round. The above formula indicates that only the part of the spectrum that is selected by $H(k + k_n)$ is updated.

5. Update all angle spectrum

Repeat steps (2)~(4) until the low-resolution image of all illumination angles is updated, which can be regarded as completing an iterative process.

6. Iteration to convergence

Repeat iteration steps (2)~(5) until the reconstruction algorithm converges, so as to obtain the high-resolution spectrum of the object, and then go back to the airspace by inverse Fourier transform to obtain the complex amplitude of the high-resolution object and complete the reconstruction process.

3. Proposed Method

In this paper, we proposed a network model of deep multi-feature transfer network (DMFTN) Fourier ptychographic microscopy image reconstruction methods built using deep learning neural networks, which can simultaneously acquire the high-resolution intensity and phase of the sample, improve the quality of the FPM reconstruction problem, and shorten the reconstruction time. The DMFTN model is built, as shown in Figure 2, including: network model input, migration learning ResNet50 [32], Xception [33] and Densenet121 [34], down sampling feature extraction of the network framework, cascaded feature fusion strategy and upsampling reconstruction network structure.

3.1. Network Model Input

In the FPM system, hundreds of original low-resolution images are obtained simultaneously by programmable control of hundreds of LED array light sources, so the input data in the corresponding neural network is a three-dimensional image tensor with hundreds of channels. To extract features from the convolutional layer of a convolutional neural network, too many channels will cause the network parameters to increase exponentially. The convolutional layer will require hundreds or even thousands of channels, which is

difficult for the network to implement. If the number of network input channels trained is fixed to the number of LEDs, the trained network will not work on other systems with different numbers of LEDs.

Figure 2. Deep multi-feature transfer network Fourier ptychographic microscopy imaging reconstruction model.

In order to avoid these problems, the low-resolution data of FPM are synthesized in the Fourier domain, and converted into a dual-channel complex amplitude image through the inverse Fourier transform as the input of the network [34]. This process can be expressed as:

$$O_n(k) = F \left\{ \sqrt{I_{nc}(r)} \right\} H(k + k_n) + O_{n-1}(k) [1 - H(k + k_n)] \quad (10)$$

$$o(r) = F^{-1} \{ O_{LED}(k) \} \quad (11)$$

where n denotes the serial number of the image, the $(n = 1, 2, \dots, N_{LED})$. $o(r)$ is a complex form, and its intensity and phase are input to the neural network as a channel, respectively. Using Equations (10) and (11), the number of channels of input data is greatly reduced, while the spatial size is increased to contain more effective high-frequency information.

3.2. Reconstructed Network Structure of the Built Deep Multi-Feature Transfer Network

In this paper, DNN is used to reconstruct the network structure and train the model to learn the nonlinear mapping relationship between input and output. The original low-resolution intensity image is synthesized and input into the dual-channel complex amplitude through formulas (10) and (11) as the input of the network model data. After transfer learning ResNet50 [32], Xception [33] and Densenet121 [34] networks, the down sampling feature extraction is realized. The cascaded feature fusion strategy performs channel fusion on the image features extracted by the transfer learning network. Finally, the pre-upsampling reconstruction network module is used to reconstruct the high-resolution intensity and phase with the same size as the output size.

Through the training of data set, the network can achieve fast and high-quality FP reconstruction. From a functional point of view, it is mainly composed of four parts, constructing the residual network of the ResNet50 transfer learning network, the channel attention of the Xception transfer learning network, the dense connection of the DenseNet121 transfer learning network, and the pre-up sampling reconstruction network [35–38].

3.2.1. Transfer ResNet50 Base-Layer

The neural network only needs to calculate the residual between input and output, and the output is obtained by adding the residual to the input. The residual network is very easy to implement, can greatly reduce the difficulty of network training and significantly improve the reconstruction effect. The ResNet50 transfer network is mainly composed of residual structures.

As Figure 3 shows the transfer learning ResNet50 base-layer network model framework, where (a) represents the migration ResNet50 base-layer network model, (b) representation of the CONV BLOCK network structure module formed in the migration base layer and (c) represents the ID BLOCK network structure module formed in the migration-base layer.

Figure 3. Migration ResNet50 base-layer network model framework: (a) schematic diagram of base-layer network model; (b) CONV BLOCK network structure module; (c) ID BLOCK network structure module.

