



Feature-based phase retrieval wavefront sensing approach using machine learning

GUOHAO JU,^{1,2} XIN QI,¹ HONGCAI MA,¹ AND CHANGXIANG YAN^{1,3}

¹Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, Changchun 130033, China

²juguohao123@163.com

³yanxc@ciomp.ac.cn

Abstract: A feature-based phase retrieval wavefront sensing approach using machine learning is proposed in contrast to the conventional intensity-based approaches. Specifically, the Tchebichef moments which are orthogonal in the discrete domain of the image coordinate space are introduced to represent the features of the point spread functions (PSFs) at the in-focus and defocus image planes. The back-propagation artificial neural network, which is one of most wide applied machine learning tool, is utilized to establish the nonlinear mapping between the Tchebichef moment features and the corresponding aberration coefficients of the optical system. The Tchebichef moments can effectively characterize the intensity distribution of the PSFs. Once well trained, the neural network can directly output the aberration coefficients of the optical system to a good precision with these image features serving as the input. Adequate experiments are implemented to demonstrate the effectiveness and accuracy of proposed approach. This work presents a feasible and easy-implemented way to improve the efficiency and robustness of the phase retrieval wavefront sensing.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Broadly speaking, phase retrieval wavefront sensing represents one class of image-based approach that recovers the wavefront phase of an optical system in the pupil plane when only intensity measurements in the image plane are available [1,2]. Compared to other wavefront sensors (Hartmann sensor or shearing interferometry), they have several important advantages, such as low requirement for optical hardware and no special need for calibration [3,4]. Since the inception of them about three decades ago, phase retrieval wavefront sensing approaches have played an important role in detecting the wavefront aberrations of large astronomical telescopes, especially for space applications [5–9].

The conventional phase retrieval wavefront sensing approaches mainly can be classified into two general categories: iterative-transform and parametric approaches. The former is also known as the Gerchberg-Saxton (G-S) or error-reduction algorithm [1,10], which involves iterative Fourier transformation back and forth between the object and Fourier domains and application of the measured intensity data or known constraints in each domain. The latter is also named as model-based optimization algorithm or directly called phase diversity algorithm [11–14], which recovers the parameterized wavefront aberrations by establishing objective function (error metric) and then minimizing the optimization with nonlinear optimization methods. Both of these two classes of iterative approaches are time-consuming and not suitable for applications with a high requirement for efficiency. Besides, the results of these approaches partly depend on the original values needed in the iterative transformation or iterative optimization process, and they subject to the stagnation problem (and therefore the robustness is low) [9].

Artificial neural network, which belongs to machine learning, has been introduced to the area of phase retrieval wavefront sensing [15–18]. A neural network is an input-output information processor composed of parallel layers of elements or neurons, loosely modeled

on biological neurons, which possess local memory and are capable of elementary arithmetic. It can be used to learn and store a great deal of nonlinear mapping relations of the input-output model [19]. Previously, neural networks were utilized to measure the optical phase distortion induced by air turbulence in adaptive optics [15,16], which were then applied to wavefront reconstruction for the Hubble Space Telescope [17]. These attempts used a network with each pixel of the PSF as part of an input vector which was matrix-multiplied into a single “hidden” vector. The resulting vector was then fed through a nonlinear sigmoid function and matrix-multiplied to an output vector which corresponds to the aberration coefficients. The main problem for this approach is the contradiction between too much inputs and only one hidden layer. For example, for a pair of 32×32 PSF images, we need about 1800 ($2 \times 32 \times 32$) neurons in the input layer. The intensity of each pixel is not independent from the others for a certain PSF pattern, and the final wavefront phase is determined by the intensities of all pixels. On the other hand, this type of machine learning model considers each pixel independently, and it is the task of the hidden layer to build the inherent relations between the intensities of different pixels. Taking these two aspects into account, we can deduce that it is very hard to use this kind of machine learning model (with only one hidden layer) to establish the precise nonlinear mapping between the inputs and outputs. If we decrease the size of PSF images (in references [15–17] 16×16 PSF images are used), then the capture range are restricted (generally a larger wavefront error corresponds to a PSF with a larger size).

Recently, another type of model, the convolutional neural network (CNN), has been introduced to image-based wavefront sensing [18]. In machine learning, CNNs are a class of deep, feed-forward artificial neural networks which can directly consider a group of pixels rather than each pixel independently. A CNN consists of an input and an output layer, as well as multiple hidden layers, including convolutional layers, pooling layers, fully connected layers and normalization layers [20]. While CNN needs relatively little pre-processing and is independent from prior knowledge and human effort in feature design, an enormous amount of efforts are needed to train this multi-layer neural network with so much number of inputs. The training of multi-layer neural network suffers from a series of problems, such as vanishing gradient and exploding gradient [21,22]. The training result depends heavily on personal experience, since some parameters needed in the training process (such as the number of node in each layer and the learning rate) are determined by experience. On the other hand, the accuracy of CNN for image-based wavefront sensing is still not very high at present. We can see from Fig. 4 of reference [18] that the root-mean-square wavefront error (RMS WFE) values for wavefronts synthesized from CNN predicted coefficients compared to the true wavefront is about 0.2 waves (maybe the emphasis in this reference is on capture range, not fitting accuracy). Besides, no experiment is performed to demonstrate its feasibility in practical environment.

To greatly simplify the structure of the neural network for machine learning while maintaining a good non-linear fitting accuracy, a feature-based method is proposed in contrast to the intensity-based methods mentioned above which directly deal with the gray-scale pixels of the PSF images. Specifically, the geometric features of a pair of PSF images (usually collected from in-focus and defocus image planes) are extracted to serve as the input of the neural network (no longer the intensities of pixels). This process of feature extraction can be seen as an effective compression of the PSF image data. Tchebichef moments, which are orthogonal in the discrete domain of the image coordinate space, are introduced as the geometric features [23–25]. The back-propagation artificial neural network, which is one of the most wide applied and easy implemented neural network models, is utilized as the nonlinear fitting tool [26,27]. Tchebichef moments can effectively characterize the intensity distribution of the PSF, and they do not involve numerical approximation of continuous integrals and coordinate space transformation. The back-propagation artificial neural network with Tchebichef moments serving as the input has a simple structure and is very convenient

to train. Once well trained, it can estimate the aberrations to a good precision with high efficiency and robustness (completely free of stagnation problem). The effectiveness and accuracy of this feature-based approach using machine learning is fully verified by experiment. Some other discussions concerning this approach are also presented.

This paper is organized as follows. In Section 2, we introduce the principle of a wide-applied machine learning tool, back-propagation artificial neural network. Then we continue to describe the Tchebichef moments and present the feature-based phase retrieval approach using machine learning in Section 3. Experimental validations and discussions on the proposed approach is presented in Section 4. In Section 5, we conclude the paper.

