

Performance Characterization of Deep-Phase-Retrieval Shack-Hartmann Wavefront Sensors

Manting Zhang¹ and Youming Guo¹

Abstract—Shack-Hartmann wavefront sensor (SHWFS) is the most popular wavefront sensor in adaptive optics systems. The Deep-Phase-Retrieval Wavefront Reconstruction (DPRWR) method, which was proposed by our group previously, is a kind of deep learning-based wavefront reconstruction method. It can extract more information from the SHWFS images to accurately obtain more Zernike mode coefficients. However, the application limits, performance upper bound, and noise immunity have not been investigated in detail in previous reports. In this paper, sub-aperture spot sampling, bit depth, number of reconstructed mode coefficients, and noise intensities are analyzed by simulations and experiments to investigate the influence of changes in these parameters on the performance of DPRWR. This work aims to optimize the configuration of DPRWR for better measurement accuracy, spatial resolution, and robustness.

Index Terms—Adaptive optics, Shack-Hartmann, wavefront reconstruction, deep-phase-retrieval.

I. INTRODUCTION

ADAPTIVE optics (AO) is to solve the image degradation caused by atmospheric turbulence. At present, almost all ground-based telescopes in the world are equipped with AO systems [1], [2]. AO system mainly consists of two modules, first measuring the aberrated wavefront phase, then compensating the measured wavefront in reverse by employing a deformable mirror (DM) or a spatial light modulator (SLM) for aberration correction. Therefore, the accuracy of the wavefront measurement determines the upper bound of phase correction accuracy.

Manuscript received 30 January 2023; revised 8 March 2023; accepted 9 March 2023. Date of publication 20 March 2023; date of current version 6 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 12173041, in part by the Youth Innovation Promotion Association, Chinese Academy of Sciences under Grant 2020376, and in part by the Frontier Research Fund of Institute of Optics and Electronics, Chinese Academy of Sciences under Grant C21K002. (*Corresponding author: Youming Guo.*)

Manting Zhang is with the Key Laboratory on Adaptive Optics, Chinese Academy of Science, Chengdu 610209, China, with the Institute of Optics and Electronics, Chinese Academy of Science, Chengdu 610209, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangmanting20@mails.ucas.ac.cn).

Youming Guo is with the Key Laboratory on Adaptive Optics, Chinese Academy of Science, Chengdu 610209, China, and also with the Institute of Optics and Electronics, Chinese Academy of Science, Chengdu 610209, China (e-mail: guoyouming@ioe.ac.cn).

Digital Object Identifier 10.1109/JPHOT.2023.3256082

The major wavefront measurement methods can be divided into two classes. One group of methods is the wavefront sensorless method, such as phase diversity, which iteratively estimates wavefront from two degraded images acquired by focal cameras without additional wavefront sensors [3], [4]. This approach is capable of precisely measuring low-order aberrations through iterative solutions [5]. Due to its quite time-consuming calculation process, it is difficult to meet the real-time correction requirements of AO systems. Although subsequent work proposed a novel real-time non-iterative phase-diversity wavefront sensing by using a phase diversity convolutional neural network, it still focuses on low-order aberrations measurement [6].

The other class of methods utilizes wavefront sensor measure wavefront. In real-time AO systems, the most widely applied wavefront sensor is the Shack-Hartmann wavefront sensor (SHWFS), which has a simple structure consisting of a microlens array for wavefront segmentation and a CCD array for spot position measurement. Each sub-lens corresponds to a sub-aperture spot on the sensor. The traditional wavefront reconstruction algorithm, called the slope-based method, calculates the wavefront mode coefficients by multiplying the sub-aperture spots displacements vector with a reconstruction matrix generally obtained by singular value decomposition [7], [8], [9]. Nevertheless, when the mode number continues to increase, the condition number of the reconstructed matrix would increase quickly. Thus the number of reconstructed mode coefficients in this approach is typically about 0.7 to 0.8 times the number of sub-apertures [10], [11], [12]. In addition, due to the similarity of the slope vectors, modal coupling and modal confusion errors are likely to occur, leading to a decrease in the accuracy of wavefront measurements [13].

To further improve the wavefront measurement spatial resolution of SHWFS. Phase retrieval by SHWFS is proposed. Li et al. proposed to place the detector in the microlens' defocused plane and reconstruct the wavefront aberration by a stochastic parallel gradient descent algorithm. This method improves the measurement accuracy of SHWFS sensors for higher-order aberrations [14]. Vieggers et al. proposed the SABRE-M method to reconstruct the wavefront aberration using the first- and second-order moments of the sub-aperture image. The simulation results showed that the measurement accuracy of the first-order moment method could be achieved using half of the sub-aperture sampling rate [15]. However, such conventional

phase retrieval by SHWFS requires iterative computation, which is time-consuming and tends to fall into local optima [16].

In recent years, the neural network has led to new solution paths, from “solving ill-posed inverse problems” to “learning pseudo-inverse mapping” [17]. Researchers have proposed to establish a nonlinear mapping relationship between slope and aberration by using simple artificial neural networks [18], [19], [20]. These methods are capable to reduce the wavefront measurement error compared to the slope-based method, yet the network is difficult to converge because of large slope measurement errors under strong turbulence. To cope with strong turbulence or strong noise, some researchers have proposed methods to compute the spot center-of-mass position or classify and identify sub-aperture spots with the aid of neural networks [21], [22], [23].

Furthermore, Zernike mode coefficients can be reconstructed by convolutional neural networks (CNN) directly from the SHWFS image. For instance, Hu et al. proposed the learning-based Shack-Hartmann wavefront sensor and achieved the measurement of the 2nd-120th Zernike mode coefficients [24]. In our previous works, the Deep-Phase-Retrieval Wavefront Reconstruction method (DPRWR) was proposed, which can achieve high speed and high spatial resolution wavefront measurement and obtain the prediction of the 2nd-299th Zernike mode coefficients by simulation [25].

The slope-based wavefront reconstruction method recovers the wavefront by calculating the tilt of the SHWFS sub-aperture spot center of mass. When the spatial frequency of the aberration is less than or equal to the SHWFS sub-aperture spatial sampling rate, the wavefront within the sub-aperture can be approximated as a plane wave. In this case, there is almost only tilted aberration in the sub-aperture. Thus, the reconstruction error of the conventional slope-based wavefront method is small. When the aberration spatial frequency is much larger than the SHWFS sub-aperture spatial sampling rate, there are defocused and higher-order aberrations in the sub-aperture wavefront. The shape of the sub-aperture spot diffusion function changes and the traditional slope-based method can neither accurately extract the wavefront tilt nor extract the higher-order aberration information within the sub-aperture. However, DPRWR can reconstruct the higher-order aberrations based on the sub-aperture spot position and morphological features. Thus the higher accuracy wavefront reconstruction can be achieved by DPRWR. Compared with conventional phase retrieval by SHWFS, although the DPRWR method takes a long time to obtain the dataset and train the network, the wavefront reconstruction for a single frame of SHWFS image is estimated to take only about 0.8 ms by the trained network [25]. DPRWR can meet the need for a real-time AO system.

However, the effects of SHWFS configuration parameters on the DPRWR performance are not clear for configuration optimization. The configuration constraints, spatial resolution limits, and robustness of the DPRWR are not investigated.