3.2.2. Transfer Xception Base-Layer

Different weights are applied to the feature information of different channels in the feature map to improve the utilization rate of effective information. Since different channels of the feature map correspond to different convolution kernel operations, by weighting specific channels, useful information can be more effectively combined, thereby improving the performance of the network. The Xception migration network is mainly composed of channel attention structure.

As Figure 4 shows the transfers learning Xception base-layer network model framework, where (a) represents the migration Xception base-layer network model, (b) represents the RSCONV BLOCK network structure module formed in the migration base-layer and (c) represents the RSID BLOCK network structure module formed in the migration base layer.

3.2.3. Transfer DenseNet121 Base-Layer

Dense connection is established between multiple modules of the neural network, which greatly increases the number of paths of data flow. By introducing dense connections, each module is directly connected to the input and output, and each module is also connected to each other. The image features are reused to the greatest extent. The DenseNet121 migration network is mainly composed of dense connection block structure.

As Figure 5 shows the transfer learning DenseNet121 base-network model framework: where (a) represents the migrated DenseNet121 base-layer network model, (b) represents the DENSE BLOCK network structure module formed in the migration base layer and (c) represents the Tslayer network structure module formed in the migration base layer.

Figure 4. Migration Xception base network layer network model framework: (a) schematic diagram of the base network layer network model; (b) RSCONV BLOCK network structure module; (c) RSID BLOCK network structure module.

Figure 5. Migration of DenseNet121 base-layer network model framework: (a) schematic diagram of base-layer network model; (b) DENSE BLOCK network structure module; (c) Tslayer network structure module.

3.2.4. Cascade Feature Fusion

In transfer learning, feature fusion can be used to connect multiple transfer network models together [34]. The purpose of feature fusion is to combine the features extracted from the image into a more discriminative feature than the input feature.

The feature fusion algorithm can fuse the output multiple feature maps to obtain the fused feature map, thus connecting multiple networks together, which is the fusion point. Multiple networks begin to learn independently before the fusion point, so when the feature fusion method is introduced, the network fuses the features of independent learning at the fusion point, and finally starts to learn together.

In this paper, the single feature extracted from three sub-network structures (ResNet50, Xception, DenseNet121) is trained for feature fusion. The feature fusion function retains the results of three feature maps, which makes the number of channels after fusion become the sum of the number of channels in the original feature map. The formula is:

$$y = f(x^a, x^b, x^c) \quad (12)$$

$$y_{i,j,2d} = x_{i,j,d}^a, y_{i,j,2d-1} = x_{i,j,d}^b, y_{i,j,2d+1} = x_{i,j,d}^c \quad (13)$$

where x^a , x^b and x^c denote the characteristics of images obtained by different transfer neural networks, y represents the fused features. $y \in \mathbb{R}^{H \times W \times D}$, H , W and D represent the length, width and channel number of feature vector.

3.3. Pre-Upsampling

After feature fusion, the upsampling module is composed of convolution layer and pixel recombination layer to sample the feature map to the same spatial size as the network input. Through a convolution layer and a pixel recombination layer, the space size can be doubled, and the pixels are continuously convoluted upward until the pre-upsampling reconstruction module receives the dual-channel feature map output. The low-resolution image is up-sampled at the front end of the neural network. As shown in Figure 6, the blurred image can be made clear, and the operations of feature extraction, mapping and reconstruction are completed in the high-resolution space.

Figure 6. Pre-upsampling reconstruction module.

4. Experiment

The DMFTN reconstruction method proposed in this paper is verified on the simulation and experimental data sets. Using the simulation experimental data, the reconstruction results of the FPM reconstruction method under deep learning are compared and evaluated with the iterative phase recovery reconstruction alternating projection G-S [34] method, the latest Zhang [19] and the Zuo's AS [39] method.

4.1. Experimental Environment

The test experimental platform is a computer with AMD Ryzen9 5900X CPU, 3.70 GHz, 64-bit operating system, 64G of memory and Window10 operating system. The programming language used to implement the method in this paper is python3.7, using the TensorFlow open-source framework; the G-S algorithm is run under the Matlab2018b environment. The Zhang et al. deep learning method uses the Keras open-source framework. The experimental reconstruction results were compared using the values of peak signal to noise rate (PSNR) and structural similarity (SSIM). Simulation and pre-processing of the dataset were implemented through MATLAB. For training the network, a learning rate of 4×10^{-5} , Adaptive moment estimation (Adam) optimizer was used to implement gradient descent, the size of the input and truth data was 192×192 pixels, the loss function was Mean Square Error (MSE), and 200 epochs were trained on a graphics card with a training batch size is 4, the model is saved and the experiments are performed on the test experimental platform.