2. Introduction of the back-propagation artificial neural network

Artificial neural networks are one class of the most widely applied machine learning tools. They are computing systems made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs. Such systems learn (progressively improve performance) to do tasks by considering examples, generally without task-specific programming. They have found most use in applications difficult to express in a traditional computer algorithm using rule-based programming. Neural networks have been widely applied in pattern recognition, intelligent control and some other areas.

An artificial neural network contains a collection of connected units called artificial neurons, which is analogous to axons in a biological brain. Each connection (synapse) between neurons can transmit a signal to another neuron. The receiving (postsynaptic) neuron can process the signal(s) and then signal downstream neurons connected to it. Specifically, the output of a neuron, y_k , can be expressed as

$$y_k = \varphi \left(\sum_{i=1}^m x_i w_{ik} - b_k \right), \quad (1)$$

where $\varphi(\cdot)$ is called activation (or transfer) function, m is the number of the input neurons, w_{ik} is the weight for the i th input signal, x_i , b_k is the bias (or offset) which used to properly shift the results of this linear transformation. The purposes of the activation function are introducing nonlinearity to the neural networks and bounding the value of the neuron so that the neural network is not paralyzed by divergent neurons. A common example of activation function is the sigmoid (or logistic) function, which is shown in Eq. (2).

$$\varphi(z) = \frac{1}{1 + \exp(-z)}. \quad (2)$$

The mathematical model of an artificial neuron is illustrated in Fig. 1.

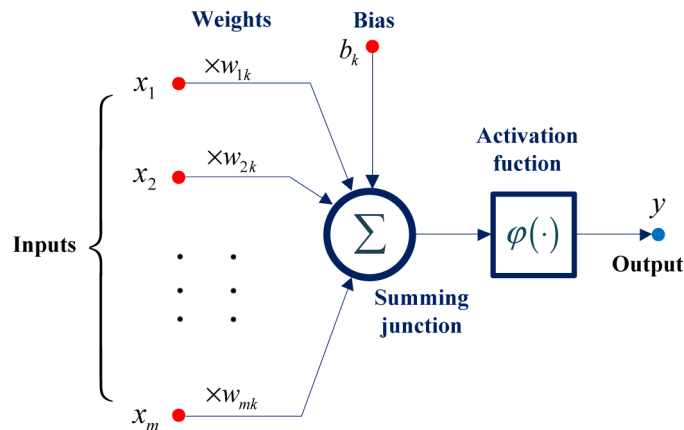


Fig. 1. The mathematical model of an artificial neuron.

The process of adjusting the weights such that it can learn the appropriate mapping relations between the inputs and the outputs is called learning or training. The training begins with random weights, and the goal is to adjust them so that the error will be minimal. The back-propagation algorithm is one of the most widely used algorithms for training the neural networks. The sketch of back-propagation algorithm for a neural network with three layers is shown in Fig. 2, where each circular node represents an artificial neuron.

The back-propagation algorithm can be divided into the following two steps:

- (1) Forward propagation of operating signal. The input signal is propagated from the input layer, via the hide layer, to the output layer. During this process, the weight and offset values of the network are maintained constant and the status of each layer of neuron will only exert an effect on that of next layer of neuron.
- (2) Back propagation of error signal. The error signal is propagated from the output end to the input layer in a layer-by-layer manner and the weights of network are regulated by the error feedback. The difference between the real output and expect output of the network is defined as the error signal. Continuous modification of weight and offset values will be applied to make the real output of network closer to the expected one.

In this algorithm, gradient descent method is usually used to adjust the weight of each neuron according to the error between the current output and the desired output. Once the network is well trained, we can save the values of the weights and offsets of the network. Then we can use them to directly obtain a set of outputs when a set of inputs is available, without the need for any iterative computation or optimization process. Note that any supervised learning involves back-propagating error to correct weights. On the other hand, back-propagation is not unique to networks that are trained via supervised learning. The term "back-propagation neural network" in this paper indicates the neural networks are trained via supervised learning which includes back-propagating error in the model.

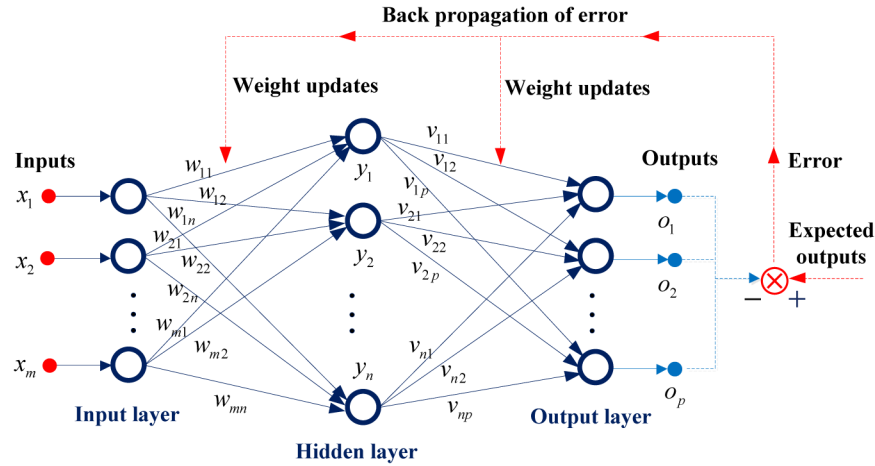


Fig. 2. Back-propagation algorithm for a neural network with three layers.

3. Feature-based phase retrieval approach using machine learning

In this section, we first introduce the discrete orthogonal Tchebichef moments which is use to represent the features of the PSF images and effectively compress the image data. Then the feature-based phase retrieval approach using artificial neural network as well as its application procedure are presented.

3.1 Tchebichef moment features

In this paper, the Tchebichef moments are introduced to extract or represent the features of the point spread functions (PSFs) at the in-focus and defocus image planes. Moments with orthogonal basis functions are powerful feature descriptors due to its property of minimal information redundancy in a moment set [25]. Compared to the other continuous orthogonal moments, such as Zernike and Legendre moments, Tchebichef moments which belong to the class of discrete orthogonal moments have several important advantages in feature representation. Specifically, they do not involve numerical approximation of continuous integrals and coordinate space transformation. Therefore, they have a higher feature representation capability [23,24].