In this paper, we evaluate the influence of spot sampling rate, bit depth, number of reconstructed modes, and noise on the performance of DPRWR. The SHWFS with a higher spot sampling rate means that there are more pixels of each sub-aperture spot

and the spot shape would be clearer. Besides, the higher the bit depth, the more it can reflect the grayscale details of the spot. How these input features after passing through the network layers impact the output need to be analyzed and verified by specific experiments, which are done in Sections III-A and III-B of this paper. A relevant experimental analysis of the performance upper bound related to the number of reconstructed modes is done in Section III-C. In addition, low signal-to-noise SHWFS images due to strong background interference can severely limit the working capability and the range of use of AO systems [26]. How to improve the anti-interference ability of wavefront sensing methods has been continuously investigated [27]. Therefore, we compare the robustness of DPRWR under four noise processing schemes designed based on neural network characteristics in Section III-D. Finally, an AO system using SLM was built to verify the above analysis, which is described in detail in Section IV.

II. METHODS

The simulation of SHWFS images is carried out on the Matlab platform, and its principle is described in detail in Section II-A. The SHWFS configuration parameters are consistent in simulations and experiments. The DPRWR network structure is written based on the Keras framework and is described in detail in Section II-B. To reduce the network training time, we use NVIDIA GTX 1080Ti to help speed up the gradient computation process.

A. SHWFS Simulation Principle

Based on optical diffraction theory, the imaging process of SHWFS is simulated, the optical field modulated by atmospheric turbulence focusing on the focal plane of the microlens array. The SHWFS used in this simulation is composed of 91 microlenses, as Fig. 1(a). Each sub-aperture spot image can be obtained by utilizing Fraunhofer diffraction integral equation [28].

$$F(x_2, y_2) = \left| \frac{e^{i\frac{\pi(x_2^2+y_2^2)}{\lambda f}}}{i\lambda f} \cdot \int \int U_0(x_1, y_1) \cdot e^{-i\frac{2\pi}{\lambda f}(x_2x_1+y_2y_1)} dx_1 dy_1 \right|^2 \quad (1)$$

$U_0(x_1, y_1)$ is the complex amplitude of the sub-aperture source optical field,

$$U_0(x_1, y_1) = A(x_1, y_1) e^{i\varphi(x_1, y_1)} \quad (2)$$

Where $A(x_1, y_1)$ is the sub-lens pupil function, which is generally a fixed amplitude distribution conforming to the actual conditions, and it conforms to the uniform distribution in this simulation. $\varphi(x_1, y_1)$ is the phase function conforming to the Kolmogorov turbulence [29],

$$\varphi(x_1, y_1) = \sum_{k=1}^K a_j(k) M_k(x_1, y_1) \quad (3)$$

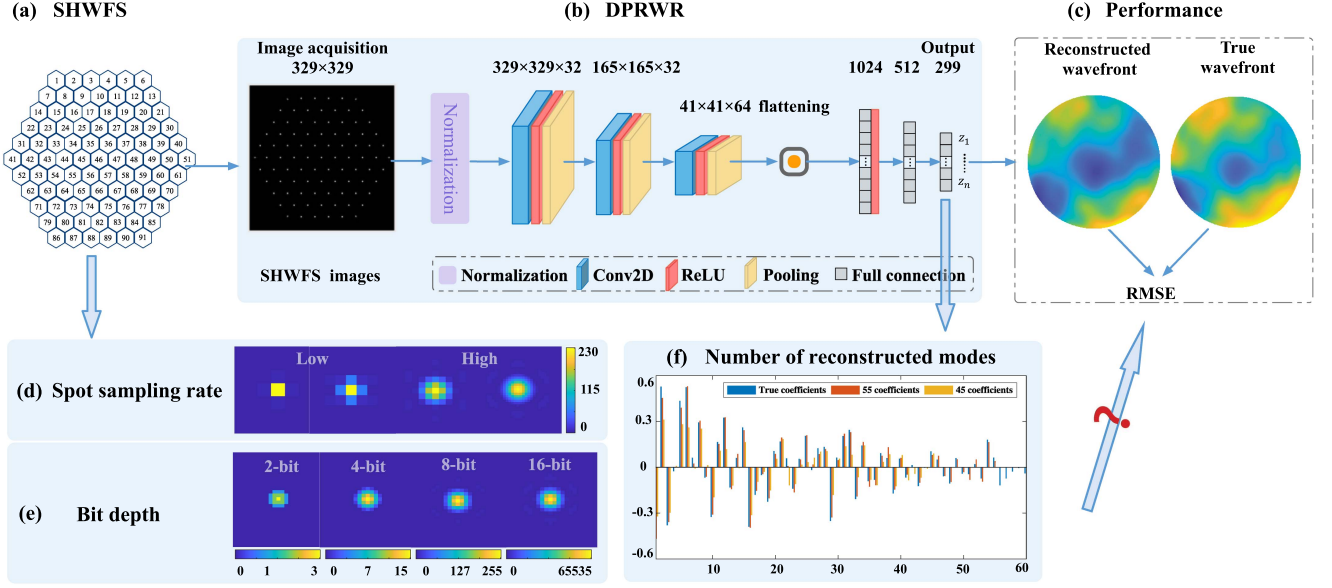


Fig. 1. Schematic diagram of SHWFS and DPRWR. (a) Microlens array of SHWFS with 91 sub-apertures. (b) The structure of the DPRWR. (c) The reconstructed wavefront and the true wavefront. (d) Sub-apertures images with different spot sampling rates. (e) Sub-apertures images with different bit depths. (d) Different numbers of reconstructed mode coefficients.

Where $M_k(x_1, y_1)$ is the k -order Zernike polynomial, corresponds to the Zernike mode coefficient, and K is the highest Zernike mode frequency.

The (x_1, y_1) is the source optical field coordinates in front of the microlens, (x_2, y_2) is the microlens focal plane coordinates, f is the microlens focal length, and λ is the light wave wavelength. To facilitate the numerical solution, the source light wave plane and focal plane need to be discretized and sampled, and the integral values can be solved by the Fourier transform. The transformed equation is as follows.

$$F(m_2, n_2) = \left| \frac{e^{i\frac{\pi(m_2^2\delta_2^2 + n_2^2\delta_2^2)}{\lambda f}}}{i\lambda f} \cdot \mathcal{F}(U_0(m_1, n_1)) \right|^2 \quad (4)$$

The (m_1, n_1) , (m_2, n_2) are the coordinates of the sampled source optical field plane and microlens focal plane, respectively. δ_1 , δ_2 are the sampling intervals respectively corresponding to the input plane and output plane. They satisfy the following condition.

$$\delta_2 = \frac{\lambda f}{N\delta_1} \quad (5)$$

In the following simulations, we need to generate the spot images under different sensor parameters. The spot sampling rate can be adjusted by altering the N or δ_1 value according to (5). Images with different bit depths can be simulated by multiplying the normalized focal plane spot intensity $F(m_2, n_2)$ with different gray levels. The below Fig. 2 show the simulation process of the training datasets.

B. The Structure of the DPRWR

Fig. 1(a) is the schematic diagram of the SHWFS microlens array. Fig. 1(b) shows the network structure of DPRWR. The

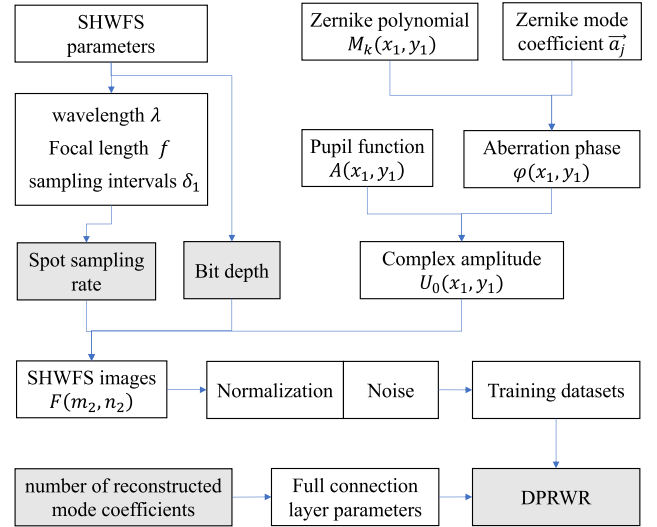


Fig. 2. A flowchart map of the simulations of the SHWFS images.

