Article

Fourier Ptychographic Reconstruction Method of Self-Training Physical Model

Xiaoli Wang [1,2], Yan Piao [1,*], Yuanshang Jin [2], Jie Li [2], Zechuan Lin [2], Jie Cui [2] and Tingfa Xu [3]

1   Information and Communication Engineering, Electronics Information Engineering College, Changchun University of Science and Technology, Changchun 130022, China
2   Electronics Information Engineering College, Changchun University, Changchun 130022, China
3   School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China
*   Correspondence: piaoyan@cust.edu.cn

Abstract: Fourier ptychographic microscopy is a new microscopic computational imaging technology. A series of low-resolution intensity images are collected by a Fourier ptychographic microscopy system, and high-resolution intensity and phase images are reconstructed from the collected low-resolution images by a reconstruction algorithm. It is a kind of microscopy that can achieve both a large field of view and high resolution. Here in this article, a Fourier ptychographic reconstruction method applied to a self-training physical model is proposed. The SwinIR network in the field of super-resolution is introduced into the reconstruction method for the first time. The input of the SwinIR physical model is modified to a two-channel input, and a data set is established to train the network. Finally, the results of high-quality Fourier stack microscopic reconstruction are realized. The SwinIR network is used as the physical model, and the network hyperparameters and processes such as the loss function and optimizer of the custom network are reconstructed. The experimental results show that by using multiple different types of data sets, the two evaluation index values of the proposed method perform best, and the image reconstruction quality is the best after model training. Two different evaluation indexes are used to quantitatively analyze the reconstruction results through numerical results. The reconstruction results of the fine-tuning data set with some real captured images are qualitatively analyzed from the visual effect. The results show that the proposed method is effective, the network model is stable and feasible, the image reconstruction is realized in a short time, and the reconstruction effect is good.

Keywords: fourier ptychographic microscopy; reconstruction; SwinIR; self-training physical model

1. Introduction

The microscope is an optical instrument consisting of single or multiple lenses. It is an important part of the production, life, and research fields such as clinical medicine and industrial inspection. In 2013, Zheng Guo 'an et al. proposed a new imaging technique called Fourier ptychographic microscopy (FPM) [1]. FPM is a computational imaging technology with a large field of view and high resolution. Fourier ptychographic microscopy combines synthetic aperture [2–11], phase retrieval [12–23], optimization theory [24–36], and other concepts to break many limitations of traditional microscopes. Different from the traditional microscope, Fourier ptychographic imaging is a series of low-resolution intensity images collected by Fourier ptychographic microscopy system. Combined with a reconstruction algorithm, high-resolution intensity and phase images are reconstructed from the collected low-resolution images. The imaging process is as follows: the samples are illuminated successively with different LED incident angles, a set of low-resolution (LR) images with different spatial spectra are captured accordingly, and then these spectral sub-apertures obtained from the low-resolution images are spliced together in the Fourier domain to reconstruct the entire high-resolution (HR) spatial spectrum. In short, only

the low-resolution intensity image corresponding to the high-resolution spatial spectrum sub-aperture image is obtained, and the low-resolution intensity image is used to restore the high-resolution spatial spectrum.

The traditional G-S (Gerchberg–Saxton) method [37] generates phase information by changing the phase gradient caused by oblique incident illumination, thereby achieving synthetic aperture and reconstructing high-resolution complex amplitude images. The iterative update process of this reconstruction method uses the traditional gradient descent method to obtain the global optimal solution. Since FPM technology was proposed in 2013, researchers have proposed many improvements to FPM reconstruction algorithms, such as Gauss–Newton [38], Wirtinger flow [39] (Wirtinger flow optimization), adaptive step [40] (A-S, adaptive step-size), and other optimization algorithms. However, the traditional reconstruction algorithm of Fourier ptychographic microscopy requires more iterative updates, which cannot escape the idea of gradient descent, and the reconstruction effect is poor. Researchers have focused their research on neural networks and deep learning. By using neural network algorithms, the computing power of computational imaging can be improved. For example, the neural network modeling method proposed by Jiang et al. [41] uses back propagation as a tool to solve the global optimal solution; the deep learning method [42–45] has the advantages of high processing performance and fast operation speed. It can be used in Fourier laminated microscopic computational imaging technology to solve the problem of image reconstruction quality.

The author proposed an FPM reconstruction algorithm based on the SwinIR physical model [46] at the 2022 ICAITA International Conference. The algorithm completes the reconstruction by screening the appropriate network parameters and using the trained weights. At the same time, fast and high-quality target image reconstruction is realized. The results show that it has strong robustness to noise, fast reconstruction speed, and high reconstruction quality.

At present, deep learning and machine learning have developed rapidly, and have been studied in many aspects such as machine crown segmentation [47], music [48], life prediction [49], and image classification [50]. The main research of this paper is based on transfer learning to introduce the network with a better imaging effect in the field of super-resolution into Fourier laminated microscopic reconstruction, which improves the reconstruction speed and quality, and enhances the robustness of real images and ideal images with noise. At the same time, Zuo et al. [51] proposed a new non-interference method for quantitative phase imaging. By obtaining quantitative phase information only through intensity measurement, the best spatial resolution, higher signal-to-noise ratio, and better image quality can be achieved. Michael Kovalev et al. [52] proposed a simple method to reconstruct the spatial parameters of laser beams based on the intensity transfer equation. A single CMOS camera is used to register the cross-sectional light intensity distribution in multiple planes.

In this article, we use the SwinIR network [53] in the field of super-resolution as the main network of Fourier stack microscopic reconstruction. However, unlike SwinIR, in order to ensure that our FPM data set can adapt to the SwinIR model, we customize the training process and modify its input into a two-channel input. The end-to-end processing is realized by the Fourier stack microscopic imaging algorithm (STPM-FPM) of the self-training physical model, and the data set is established to train the network. Finally, the results of high-quality Fourier microscopic reconstruction are realized. Simulation experiments show that the self-training SwinIR network structure is superior to the traditional algorithm in terms of reconstruction quality and speed, and STPM-FPM can use less low-resolution images for image reconstruction, which greatly improves the time resolution of FPM imaging. It also has better robustness than other algorithms for external interference factors such as noise.

2. Principle of FPM

2.1. Fourier Ptychographic Microscopy Device

The biggest difference between the Fourier ptychographic microscopy device and the traditional microscope is that the Fourier ptychographic microscope uses an LED array illumination light source instead of the microscope light source, that is, reposed on the single hardware modification of the traditional bright field microscope to achieve the goal of a large field of view and high resolution. This imaging mode is implemented in the same optical settings, no moving parts, only need to select the appropriate LED to open. In addition, this experimental device is equipped with a DMK 33UX264 camera (3.45 um, 2448 × 2048), which is used to digitally image the sample. In addition, the microscope uses the RX50 series upright bright field microscope for optical microscopy imaging, which can be directly viewed, and can also collect actual real pictures through the camera. Figure 1 shows the hardware structure of the Fourier ptychographic microscopy.
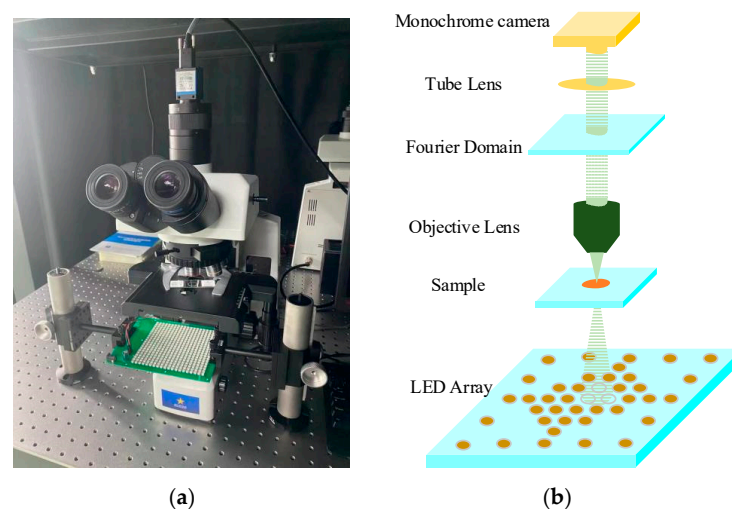

(**a**)   (**b**)

Figure 1. Fourier Ptychographic Microscopy Device Diagram. (**a**) Real FPM System; (**b**) FPM System Diagram.

FPM consists of an eyepiece, optical path selection rod, *Y*-axis mobile handwheel, mirror group, loose and tight adjustment handwheel, adjustment light wheel, collecting light microscope, *X*-axis mobile handwheel, mechanical platform, a DMK 33UX264 camera, an LED lamp holder, an LED lamp plate (20 × 20), and other components. The function of the LED lamp board bracket is to clamp the LED lamp board and facilitate the adjustment of the LED height. The function of the LED light board is to analyze the instructions sent by the Matlab program through the serial port, light the LED lights at a specific position, and realize the collection of FPM data by successively lighting the LED lights at a specific position. The built-in RGB three-color LED lamp beads can be lit in turn by three RGB colors. The black and white camera is used to collect the image and synthesize the color image of the sample.

2.2. Fourier Ptychographic Microscopy Imaging Process

In the FPM system, the LED array with multi-angle illumination is used as the illumination light of the system. This section analyzes the transfer process of the sample information of the FPM system under more general illumination, as shown in Figure 2. Suppose the incident light from the $E^m(x_0, y_0)$ LED:

$$E^m(x_0, y_0) = \exp[i2\pi(\mu^m x_0 + v^m y_0)] \tag{1}$$

where (μ, v) denotes that the illumination wave vector is determined by the illumination angle of the incident light; (x_0, y_0) refers to the LED light source position.

$$\mu^m = \frac{n \sin \varphi_x^m}{\lambda}, v^m = \frac{n \sin \varphi_y^m}{\lambda} \tag{2}$$

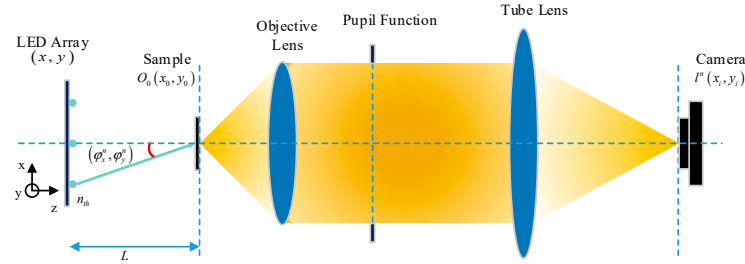where n is the refractive index of the medium in the illumination path.

Figure 2. Imaging schematic of Fourier ptychographic microscopy system.