The mathematical basis that leads to a definition for discrete orthogonal moments of an image intensity distribution are first presented below. If $\{t_n(x)\}$ is a set of discrete orthogonal polynomials, i.e., it satisfies the following condition

$$\sum_{x=0}^{N-1} t_m(x) t_n(x) = \rho(n, N) \delta_{mn}, \quad (3)$$

where $m, n = 0, 1, 2, \dots, N-1$, and $\rho(n, N)$ is the squared norm of the polynomial set $\{t_n(x)\}$, then any bounded function $f(x, y)$, $0 \leq \{x, y\} \leq N-1$, has the following polynomial representation

$$f(x, y) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} T_{mn} t_m(x) t_n(y), \quad (4)$$

where the coefficients are given by

$$T_{pq} = \frac{1}{\rho(p, N)\rho(p, N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} t_p(x) t_q(y) f(x, y), \quad (5)$$

with $p, q = 0, 1, 2, \dots, N-1$. It can be recognized that if we define a discrete orthogonal moment functions as in Eq. (5) with $\{t_n(x)\}$ as the basis set, then the image can be reconstructed from the moments according to Eq. (4). This indicates that discrete orthogonal moments can effectively represent the features of the intensity images and the original image data are effectively compressed.

The classical discrete Tchebichef polynomials satisfy the property of orthogonality presented in Eq. (3), with

$$t_n(x) = n! \sum_{k=0}^n (-1)^{n-k} \binom{N-1-k}{n-k} \binom{n+k}{n} \binom{x}{k}. \quad (6)$$

In this case, according to Eq. (3), $\rho(n, N)$ can be obtained as

$$\rho(n, N) = \frac{N(N^2-1)(N^2-2^2)\cdots(N^2-n^2)}{2n+1}. \quad (7)$$

However, the value of $t_n(x)$ grows as N^n , and the value of the moment T_{pq} grows as $N^{-(p+q)}$, making it not suitable for feature representation. To solve this problem, the Tchebichef polynomials are usually scaled with a factor of N^n , and $\rho(n, N)$ are scaled with the factor of N^{2n} correspondingly. In this case, there will not be large variations in the dynamic range of values of the moments computed with Eq. (5).

It seems that the scaled Tchebichef polynomials presented above are very complicated to calculate and program. In effect, they can be conveniently calculated and programmed with the following recurrence formula,

$$\tilde{t}_n(x) = \frac{2n-1}{n} \tilde{t}_1(x) \tilde{t}_{n-1}(x) - \frac{n-1}{n} \left[1 - \frac{(n-1)^2}{N^2} \right] \tilde{t}_{n-2}(x), \quad (8)$$

where $n = 2, 3, \dots, N-1$, and

$$\tilde{t}_0(x) = 1, \quad (9)$$

$$\tilde{t}_1(x) = (2x+1-N)/N. \quad (10)$$

Therefore, the discrete orthogonal Tchebichef moments not only have a superior capacity for image feature representation, but are very convenient for application.

3.2 Feature-based phase retrieval approach

The sketch map of the feature-based phase retrieval wavefront sensing approach is shown in Fig. 3. Specifically, the problem of phase retrieval wavefront sensing is converted to a feature-based nonlinear fitting problem. The discrete orthogonal Tchebichef moments are utilized to extract or represent the features of the point spread functions (PSFs) at the in-focus and defocus image planes. This feature extraction process can be seen as a data compression process, which can efficiently reduce the effective data. The two-dimensional intensity image data is converted to a one-dimensional feature vector. Then this feature vector is utilized as the input of the neural network, which is a powerful nonlinear fitting tool and can establish the nonlinear mapping between these image features and the corresponding wavefront

aberration coefficients. Compared to the conventional iterative phase retrieval approaches, this approach can directly output the aberration coefficients of the optical system without the need for the time-consuming iterative transformation or optimization process. Compared to the current intensity-based neural-network approaches, this feature-based approach has a far low computational load on the hardware. The neural network with the image features serving as the input has a simple structure and is very convenient to train and implement.

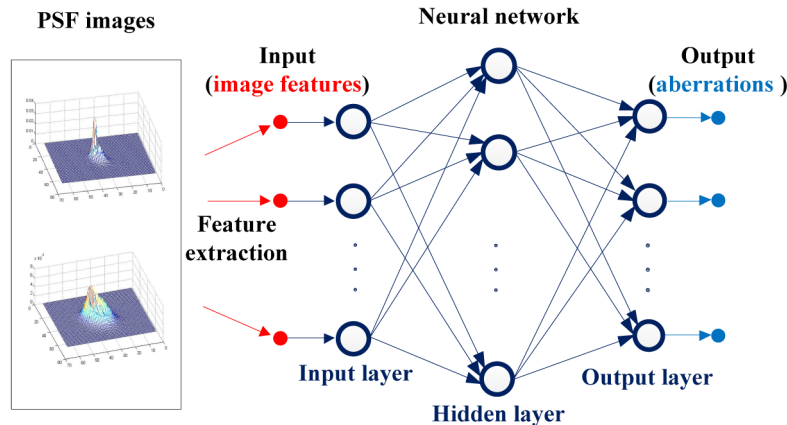


Fig. 3. Sketch map of the feature-based phase retrieval wavefront sensing approach using machine learning.

Note that in this approach, a pair of PSF images obtained at different focal planes are also needed, for the mathematical mapping from the set of all possible pupil phase screens to the set of all possible intensity distributions is a many-to-one mapping. Therefore, to invert this mapping and guarantee the uniqueness of the solution for wavefront phase, a pair of PSF images with a known defocus diversity between them are needed here.

The application procedure of the feature-based phase retrieval wavefront sensing approach is presented below (also illustrated in Fig. 4):

- (1) Determine the system parameters needed in phase retrieval wavefront sensing, mainly including wavelength, aperture size, focal length, pixel size of the detector, and the defocusing length (used to obtain a pair of PSF images at different focal planes). The geometric of the pupil should also be precisely determined. These parameters are the premise for us to generate the data set needed for training the network.
- (2) Generate the data set for training the neural network under the specified system parameters. Specifically, within certain range of the wavefront aberration coefficients, a set of aberration coefficients is randomly introduced; a pair of in-focus and defocus PSF images can be generated using this set of aberration coefficients according to the principle of Fourier optics. An error in the defocus distance is considered in this process to simulate the actual defocus error and a proper level of noise is introduced to the generated PSF images to simulate the practical noisy condition. After appropriate pretreatment, the discrete orthogonal Tchebichef moments of the pair of PSF images are extracted. The extracted features serve as one column of input matrix and the corresponding aberration coefficients serve as one column of the output matrix. This process is illustrated in Fig. 4. After a large number of the repetition of the above process, the input data set and output data set can be generated.
- (3) Properly select the number of neurons in each layer and train the neural network with the input data set and the corresponding output data set.

- (4) Apply the trained neural network to determine the wavefront aberration coefficients with the PSF images actually collected from the optical system. Note that the image features should be extracted first before they can be handled with the neural network.