SHWFS images are the input of the DPRWR network, followed by 3 sets of sub-networks with the same structure, in the order of convolution-ReLU-pooling layer, for extracting highly abstracted features. These features are then flattened to the sample label space through 3 full connection layers. The number of neurons in the last full connection layer is determined by the number of reconstructed mode coefficients. The reconstructed wavefront (Fig. 1(c)) can be built by mode coefficients and Zernike polynomials, based on (3). The error of the reconstructed wavefront and true wavefront indicate the performance of the DPRWR method. Due to the orthogonality of the Zernike polynomial, the root mean square error (RMSE) of the wavefront is equal to the RMSE of the Zernike mode coefficients, which can

TABLE I
THE SENSOR PIXEL SIZE AND THE HORIZONTAL PIXEL COUNT OF
SUB-APERTURE SPOT IMAGE AT DIFFERENT FWHM

FWHM (pixel)	0.94	0.96	0.98	1.5	1.9	2.8	4.7	6.5
Pixel size (μm)	89	45	33.7	17.8	13.5	9	5.4	3.8
horizontal pixel count	3	6	8	15	19	29	48	68

be calculated by the following equation.

$$\text{RMSE} = \sqrt{\sum_{k=1}^K (a_j(k) - \widehat{a_j(k)})^2} \quad (6)$$

The $\widehat{a_j(k)}$ is the output label of the DPRWR network, and the $a_j(k)$ is the true Zernike mode coefficient.

III. SIMULATION RESULTS

To investigate the reconstruction performance of DPRWR in the following four aspects, different spot sampling rates (Fig. 1(d)), bit depths (Fig. 1(d)), numbers of reconstructed modes (Fig. 1(e)), and noise, we generated different datasets in each simulation. Each dataset contains 10000 SHWFS images, which are divided into a training set, validation set, and test set according to the ratio of 8:1:1. The labels of the training sets are random Zernike mode coefficients obeying a turbulence intensity $D/r_0 = 10$. The network trained for 50 epochs converges well. Details will be described in the following subsections. Each RMSE in the graphs and tables is the average of 100 measurement errors rather than a single measurement error.

A. Influence of the Spot Sampling Rate

The full width at half height of the SHWFS central sub-aperture spot under parallel light is represented as the spot sampling rate. According to the diffraction limit theory, $d = 1.22\lambda f/D$, the wavelength, the focal length and the diameter of the microlens determine the diameter of the Airy spot. The number of pixels taken up by the Airy spot on the sensor is determined by both the Airy spot diameter and the sensor sampling interval, which can be described by the FWHM of the spot. Therefore, the variations of light wave wavelength, lens focal length, lens diameter, and sensor sampling interval are all reflected in the variations of spot FWHM.

In this simulation, 8 noiseless datasets with different FWHMs are generated. The bit depth of all datasets is 8-bit and the number of reconstructed coefficients is 299. The sensor pixel sizes and the horizontal pixel counts of sub-aperture spot images corresponding to eight different FWHM are shown in Table I. A larger FWHM means a smaller pixel size, greater sub-aperture image size, and higher spot sampling rate. However, considering the memory size of the training datasets in this simulation, all the sub-aperture spot images with different FWHMs are cropped or filled with 0 to 21×21 . So the size of the SHWFS image in all datasets is 241×241 pixels. Therefore, each network corresponding to the 8 datasets has the same topology. In addition, for comparison with the slope-based method, eight reconstructed

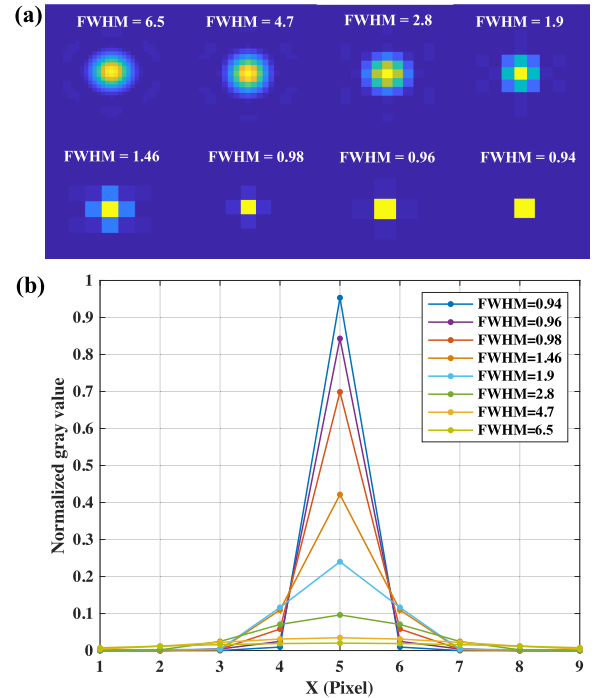


Fig. 3. Comparison of spot images with different spot sampling rates. (a) Central sub-aperture spot image. (b) Each broken line corresponds to the gray values of the row pixels where the maximum value of the sub-aperture spot image with different sampling rates is located in.

matrices under different spot sampling rates have been simulated for the test.

The sub-aperture spot images at 8 different spot sampling rates are compared in Fig. 3. It can be found that when FWHM is less than 0.94 pixels, the center pixel gathers almost more than 90% spot intensity, and as the FWHM decreases, more energy is gathered in a central pixel. Conversely, when the FWHM increases, the spot energy is dispersed within multiple pixels, and the spot morphology is more detailed.

The measurement errors of DPRWR and the slope-based method for dynamic aberrations with different D/r_0 at eight different spot sampling rates are compared in Fig. 4 and Table II. Based on the table, it can be calculated that when the FWHM increases from 0.94 to 0.98 pixels, the measurement errors of the DPRWR method and the slope method decrease by 0.1524λ and 0.5278λ , respectively. When the spot sampling rates rise 2.2 times from 1.46 to 4.7 pixels, the error of DPRWR only decreases by 0.0096λ , while the error of the slope-based method is almost unchanged.

These changes are consistent with Fig. 4. With the gradual increase of FWHM, the reconstruction errors of both methods firstly decrease sharply when the FWHM is less than 1 pixel, and then tend to be flatter when the FWHM is larger than 1.46 pixels, either in strong or weak turbulence aberration.