For the LED array structure in Figure 2, the illumination angle of each LED is calculated as follows:

$$
\begin{aligned}
\sin \varphi_x^m &= \frac{x_0^{center} - x^m}{\sqrt{\left(x_0^{center} - x^m\right)^2 + \left(y_0^{center} - y^m\right)^2 + L^2}}, \\
\sin \varphi_y^m &= \frac{y_0^{center} - y^m}{\sqrt{\left(x_0^{center} - x^m\right)^2 + \left(y_0^{center} - y^m\right)^2 + L^2}}
\end{aligned} \tag{3}
$$

In Equation (3), $(x_0^{center}, y_0^{center})$ represents the center coordinates of the unanalyzed area on the object surface, (x^m, y^m) represents the coordinates of the mth LED, and L represents the distance from the LED array to the sample. By combining with Formula (2), the spectral information transmitted to the focal plane behind the objective lens can be obtained as follows:

$$
\begin{aligned}
U_0^m\left(f_x, f_y, 0\right) P\left(f_x, f_y\right) &= \mathcal{F}\{O_0(x_0, y_0, 0) E^m(x_0, y_0)\} P\left(f_x, f_y\right) \\
&= U_0\left(f_x - \mu^m, f_y - v^m, 0\right) P\left(f_x, f_y\right) \\
&= U_0\left(f_x, f_y, 0\right) P\left(f_x + \mu^m, f_y + v^m\right)
\end{aligned} \tag{4}
$$

where $\mathcal{F}$ represents Fourier transform.

Finally, the expression of the low-resolution image acquired by the FPM system can be obtained:

$$I^m(x_i, y_i) = ||\mathcal{F}^{-1}\{U_0\left(f_x, f_y, 0\right) P\left(f_x + \mu^m, f_y + v^m\right)\}||^2 \tag{5}$$

It can be seen that in the case of oblique incident illumination, the spectrum of the sample has shifted as a whole, which can be equivalent to the interception of the pupil function of the system at various locations of the sample spectrum, and the acquisition of information exceeding the original passband is realized. At another angle, the illumination of different LEDs in the FPM system can be equivalent to the pupil function (sub-aperture) at different positions on the sample spectrum. A series of low-resolution images with different information are taken by the camera in turn and iterated in the frequency domain to update the corresponding spectrum information in different sub-apertures, expand the sample spectrum range, and recover the high-frequency complex amplitude information limited by the original objective resolution, which is also the basis for FPM technology to realize the images of super-resolution reconstruction.

Since the zero-frequency information related to the background light is located in the center of the spectrum, we divide the captured image into a bright field image and a dark field image. The way to distinguish these is to detect whether the spectrum passing through the system contains zero-frequency information. It is expressed by the formula:

$$NA_{ill} = n\sqrt{\sin^2 \varphi_x + \sin^2 \varphi_y} \tag{6}$$

Among them, NA_{ill} is the lighting NA. If $NA_{ill} \leq NA_{obj}$, the image is a bright field image; if $NA_{ill} > NA_{obj}$, the image is a dark field image. Figure 3 shows the imaging diagram of bright field and dark field images.
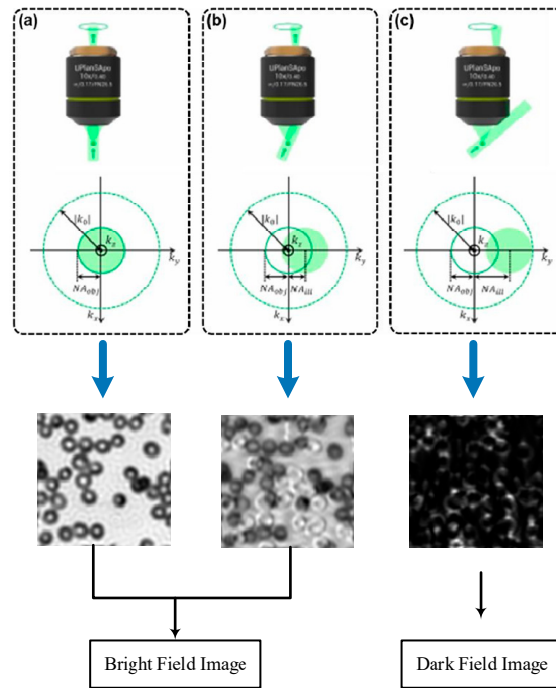

Figure 3. Imaging diagram of bright field image and dark field image. (**a**) LED direct incidence lighting (**b**) LED small angle tilt lighting (**c**) LED large angle tilt lighting.

For the collected bright field image and dark field image, the gray mean value is not an order of magnitude at all. Therefore, for the weak signal acquisition of the dark field image, it is often needful to consider the dynamic range of the camera used in the FPM system, so as to prevent the signal from being approximated to 0 when it is too weak. In the process of collecting images, in order to ensure the integrity of the dark field image information, a 16 bit camera or LED array is often used to increase the exposure time.

2.3. Fourier Ptychographic Reconstruction Process

The traditional FPM reconstruction algorithm updates the spectrum of the sample by alternating iterations in the spatial and frequency domains. Taking the most traditional G-S method as an example, the reconstruction process is based on the gradient descent algorithm to minimize the loss function. From the FPM imaging process in the previous section, we can obtain the complex amplitude corresponding to each LED array illumination Bash transmitted to the image plane, $U_{est}^l(x,y)$, and $l = 1,2, \ldots \ldots$ N, N is the number of LED arrays. The reconstruction process is as follows: the loss function is constructed by combining the low-resolution light intensity image actually collected by the FPM system, and the loss function converges to the minimum value through iterative updates. The high-resolution complex amplitude reconstruction process of the whole sample is regarded as a nonlinear minimization problem based on the loss function. The traditional FPM technology often adopts an iterative update strategy based on the minimization of mean squared error (MSE). The expression of the loss function can be expressed as follows:

$$Loss = \sum_{l}^{N_{LED}} \left|\left| \left|U_{est}^l(x,y)\right| - \sqrt{I^l(x,y)} \right|\right|^2 + \tau R(O(x_0,y_0))$$

$$= \sum_{l}^{N_{LED}} \left|\left| \left|\mathcal{F}^{-1}\left\{ P(k_x,k_y) \cdot \tilde{O}(k_x - \mu_l, k_y - v_l) \right\}\right| - \sqrt{I^l(x,y)} \right|\right|^2 + \tau R(O(x_0,y_0))$$

(7)

Here, $U_{est}^l(\mathrm{x,y})$ represents the low-resolution complex amplitude obtained under the illumination of the first LED array. $I^l(x,y)$ represents the low-resolution light intensity image actually collected corresponding to the lth LED illumination; $P(k_x, k_y)$ represents the pupil function of the system; $O(x_0, y_0)$ represents the complex amplitude function representing the high resolution of the sample; $\mathcal{F}^{-1}$ represents inverse Fourier transform; $\widetilde{O}(k_x, k_y)$ represents the high-resolution spectrum; and (μ_l, v_l) represents the illumination wave vector represented by the lth LED array. Considering that the nonlinear effect of the model will lead to overfitting of the reconstruction results, a regular term (μ_l, v_l) is added to the reconstructed complex amplitude. τ is the regularization weight.

Taking the relatively simple first-order gradient descent algorithm and the lth LED array illumination as an example, using the nonlinear optimization method to solve the high-resolution complex amplitude in Equation (7), the first derivative of the high-resolution complex amplitude spectrum to be solved is conducted for Equation (7):

$$G_{\widetilde{O}(k_x,k_y)} = P^*(k_x, k_y) \cdot \mathcal{F}\left\{ \frac{U_{est}^l(x,y)}{|U_{est}^l(x,y)|} \left[\left| U_{est}^l(x,y) \right| - \sqrt{I^l(x,y)} \right] \right\} \tag{8}$$

By minimizing the loss function through the traditional gradient descent method, the formula for solving the sample spectrum of FPM imaging optimization is updated:

$$\begin{aligned} &\widetilde{O}_{new}(k_x - \mu_l, k_y - v_l) \\ &= \widetilde{O}(k_x - \mu_l, k_y - v_l) + \alpha \frac{P^*(k_x, k_y)}{\left\| P(k_x - k_y) \right\|_{max}^2} \mathcal{F}\left\{ \frac{U_{est}^l(x,y)}{|U_{est}^l(x,y)|} \sqrt{I^l(x,y)} - U_{est}^l(x,y) \right\} \end{aligned} \tag{9}$$

where α is the step length of updating the sample spectrum, and $\frac{U_{est}^l(x,y)}{|U_{est}^l(x,y)|} \sqrt{I^l(x,y)}$ is the replacement process of spatial information in Formula (9), that is, the estimated low-resolution complex amplitude is updated by using the actual collected light intensity image. The participating pupil function means that, for each LED, only a part of the spectrum is updated, which is equivalent to the constraint of the update range introduced by the pupil function in the frequency domain.

This algorithm replaces and projects the collected low-resolution images and pupil functions as spatial and frequency domain constraints, so this method is also called the AP algorithm (alternating projection). When $\alpha = 1$, the iterative update process of the above formula is the G-S algorithm commonly used in traditional FPM.

In each iteration, the frequency domain constraint is updated once. Therefore, due to the updated frequency domain constraints, the spectral radius reconstructed by FPM technology only depends on the numerical aperture of the objective lens NA_{obj} and the maximum illumination of the LED array NA (NA_{ill}^{max}): $\frac{NA_{obj} + NA_{ill}^{max}}{\lambda}$. Therefore, the reconstructed complex amplitude resolution is:

$$\varepsilon_{rec} = \frac{\lambda}{2\left(NA_{obj} + NA_{ill}^{max}\right)} \tag{10}$$

The traditional G-S algorithm only considers the low-pass nature of the pupil function and cannot deal with the aberration problem. Therefore, the process of spectrum updates and the update process of the pupil function of the system must be synchronized, as shown in Equation (11):

$$\begin{aligned} &P_{new}(k_x, k_y) = P(k_x, k_y) \\ &+ \beta \frac{\widetilde{O}^*(k_x - \mu_l, k_y - v_l)}{\left\| \widetilde{O}(k_x - \mu_l, k_y - v_l) \right\|_{max}^2} \cdot \mathcal{F}\left\{ \frac{U_{est}^l(x,y)}{|U_{est}^l(x,y)|} \sqrt{I^l(x,y)} - U_{est}^l(x,y) \right\} \end{aligned} \tag{11}$$

In Formula (11), β is the step size for updating the pupil function. Combined with Equation (9), the FPM optimization algorithm with aberration correction can be obtained.

It is worth noting that β < α. The above update formulas are several commonly used aberration-free reconstruction algorithms (EPRY).

2.4. Evaluation Indicators

FPM technology is a computational imaging method, and its reconstruction results directly reflect the quality of the reconstruction network or algorithm. In the field of images, the evaluation of image quality can generally be divided into two aspects, subjective evaluation and objective evaluation. Subjective evaluation means that people observe the reconstruction results with the naked eye and give intuitive qualitative evaluation through vision. This evaluation method is easily affected by subjective factors. Although it is simple and fast, it cannot be quantitatively and accurately evaluated. It is also easily affected by individual differences and has poor operability. Therefore, in the actual judgment of reconstruction quality, objective evaluation is often used to judge the advantages and disadvantages of different reconstruction algorithms, and different algorithms are quantitatively compared. This summary introduces two objective evaluation methods commonly used in Fourier ptychographic microscopy reconstruction.