While the pretreatment before feature extraction is not included in Fig. 4, this process should be taken seriously. The pretreatment mainly includes smooth de-noising, intensity normalization and sub-pixel translation. The purpose of intensity normalization is to make sure that the same wavefront aberration coefficients correspond to the same intensity distribution (such that it is not affected by the intensity of the light source). The purpose of sub-pixel image translation is to guarantee that the same wavefront aberration coefficients (do not include tip and tilt terms) correspond to the same position of the PSF in the image (in other words, the effects of tip-tilt terms on the position of the PSF in the image should be eliminated). These two aspects are important for guaranteeing the accuracy of the nonlinear mapping between the aberration coefficients and the PSF intensity distribution established by the neural network.

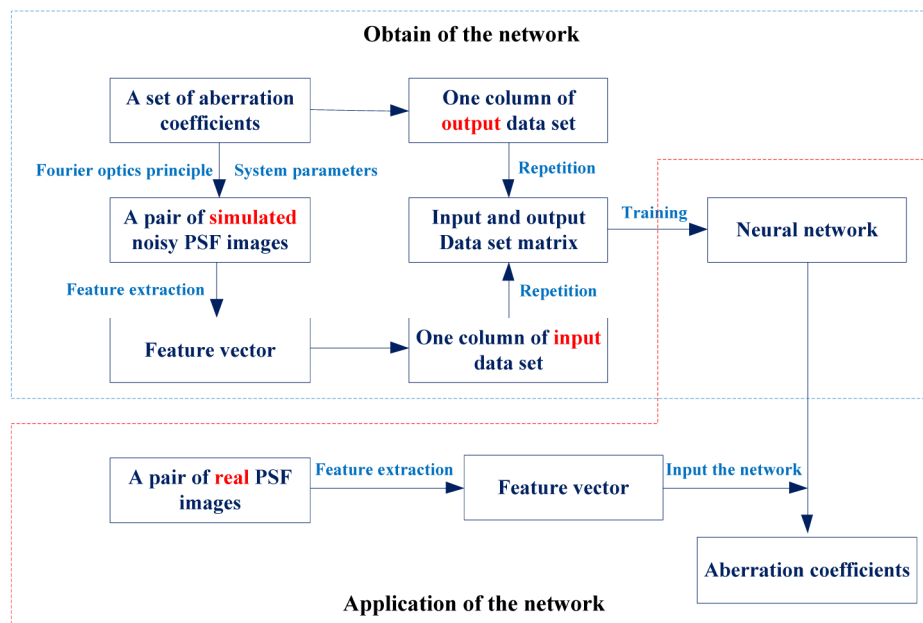


Fig. 4. Application procedure of the feature-based phase retrieval wavefront sensing approach using machine learning.

4. Experimental validation

In this section, an experiment will be performed to validate the effectiveness of the proposed approach. We first introduce the experimental setup and obtain a suitable neural network which takes into consideration the image noise and the error in the defocus distance. The training result can also show the non-linear fitting accuracy of the neural network. Then we apply the neural network to the collected PSF images and recover the wavefront aberrations. The accuracy of the experimental results are demonstrated and analyzed. Besides, some other discussions concerning the contradiction between the accuracy and capture range of the proposed approach are also presented.

4.1 Experimental setup and obtain of the neural network

The sketch and physical map of the optical system used in the experiment are shown in Fig. 5(a) and Fig. 5(b), respectively. The interferometer (PhaseCam 4020) in Fig. 5 performs two major roles. On one hand, it can directly measure the aberrations of the optical system which

is composed only by one lens (the aberrations of this system are introduced by slightly tilting or translating the lens (while the aperture stop stays unchanged) so that an off-axis field is used). The results of interferometer will be compared to results of the proposed approach to demonstrate the accuracy of the proposed approach. On the other hand, the focus of interferometer is used to serve as point light source. The beam passes through the system two times and a PSF can be obtained at the detector. By using the one-dimensional precision translation stage we can obtain a pair of PSFs with a known defocus diversity between them. The focal length of the lens is 180mm, the diameter of the aperture stop is 8.5mm, the defocus distance is 2mm, the wave length is 0.6328 μm , and the pixel size of the detector is 5.5 μm .

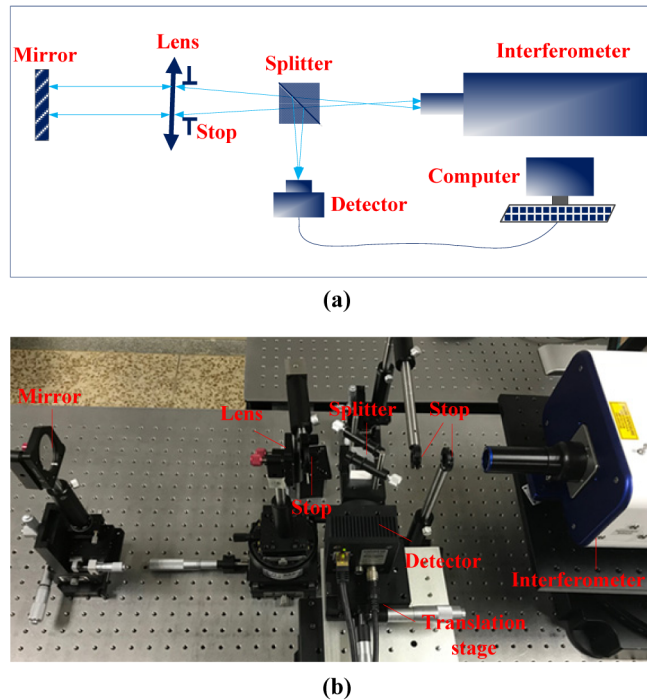


Fig. 5. The sketch (a) and physical map (b) of the optical system used in the experiment.

Then we will obtain a proper neural network for recovering the wavefront of the experimental optical system. To this end, 100000 sets of aberration coefficients (4th~9th Fringe Zernike coefficients corresponding to focus, astigmatism, coma and spherical aberration) are randomly generated in certain ranges which are shown in Table 1. These aberration coefficients constitute the output data set. The aberration ranges in Table 1 are partly determined according to the aberration property of the optical system used in the experiment. 100000 pairs of PSF images are computed with the specified optical parameters of the system according to Fourier optics. Meanwhile, an error in the defocus distance within the range of $[-0.1\text{mm}, 0.1\text{mm}]$ is considered in this experiment. A noise level of 50dB is added to the PSF images to simulate the practical noisy condition. The low order ($p, q = 0, 1, 2, 3, 4$) Tchebichef moments of each pair of PSF images are extracted, which constitute the input data set. Then the neural network can easily be trained using the neural network fitting tool of Matlab (the number of nodes in the hidden layer is selected as 30). In this process, the data set is separated into three parts, i.e., training set, validation set and test set. The training set is used for learning, which is to fit the weights of the network; the validation set is used for tuning the final architecture of the network; the test set is only used

for assess the performance of the network. The ratio between them in this work is 40%: 30%: 30% (i.e. 40000 sets for training, 30000 sets for validation, and 30000 sets for test).