Differently, the error of the slope-based method is almost more than four times that of DPRWR. When the FWHM continues to increase from 1.46 to 4.7 pixels, the measurement error of DPRWR decreases steadily, while the RMSE of the slope-based method generally remains stable. The reason for this may be

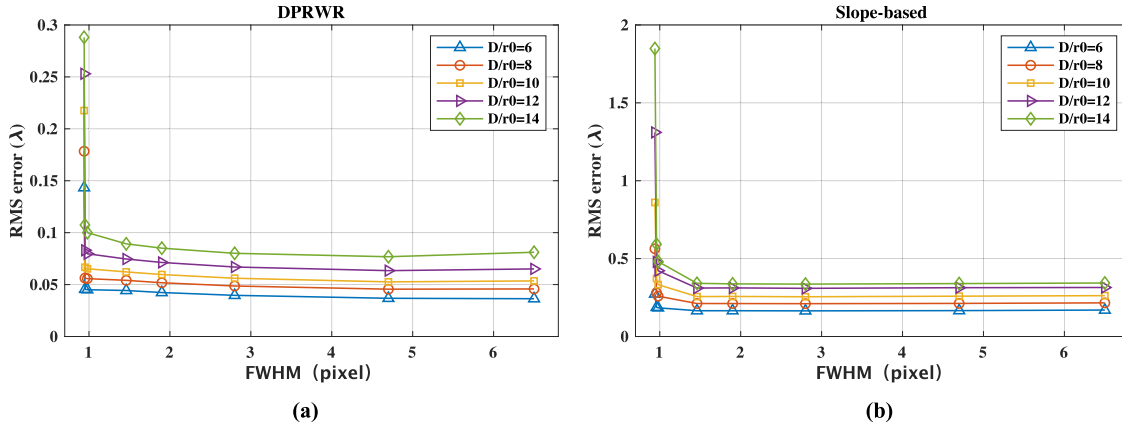

 Fig. 4. RMSE of the two methods for dynamic aberrations with different D/r_0 . (a) RMSE of DPRWR. (b) RMSE of the slope-based method.

 TABLE II
 RMSE ON TEST SETS WITH DIFFERENT FWHM AT $D/r_0 = 10$

FWHM (pixel)	0.94	0.96	0.98	1.46	1.9	2.8	4.7	6.5
RMSE of DPRWR (μm)	0.2176	0.0641	0.0652	0.0622	0.0596	0.0561	0.0526	0.0535
RMSE of slope-based (μm)	0.8605	0.3725	0.3327	0.2564	0.2571	0.2552	0.2590	0.2623

 TABLE III
 RMSE ON TEST SETS WITH DIFFERENT BIT DEPTHS AT $D/r_0 = 10$

Bit depth (bit)	1	2	4	6	8	10	12	14	16
RMSE of DPRWR (μm)	0.3887	0.0666	0.0618	0.0593	0.0596	0.0561	0.0603	0.0594	0.0587
RMSE of slope-based (μm)	0.3002	0.2995	0.2872	0.2710	0.2570	0.2454	0.2417	0.2402	0.2396

that as the number of pixels of the spot increases further, the CNN can still extract useful information for wavefront measurement, and the slope features are more accurate. Yet linear mapping in a slope-based method no longer meets more precise reconstruction.

B. Influence of the Bit Depth

We evaluated the reconstruction performance for nine different bit depths of 1, 2, 4, 6, 8, 10, 12, 14, and 16, corresponding to nine datasets with FWHM of 1.9, 299 mode coefficients, and noise-free. The same network topology and training parameters were used in these 9 cases.

The average errors of the two methods at different bit depths are shown in Table III. Then the influences of aberration with different D/r_0 on the relationship between bit depth and reconstruction performance are analyzed, and the results are shown in Fig. 5. As seen in Fig. 5, at $D/r_0 = 10$, the RMSE of the slope-based method shows a steady decline with increasing bit depth. RMSE reduces by 20% from 1-bit to 16-bit, namely 0.0606λ . However, the RMSE of DPRWR sharply drops by 80%, namely 0.3220λ , with the bit depth increasing from 1-bit to 2-bit. Then it slowly decreases by 0.0105λ with the bit depth

increasing from 2-bit to 10-bit. In general, increasing the bit depth can improve the measurement accuracy to some extent for both methods. Especially, the impact of the distorted information of the binary image on the DPRWR performance is significant. When bit depth changes from 10-bit to 16-bit, the RMSE of DPRWR remains stable. This performance characterization is similar to what FWHM performs.

Comparing Tables II and III, when FWHM is 1.9 pixels in Table II and bit depth is 8-bit in Table III, the configurations of SHWFS are the same, and RMSEs of the two methods are 0.0596λ , 0.2570λ respectively. When FWHM increases from 1.9 pixels to 6.5 pixels in Table II, the RMSE decreases from 0.0596λ to 0.0535λ , while the RMSE is 0.0587λ in Table III when the bit depth increase from 8-bit to 16-bit. For the DPRWR, improving spot sampling further is more effective than bit depth. For the slope-based method, conversely improving bit depth further is more effective than spot sampling. RMSE at 6.5 pixels is 0.2623λ in Table II, while RMSE at 16-bit is 0.2396λ in Table III.

A high bit depth sensor means that the light wave energy arriving at sensors can be characterized by grayscale values with a higher sampling rate. It will help the neural network to extract more features and to improve the fitting ability of the network, while this characterization does not exist when the bit depth is

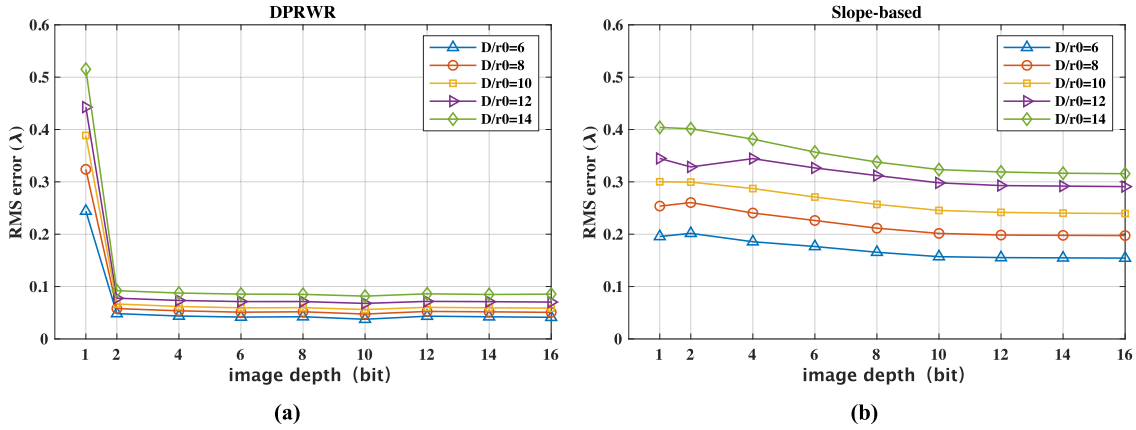


Fig. 5. RMSE of the two methods for dynamic aberrations with different D/r_0 at different bit depths. (a) RMSE of DPRWR. (b) RMSE of the slope-based method.

TABLE IV
DPRWR FULL CONNECTION LAYER PARAMETERS AT DIFFERENT NUMBERS OF RECONSTRUCTED MODES

number	45	55	66	91	135	209	377	464	561	666	780
1 st layer	1024	1024	1024	1024	1024	1024	1024	1024	1024	1024	1024
2 nd layer	512	512	512	512	512	512	512	464	561	666	780
3 rd layer	45	55	66	91	135	209	377	-	-	-	-

higher than 10-bit. Nevertheless, for a single feature, spot slope, it is more strongly correlated with the bit depth than the spot sampling rate.

C. Influence of the Number of Reconstructed Modes

Zernike polynomials are orthogonal functions in the circular domain and aberrations can be represented by linear combinations of different modes. High-order modes imply high spatial frequencies of aberrations. In this simulation, all datasets were generated by the 780 random Zernike mode coefficients, but the label used for train contains different numbers of mode coefficients, separately, 45, 55, 66, 91, 135, 209, 377, 464, 561, 666, and 780. By the way, the process of counting RMSE between true coefficients and reconstructed coefficients slightly differs from the previous 3 subsections. For instance, when the number of reconstructed mode coefficients is 45, the first 45 of 780 reconstructed coefficients are the network output labels, and the last 735 coefficients are set to 0. It means that the reconstruction error contains the prediction error coming from reconstructed modes and the error induced by the modes that fail to reconstruct but actually exist. Furthermore, the networks used for eleven datasets are different in the number of neurons of full connection layers, as shown in Table IV. In addition, other parameters are FWHM of 1.9, a bit depth of 8-bit, $D/r_0 = 10$, and a size of 241×241 pixels.