2.4.1. Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Rate (PSNR) is the most normally and extensively used objective evaluation index. It is an image quality evaluation grounded on the error between the corresponding pixels, that is, grounded on the sensitive error. PSNR is an engineering term for the ratio of the maximum possible power of a signal to the destructive noise ratio that affects its representation accuracy. Because many signals have a very wide dynamic range, PSNR is often expressed in logarithmic decibels (dB). In the field of image evaluation, its mathematical definition is:

$$
\begin{aligned}
MSE &= mean\left((I_1 - I_2)^2\right), \\
PSNR &= -10 * \log(MSE)
\end{aligned}
\tag{12}
$$

Among them, MSE is mean squared error (MSE), which is also a common loss function which represents the original image, represents the reconstructed image, and is the same size. The value of PSNR is generally 0–100 dB. The larger the value of PSNR, the smaller the distortion and the better the image quality.

2.4.2. Structural Similarity

Although traditional evaluation methods such as PSNR can objectively measure the difference between the original image and the reconstructed image, there is often a certain gap between these evaluation indicators and the subjective evaluation of people, and it is easy for it to appear that the image evaluation index with good subjective quality is low. Structure similarity (SSIM) is a full-reference image quality evaluation index. It measures the similarity of images from three aspects: brightness, contrast, and structure, which is more in line with a human subjective evaluation index. The input of SSIM is two images, one is an uncompressed undistorted image (grand truth), and the other is a restored image. The calculation is as follows:

$$
SSIM = [l(I_1, I_2)]^\alpha [c(I_1, I_2)]^\beta [s(I_1, I_2)]^\gamma,
$$

$$
l(I_1, I_2) = \frac{2\mu_1\mu_2 + C_1}{\mu_1^2 + \mu_2^2 + C_1},
$$

$$
c(I_1, I_2) = \frac{2\sigma_1\sigma_2 + C_2}{\sigma_1^2 + \sigma_2^2 + C_2}
\tag{13}
$$

$$
s(I_1, I_2) = \frac{\sigma_{12} + C_3}{\sigma_1\sigma_2 + C_3} \#(13)
$$

Among them is the brightness comparison of two images; it is the contrast comparison of two images; and it is the structure comparison of two images. α, β, and γ are weight

parameters, α, β, γ > 0; μ_1 and μ_2 are the average values of the two images; σ_1 and σ_2 are the standard deviation of two images. σ_{12} is the covariance of two images. and C_1, C_2, and C_3 are constants used to maintain the stability of the calculation. In practical application, α = β = γ = 1 and $c_3 = \frac{C_2}{2}$, and the updated SSIM expression is:

$$SSIM = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)} \tag{14}$$

where $C_1 = (0.01 * L)^2$, $C_2 = (0.03 * L)^2$, and L is the dynamic range of pixel value. The value of SSIM is generally 0–1. The larger the value of SSIM, the smaller the gap between the output image and the undistorted image. When the two images are consistent, SSIM = 1. At the same time, SSIM has symmetry, that is, SSIM(x, y) = SSIM(y, x).

SSIM can also use the sliding window to divide the image into blocks. Let the total number of blocks be N. Considering the influence of window shape on the block, the mean, variance, and covariance of each window are calculated by Gaussian weighting. Then, the structural similarity SSIM of the corresponding block is calculated. Finally, using the average value as the structural similarity measure of the two images, the average structural similarity is MSSIM.

$$MSSIM(X, Y) = \frac{1}{N} \sum_{K=1}^{N} SSIM(x_k, y_k) \tag{15}$$

3. Self-Training SwinIR Network Structure

3.1. Overall Network Structure

Considering the inverse problem of recovering vectorized image *a* from measurements:

$$b = S\{a\} + \varepsilon \tag{16}$$

Among them, *S* is *a* forward measurement operator, and ε represents the noise generated by imaging. For image reconstruction applications suitable for deep learning methods such as image denoising and super-resolution direction, the forward model *S* is unknown [3]. At present, the deep learning method for such problems is to train a DNN as a reverse model. The training is performed in a supervised manner, so the data sets need to have (*a*, *b*) pairs of images.

For many applications of computational imaging, such as FPM, the forward model S is partially or completely known. Therefore, the forward imaging process is implemented using open source machine learning libraries such as TensorFlow and PyTorch. A method known as physics-based learning allows the measurement *b* to estimate the image *a* by means of unsupervised back propagation. Since the results of each image reconstruction can be estimated independently, the training set is not required for training. Based on the method of physics learning, Jiang et al. [41] introduced FPM and SUN et al. [43] improved it. The physics-based learning method combines the interpretability of the model and the automatic differentiation of the network. In addition, it does not require a separate reconstruction model to reconstruct the output image.

This paper introduces the SwinIR network that has performed well in the field of super-resolution recently and uses SwinIR as a physical model to reconstruct our data. Figure 4 shows the structure diagram of the SwinIR physical model. The SwinIR physical model includes a shallow feature extraction module, a deep feature extraction module, and a reconstruction module [2]. It is worth mentioning that the feature map of the image after the deep feature extraction module and the feature map obtained by the shallow feature extraction module are cascaded. With feature fusion, the final output to the reconstruction module, the reconstruction result is obtained [2].
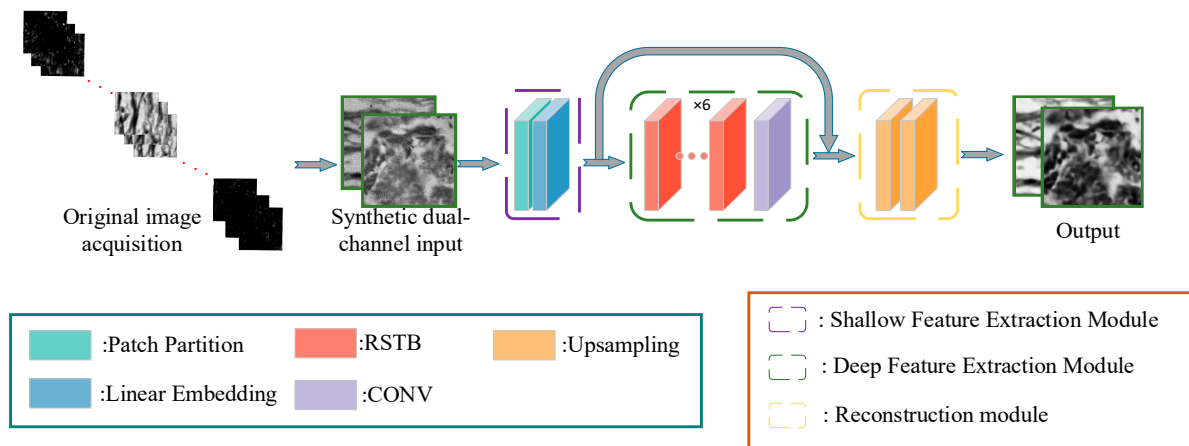
Figure 4. SwinIR physical model.

3.2. Subnetwork Structure

In the SwinIR network, the original image is pre-processed and collectively referred to as a dual-channel image, which is transmitted to the SwinIR network as intensity and phase, respectively. The reconstruction result is obtained after the shallow feature extraction module, the deep feature extraction module, and the reconstruction module.

3.2.1. Shallow Feature Extraction Module

This module mainly uses the convolution layer to extract the features of the image in order to retain the low frequency information. Given a low-resolution image I_{LQ} as input:

$$I_{LQ} \in R^{H*W*C_{in}} \tag{17}$$

H is the height of the image, W is the width of the image, and C_{in} is the number of input channels of the image. Therefore, the extracted features F_0 are:

$$F_0 = H_{SF}(I_{LQ}),$$
$$F_0 \in R^{H*W*C} \tag{18}$$

$H_{SF}(\cdot)$ is a 3×3 convolutional layer, and C is the feature channel number.

3.2.2. Deep Feature Extraction Module

The module mainly reconstructs the image loss high-frequency features based on the residual SwinTransformer block (RSTB). Each RSTB block uses multiple Swin Transformer layers to implement a local attention mechanism and cross-window interaction mechanism (Cross-Windows Interaction), and a 3×3 convolution layer is placed at the end for feature enhancement.

The deep feature extraction module of the SwinIR network mainly consists of six RSTB (Residual Swin Transformer Block) blocks. Each RSTB block uses multiple Swin Transformer layers (STL) for the local attention mechanism and cross-window interaction. At the same time, a residual connection is added to each RSTB module. The purpose of adding residual connections is to enable different levels of features to be fused together to enhance the details of the image lost in one-layer feature extraction. The structure of the RSTB is shown in Figure 5.
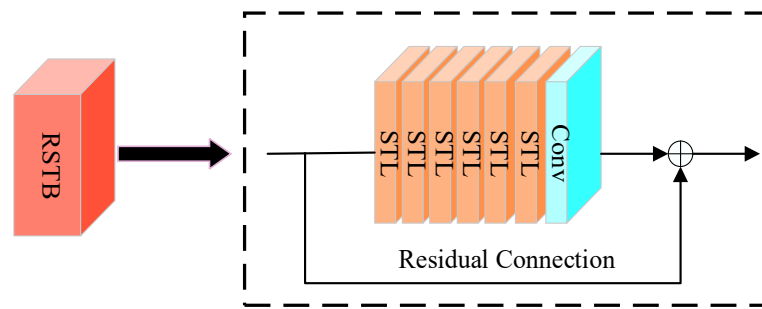
Figure 5. Residual Swin Transformer Block (RSTB).

In RSTB, there are six STL modules and a 3×3 convolution layer. The STL module is based on the traditional standard multi-head attention mechanism, but different from the STL module, the local window attention and cyclic shift are added. The local window attention is used to reduce the computational burden caused by the global attention of the traditional transformer, and the calculation of attention is limited to each window, which is still the original multi-head self-attention in each window. Generally, a 3×3 convolution layer is used to extract local adjacent information to reduce the computational cost of the network.

Two residual connections are used in the STL layer, and the nonlinear problem is solved by a multi-layer perceptron (MLP). The LayerNorm layer is added before MSA (Multi-head Self-Attention) and MLP to ensure the stability of the data feature distribution, which can use a larger learning rate and a faster convergence speed. At the same time, the LayerNorm layer also has a certain anti-overfitting effect, which makes the training process more stable and avoids the data falling into the saturation area of the activation function to reduce the problem of gradient disappearance. The structure of the STL is shown in Figure 6.
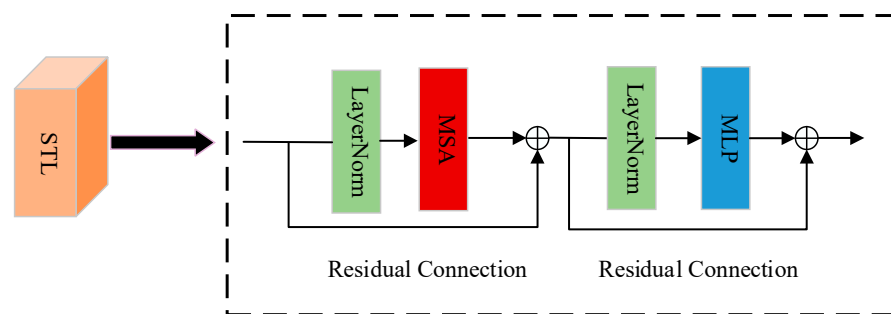
Figure 6. Swin Transformer Layer (STL).