Table 1. Ranges of different aberration coefficients for generating the data set

Coefficient	C4	C5	C6	C7	C8	C9
Range	[-0.5,0.5]	[-0.7,0.7]	[-0.7,0.7]	[-0.3,0.3]	[-0.3,0.3]	[-0.1,0.1]

These Fringe Zernike coefficients are in λ ($\lambda = 632.8\text{nm}$)

The training result is shown in Fig. 6, which provides the distribution of the error between the targets and the actual outputs of the network in the form of histogram. We can roughly recognize that for most of the cases the fitting error is within 0.025 waves. More specifically, the root mean square errors between the targets and outputs of the network in the training set, validation set and testing set are 0.0088 waves, 0.0089 waves, and 0.0089 waves, respectively (the mean squared errors between the targets and outputs are $7.75\text{e-}5$, $7.94\text{e-}5$, and $7.87\text{e-}5$, respectively). This partly demonstrates the accuracy of the neural network with Tchebichef moment features serving as the input in the presence of image noise and the error in defocus distance.

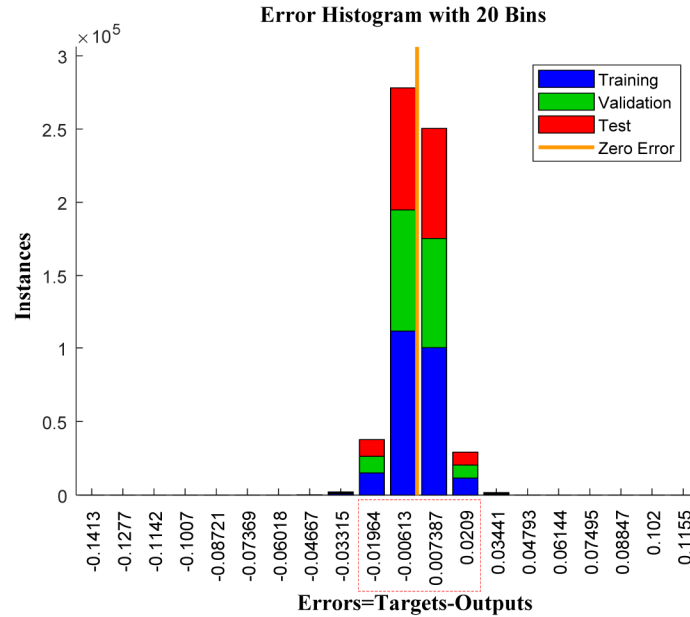


Fig. 6. Distribution of the error between the targets and the actual outputs of the network in the form of histogram in the presence of image noise and the error in the defocus distance.

4.2 Application of the neural network and result analysis

Then we will apply the obtained neural network to the real PSF images collected from the experimental optical system. After proper pretreatment mentioned in Section 3, the Tchebichef moments of the PSF images are extracted, which will serve as the input. The neural network can directly output the aberration coefficients of optical system corresponding to each pair of PSF images.

The effectiveness of the proposed approach is validated using the following two methods. On one hand, we use the recovered aberration coefficients to re-generate a pair of PSF images (in-focus image and defocus image) according to Fourier optics. The effectiveness of the proposed approach can be qualitatively validated by comparing this pair of generated PSF images with those real PSF images collected from the optical system. In this experiment, we collect 20 pairs of PSF images, and 20 sets of aberration coefficients are obtained with the

proposed approach. Then 20 pairs of PSF images are regenerated. The comparison between the collected PSF images and regenerated PSF images is shown in Fig. 7.

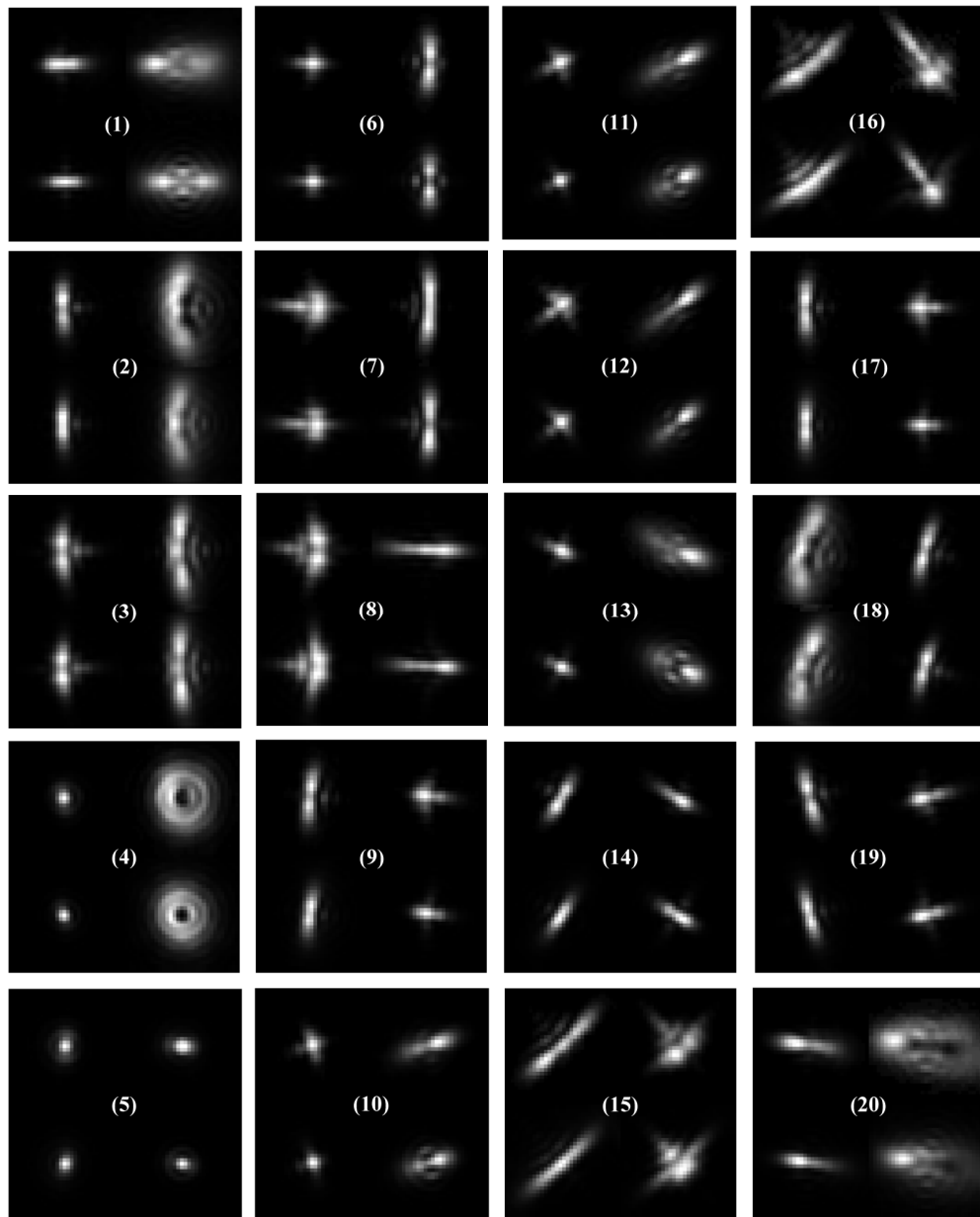


Fig. 7. Comparison between the 20 pairs of real PSF images and the 20 pairs of regenerated PSF images. In each of the 20 images in this figure, the upper two PSFs are collected from the optical system (at different focal planes) and the two PSF images below are generated with the recovered aberration coefficients. We can recognize that the regenerated PSF images bear strong similarities with those real collected, which qualitatively demonstrate the accuracy of the recovered aberration coefficients.