Fig. 6 shows the relationship between reconstructed mode numbers and the DPRWR performance at different spot sampling rates. We find that the RMSE decreases first and then increases as the number of reconstructed modes rises. This characteristic is the same at different spot sampling rates. It performs more clearly when FWHM is 0.98 pixels (as shown

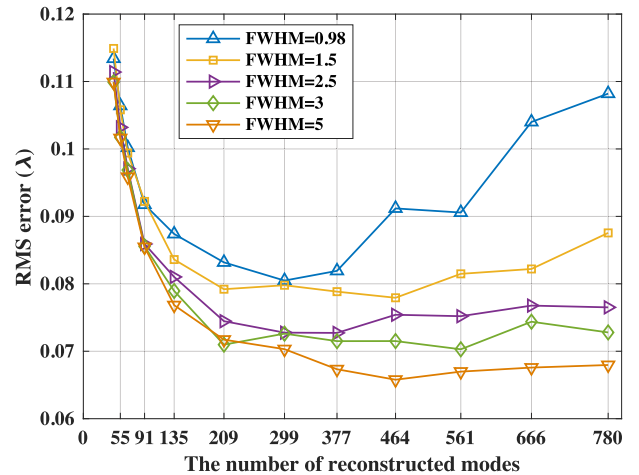


Fig. 6. Influence of spot sampling rate on the relationship between the number of reconstructed modes and RMSE of DPRWR.

in the blue curve in Fig. 6. At this time the best reconstructed mode number is at or near 299. The lower-order mode errors and higher-order mode errors when the numbers are 135, 299, and 464 respectively are shown in Table V. In this Table, it can be found that both the measured mode error ($3 \sim 135$, $3 \sim 299$) of the network and the unmeasured mode error ($136 \sim 299$, $300 \sim 464$) are lower when the reconstructed mode number is 299, comparing to 135, 464. In summary, with other configuration parameters unchanged, there must be an optimal value for the reconstructed mode numbers of the DPRWR. And at this optimum, it is more accurate in measuring low-order aberrations and high-order unmeasured aberrations. In other

TABLE V
 COMPARISON OF LOW-ORDER MODE ERROR AND HIGH-ORDER MODE ERROR WITH 135, 299 AND 464 RECONSTRUCTED MODES

Numbers of modes	Original WF (λ)	Network WF (λ)	RMS error(λ) (780)	RMS error(λ) (3~135)	RMS error(λ) (136~299)	RMS error(λ) (3~299)	RMS error(λ) (300~464)
135	0.6214	0.3739	0.08673	0.06048 \uparrow	0.04901 \uparrow	-	-
299	0.6214	0.3739	0.08048	0.05788	0.04090	0.07096	0.02830
464	0.6214	0.3739	0.09046	-	-	0.07953 \uparrow	0.03671 \uparrow

 TABLE VI
 FOUR DPRWR SCHEMES TO COPE WITH NOISE

	Training set	Test set
Scheme 1	No noise,no pre-processing	With noise,no pre-processing
Scheme 2	No noise,no pre-processing	With noise, pre-processing
Scheme 3	With noise,no pre-processing	With noise, no pre-processing
Scheme 4	With noise, pre-processing	With noise, pre-processing

words, there is an upper limit for the spatial resolution of the DPRWR to guarantee the minimum wavefront error. When the number of reconstructed modes is relatively small, the network is prone to misclassify the high-order modes and low-order modes, thus increasing the measurement accuracy of the low-order modes. In this case, increasing the reconstructed mode number can improve the measurement accuracy of high-order modes and at the same time reduce the misclassification of low-order modes. When the reconstructed mode number is relatively large, it is difficult for DPRWR to accurately discriminate between the high-order modes and low-order modes from the currently available information, so the performance both degrade.

Another point worth noting in Fig. 6 is that RMSE increases more and more slowly with the increase of mode number as the FWHM goes from 0.98 to 4.7 pixels and the number of reconstructed modes is more than 299. When it is less than 299, the change in RMSE is greater. This means that the risk of reconstructing more modes is much lower. Therefore, there is a special recommendation for the selection of the reconstructed mode numbers: the appropriate range of mode numbers is 299 to 561, which is 3.28 to 6.16 times the current sub-aperture number 91.

D. Influence of Noise

The exposure time of the sensor and target brightness indirectly affect the performance of the DPRWR by affecting the image signal-to-noise ratio. Therefore, in this section, we investigate the robustness of the DPRWR and a suitable scheme to cope with the variation of image signal-to-noise ratio caused by the sensor exposure time, target brightness, and so on.

We design four DPRWR schemes in Table VI for noise suppression. The main differences between these 4 schemes are the different preprocessing processes of the training data, including with or without noise and with or without noise reduction processing. The specific noise process refers to subtracting a

 TABLE VII
 GRAYSCALE RANGE OF NOISE WITH DIFFERENT PSNR ADDED TO A SUB-APERTURES SPOT IMAGE WITH $D/R_0 = 10$ WHEN THE GRAYSCALE LEVEL IS 256

PSNR of training set	5	15	25	30	50	100
Grayscale range	0~20	0~5	0~2	0~1	0~0.1	0~0.05

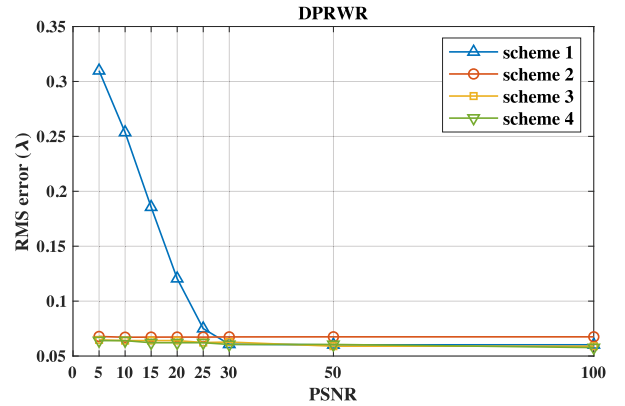


Fig. 7. Reconstruction performance of DPRWR using four noise suppression schemes under different PSNR noise. (In schemes 1 and 2, there is noise-free on the training set, and the horizontal axis is the PSNR of noise on the test set. In schemes 3 and 4, it is the PSNR of noise added to the training set, the same in the test set.)

 TABLE VIII
 RMSE OF DPRWR IN SCHEME 3 AT DIFFERENT PSNR OF THE TRAINING SET AND TEST SET

PSNR of the test set	PSNR of training set				
	5	15	25	30	50
5	0.06469	0.06593	0.06487	0.06538	0.06179
15	0.06429	0.06406	0.06245	0.06303	0.05924
25	0.06453	0.06405	0.06231	0.06292	0.05910
30	0.06444	0.06387	0.06214	0.06274	0.05894
50	0.06447	0.06391	0.06218	0.06277	0.05897

fixed gray value from the SHWFS image, and the value is related to the actual noise intensity, which may be the median or first quartile of the grayscale range of the pure noise image. Other configuration parameters are the same, with an FWHM of 1.9, a bit depth of 8-bit, and a size of 241x241 pixels.