Extracting deep features F_{DF} from shallow features F_0 occurs block by block.

$$F_{DF} = H_{DF}(F_0),$$

$$F_{DF} \in R^{H*W*C}$$

(19)

$H_{DF}(\cdot)$ is a deep feature extraction module that contains six Residual Swin Transformer blocks (RSTB) and a 3×3 convolutional layer.

The RSTB module includes a Swin Transformer layer (STL) and a convolutional layer. Given the input features of the ith module, the role of the STL layer is to extract intermediate features $F_{i,1}, F_{i,2} \ldots \ldots F_{i,L}$, namely:

$$F_{i,j} = H_{STL_{i,j}}(F_{i,j} - 1), \ j = 1, 2, \ldots \ldots, L$$

(20)

$H_{STL_{i,j}}(\cdot)$ is the jth STL layer in the ith RSTB.

The output of the RSTB module is:

$$F_{i,out} = H_{CONV_i}(F_{i,L}) + F_{i,0} \tag{21}$$

$H_{CONV_i}(\cdot)$ is the convolution layer in the ith RSTB.

3.2.3. Feature Fusion

In the whole network, the extracted low-frequency information and high-frequency information are connected by the residual method for feature fusion, as shown in Figure 7.
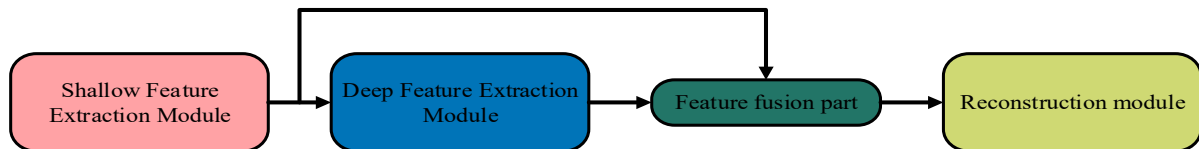
Figure 7. Feature fusion diagram.

After the deep feature extraction, the obtained feature map may lose some edge details. Therefore, the feature fusion with the residual as the main structure can fully collect and supplement the complete image information, so as to obtain a better input.

3.2.4. Reconstruction Module

In this module, the PixelShuffle convolutional layer is used to upsample the input features to increase the image size. The generated feature map passes through the reconstruction module to obtain the output of the network, thereby completing the reconstruction.

4. Experimental Results and Analysis

4.1. Construction of Data Sets

4.1.1. Dual-Channel Synthetic Input

The data set used in this paper is a series of low-resolution images collected by the FPM system, synthesized in the Fourier domain, converted into dual-channel data through the inverse Fourier transform, and then used as the input of the network. The whole process of synthesizing double inputs can be expressed by the following formula:

$$O_n(k) = \mathcal{F}\left\{\sqrt{I_{nc}(r)}\right\} H(k + k_n) + O_{n-1}(k)[1 - H(k + k_n)] \tag{22}$$

$$o(r) = \mathcal{F}^{-1}\left\{O_{N_{LED}}(k)\right\} \tag{23}$$

Here, *n* is the sequence number of the image, *n* = 1,2,..., N_{LED}. This operation can be regarded as a half-iteration process of the traditional FPM reconstruction method. *O*(*r*) is a complex number, so its intensity and phase can be transported to the SwinIR physical model as a channel. The advantage of using the above method is that after processing the data, the number of channels is greatly reduced, and the space size is increased, which can be sampled and compared at the front end of the super-resolution network, and it contains rich and effective high-frequency information. By using the synthesized dual-channel data, we can make our reconstruction better, which is equivalent to adding a residual structure to the neural network structure. The schematic diagram using a synthetic input is shown in Figure 8.
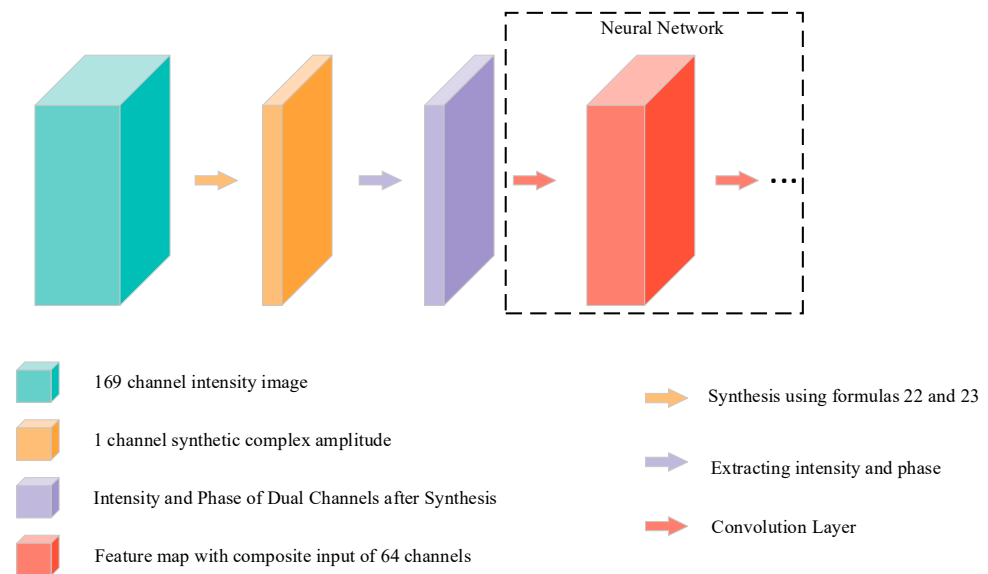
Figure 8. Synthetic input diagram.

4.1.2. Construction of Real Data Set

For the field of image super-resolution, the most common method of establishing a data set is to collect a large number of images with better accuracy, and downsample these images to obtain the corresponding true values and inputs. In Fourier ptychographic microscopy, the number of low-resolution images is closely related to the size of the LED array under the FPM system, but the original data are at least hundreds of images. If these images correspond to the neural network, there will be a multi-dimensional tensor with a certain number of channels, and the convolutional layer in the neural network needs to extract the information of the image from the input. These channels will greatly increase the number of network parameters, resulting in a too large network size and reconstruction failure.

At present, there are two ways to build FPM data sets. The first one is to directly collect the low-resolution image of FPM to generate the input of the whole network model, and then use the high-resolution complex amplitude obtained by the traditional FPM reconstruction algorithm as the output. This method of directly collecting low-resolution images and using them as input is simple and intuitive, and can cover any error encountered in actual imaging. However, this method has a series of shortcomings: (1) The way to obtain the true value is obtained by the traditional FPM reconstruction algorithm. However, the traditional reconstruction algorithm cannot escape the idea of the gradient descent algorithm in iterative updates, which has great limitations. (2) The number of data sets to be collected is too large to obtain such a large data set. The second method is to directly use the simulation image and generate the simulation results through the Fourier stack microscopic imaging model. In order to avoid the first method of collecting data sets, this paper uses the second method to construct FPM data sets.

We directly select the image as the truth value, and then generate the input corresponding to the true value through the Fourier ptychographic microscopy imaging process. Firstly, 400 high-resolution images are collected, and then 1600 sets of complex amplitudes are obtained by random combinations of phase and amplitude as the truth value of the whole network. Using these complex amplitudes, the input of the network is obtained by Fourier ptychographic microscopy imaging and the synthesis of dual-channel images. The pre-processed image is randomly cropped to obtain 25,600 sets of input data and true values. The process is shown in Figure 9.
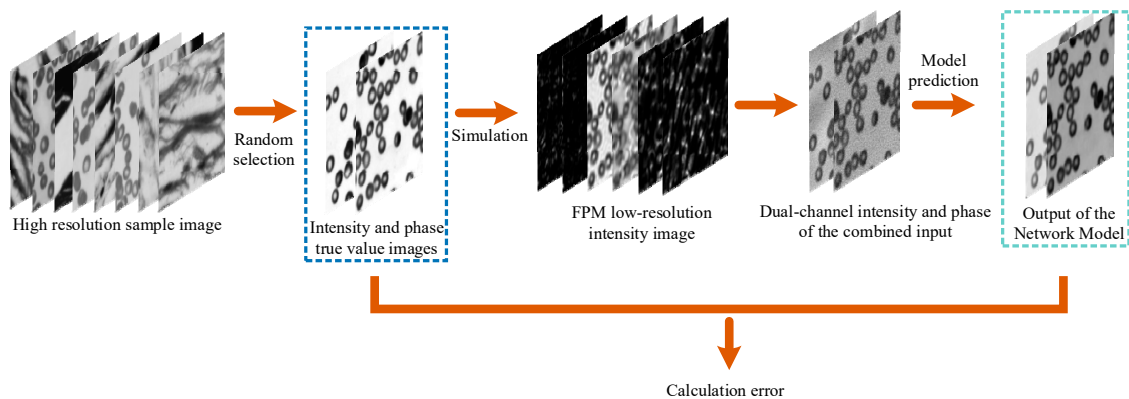
Figure 9. The construction process of real data set.

4.1.3. Experimental Data Preparation

The method based on deep learning needs to build data sets for network training and testing. In this paper, 3.1 data sets are used to build the network. The low-resolution images collected by FPM are still preprocessed and synthesized into dual-channel images, which are sent to the network as intensity and phase, respectively. In the experiment, the FPM system has an objective lens with a numerical aperture (NA) of 0.13, and a 2560×2560 pixel (6.5 um pixel size) CMOS camera is used to collect and record light intensity images. The planar array is used as a 13×13 programmable control light source element LED, and the illumination wavelength is 505 nm to provide illumination at 100 mm below the sample.

In this paper, the data sets are divided into a training set and a test set. The training set is used to train the neural network. The network calculates the error between the truth value and the network output by calculating the value of the loss function, and updates the hyperparameters of the network by back propagation and gradient descent. The test set is used to quantitatively evaluate the effect of network training during the training process. In order to prevent overfitting, the process of back propagation is turned off when using the test set.

In addition, in order to enhance the ability of the reconstructed network model to process real collected data, a data set with a mixture of real collected data and simulated data sets is built. In the mixed data set, the original low-resolution images of 50 groups of samples were first collected by a Fourier ptychographic microscopy, and the results reconstructed by the traditional phase recovery method were used as the true values. For speeding up the training of the network and achieving better training results, the images in the mixed data set are cut into fixed-size image blocks before network training, and more training data samples are obtained by reducing the cutting step size. After the above operation, 500 sets of low-resolution complex amplitude images and the corresponding truth values were finally obtained, in which the ratio of training set to test set was 9:1. The physical model training flow chart is shown in Figure 10.