Table 2. Comparison between the astigmatic ($C5/C6$) and coma ($C7/C8$) aberration coefficients measured by interferometer (A) and those recovered using the proposed approach (B)

Index	Type	$C5$	$C6$	$C7$	$C8$	Mean Error
1	A	0.4095	0.0206	-0.0311	0.0007	0.0214
	B	0.3741	0.0068	-0.0114	0.0002	
2	A	-0.3593	-0.0282	-0.1206	0.0283	0.0333
	B	-0.4156	-0.0064	-0.0982	0.0115	
3	A	-0.5594	-0.0379	-0.1039	0.0106	0.0176
	B	-0.5382	-0.0306	-0.1110	-0.0157	
4	A	-0.1025	-0.0277	-0.0752	0.0061	0.0233
	B	-0.0873	-0.0488	-0.0380	-0.0046	
5	A	0.1147	-0.0230	0.0661	0.0021	0.0241
	B	0.0727	-0.0381	0.0507	-0.0075	
6	A	-0.3747	0.0339	0.0487	-0.0019	0.0204
	B	-0.3541	0.0080	0.0585	-0.0236	
7	A	-0.5887	0.0464	0.0993	0.0038	0.0291
	B	-0.5482	0.0309	0.1123	-0.0327	
8	A	0.6683	-0.0535	0.1154	-0.0075	0.0301
	B	0.6206	-0.0624	0.1510	-0.0112	
9	A	0.4222	-0.0926	-0.0993	0.0174	0.0413
	B	0.3409	-0.0912	-0.0855	0.0145	
10	A	0.2765	0.1906	0.0682	0.0485	0.0432
	B	0.2122	0.1381	0.0764	0.0263	
11	A	0.2189	0.2851	0.0759	0.0594	0.0506
	B	0.1567	0.2103	0.0889	0.0348	
12	A	-0.0929	0.4374	0.0954	0.0730	0.0500
	B	-0.0615	0.3436	0.0927	0.0580	
13	A	0.2273	-0.2208	0.0704	-0.0460	0.0244
	B	0.1819	-0.2204	0.087	-0.0398	
14	A	0.1768	-0.3173	0.0638	-0.0637	0.0137
	B	0.1538	-0.3269	0.0577	-0.0539	
15	A	-0.0505	-0.6406	0.0641	-0.1176	0.0241
	B	-0.0740	-0.6062	0.0832	-0.1024	
16	A	-0.1815	-0.7597	0.0776	-0.1459	0.0536
	B	-0.1924	-0.6551	0.0868	-0.1645	
17	A	0.4486	-0.0006	-0.0875	0.0127	0.0413
	B	0.3679	-0.0103	-0.0752	0.0039	
18	A	0.4360	-0.2481	-0.1095	0.0206	0.0283
	B	0.3909	-0.2436	-0.1027	0.0537	
19	A	0.3914	0.2134	-0.1043	-0.0144	0.0157
	B	0.3715	0.2002	-0.0869	-0.0253	
20	A	0.4314	-0.1113	-0.0851	0.0357	0.0301
	B	0.4507	-0.1260	-0.1399	0.0419	

In Fig. 7, there are 20 images, and each image includes the collected in-focus (up-left) and defocus (up-right) PSF images as well as the generated in-focus (down-left) and defocus (down-right) PSF images. We can easily recognize that the collected PSF images bear strong similarities with those regenerated PSF images, which can qualitatively validate the effectiveness of the proposed approach.

On the other hand, the interferometer can also directly measure the wavefront of the optical system, the result of which will serve as the reference data to quantitatively evaluate the accuracy of the neural network. Note that the focus of the interferometer and the detector used to collect PSF images are generally located at different axial positions (or focal planes), and therefore the interferometer cannot validate the accuracy of those rotationally symmetric aberrations (focus and spherical aberration) which are sensitive to axial position. However, the non-rotationally symmetric aberrations (astigmatism and coma) are not sensitive to this position (the non-path error induced by the splitter should be considered and calibrated in advance). The comparison between the astigmatism ($C5/C6$) and coma ($C7/C8$) measured by interferometer and those recovered using the proposed approach (corresponding to the 20 pairs of PSF images in Fig. 7) are shown in Table 2.

In Table 2, for a better shown of the accuracy of the proposed approach, if the deviation between the 4 measured coefficients and recovered coefficients is larger than 0.03 waves, we mark it with red color and if the deviation is larger than 0.05 waves, we further make it in bold. We can see that for most cases the error of each recovered coefficient is smaller than 0.03 waves. Besides, “Mean Error” represents the root mean square error between the 4 measured coefficients using interferometer and the 4 recovered coefficients using the proposed approach. The average value of the “Mean Error” for the 20 sets of experiment is 0.031 waves, which can definitely demonstrate the effectiveness and accuracy of the proposed approach.

We should also explain why the magnitude of $C8$ is usually smaller than $C7$ in the experiment. The optical system is composed by only one lens and an aperture stop. This lens can be decentered in two directions while it can only be tilted in one direction (only the azimuth angle can be adjusted and pitching angle cannot be adjusted) in this experiment. The coma of the system is sensitive to the tilt of the lens and the change of azimuth angle mainly introduces a coma in x direction which corresponds to 7th Fringe Zernike coefficients. The decenter of lens relative to the aperture stop can introduce a coma in y direction. On the other hand, the training of neural network does not take this fact into consideration and there are several cases which have a relatively large coma in y direction (15th and 16th experiment in Table 2). Therefore, the experiment is enough for us to validate the effectiveness of the proposed approach.

4.3 Other discussions

This subsection further presents some discussions concerning the contradiction between the accuracy and capture range of the proposed approach. We can see from Table 2 that the accuracy of the recovered coma aberration coefficients ($C7/C8$) is higher than the recovered astigmatic aberration coefficients ($C5/C6$). The main reason for this is that the range of astigmatism considered in Table 1 is larger than the range of coma. A larger capture range can generally means a sacrifice in non-linear fitting accuracy.