4.2. Construction of the Dataset

4.2.1. Synthesis of Simulated Experimental Data

The photographer and street map are used as amplitude and phase images, respectively. The process of simulating experimental data is shown in Figure 7. In the simulation process, the NA of the objective lens is set to 0.13, the wavelength is set to $0.505 \mu\text{m}$, and the simulated planar array light source is 13×13.169 low-resolution intensity images with 48×48 pixels were obtained by using the FPM imaging process. The dual-channel complex amplitude image was reconstructed as the model input through the traditional iterative phase reconstruction method. In order to simulate the influence of noise on the

reconstruction results in the image acquisition process of the imaging system, the complex amplitude image input model with noise is reconstructed by adding Gaussian noise while obtaining low resolution in the FPM imaging process.

Figure 7. Process of simulating experimental data.

4.2.2. Experimental Data Set Construction

In this paper, the simulation method is used to construct the data set [35], and the simulation image is directly used as the true value. The corresponding network model input data is generated by the simulation of the FPM imaging model, as shown in Figure 8. A total of 400 high-resolution images of various samples were collected, such as penile tissue slices, blood cell smears and alveolar tissue slices, etc. Through the random combination of phase and intensity, 1600 complex amplitudes were generated as the true value of the network output. These complex amplitudes are used to obtain low resolution intensity images through the FP imaging process, and the input method is synthesized to generate the input dual-channel intensity and phase of the network. In the process of image acquisition, the influence of natural light on the quality of image reconstruction provided by the experimental light source is considered. Therefore, Gaussian noise with mean of 0 and standard deviation is added to the input data to simulate the noise in the actual imaging process. The data generated by the simulation are randomly cut, and 25,600 groups of input and true value data are finally obtained. Among them, 90% are used as the network model training data set with the number of training sets being 23,040; 10% as the model test data set, and the number of test sets is 2560.

Figure 8. Construction of the dataset and training process of the network model.

On the basis of the above simulation data set, in order to improve the processing effect of the network for real system acquisition data, a fine-tuning data set composed of real experimental equipment acquisition images is established. The original data of 50 groups of samples of Fourier laminated microscopy were collected, and the reconstructed image was used as the true value. Then, 450 groups of input low resolution images and corresponding true value images were obtained by random clipping, including 400 training sets and 50 test sets. In order to prevent the over-fitting of the network training process, the fine-tuning data set is randomly added to the simulation data set for training.

4.3. Evaluation Method of Reconstruction Results

The Mean Square Error (MSE) and accuracy index are used to measure the stability of the network model. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used to evaluate the quality of image reconstruction.

4.3.1. Evaluation Indicators of the Model

Loss function: The MSE loss function is used to evaluate the error between the predicted value and the real value of the sample. The stability of the model training is evaluated by analyzing the loss function value of the training sample value and the iteration number curve.

$$MSE(y, y') = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \quad (14)$$

where y denotes the sample true value, y' denotes the sample predicted value and n denotes the number of iterations.

Accuracy: The number of correctly predicted samples divided by the number of all samples, which is generally used to evaluate the global accuracy of the model. The higher the accuracy, the closer the training accuracy curve is to the test accuracy curve and the better the model stability.

4.3.2. Reconstructed Image Evaluation Index

Peak signal-to-noise ratio (PSNR): An objective criterion for evaluating the image, which indicates the quality of the output image compared with the original image after the image is processed. The larger the PSNR value, the better the image quality and the less the image distortion.

$$PSNR = 20 \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (15)$$

where, MAX_I denotes the maximum value of the image color, and the 8-bit sampling point is expressed as 255; MSE is the mean square error between the original image and the reconstructed image, and its definition can be expressed as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \quad (16)$$

where, I denotes the original image, K denotes the reconstructed image and the size of the image is the same. m, n are the height and width of the image, respectively.