4.2. The Selection of Network Hyperparameters under Noise Data

In the field of deep learning, how to find the parameters that make the loss function value as small as possible is the main problem. The process of problem solving is called optimization. Unfortunately, this problem is difficult to solve, because the parameter space is very complex and the minimum value cannot be obtained by mathematical formula, and in the deep neural network, the size of parameters is very huge, resulting in the optimization problem being more complex. It is worth noting that the objectives of optimization and deep learning are different. The objective function of the optimization algorithm is usually a loss function based on the training data set. Its goal is to reduce the training error, and the purpose of deep learning is to reduce the generalization error.
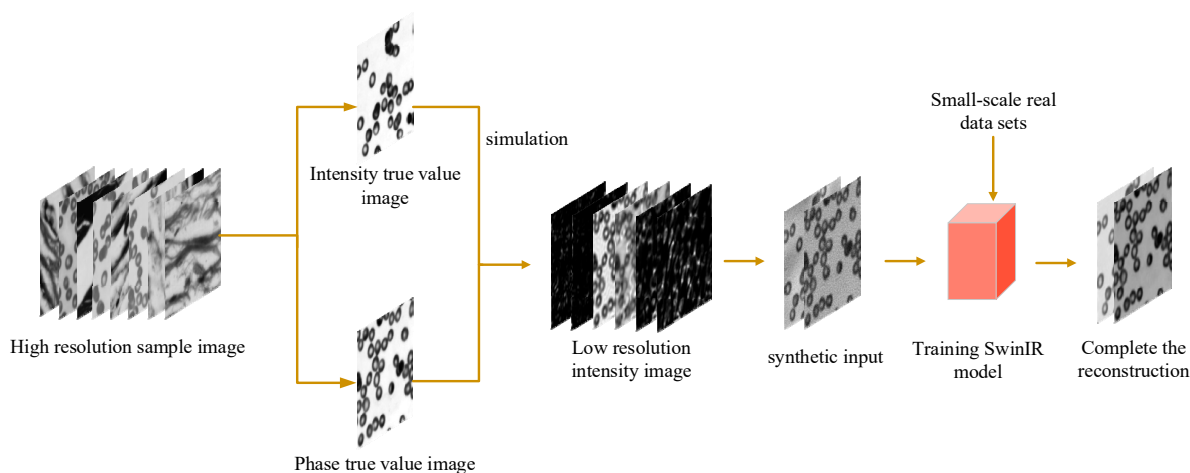
Figure 10. Physical model training flow chart.

So far, there are many ways to find the optimal parameters. For example, the stochastic gradient descent (SGD) method uses the gradient of the parameters, updates the parameters along the gradient direction, and repeats this step to gradually approach the optimal parameters. Small-batch stochastic gradient descent uses a small batch of randomly uniformly sampled samples to calculate the gradient. Usually, the time consumption of the small-batch stochastic gradient in each iteration period is between that of gradient descent and stochastic gradient descent. The momentum method uses the idea of the average exponentially weighted movement. It makes a weighted average of the gradient of the past time step, and the weight decays exponentially according to the time step. The momentum method makes the independent variable update of the adjacent time step more consistent in the direction.RMSProp adjusts the learning rate by using a small batch of random gradients to exponentially weighted moving averages of the square of the elements. The AdaDelta algorithm replaces the learning rate of the RMSProp algorithm by using the exponential weighted average moving term of the square of the independent variable update amount to achieve better results in finding the optimal parameters. Based on basis of RMSProp algorithm, Adam algorithm also makes exponential weighted moving average for small batch stochastic gradient, and uses deviation correction to find the optimal parameters faster and more accurately.

In the custom training SwinIR code, we use AdamW as the training optimizer. The AdamW optimizer directly adds the gradient of the regularization term to the back propagation, eliminating the step of manually adding the regularization term to the loss, which improves the efficiency and time of the program. The initial learning rate is set to 2×10^{-3}, and the multi-step attenuation strategy is used to update the learning rate by 0.1 times in the case of 30, 50, and 80 iterations. At the same time, in order to present the optimal reconstruction results, the selection of loss function is very important during training.

For making the model fit the training data during the training process to minimize the error, in the process of selecting the loss functions, we refer to the current loss function that has a better effect on Fourier ptychographic microscopic reconstruction, and select L2-loss, SSIM, and combined loss function L2-loss and SSIM as our comparison loss function. L2-loss as a loss function is currently the most popular choice for FPM reconstruction networks based on neural networks and deep learning directions. It can prevent overfitting and improve the generalization capability of the model. The L2-loss loss function is also called the square loss function, L2 loss, mean squared error (MSE), secondary loss, etc. Its mathematical expression is as follows:

$$MSE = mean\left((I_1 - I_2)^2\right) \tag{24}$$

I_1 represents the original image, I_2 represents the reconstructed image, and the size of the I_1 and I_2 images is the same. In addition, the basic principle of SSIM as a loss function refers to the evaluation index in Section 2.4, and the loss value is calculated from three aspects: brightness, contrast, and structure.

For verifying the impact of different loss functions on the reconstructed results, the three loss functions are used as the loss functions of the network respectively, and the Gaussian noise data set with a mean value of 0 and a standard deviation of 3×10^{-4} is trained. Visual reconstruction results, as shown in Figure 11, are the reconstruction results of different kinds of loss functions under the same noise data set. Table 1 shows the indexes of reconstruction results of varied loss functions. From the comparison of visual effects and numerical results, it can be seen that L2-loss is optimal as the main loss function of the network.
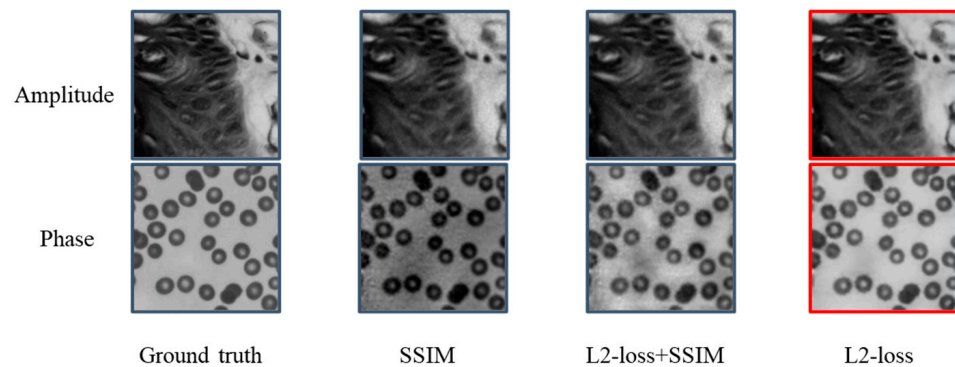
Figure 11. Reconstruction results of different loss functions under the same data set.

Table 1. Evaluation index of different loss function reconstruction results.

	L2-Loss PNSR/SSIM	SSIM PSNR/SSIM	L2-Loss + SSIM PSNR/SSIM
Amplitude	36.000/0.970	32.666/0.950	34.121/0.955
Phase	23.414/0.963	13.336/0.803	21.988/0.899

The epoch loss results of different loss functions on the same training set are shown in Figures 12–14. Among them, the abscissa is the number of the epoch, and the ordinate epoch loss represents that there are multiple iterations in an epoch. The loss values of each iteration in an epoch are added. When an epoch training is completed, the accumulated loss value is divided by the number of iterations, and finally the epoch loss value of an epoch is obtained. In order to show the gap between epoch loss, the ordinate of the right image is set to "log".
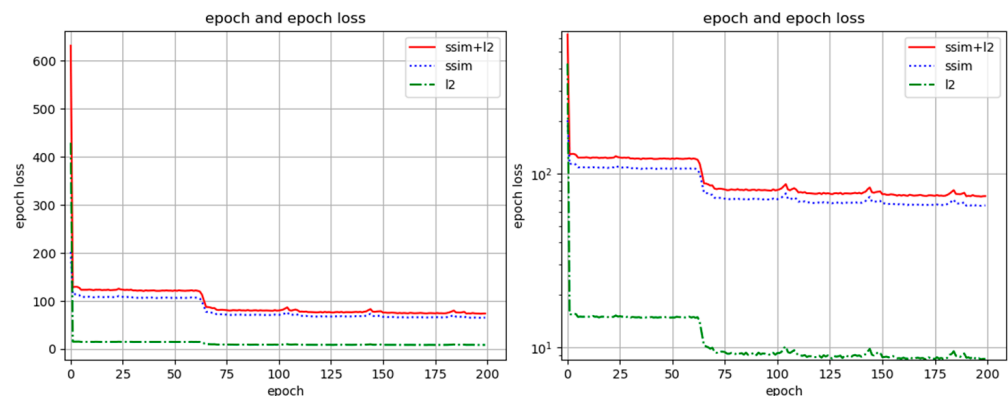
Figure 12. L2 loss function training results on 3×10^{-4} noise data set.
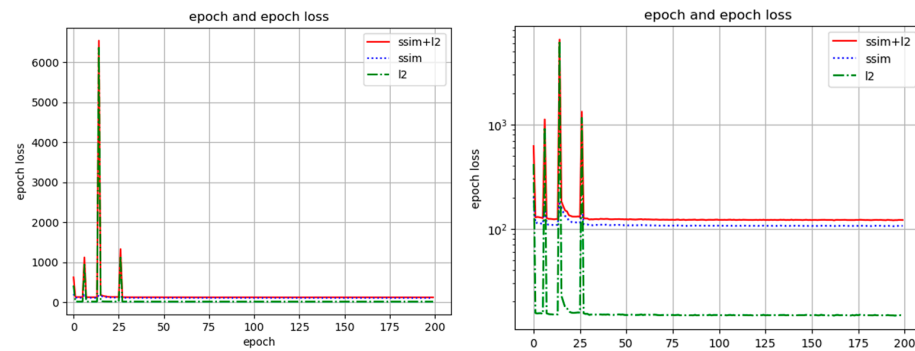
Figure 13. The training results of SSIM loss function on 3×10^{-4} noise data set.
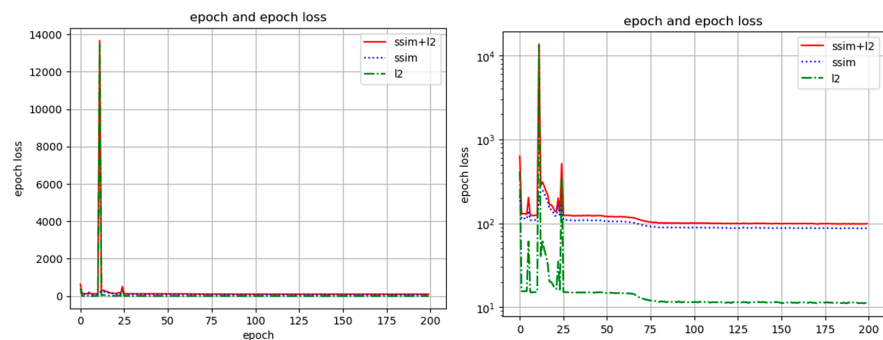
Figure 14. Training results of L2-loss and SSIM loss function on 3×10^{-4} noise dataset.