Here we will present some suitable simulations to demonstrate this statement. Another case with smaller ranges of aberration coefficients ($C4-C9$) are considered, which are shown in Table 3.

Table 3. Another case with a different range of aberration coefficients

Coefficient	C4	C5	C6	C7	C8	C9
Range	[-0.25,0.25]	[-0.35,0.35]	[-0.35,0.35]	[-0.15,0.15]	[-0.15,0.15]	[-0.1,0.1]

These Fringe Zernike coefficients are in λ ($\lambda = 632.8\text{nm}$)

For this case, 100000 sets of aberration coefficients are randomly generated for training the neural network, as mentioned before. Apart from the range of aberration coefficients, other conditions are the same for these two cases (those shown in Table 1 and Table 3), including the noise level and range of error in defocus distance. We can then obtain another neural network corresponding to this case, and the fitting error of each network can be

illustrated with a histogram, which shows the error distribution in detail. The result is shown in Fig. 8.

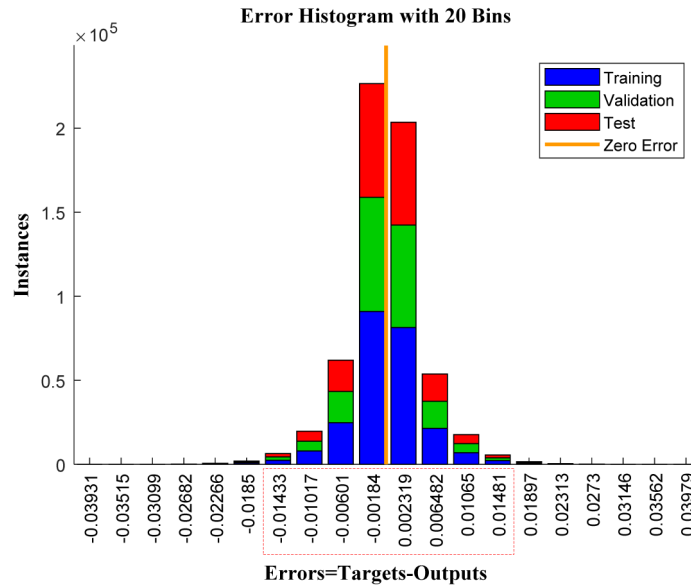


Fig. 8. Distribution of the error between the targets and the actual outputs of the network in the form of histogram for the case shown in Table 3.

We can recognize that the fitting error shown in Fig. 8 is much smaller than that shown in Fig. 6. Specifically, the root mean square errors between the targets and outputs of the network in the training set, validation set and testing set are 0.0053 waves, 0.0054 waves, and 0.0054 waves, respectively (the mean squared errors between the targets and outputs are $2.82\text{e-}5$, $2.87\text{e-}5$, and $2.89\text{e-}5$, respectively). Besides, we will also apply this new neural network to the images obtained in the experiment to further demonstrate its accuracy in practical condition. The 4th, 5th, 10th, 11th, and 13th sets of PSF images in Table 2 are used here, for these cases are within the capture range of the new neural network (shown in Table 3). The results are shown in Table 4, where the line C represents the recovered aberration coefficients with the new network.

Table 4. Demonstration for accuracy of the new neural network with a smaller capture range

Index	Type	C5	C6	C7	C8	Mean error
4	A	-0.1025	-0.0277	-0.0752	0.0061	0.0151
	C	-0.0820	-0.0202	-0.0871	0.0233	
5	A	0.1147	-0.0230	0.0661	0.0021	0.0233
	C	0.0912	-0.0291	0.0308	0.0204	
10	A	0.2765	0.1906	0.0682	0.0485	0.0218
	C	0.2496	0.1768	0.0464	0.0712	
11	A	0.2189	0.2851	0.0759	0.0594	0.0168
	C	0.1898	0.2729	0.0648	0.0617	
13	A	0.2273	-0.2208	0.0704	-0.0460	0.0191
	C	0.2056	-0.2151	0.0962	-0.0625	

The Fringe Zernike coefficients (C5~C8) are in λ ($\lambda = 632.8\text{nm}$)

Comparing Table 4 with the corresponding cases in Table 2, we can recognize a parent decrease in "Mean Error". Besides, the magnitude of this decrease is much larger than the deviation between Fig. 8 and Fig. 6 (these two figures are obtained through software

simulation), which further indicates that the new trained network is more robust to non-ideal practical measurement environment for those cases within its capture range.

On the other hand, while the new trained neural network (with a smaller capture range) is more accurate for those cases within its capture range, the accuracy of it may decrease to a large extent for those cases that go beyond its capture range. The results of applying the new neural network to several cases in Table 2 beyond its capture range (the 2th, 3th, 7th, 8th, 15th, 16th, 17th, 18th, 19th, and 20th cases) are shown in Table 5. We can draw the following conclusions from Table 5:

- (1) On one hand, the new neural network still has the ability of recovering the aberration coefficients for the cases beyond its capture range (defined by Table 3). We can see that the network can still output a rough solution for the cases in Table 5.
- (2) On the other hand, the accuracy of the network can no longer be guaranteed for the cases beyond its capture range. It appears to vary considerably among these individual cases. On the whole, this accuracy shown in Table 5 is much lower than that shown in Table 2.

Table 5. The results of applying the new neural network to some cases beyond its capture range

Index	Type	C5	C6	C7	C8	Mean Error
2	A	-0.3593	-0.0282	-0.1206	0.0283	0.0389
	C	-0.4105	-0.0141	-0.1759	0.0153	
3	A	-0.5594	-0.0379	-0.1039	0.0106	0.0376
	C	-0.4981	-0.0184	-0.1397	-0.0045	
7	A	-0.5887	0.0464	0.0993	0.0038	0.0585
	C	-0.4962	0.0912	0.0489	-0.0207	
8	A	0.6683	-0.0535	0.1154	-0.0075	0.0443
	C	0.6100	-0.0340	0.1548	-0.0576	
15	A	-0.0505	-0.6406	0.0641	-0.1176	0.0487
	C	-0.0279	-0.7000	0.0069	-0.1642	
16	A	-0.1815	-0.7597	0.0776	-0.1459	0.0823
	C	-0.1566	-0.6521	0.0369	-0.2611	
17	A	0.4486	-0.0006	-0.0875	0.0127	0.1358
	C	0.4270	-0.0131	0.1601	0.1213	
18	A	0.4360	-0.2481	-0.1095	0.0206	0.1943
	C	0.2754	-0.1960	0.2115	0.1603	
19	A	0.3914	0.2134	-0.1043	-0.0144	0.0429
	C	0.4235	0.2521	-0.0622	0.0409	
20	A	0.4314	-0.1113	-0.0851	0.0357	0.0222
	C	0.4094	-0.1305	-0.0542	0.0483	

The Fringe Zernike coefficients (C5~C8) are in λ ($\lambda = 632.8\text{nm}$).