Structural similarity (SSIM) [40]: it is an evaluation index which is more consistent with human subjective evaluation and can be used to measure the similarity between digital images, which measures the image similarity in terms of brightness, contrast and structure, respectively, and is calculated as follows:

$$SSIM(X, Y) = I(X, Y)c(X, Y)s(X, Y) \quad (17)$$

where, $l(X, Y)$ comparing the brightness of two images, $c(X, Y)$ comparing the contrast of two images, $s(X, Y)$ comparing the structure of two images, are expressed as:

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (18)$$

$$c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (19)$$

$$s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \quad (20)$$

where, μ_X and μ_Y denote the mean of the two images, σ_X and σ_Y denote the standard deviation of the two images, σ_{XY} are the covariance of the two images, C_1, C_2, C_3 are the constants of the calculation, in use, the parameters are generally set to $C_3 = C_2/2$, to obtain the simplified form

$$SSIM = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)} \quad (21)$$

where $C_1 = (0.01L)^2$, $C_2 = (0.03L)^2$, L is the dynamic range of the pixel value, SSIM takes the value range $[0, 1]$, the larger the value, the smaller the image distortion.

4.4. Experimental Results and Analysis

4.4.1. Network Model Evaluation

In this paper, the stability of the network model is measured by the fitting curve of the loss function and accuracy index value of the training and test data output under different epochs, as shown in Figure 9. Figure 9a shows the curve of the number of training iterations and the loss function. As the number of iterations increases, the closer the fitting curve of the reconstruction network training loss function and the reconstruction network prediction loss function is to the model, the better the stability is. The prediction curve does not shake violently, and the model does not appear gradient explosion. Figure 9b is the curve of the number of training iterations and the accuracy rate. The closer the fitting curve of the reconstruction network training accuracy rate and the reconstruction network prediction accuracy rate is, the higher the predicted result value, and the better the reconstruction effect.

Figure 9. Evaluation of stability of model fitting curve. (a) Plot of training iterations vs. loss function; (b) plot of training iterations vs. accuracy.

4.4.2. Comparison of Reconstruction Performance under Noise Conditions

Considering the existence of Gaussian noise in the image acquisition process of Fourier ptychographic microscopy imaging system, the mean value is 0 and the standard deviation is Gaussian noise in the simulation test set. The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used to compare the reconstruction quality. The DMFTN method is compared with the traditional alternating projection method and the neural

network method. As shown in Figure 10, Table 1 is the index value of different image reconstruction results under the same noise.

Figure 10. Reconstruction results under the same noise.

Table 1. Index values of different image reconstruction results under the same noise.

Add the Noise Level. 3×10^{-4}		Low Resolution (PSNR/SSIM)	G-S H-Method (PSNR/SSIM)	Zhang's Method (PSNR/SSIM)	DMFTN (INPUT) (PSNR/SSIM)	DMFTN (OUTPUT) (PSNR/SSIM)
Group 1 Images	Intensity	21.25/0.723	24.02/0.761	26.27/0.922	23.91/0.771	34.79/0.979
	Phase		13.93/0.278	22.52/0.786	12.12/0.083	25.99/0.965
Group 2 Images	Intensity	24.76/0.844	22.62/0.592	20.03/0.787	23.13/0.613	35.73/0.981
	Phase		4.29/0.229	16.01/0.877	3.51/0.016	33.43/0.976
Group 3 Images	Intensity	24.77/0.840	28.94/0.554	22.77/0.920	23.66/0.678	36.54/0.979
	Phase		10.93/0.264	21.52/0.932	9.19/0.015	24.84/0.932

In order to verify the universality and robustness to noise of the reconstruction methods in this chapter, the reconstruction results of each method with different noise added under the same image are shown in Figure 11, and Table 2 shows the index values of the reconstruction results of the same image under different noise.

From the above experimental results, it can be seen that the reconstruction quality index values of DMFTN under different noise conditions are better than those of GS and Zhang et al. The DMFTN reconstruction results contain more details and less errors, while the GS method has poor phase reconstruction results, and the strength reconstruction results of Zhang et al. have artifacts.

Noise		Ground truth	Low Resolution	GS Method	Zhang's Method	DMFTN (INPUT)	DMFTN (OUTPUT)
1×10^{-4}	Intensity
	Phase
2×10^{-4}	Intensity
	Phase
3×10^{-4}	Intensity
	Phase

Figure 11. Reconstruction results under the different noise.

Table 2. Index values of the same image reconstruction results under different noise.