It can be seen from the results that whether L2-loss, SSIM, or SSIM and L2-loss are used as the loss function, the convergence speed and epoch loss value of L2-loss are the lowest. At the same time, in order to verify the correctness of using L2-loss as a loss function, we train it on data sets with a mean value of 0 and standard deviation of 2×10^{-4} and a mean value of 0 and standard deviation of 1×10^{-4}. The training epoch and epoch loss curves are shown in Figure 15.

(**a**) The training results of L2 loss function on 2×10^{-4} noise data set

(**b**) The training results of L2 loss function on 1×10^{-4} noise data set
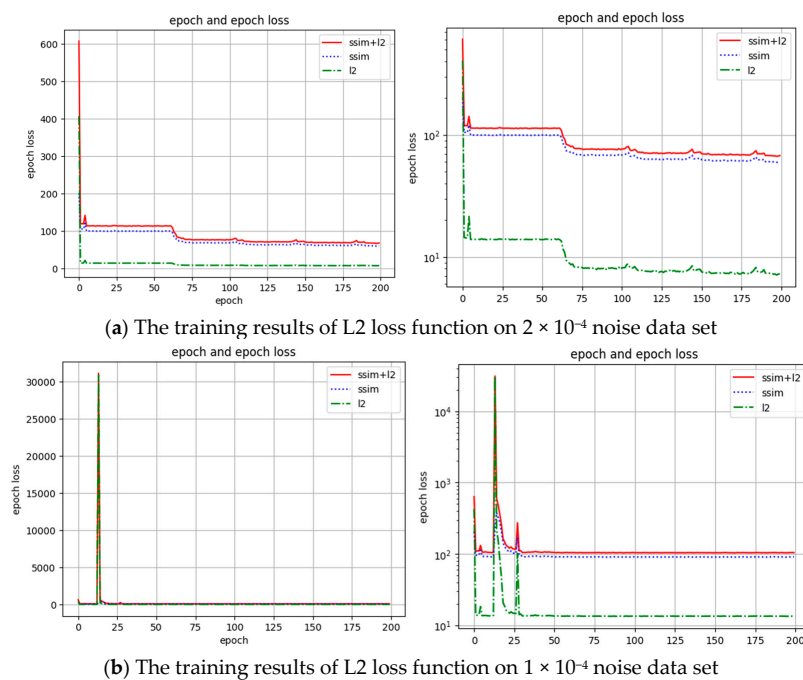
Figure 15. The training curve of L2-loss as the main loss function under 2×10^{-4} and 1×10^{-4} noise data sets.

It can be seen from Figure 15 that, on different data sets, L2-loss has the fastest convergence speed and the smallest epoch loss value as a loss function, and also has a stronger performance than the other two loss functions. Therefore, in the self-training SwinIR network, L2-loss will be used in the next experiment.

4.3. Comparison and Analysis of Reconstruction Methods under Different Noise

4.3.1. Comparison Results and Analysis of Reconstruction Methods under the SAME Noise

The true value image used in this experiment is shown in Figure 16, Including the first behavior amplitude image and the second behavior phase image. The images used in this paper are 192×192.
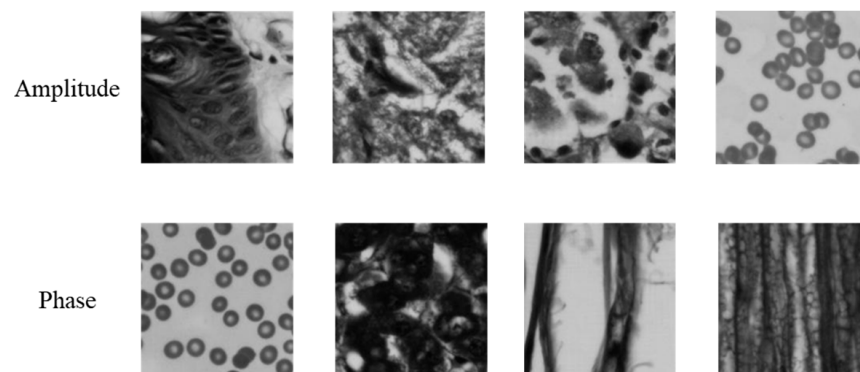
Figure 16. Truth image.

Considering the process of real reconstruction of the Fourier ptychographic microscopy imaging system, the noise effect that may occur in the actual acquisition process is simulated, because the noise source in the actual acquisition is mainly from the intensity fluctuation of the LED component, and Gaussian noise is caused by sensor noise caused by different brightness light sources in the process of image acquisition. Therefore, in the subsequent experiments, we use Gaussian noise with different standard deviations as the main interference condition of our experiment, based on which we verify the robustness and anti-interference of the experimental method in this paper.

Under the data set with Gaussian noise, the Fourier ptychographic microscopy reconstruction results are obtained by using the STPM-FPM method in this paper. The traditional phase recovery G-S method, A-S method, Jiang et al.'s method based on neural network reconstruction, and Zhang et al.'s method based on deep convolutional neural network are used as comparison algorithms to reconstruct the data set with Gaussian noise. Three sets of images are randomly selected from the test set. The reconstruction results of different reconstruction algorithms under the same noise (3×10^{-4}) are shown in Figure 17.

At the same time, PSNR and SSIM are used to compare the advantages and disadvantages of each reconstruction algorithm. The reconstructed results in Figure 17 are used as the evaluation index images of each method to obtain the evaluation index values. The results are shown in Table 2.

Table 2. Algorithm comparison results of multiple sets of images under 3×10^{-4} noise.

Number of Classes	Method	Amplitude		Phase	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM
1	G-S	25.538	0.739	7.090	0.186
	A-S	25.616	0.742	7.117	0.189
	Zhang	23.807	0.935	21.255	0.900
	Jiang	25.126	0.812	19.467	0.886
	STPM-FPM	36.000	0.970	23.414	0.963

Table 2. *Cont.*

Number of Classes	Method	Amplitude		Phase	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM
2	G-S	24.111	0.764	18.379	0.435
	A-S	24.097	0.766	18.439	0.439
	Zhang	22.590	0.939	9.200	0.529
	Jiang	27.202	0.800	28.273	0.698
	STPM-FPM	32.761	0.958	30.922	0.955
3	G-S	23.945	0.718	7.015	0.178
	A-S	23.985	0.718	7.013	0.181
	Zhang	22.101	0.936	22.957	0.922
	Jiang	17.888	0.753	12.654	0.679
	STPM-FPM	34.524	0.974	24.887	0.957
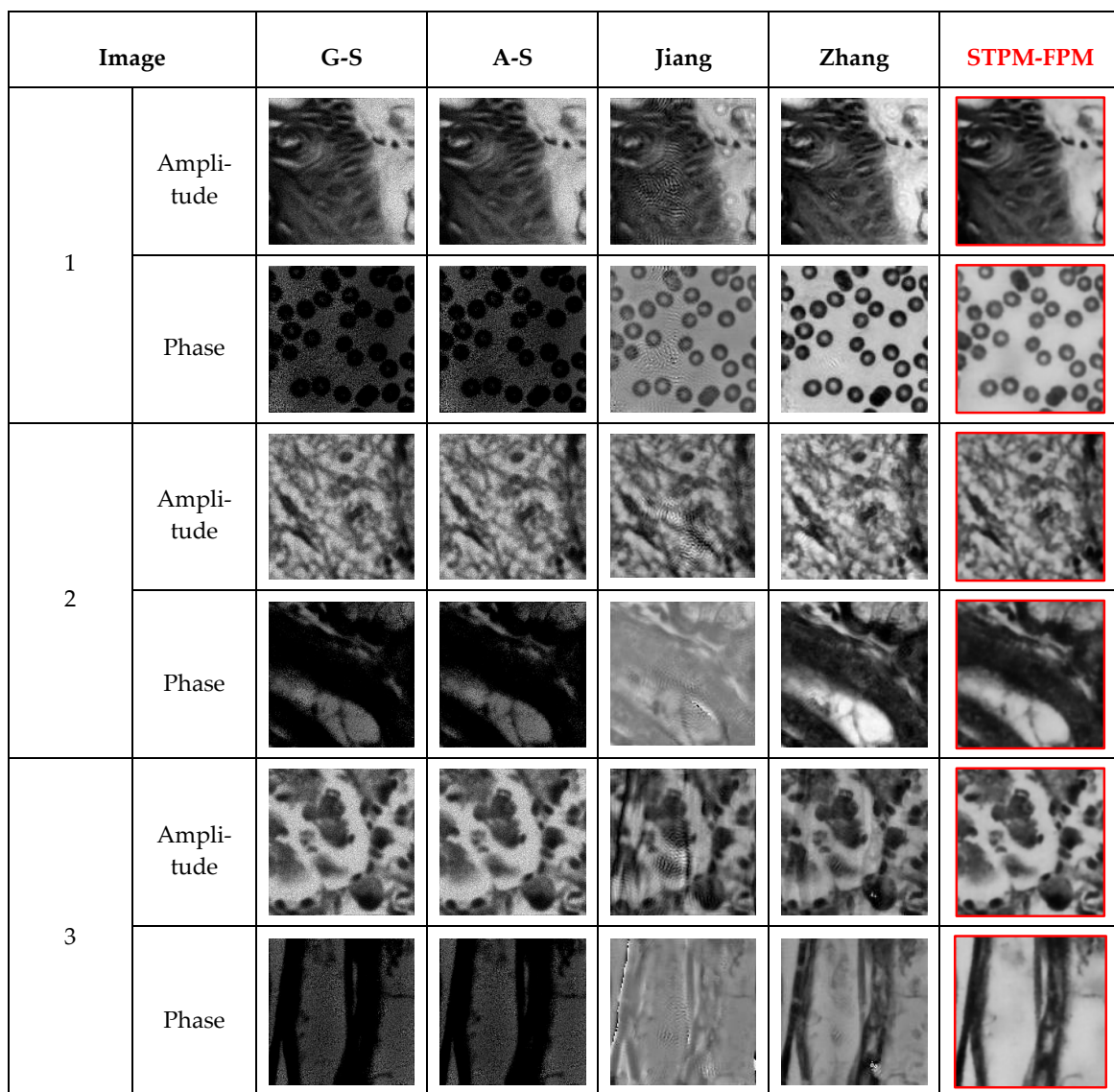
Figure 17. Reconstruction results of the same reconstruction algorithm under the same noise (3×10^{-4}).

From Figure 17 and Table 2, it can be seen that the Fourier ptychographic microscopy reconstruction method using the self-training physical model has a good reconstruction effect and reconstruction index value. The reconstructed image results have higher clarity

than the phase reconstruction results of G-S and A-S methods. Compared with the methods of Jiang et al. and Zhang et al., the experimental results of this paper have no obvious artifacts, and contain more texture details and have reduced error. In Table 2, the result of the red part is the best, and the blue part is the second best. For the reconstruction index value, the proposed method is better than the other four methods, while Zhang's method has sub-optimal performance.