TABLE IX
LIST OF PARAMETERS USED IN THE EXPERIMENTAL SYSTEM

SHWFS		SLM	
number of sub-apertures	91	Image size	256×256
Obscured diameter	2.86 mm	Bit depth	16
Focal length	10.8 mm	Others	
Pixel size	8.4×8.4 μm	Wave length	650 nm
Bit depth	8 bits	Focal length of L2,L3	500 cm
image size	329×329 px	Focal length of L4	300 cm

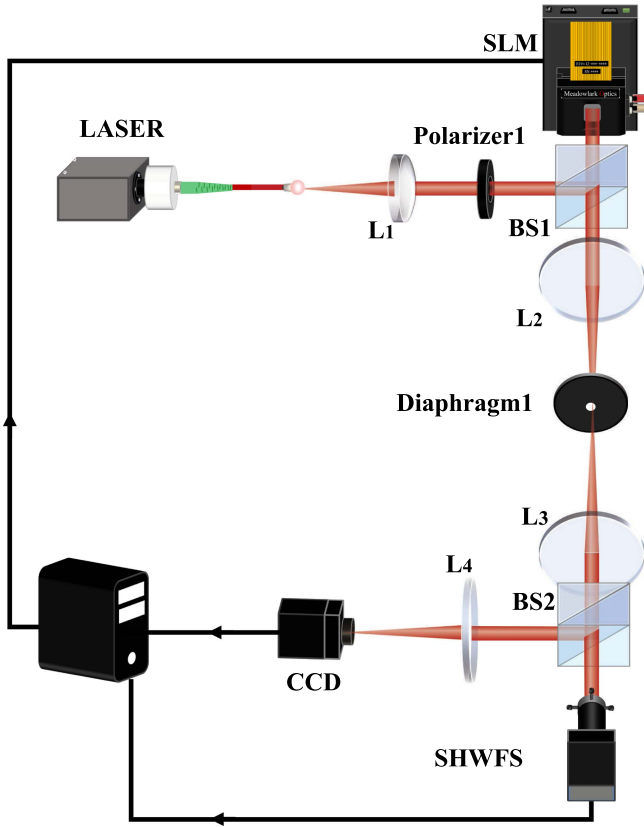


Fig. 8. Schematic diagram of the experimental optical bench.

In this simulation, we compare the effects of the 4 noise processing schemes (Table VI) on the wavefront measurements under Gaussian noise with different peak signal-to-noise ratios (PSNR). The details of the four schemes are introduced in II(E). The grayscale dynamic range of the added noise is given in Table VII when the gray level is 256, and Fig. 7 shows the RMSE of DPRWR when the training sets are added with different PSNR noise.

As seen from the figure, when the PSNR is greater than 30, the noise intensity is weak and the overall effect of the noise on the wavefront measurement is not substantial regardless of whether additional noise suppression pre-processing is taken. When the PSNR is less than 30, schemes 2, 3, and 4 all

significantly improve the performance of the DPRWR against noise interference. Among them, the RMSE of scheme 3 is the smallest. This means that either adding noise to the training set or subtracting the threshold on the data set can suppress the noise to some extent. But the most effective method among these three schemes is scheme 3, adding noise to the training set and no image denoising processing. The neural network is capable to learn a strong reconstruction performance directly from the noisy image.

In order to find the optimal intensity of the noise signal added to the training set in scheme 3, the RMSEs measured by the DPRWR method on test data with different PSNR of 5, 15, 25, 30, and 50, respectively, are shown in Table VIII. Under the current parameter conditions, when the PSNR of the training set is about 50, the DPRWR method has the smallest RMSE both in strong noise and weak noise compared to the other four cases. At this time, the RMSE is 0.0618λ when the test data PSNR is 5. Compared with the RMSE of 0.0597λ when both the training set and the test set are free of noise, RMSE only rises by 3.5%, indicating that scheme 3 can effectively help DPRWR suppress the noise when the PSNR ratio is higher than 5.

IV. EXPERIMENTAL VERIFICATION

A. Experimental Optical Bench

After completing the above simulation analysis, we conducted an experiment on DPRWR to further verify its performance characteristics. The experimental system is shown in Fig. 8, and the relevant parameters are listed in Table IX. The incident laser is modulated to a fixed polarization state by a polarizer and then passes through a collimator lens L1 and an aperture diaphragm to obtain a parallel light, which is split to the SLM by a beam splitter. In order to eliminate the multi-order diffracted light (including zero-order light) caused by the SLM pixel structure and the high-order spot, a 4f system (including L2 and L3) and a diaphragm are built for remaining a secondary order spot in this experiment. The emergent light from lens L3 is divided into two beams through the beam splitter. One beam arrives at the SHWFS forming a spot array image, which is subsequently acquired by the computer as a training data set. The other beam passes through lens L4 forming a focal PSF image on the CCD, which is used to observe the restored spot modulated by SLM and to analyze the wavefront reconstruction performance of DPRWR.

In this environment, the noise grayscale range of the acquired sub-apertures spot images is about 0~2. Based on the analysis in Section III-D, we adopt scheme 3 to eliminate the effect of noise. Each sub-aperture of the training set and test data are normalized to remove the effect of laser power variation.

In addition, due to the DPRWR performance variation at different turbulence intensities is basically the same in the simulation analysis, we only selected one turbulence intensity for verification in this experiment. The D/r_0 of all the training sets is 10.

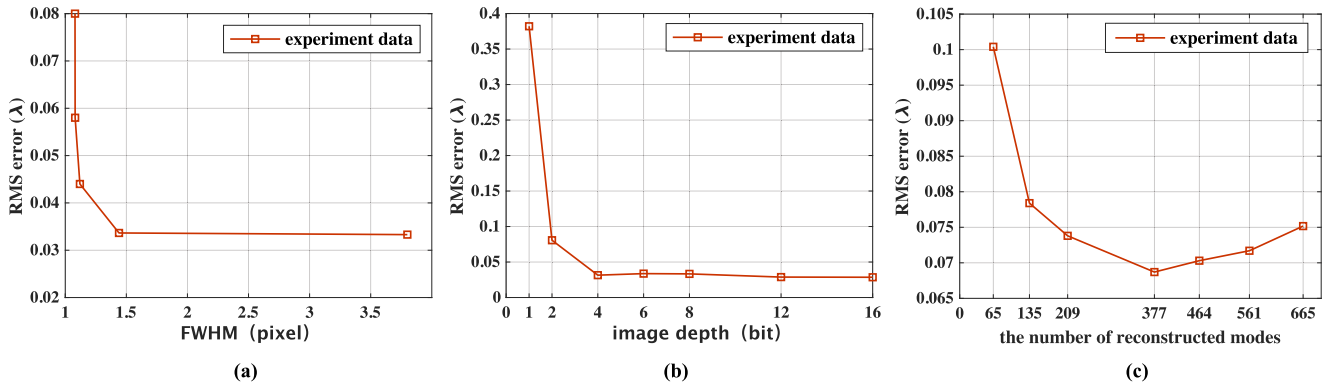


Fig. 9. RMSE obtained by experimental measurement at different configuration parameters. (a) RMSE of different FWHM. (b) RMSE of different bit depth. (c) RMSE of different numbers of reconstructed mode coefficients.