Therefore, we can see that there is a contradiction here between the capture range and accuracy of the neural network, i.e., to guarantee the robustness of the neural network and avoid the cases where the actual wavefront errors go beyond the capture range, we should increase the specified ranges of the aberration coefficients for training the neural network, while this can decrease the fitting accuracy of it.

One possible solution to this contradiction is using two kinds of neural network, one for rough classification, and another one kind for fine fitting. Specifically, on one hand, we can use a neural network with a very large capture range to determine the rough RMS error and the dominant kind of aberrations for a certain case. On the other hand, we can train a series of neural networks corresponding to different ranges of RMS error or different kinds of the

dominant aberrations, and then we select a proper one for fine fitting according to the result of rough classification.

5. Conclusion

This paper proposes a novel phase retrieval mechanism using machine learning, where the phase retrieval problem is converted to a feature-based nonlinear fitting problem. The discrete orthogonal Tchebichef moments which do not involve numerical approximation of continuous integrals and coordinate space transformation are introduced to extract the features of the in-focus and defocus PSF images. Then the back-propagation neural network is utilized as the nonlinear fitting tool to establish the input-output mapping between the image features and the wavefront aberration coefficients of the optical system. Once well trained, the neural network can directly output the aberration coefficients of the optical system to a good precision with these image features serving as the input. Compared to the conventional iterative phase retrieval approaches, this feature-based approach is lower in computational load and higher in efficiency and robustness (free of stagnation problem). Compared to the current intensity-based phase retrieval approaches using machine learning which are complicated in structure and very hard to train, this feature-based approach is more convenient to train and more practical to use for average researchers in the area of optics. Besides, adequate experiments are implemented to demonstrate the effectiveness and accuracy of proposed approach while the current image-based wavefront sensing methods using machine learning are lacking in experimental validation.

While the accuracy of this feature-based approach may be still a little lower than the conventional iterative phase retrieval approaches, it is acceptable for general cases. On the other hand, considering that the traditional iterative phase retrieval approaches are susceptible to the stagnation problem, especially when a bad initial value is selected, this feature-based approach can be used for providing a perfect initial value for the traditional iterative approaches, and a more accurate result can then be obtained easily. Besides, we point out and discuss the contradiction between the capture range and fitting accuracy of the proposed method. We also propose a possible solution to this problem, which is our future work.

This work presents a feasible and easy-implemented way to improve the efficiency and robustness of the phase retrieval wavefront sensing and contributes to the application of machine learning methods to the area of image-based wavefront sensing.

6. Funding

National Natural Science Foundation of China (NSFC) (61705223); National Key Research and Development Program (2016YFE0205000).

References

1. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," *Optik (Stuttg.)* **35**, 237–246 (1972).
2. J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**(15), 2758–2769 (1982).
3. P. Groot, "Phase-shift calibration errors in interferometers with spherical Fizeau cavities," *Appl. Opt.* **34**(16), 2856–2863 (1995).
4. J. Vargas, L. González-Fernandez, J. A. Quiroga, and T. Belenguer, "Calibration of a Shack-Hartmann wavefront sensor as an orthographic camera," *Opt. Lett.* **35**(11), 1762–1764 (2010).
5. J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble Space Telescope characterized by using phase-retrieval algorithms," *Appl. Opt.* **32**(10), 1747–1767 (1993).
6. C. Roddier and F. Roddier, "Combined approach to the Hubble Space Telescope wave-front distortion analysis," *Appl. Opt.* **32**(16), 2992–3008 (1993).
7. D. Redding, P. Dumont, and J. Yu, "Hubble Space Telescope prescription retrieval," *Appl. Opt.* **32**(10), 1728–1736 (1993).
8. C. Roddier and F. Roddier, "Wavefront reconstruction from defocused images and the testing of ground-based optical telescopes," *J. Opt. Soc. Am. A* **10**(11), 2277–2287 (1993).
9. B. H. Dean, D. L. Aronstein, J. S. Smith, R. Shiri, and D. S. Acton, "Phase Retrieval Algorithm for JWST Flight and Testbed Telescope," *Proc. SPIE* **6265**, 626511 (2006).

10. D. L. Misell, "An examination of an iterative method for the solution of the phase problem in optics and electron optics," *J. Phys. D* **6**(18), 2200–2225 (1973).
11. R. A. Gonsalves and R. C. Hidlaw, "Wavefront sensing by phase retrieval," *Proc. SPIE* **207**, 32–39 (1979).
12. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Opt. Eng.* **21**(5), 829–832 (1982).
13. R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Am. A* **9**(7), 1072–1085 (1992).
14. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," *Appl. Opt.* **32**(10), 1737–1746 (1993).
15. D. G. Sandler, T. K. Barrett, D. A. Palmer, R. Q. Fugate, and W. J. Wild, "Use of a neural network to control an adaptive optics system for an astronomical telescope," *Nature* **351**(6324), 300–302 (1991).
16. J. R. P. Angel, P. Wizinowich, M. Lloyd-Hart, and D. Sandler, "Adaptive optics for array telescopes using neural-network techniques," *Nature* **348**(6298), 221–224 (1990).
17. T. K. Barrett and D. G. Sandler, "Artificial neural network for the determination of Hubble Space Telescope aberration from stellar images," *Appl. Opt.* **32**(10), 1720–1727 (1993).
18. S. W. Paine and J. R. Fienup, "Machine learning for improved image-based wavefront sensing," *Opt. Lett.* **43**(6), 1235–1238 (2018).
19. L. D. Harmon, "Studies with artificial neurons. I. Properties and functions of an artificial neuron," *Kybernetik* **1**(3), 89–101 (1961).
20. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.
21. L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Trans. Signal. Inf. Process.* **3**, e2 (2014).
22. I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," *International Conference on International Conference on Machine Learning* **38**, III-1139 (2013).
23. R. Mukundan, S. H. Ong, and P. A. Lee, "Image analysis by Tchebichef moments," *IEEE Trans. Image Process.* **10**(9), 1357–1364 (2001).
24. A. Kumar, R. Paramesran, C.-L. Lim, and S. C. Dass, "Tchebichef moment based restoration of Gaussian blurred images," *Appl. Opt.* **55**(32), 9006–9016 (2016).
25. R. Mukundan, "Some computational aspects of discrete orthonormal moments," *IEEE Trans. Image Process.* **13**(8), 1055–1059 (2004).
26. A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer* **29**(3), 31–44 (1996).
27. A. T. C. Goh, "Back-propagation neural networks for modeling complex systems," *Artif. Intell. Eng.* **9**(3), 143–151 (1995).