Same IMAGE		Noise Level	Low Resolution (PSNR/SSIM)	I-S Method (PSNR/SSIM)	Zhang's Method (PSNR/SSIM)	DMFTN (INPUT) (PSNR/SSIM)	DMFTN (OUTPUT) (PSNR/SSIM)
Group 1 Images	Intensity	1×10^{-4}	24.42/0.827	28.85/0.848	30.38/0.861	28.76/0.937	35.12/0.981
	Phase			7.97/0.223	26.11/0.928	5.99/0.021	26.63/0.45
Group 2 Images	Intensity	2×10^{-4}	22.24/0.718	25.12/0.722	27.62/0.836	25.77/0.737	34.99/0.979
	Phase			7.43/0.205	25.72/0.923	6.17/0.118	26.35/0.944
Group 3 Images	Intensity	3×10^{-4}	20.11/0.622	23.26/0.643	21.47/0.861	23.91/0.662	34.79/0.979
	Phase			7.27/0.197	18.78/0.92	5.96/0.013	26.08/0.952

4.4.3. Reconstruction Performance with Reduced Amount of Acquired Image Data

In order to evaluate the performance of this method when the amount of image data is small, the reconstruction performance is tested when the amount of original collected image data is 9, 25, 49, 81, 121 and 169. The simulated light source of the image test data set is 9, 25, 49, 81, 121 and 169, and the Gaussian noise with the mean value of 0 and the standard deviation of 1×10^{-4} is added.

The test results are compared with the performance of Zhang et al.'s reconstruction methods, as shown in Figure 12, Figure 12a is the PSNR curve of the reconstruction results, and Figure 12b is the SSIM curve of the reconstruction results. The red and yellow represent the PSNR and SSIM curves of the reconstructed intensity and phase quality, and the blue and green represent the PSNR and SSIM curves of the reconstructed intensity and phase quality by Zhang et al.

Figure 12. Reconstruction quality curve with reduced number of acquired images: (a) PSNR plot of the reconstruction results; (b) SSIM plot of the reconstruction results.

It can be seen from the reconstructed quality curve that the intensity and phase PSNR curves of the reconstructed results of this method are higher than those of Zhang et al. The SSIM curve reconstructed by this method is higher and closer to the real value. Therefore, the use of deep learning method can not only improve the reconstruction quality, but also greatly improve the speed of image acquisition.

4.4.4. Reconstruction Results on the Actual Dataset

In order to verify the validity of Fourier ptychographic microscopy imaging reconstruction method based on deep multi-feature transfer network on real data, the test data set of fine-tuning data set constructed by equipment is used to compare the results of DMFTN with traditional phase recovery reconstruction and Zhang et al. Since the original low-resolution image can only be collected under experimental conditions, the true value of complex amplitude reconstruction results does not exist. Figure 13 shows the reconstruction results of DMFTN, GS and Zhang on the test set. It can be seen that the DMFTN performs well on real data, and obtains the optimal reconstruction results. The error of image background is less and the texture details are clearer.

4.4.5. Reconstruction Time Comparison

The reconstruction process of FPM is the reconstruction hundreds of multi-angle illumination images to increase the number of acquired images to improve the redundancy of the reconstructed data and obtain high-resolution images, the FPM needs to consume a large amount of time resolution while improving the spatial resolution, resulting in the acquisition of the original image and the reconstruction process consuming a lot of time. Therefore, in the reconstruction process, not only the reconstruction quality but also the reconstruction speed should be considered.

When the image pixel size is 192×192 , the DMFTN reconstruction method proposed in this paper is compared with the traditional phase recovery method (GS), adaptive step reconstruction method (AS) and the convolution neural network reconstruction method proposed by Zhang et al. Since DMFTN is based on deep learning without iterations, it has the least reconstruction time, as shown in Table 3.

Figure 13. Reconstruction results on the actual dataset.

Table 3. Comparison of reconstruction time of four methods.