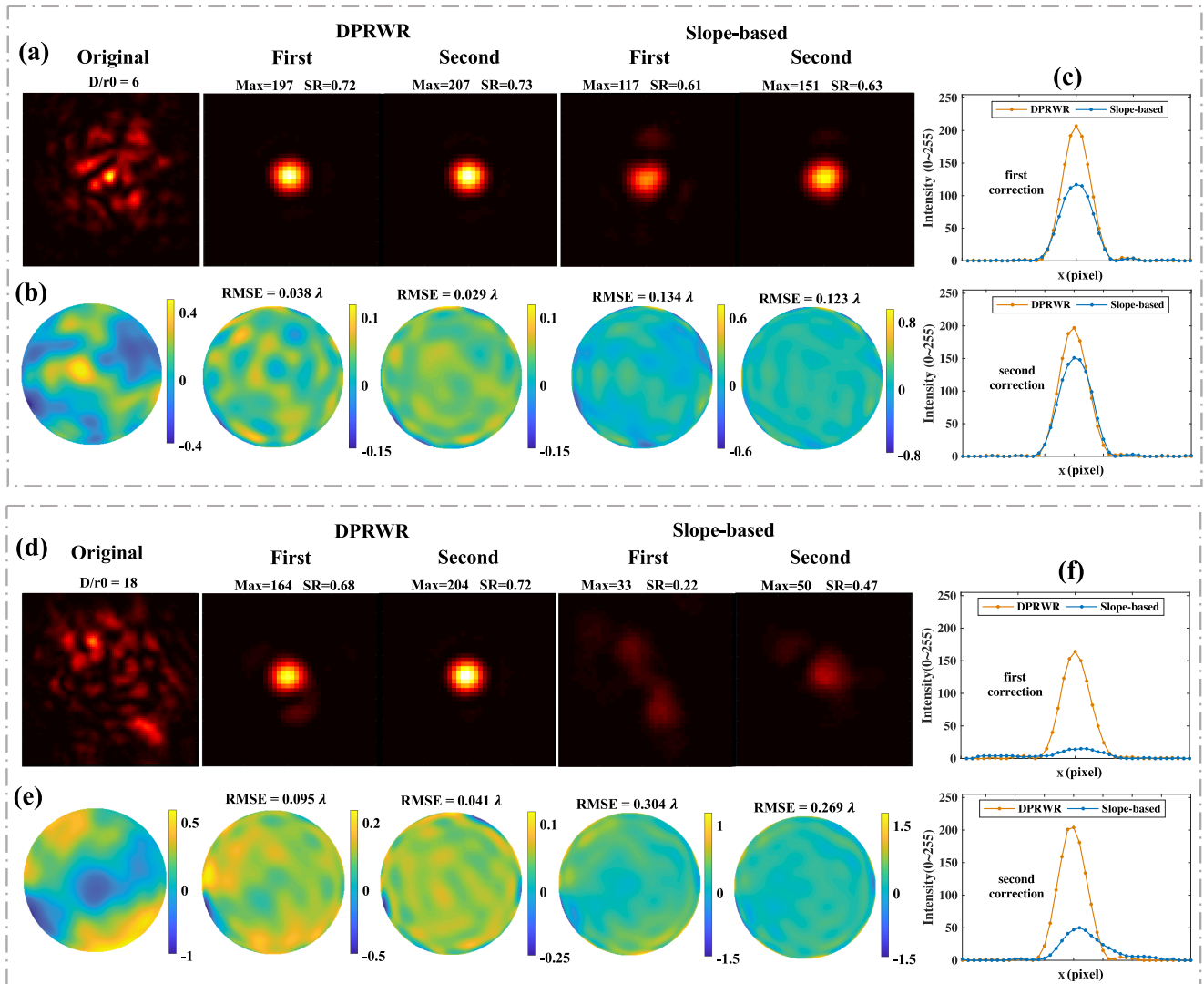


Fig. 10. Comparison of closed-loop iterative correction results measured by DPRWR and slope-based method, respectively for two wavefronts of RMS $0.16 \mu\text{m}$ and $0.37 \mu\text{m}$. (a), (d) Are the focal PSF images. (b), (e) Is the corresponding reconstructed wavefront phase screens. (c), (f) Are the intensity of the pixels in the row where the maximum value of each image is in (a), (d).

B. Experimental Results

To verify the relationship between spot sampling rate and reconstruction performance, the 10000 sub-apertures spot images acquired were downsampled four times. The FWHM of the original spot and downsampled modulation-free spots are 3.7, 1.44, 1.12, 1.08, and 1.06 pixels in this order. The RMSEs in these five cases are shown in Fig. 9(a). It is found that the performance can be significantly improved by increasing spot sampling rates when the spot FWHM is 1.06 to 1.12 pixels, while it no longer works at an FWHM of 1.44~3.7. This characteristic is consistent with the simulation results in Fig. 4(a). The RMSE of experimental measurement at different bit depths is shown in Fig. 9(b). As the bit depth increases, the RMSE first decreases sharply and then remains almost constant, as Fig. 5(a) shows.

Differently, the minimum requirement for configuration (FWHM>1.44 pixel, bit depth>4-bit) in the experiment is slightly higher than that of simulation (FWHM>0.98 pixels, bit depth>2-bit) in order to ensure precision wavefront measurement because of the noise interference in the experimental data.

To verify the effect of reconstructed mode numbers on detection performance, each of the 10000 aberrated phase screens was generated by 665 Zernike modes. The FWHM of modulation-free spots is 3.8 pixels. Seven different networks differ in the number of neurons of the last full connection layer, which are 65, 135, 209, 377, 464, 561, and 665, respectively. As shown in Fig. 9(c), It is proved that the special resolution upper limit of DPRWR exists. Under these experimental conditions, the optimal reconstructed mode number is near 377. This conclusion is basically consistent with the simulation results.

Based on the above experimental conclusions, as well as taking into account the economy and convenience, we selected a suitable SHWFS parameter configuration with an FWHM of 3.8, a bit depth of 8, and a reconstructed mode number of 377 and then conducted two closed-loop iterative correction experiments respectively under $D/r_0 = 6$ and $D/r_0 = 18$.

The corrected focal spot and reconstructed wavefront are shown in Fig. 10 below. When the D/r_0 of the aberration is 6, a great recovery spot can be obtained by one correction in the DPRWR method. After the second correction, the Strehl ratio (SR) of the spot recovered by DPRWR is 0.73 and the RMSE of the reconstructed wavefront is 0.029λ , while the SR is 0.63 and the RMSE is 0.123λ in the slope-based method.

When the D/r_0 of the aberration is 18, the original spot diffuses severely. For this strong aberration, it only takes two iterations in the DPRWR method to obtain a great recovery spot, yet the quality of the spot recovered by the slope-based method is much worse. The SRs of the spots secondly corrected by two methods are 0.72 and 0.47 respectively. The RMSEs are 0.041λ and 0.269λ respectively. Compared to the slope-based method, the intensity of the central pixel in the recovery spot recovered by DPRWR is much higher.

We find that both for strong and weak aberrations, only two rounds of iterative correction in DPRWR can achieve a focal spot with a high SR. Yet more rounds of iterative correction are considered in the slope-based method for obtaining a better-quality spot image.

V. CONCLUSION AND DISCUSSION

In this paper, we explore the relationship between SHWFS configuration parameters, network structure parameters (spot sampling rate, bit depth, number of reconstructed modes, noise) and reconstruction performance of the CNN-based DPRWR method. We gain a clear understanding of the application limits and resolution upper bound of DPRWR. To further optimize the DPRWR, we find a suitable SHWFS configuration and noise suppression scheme adapted to the neural network. This study provides a theoretical and experimental basis for the parameter selection of DPRWR to advance its future application to the telescope system.