4.3.2. Comparison and Analysis of Reconstruction Methods under Different Noises

In the process of real reconstruction, the interference encountered in the experiment is different, so we simulate the reconstruction results and reconstruction indexes of the same set of images under different noise conditions. Among them, the noise used is 1×10^{-4}, 2×10^{-4}, and 3×10^{-4} respectively. Figure 18 shows the reconstruction results of the same set of images under different noise conditions. Table 3 is the evaluation index value of the same set of images under varied noise conditions.
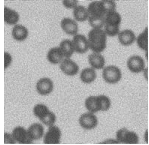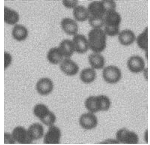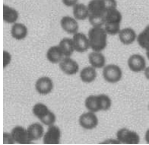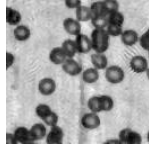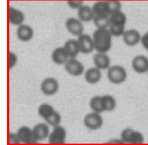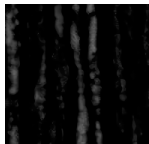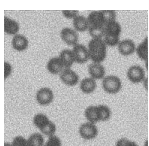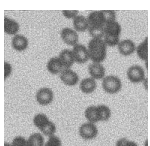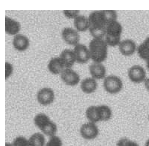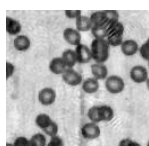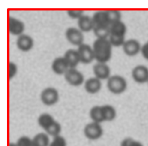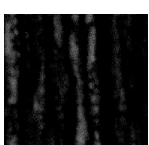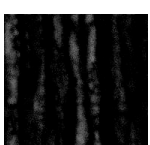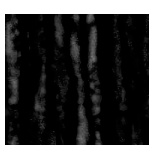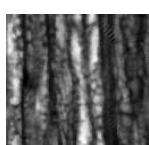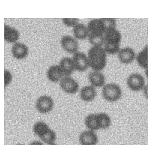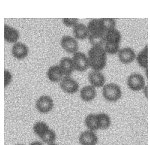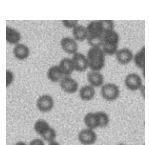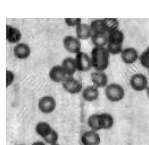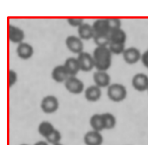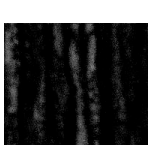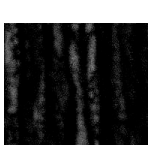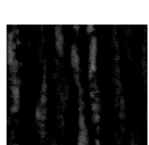
Image		G-S	A-S	Jiang	Zhang	STPM-FPM
1×10^{-4}	Ampli-tude					
	Phase					
2×10^{-4}	Ampli-tude					
	Phase					
3×10^{-4}	Ampli-tude					
	Phase					

Figure 18. Reconstruction results of the same group of images under different noise conditions.

Table 3. Reconstruction index of the same group of images under different noise conditions.

Noise Size	Method	Amplitude		Phase	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM
1×10^{-4}	G-S	27.978	0.794	12.052	0.228
	A-S	27.970	0.790	12.046	0.227
	Zhang	21.067	0.845	23.605	0.936
	Jiang	18.209	0.767	19.909	0.854
	STPM-FPM	33.861	0.971	26.093	0.951
2×10^{-4}	G-S	24.492	0.650	12.295	0.246
	A-S	24.468	0.647	12.280	0.246
	Zhang	21.083	0.849	23.534	0.935
	Jiang	18.209	0.767	19.909	0.854
	STPM-FPM	34.649	0.965	26.787	0.947
3×10^{-4}	G-S	22.693	0.562	12.497	0.262
	A-S	22.683	0.562	12.490	0.261
	Zhang	20.334	0.838	24.100	0.933
	Jiang	18.209	0.767	19.909	0.854
	STPM-FPM	36.949	0.982	25.889	0.943

According to Figure 18, STPM-FPM is visually superior to the other three reconstruction methods in this paper. Moreover, the STPM-FPM method has strong robustness to different noise data sets, which proves the generalization of the network and improves the quality of reconstruction.

Table 3 uses the method of the deep learning network to analyze the objective evaluation results and study the noise resistance of various methods. From the quantitative analysis results in Table 3, it can be clearly seen that the anti-noise performance of STPM-FPM is obviously due to other methods. The result is consistent with that in Figure 18.

4.3.3. Time Comparison Results of Reconstruction Methods

This paper adopts a network structure based on deep learning, so there is no need for iterative updates and less reconstruction time. At the same time, we compare the time of the STPM-FPM reconstruction image. Among them, the methods used are still based on the traditional methods, G-S, A-S, Jiang et al. based on neural networks, and Zhang et al. based on deep learning, as shown in Table 4.

Table 4. Time comparison of different reconstruction algorithms.

Reconstruction Method	Iteration Times	Reconstruction Time
G-S	50	2.170 s
A-S	50	2.744 s
Jiang et al.	20	33.503 s
Zhang et al.	20	204.304 s
STPM-FPM	0	0.139 s

4.3.4. Real Image Analysis

For verifying the advantages and disadvantages of the reconstructed results of the STPM-FPM method in this paper under the data collected on real devices, this section uses a fine-tuned data set with small samples to test the reconstruction results, and the reconstruction results are compared and analyzed. Four real collected intensity images of blood cell smear, penis tissue section, alveolar tissue section and testis tissue section were selected as reconstructed contrast images. The traditional reconstruction methods G-S, A-S, Jiang et al.'s method based on neural networks, and Zhang et al.'s method based on deep convolutional neural network are selected for comparison and analysis. In the process of

using Fourier ptychographic microscopy to collect real data, only the low-resolution image of the original image sample can be collected, and there is no true value of the intensity image and the phase image. Therefore, this experiment can only be qualitatively evaluated by comparing with the reconstruction results of the four methods.

As shown in Figure 19, according to the results, under the premise of adding some real collected data, the experimental method in this paper still has better reconstruction performance and visual effect. Compared with the traditional algorithm, the STPM-FPM algorithm has a more obvious effect of removing artifacts. At the same time, it has better reconstruction result clarity from the naked eye. The background error of the reconstructed image is relatively less, and there are clearer texture details, which proves the generalization and robustness of STPM-FPM in the real acquisition data set. Thanks to the end-to-end processing based on the deep learning method and the acceleration of the image processor, the operation rate of the proposed method is higher than that of the other three comparison algorithms, which is helpful for the practical application of Fourier ptychographic imaging under high temporal resolution.
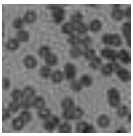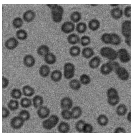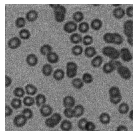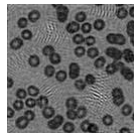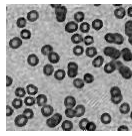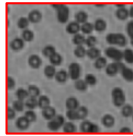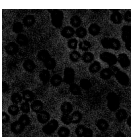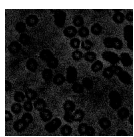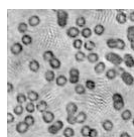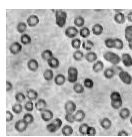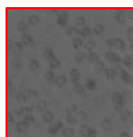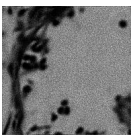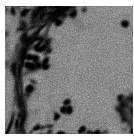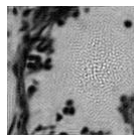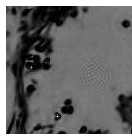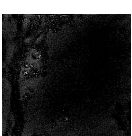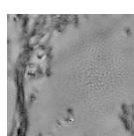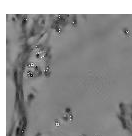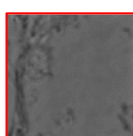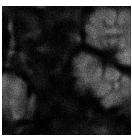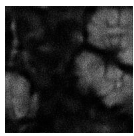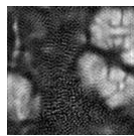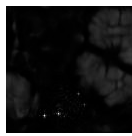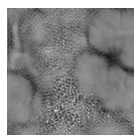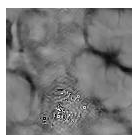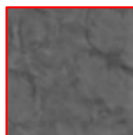
Figure 19. *Cont.*

Figure 19. Reconstruction results under real data acquisition.

5. Conclusions

According to this article, we propose a self-training physical model of the Fourier ptychographic microscopy reconstruction method (STPM-FPM). The reconstruction method trains the weight of the physical model with a custom training code, and then completes the reconstruction. When training the network, the effects of different loss functions on the data set collected by the Fourier ptychographic microscopy were compared, and the loss function suitable for this network was selected. In the experimental stage, the validity, generalization, and robustness of the reconstruction method in this paper were verified by using the same noise data set, a different noise data set, and a fine-tuning data set with some real data. The classical reconstruction algorithms G-S, A-S, Jiang et al.'s method based on neural networks, and Zhang et al.'s method based on deep convolutional neural networks were compared. The reconstruction qua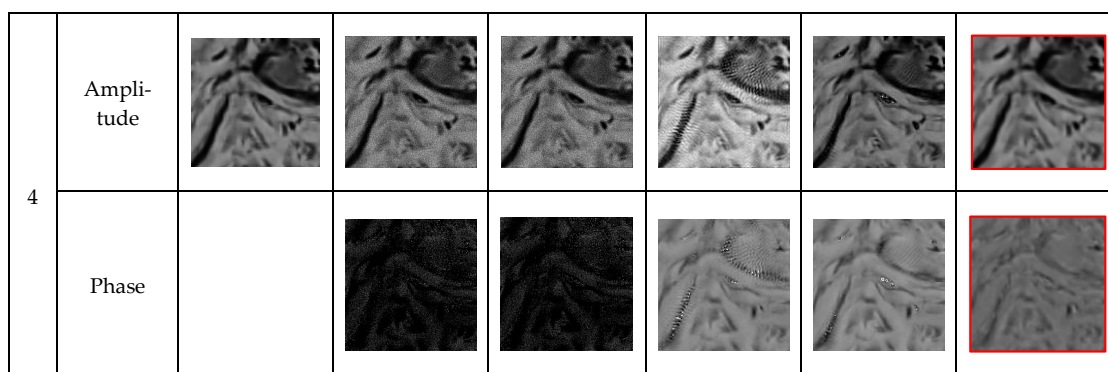lity of the reconstruction algorithm was proved through this article, and two evaluation indexes were used to quantitatively and qualitatively analyze the reconstruction results from different perspectives.

The method used in this paper is the FPM reconstruction method of a self-training physical model, which effectively solves the decisive factors of different network models for reconstruction quality. At the same time, the idea of transfer learning and deep learning can fully improve the quality of reconstructed images. Compared with other traditional and neural network-based reconstruction algorithms, this paper has a more obvious denoising effect on noise. However, the deep learning reconstruction method is affected by its own reconstruction network, such as the adjustment of hyperparameters, the selection of loss function, the structure of the network, and other factors. Therefore, how to improve the stability of the deep learning reconstruction method remains to be studied. At the same time, the cycle time of training based on the deep learning method is long, and how to lighten the model is one of the key research directions for the future.