Reconstruction Method	Number of Iterations	Reconstruction Time	Multiple
G-S Method	50	1.684 s	17
AS Method	50	1.977 s	22
Zhang's Method	20	204.5 s	48,822
Method of this article (DMFTN)	0	0.091 s	1

5. Summary

In this paper, a multi-convolutional feature fusion network Fourier ptychographic microscopy imaging reconstruction model (DMFTN) is proposed for FPM reconstruction, which realizes FPM reconstruction under deep learning. The method uses ResNet50, Xception and DenseNet121 migratory learning network frameworks, a cascaded feature fusion strategy and a pre-upsampling reconstruction network to construct a DMFTN model; the multi-convolution migratory learning network is used to extract the complementarity of feature information and improve the utilization rate of feature fusion. The pre-upsampling reconstruction network improves the details of high-resolution image reconstruction. The experiments verify that the trained FPM neural network can reconstruct higher quality intensity and phase images with faster speed; the method has higher robustness to noise and blurring and other factors, which can greatly improve the temporal resolution of FPM and promote its practical application. However, the use of deep learning to achieve FPM reconstruction will also have limitations. The limitations come from the construction of data sets. If the experimental equipment is used to obtain the data set, there is no true value of the data set. If the simulation data sets are obtained under ideal conditions, the distortion from the system is not considered.

Moreover, due to the rapid development of powerful tools such as TensorFlow and other deep learning tools optimizers and loss functions, it is possible to combine new these techniques to obtain further improvements, and to replace some layers of DMFTN with super-resolution medium and high-level architectures for better model reconstruction.

Author Contributions: Conceptualization, Y.P., T.X. and J.L.; methodology, X.W. and J.Y.; software, X.W. and J.Y.; validation, X.W. and J.Y.; investigation, J.Y.; writing—original draft preparation, X.W.; writing—review and editing, X.W., Y.P., J.Y., J.L., H.S., Y.J. and L.L.; supervision, T.X.; project administration, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Jilin Provincial Department of Education [NO: 2021JB502L06].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wang, D.; Fu, T.; Bi, G.; Jin, L.; Zhang, X. Long-Distance Sub-Diffraction High-Resolution Imaging Using Sparse Sampling. *Sensors* **2020**, *20*, 3116. [[CrossRef](#)] [[PubMed](#)]
- Jiasong, S.; Yuzhen, Z.; Qian, C.; Chao, Z. Fourier Ptychographic Microscopy: Theory, Advances, and Applications. *Acta Opt. Sin.* **2016**, *36*, 1011005. [[CrossRef](#)]
- Zheng, G.; Horstmeyer, R.; Yang, C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nat. Photonics* **2013**, *7*, 739–745. [[CrossRef](#)] [[PubMed](#)]
- Guoan, Z. Breakthroughs in Photonics 2013: Fourier Ptychographic Imaging. *IEEE Photonics J.* **2014**, *6*, 1–7.
- Zheng, G.; Ou, X.; Horstmeyer, R. Fourier Ptychographic Microscopy: A Gigapixel Superscope for Biomedicine. *Opt. Photonics News* **2014**, *25*, 26–33. [[CrossRef](#)]
- Xiong, C.; Youqiang, Z.; Minglu, S.; Quanquan, X. Apodized coherent transfer function constraint for partially coherent Fourier ptychographic microscopy. *Opt. Express* **2019**, *27*, 14099–14111.
- Gerchberg, R.W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **1972**, *35*, 237–250.
- Fienup, R.J. Phase retrieval algorithms: A comparison. *Appl. Opt.* **1982**, *21*, 2758–2769. [[CrossRef](#)]
- Ou, X.; Zheng, G.; Yang, C. Embedded pupil function recovery for Fourier ptychographic microscopy. *Opt. Express* **2014**, *22*, 4960–4972. [[CrossRef](#)]
- Bian, Z.; Dong, S.; Zheng, G. Adaptive system correction for robust Fourier ptychographic imaging. *Opt. Express* **2013**, *21*, 32400–32410. [[CrossRef](#)]
- Sun, J.; Qian, C.; Zhang, Y.; Chao, Z. Efficient positional misalignment correction method for Fourier ptychographic microscopy. *Biomed. Opt. Express* **2016**, *7*, 1336. [[CrossRef](#)] [[PubMed](#)]
- Zhang, J.; Xu, T.; Liu, J.; Chen, S.; Xing, W. Precise brightfield localization alignment for Fourier ptychographic microscopy. *IEEE Photonics J.* **2017**, *10*, 1109. [[CrossRef](#)]
- Tong, L.; Jufeng, Z.; Haifeng, M.; Guangmang, C.; Jinxing, H. An Efficient Fourier Ptychographic Microscopy Imaging Method Based on Angle Illumination Optimization. *Laser Optoelectron. Prog.* **2020**, *57*, 081106. [[CrossRef](#)]
- Ziqiang, L.; Xiao, M.; Jinxin, L.; Jiaqi, Y.; Shiping, L.; Jingang, Z. Fourier Ptychographic Microscopy Based on Rotating Arc-shaped Array of LEDs. *Laser Optoelectron. Prog.* **2018**, *55*, 071102. [[CrossRef](#)]
- Dong, S.; Shiradkar, R.; Nanda, P.; Zheng, G. Spectrum multiplexing and coherent-state decomposition in Fourier ptychographic imaging. *Biomed. Opt. Express* **2014**, *5*, 1757–1767. [[CrossRef](#)] [[PubMed](#)]
- Zhang, J.; Xu, T.; Chen, S.; Wang, X. Efficient Colorful Fourier Ptychographic Microscopy Reconstruction with Wavelet Fusion. *IEEE Access* **2018**, *6*, 31729–31739. [[CrossRef](#)]
- Bian, L.; Suo, J.; Zheng, G.; Guo, K.; Chen, F.; Dai, Q. Fourier ptychographic reconstruction using Wirtinger flow optimization. *Opt. Express* **2015**, *23*, 4856–4866. [[CrossRef](#)]
- Shaowei, J.; Kaikai, G.; Jun, L.; Guoan, Z. Solving Fourier ptychographic imaging problems via neural network modeling and TensorFlow. *Biomed. Opt. Express* **2018**, *9*, 3306.
- Zhang, Y.; Liu, Y.; Jiang, S.; Dixit, K.; Li, X. Neural network model assisted Fourier ptychography with Zernike aberration recovery and total variation constraint. *J. Biomed. Opt.* **2021**, *26*, 036502. [[CrossRef](#)]
- Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)]
- Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
- Ying, T.; Jian, Y.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
- Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*; IEEE Computer Society: Washington, DC, USA, 2016.
- Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.
- Li, J.; Chen, Z.; Zhao, X.; Shao, L. MapGAN: An Intelligent Generation Model for Network Tile Maps. *Sensors* **2020**, *20*, 3119. [[CrossRef](#)] [[PubMed](#)]
- Rivenson, Y.; Zhang, Y.; Gunaydin, H.; Da, T.; Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **2017**, *7*, 17141. [[CrossRef](#)] [[PubMed](#)]