Based on the previous analysis in this paper, we conclude that there are several important findings:

- 1) The performance of the CNN-based wavefront reconstruction method is more sensitive to the spot sampling rate compared to the bit depth, while the conventional slope-based method is exactly the opposite. Increasing the spot FWHM can further improve the DPRWR reconstruction accuracy, but for the slope-based method increasing the bit depth is more effective for performance improvement.
- 2) For both methods, the Hartmann sensor must meet this condition that the maximum light intensity focused on one pixel of the sub-aperture spot is at least less than 90%, otherwise, the detection performance will degrade drastically.
- 3) The impact of distortion of binary images on the prediction ability of neural networks is enormous. For the DPRWR, the bit depth needs at least 2-bit or more. However, for the slope method, this constraint does not exist.
- 4) The number of high-accuracy reconstructed modes of the DPRWR is about 3 to 6 times the number of sub-apertures, which is much higher than the spatial resolution of the slope-based method. Either increasing or decreasing the reconstructed mode number will reduce the reconstruction performance of DPRWR to some extent. Differently, for neural networks, the risk of taking a larger number of reconstructed modes is much lower than taking a smaller number of reconstructed modes, especially at larger spot sampling rates.
- 5) To achieve the optimal performance of DPRWR, the suitable range of FWHM is 1.46 to 4.7 pixels, of bit depth is 4-bit to 16-bit, and of reconstructed mode number is 3 to 6 times the number of sub-apertures.
- 6) The training set images with weak noise and without subtraction of the global threshold can effectively improve the generalization ability of the neural network under different levels of noise. In this experiment, the wavefront reconstruction performance of the DPRWR based on this scheme is only slightly degraded by 3.5% when dealing with strong noise at a PSNR of 5 compared to the case without noise interference.

In the future, we will further study solutions adapted to the neural network structure in more extreme cases, such as the spot deficit of partial sub-apertures, broad-spectrum light source illumination, system latency under high temporal frequency distortions [30], and so on. We plan to apply DPRWR to the telescope system to obtain better observation results.

ACKNOWLEDGMENT

The authors would like to acknowledge the help from Ao Tang in Institute of Optics and Electronics, Chinese Academy of Sciences during the writing of this manuscript.

REFERENCES

- [1] R. Tyson, *Principles of Adaptive Optics*. Boca Raton, FL, USA: CRC Press, 2011.
- [2] Y. Guo et al., "Adaptive optics based on machine learning: A review," *Opto-Electron. Adv.*, vol. 5, no. 7, 2022, Art. no. 200082.
- [3] R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Proc. SPIE*, vol. 21, pp. 829–832, 1982.
- [4] Z. Xie et al., "Restoration of degraded images using pupil-size diversity technology with stochastic parallel gradient descent algorithm," *IEEE Photon. J.*, vol. 8, no. 2, Apr. 2016, Art. no. 6900410.
- [5] B. L. Ellerbroek et al., "Comparison of Shack-Hartmann wavefront sensing and phase-diverse phase retrieval," *Proc. SPIE*, vol. 3126, pp. 307–320, 1997.
- [6] Y. Wu et al., "Sub-millisecond phase retrieval for phase-diversity wavefront sensor," *Sensors*, vol. 20, no. 17, 2020, Art. no. 4877.
- [7] D. G. Modai, "Model wave-front reconstruction with Zernike polynomials and Karhunen–Loève functions," *J. Opt. Soc. Amer.*, vol. 13, no. 6, pp. 1218–1225, 1996.
- [8] S. Thomas et al., "Comparison of centroid computation algorithms in a Shack–Hartmann sensor," *Monthly Notices Roy. Astronomical Soc.*, vol. 371, no. 1, pp. 323–336, 2006.
- [9] A. Narsireddy, P. Garcia, and C. M. Correia, "Peak-locking centroid bias in Shack–Hartmann wavefront sensing," *Monthly Notices Roy. Astronomical Soc.*, vol. 476, no. 1, pp. 300–306, 2018.
- [10] I. Han, "New method for estimating wavefront from curvature signal by curve fitting," *Proc. SPIE*, vol. 34, pp. 1232–1237, 1995.
- [11] C. Rao et al., "1.8-m solar telescope in China: Chinese large solar telescope," *Proc. SPIE*, vol. 1, 2015, Art. no. 024001.
- [12] C. Rao et al., "Instrument description and performance evaluation of a high-order adaptive optics system for the 1 m new vacuum solar telescope at Fuxian solar observatory," *Astrophysical J.*, vol. 833, no. 2, 2016, Art. no. 210.
- [13] J. Herrmann, "Cross coupling and aliasing in modal wave-front estimation," *J. Opt. Soc. Amer.*, vol. 71, no. 8, pp. 989–992, 1981.
- [14] C. Li, B. Li, and S. Zhang, "Phase retrieval using a modified Shack–Hartmann wavefront sensor with defocus," *Appl. Opt.*, vol. 53, no. 4, pp. 618–624, 2014.
- [15] M. Vieggers, E. Brunner, and O. Soloviev, "Nonlinear spline wavefront reconstruction through moment-based Shack–Hartmann sensor measurements," *Opt. Exp.*, vol. 25, no. 10, pp. 11514–11529, 2017.
- [16] C. Zuo et al., "Transport of intensity equation: A tutorial," *Opt. Lasers Eng.*, vol. 135, 2020, Art. no. 106187.
- [17] C. Zuo et al., "Deep learning in optical metrology: A review," *Light Sci. Appl.*, vol. 11, no. 1, 2022, Art. no. 39.
- [18] G. Hong et al., "Wavefront reconstruction with artificial neural networks," *Opt. Exp.*, vol. 14, no. 14, pp. 6456–6462, 2006.
- [19] T. B. Dubose, A. T. Watnik, and D. F. Gardner, "Intensity-enhanced deep network wavefront reconstruction in Shack–Hartmann sensors," *Opt. Lett.*, vol. 45, no. 7, pp. 1699–1720, 2020.
- [20] R. Swanson et al., "Wavefront reconstruction and prediction with convolutional neural networks (Conference presentation)," *Proc. SPIE*, vol. 10703, pp. 481–490, 2018.
- [21] Z. Li et al., "Centroid computation for Shack–Hartmann wavefront sensor in extreme situations based on artificial neural networks," *Opt. Exp.*, vol. 26, no. 24, pp. 31675–31692, 2018.
- [22] W. Yang, J. Wang, and B. Wang, "A method used to improve the dynamic range of Shack–Hartmann wavefront sensor in presence of large aberration," *Sensors*, vol. 22, no. 19, 2022, Art. no. 7120.
- [23] M. Zhao et al., "Centroid-predicted deep neural network in Shack–Hartmann sensors," *IEEE Photon. J.*, vol. 14, no. 1, Feb. 2022, Art. no. 6804810.
- [24] L. Hu et al., "Learning-based Shack–Hartmann wavefront sensor for high-order aberration detection," *Opt. Exp.*, vol. 27, no. 23, pp. 33504–33517, 2019.
- [25] Y. Guo et al., "Deep phase retrieval for astronomical Shack–Hartmann wavefront sensors," *Monthly Notices Roy. Astronomical Soc.*, vol. 510, no. 3, pp. 4347–4354, 2021.
- [26] T. Guo et al., "Research on polarization Hartmann wavefront detection technology," *Opto-Electron. Eng.*, vol. 48, no. 7, 2021, Art. no. 210076.
- [27] H. Chen et al., "A B-spline based fast wavefront reconstruction algorithm," *Opto-Electron. Eng.*, vol. 48, no. 2, 2021, Art. no. 200160.
- [28] D. Jason, *Numerical Simulation of Optical Wave Propagation*. Bellingham, WA, USA: SPIE, 2010.
- [29] N. A. Roddier, "Atmospheric wavefront simulation using Zernike polynomials," *Proc. SPIE*, vol. 29, pp. 1174–1180, 1990.
- [30] N. Wang et al., "Deep learning-based prediction algorithm on atmospheric turbulence-induced wavefront for adaptive optics," *IEEE Photon. J.*, vol. 14, no. 5, Oct. 2022, Art. no. 8554310.