Author Contributions: Conceptualization, X.W., Y.P., T.X. and J.L.; methodology, X.W. and Y.J.; validation, X.W.; formal analysis, X.W.; investigation, X.W.; resources, X.W.; data curation, X.W. and Y.J.; writing—original draft preparation, X.W.; writing—review and editing, X.W., Y.P., Y.J., J.L., Z.L. and J.C.; visualization, X.W. and Y.J.; supervision, T.X.; project administration, X.W. and J.L.; funding acquisition, Y.P. All authors have read and agreed to the published version of the manuscript.

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

1. Zheng, G.; Horstmeyer, R.; Yang, C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nat. Photonics* **2013**, *7*, 739–745. [CrossRef] [PubMed]
2. Hillman, T.R.; Gutzler, T.; Alexandrov, S.A.; Sampson, D.D. High-resolution, wide-field object reconstruction with synthetic aperture Fourier holographic optical microscopy. *Opt. Express* **2009**, *17*, 7873–7892. [CrossRef] [PubMed]
3. Gutzler, T.; Hillman, T.R.; Alexandrov, S.A.; Sampson, D.D. Coherent aperture-synthesis, wide-field, highresolution holographic microscopy of biological tissue. *Opt. Lett.* **2010**, *35*, 1136–1138. [CrossRef]
4. Granero, L.; Micó, V.; Zalevsky, Z.; García, J. Synthetic aperture superresolved microscopy in digital lensless Fourier holography by time and angular multiplexing of the object information. *Appl. Opt.* **2010**, *49*, 845–857. [CrossRef] [PubMed]
5. Hilaire, P.S.; Benton, S.A.; Lucente, M.E. Synthetic aperture holography: A novel approach to threedimensional displays. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **1992**, *9*, 1969–1977. [CrossRef]
6. Binet, R.; Colineau, J.; Lehureau, J.-C. Short-range synthetic aperture imaging at 633 nm by digital holography. *Appl. Opt.* **2002**, *41*, 4775–4782. [CrossRef]
7. García, J.; Zalevsky, Z.; Fixler, D. Synthetic aperture superresolution by speckle pattern projection. *Opt. Express* **2005**, *13*, 6073–6078. [CrossRef]
8. Alexandrov, S.A.; Hillman, T.R.; Gutzler, T.; Sampson, D.D. Synthetic Aperture Fourier Holographic Optical Microscopy. *Phys. Rev. Lett.* **2006**, *97*, 168102. [CrossRef]
9. Mico, V.; Zalevsky, Z.; García-Martínez, P.; García, J. Synthetic aperture superresolution with multiple off-axis holograms. *J. Opt. Soc. Am. A* **2006**, *23*, 3162–3170. [CrossRef]
10. Mico, V.; Zalevsky, Z.; García, J. Synthetic aperture microscopy using off-axis illumination and polarization coding. *Opt. Commun.* **2007**, *276*, 209–217. [CrossRef]
11. Martínez-León, L.; Javidi, B. Synthetic aperture single-exposure on-axis digital holography. *Opt. Express* **2008**, *16*, 161–169. [CrossRef] [PubMed]
12. Maiden, A.; Rodenburg, J.M. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy* **2009**, *109*, 1256–1262. [CrossRef] [PubMed]
13. Candes, E.J.; Strohmer, T.; Voroninski, V. PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Commun. Pure Appl. Math.* **2013**, *66*, 1241–1274. [CrossRef]
14. Fienup, J.R. Phase retrieval algorithms: A personal tour [Invited]. *Appl. Opt.* **2013**, *52*, 45. [CrossRef]
15. Gerchberg, R.W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **1972**, *35*, 237–246.
16. Fienup, J.R. Phase retrieval algorithms: A comparison. *Appl. Opt.* **1982**, *21*, 2758–2769. [CrossRef]
17. Gonsalves, R.A. Phase Retrieval and Diversity in Adaptive Optics. *Opt. Eng.* **1982**, *21*, 215829. [CrossRef]
18. Millane, R.P. Phase retrieval in crystallography and optics. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **1990**, *7*, 394–411. [CrossRef]
19. Refregier, P.; Javidi, B. Optical image encryption based on input plane and Fourier plane random encoding. *Opt. Lett.* **1995**, *20*, 767–769. [CrossRef]
20. Miao, J.; Sayre, D.; Chapman, H.N. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A* **1998**, *15*, 1662–1669. [CrossRef]
21. Paganin, D.; Mayo, S.C.; Gureyev, T.E.; Miller, P.R.; Wilkins, S.W. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J. Microsc.* **2002**, *206*, 33–40. [CrossRef] [PubMed]
22. Pfeiffer, F.; Weitkamp, T.; Bunk, O.; David, C. Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. *Nat. Phys.* **2006**, *2*, 258–261. [CrossRef]
23. Rodenburg, J.M.; Hurst, A.C.; Cullis, A.G.; Dobson, B.R.; Pfeiffer, F.; Bunk, O.; David, C.; Jefimovs, K.; Johnson, I. Hard-X-ray lensless imaging of extended objects. *Phys. Rev. Lett.* **2007**, *98*, 17–21. [CrossRef]
24. Deb, K.; Kalyanmoy, D. *Multi-Objective Optimization Using Evolutionary Algorithms*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
25. De Castro, L.N.; Von Zuben, F.J. Learning and optimization using the clonal selection principle. *IEEE Trans. Evol. Comput.* **2002**, *6*, 239–251. [CrossRef]
26. Hsu, C.-W.; Lin, C.-J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [CrossRef]
27. Papadimitriou, C.H.; Steiglitz, K. *Combinatorial Optimization: Algorithms and Complexity*; Prentice-Hall, Inc.: Hoboken, NJ, USA, 1982.
28. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [CrossRef] [PubMed]
29. Hopfield, J.J.; Tank D, W. Neural computation of decisions in optimization problems. *Biol. Cybern.* **1985**, *52*, 141–152. [CrossRef] [PubMed]
30. Dorigo, M.; Maniezzo, V.; Colorni, A. Ant system: Optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **1996**, *26*, 29–41. [CrossRef]
31. Wolpert, D.H.; Macready W, G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]

32. Osuna, E.; Freund, R.; Girosit, F. Training support vector machines: An application to face detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 130–136. [CrossRef]

33. Zitzler, E.; Thiele, L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Trans. Evol. Comput.* **1999**, *3*, 257–271. [CrossRef]

34. Bertsekas, D.; Tsitsiklis, J. *Parallel and Distributed Computation*; Athena Scientific: Belmont, MA, USA, 1989.

35. Deb, K.; Deb, K. Multi-objective Optimization. In *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*; Burke, E.K., Kendall, G., Eds.; Springer: Boston, MA, USA, 2014; pp. 403–449.

36. Luenberger, D.G.; Ye, Y. *Linear and Nonlinear Programming*; Springer Publishing Company, Incorporated: Berlin/Heidelberg, Germany, 2015.

37. Gerchberg, R.; Saxton, W. Phase determination for image and diffraction plane pictures in the electron microscope. *Optik* **1971**, *34*, 275–284.

38. Zuo, C.; Sun, J.; Chen, Q. Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy. *Opt. Express* **2016**, *24*, 20724–20744. [CrossRef] [PubMed]

39. 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]

40. Tian, L.; Li, X.; Ramchandran, K.; Waller, L. Multiplexed coded illumination for Fourier Ptychography with an LED array microscope. *Biomed. Opt. Express* **2014**, *5*, 2376–2389. [CrossRef]

41. Jiang, S.; Guo, K.; Liao, J.; Zheng, G. Solving Fourier ptychographic imaging problems via neural network modeling and TensorFlow. *Biomed. Opt. Express* **2018**, *9*, 3306–3319. [CrossRef] [PubMed]

42. Zhang, Y.; Liu, Y.; Jiang, S.; Dixit, K.; Song, P.; Zhang, X.; Ji, X.; Li, X. Neural network model assisted Fourier ptychography with Zernike aberration recovery and total variation constraint. *J. Biomed. Opt.* **2021**, *26*, 036502. [CrossRef]

43. Sun, M.; Shao, L.; Zhu, Y.; Zhang, Y.; Wang, S.; Wang, Y.; Diao, Z.; Li, D.; Mu, Q.; Xuan, L. Double-flow convolutional neural network for rapid large field of view Fourier ptychographic reconstruction. *J. Biophotonics* **2021**, *14*, e202000444. [CrossRef]

44. Zhang, J.; Xu, T.; Li, J.; Zhang, Y.; Jiang, S.; Chen, Y.; Zhang, J. Physics-based learning with channel attention for Fourier ptychographic microscopy. *J. Biophotonics* **2022**, *15*, e202100296. [CrossRef]

45. 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. [CrossRef]

46. Wang, X.; Piao, Y.; Jin, Y.; Li, J.; Han, Q.; Yu, J. A Fourier Ptychographic Microscopy Reconstruction Method Based on SwinIR Physical Model. *J. Phys. Conf. Ser.* **2022**, *2400*, 012008. [CrossRef]

47. Cong, P.; Zhou, J.; Li, S.; Lv, K.; Feng, H. Citrus Tree Crown Segmentation of Orchard Spraying Robot Based on RGB-D Image and Improved Mask R-CNN. *Appl. Sci.* **2022**, *13*, 164. [CrossRef]

48. Song, G.; Wang, Z. An Efficient Hidden Markov Model with Periodic Recurrent Neural Network Observer for Music Beat Tracking. *Electronics* **2022**, *11*, 4186. [CrossRef]

49. Peng, C.; Wu, J.; Wang, Q.; Gui, W.; Tang, Z. Remaining Useful Life Prediction Using Dual-Channel LSTM with Time Feature and Its Difference. *Entropy* **2022**, *24*, 1818. [CrossRef] [PubMed]

50. Novitasari, D.C.R.; Fatmawati, F.; Hendradi, R.; Rohayani, H.; Nariswari, R.; Arnita, A.; Hadi, M.I.; Saputra, R.A.; Primadewi, A. Image Fundus Classification System for Diabetic Retinopathy Stage Detection Using Hybrid CNN-DELM. *Big Data Cogn. Comput.* **2022**, *6*, 146. [CrossRef]

51. Zuo, C.; Li, J.; Sun, J.; Fan, Y.; Zhang, J.; Lu, L.; Zhang, R.; Wang, B.; Huang, L.; Chen, Q. Transport of intensity equation: A tutorial. *Opt. Lasers Eng.* **2020**, *135*, 106187. [CrossRef]

52. Kovalev, M.; Gritsenko, I.; Stsepuro, N.; Nosov, P.; Krasin, G.; Kudryashov, S. Reconstructing the Spatial Parameters of a Laser Beam Using the Transport-of-Intensity Equation. *Sensors* **2022**, *22*, 1765. [CrossRef] [PubMed]

53. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 1833–1844.