27. Sinha, A.T.; Lee, J.; Li, S.; Barbastathis, G. Lensless computational imaging through deep learning. *Optica* **2017**, *4*, 1117–1125. [[CrossRef](#)]
28. Kappeler, A.; Ghosh, S.; Holloway, J.; Cossairt, O.; Katsaggelos, A. Ptychnet: CNN based fourier ptychography, 2017. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.
29. Nguyen, T.; Xue, Y.; Li, Y.; Tian, L.; Nehmetallah, G. Deep learning approach for Fourier ptychography microscopy. *Optics Express* **2018**, *26*, 26470. [[CrossRef](#)]
30. Cheng, Y.F.; Strachan, M.; Weiss, Z.; Deb, M.; Carone, D.; Ganapati, V. Illumination Pattern Design with Deep Learning for Single-Shot Fourier Ptychographic Microscopy. *Opt. Express* **2019**, *27*, 644–656. [[CrossRef](#)]
31. Yu, J.; Li, J.; Wang, X.; Zhang, J.; Liu, L.; Jin, Y. Microscopy image reconstruction method based on convolution network feature fusion. In Proceedings of the 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 23–26 September 2021.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
33. Chollet, F. In Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
34. Huang, G.; Liu, Z.; Laurens, V.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
35. Zhang, J.; Xu, T.; Shen, Z.; Qiao, Y.; Zhang, Y. Fourier ptychographic microscopy reconstruction with multiscale deep residual network. *Opt. Express* **2019**, *27*, 8612–8625. [[CrossRef](#)]
36. Zhao, X.D.; Huang, H. Research on face recognition based on convolutional neural network. *Inf. Technol.* **2018**, *10*, 920.
37. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep Learning for Image Super-resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [[CrossRef](#)]
38. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H. Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Trans. Multimedia* **2019**, *21*, 3106–3121. [[CrossRef](#)]
39. Chao, Z.; Sun, J.; Qian, C. Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy. *Opt. Express* **2016**, *24*, 20724–20744.
40. Wang, Z. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]