Optics EXPRESS

Object-independent image-based wavefront sensing approach using phase diversity images and deep learning

QI XIN,[1,2] GUOHAO JU,[1,*] CHUNYUE ZHANG,[1] AND SHUYAN XU[1]

[1] *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*
**juguohao@ciomp.ac.cn*

Abstract: This paper proposes an image-based wavefront sensing approach using deep learning, which is applicable to both point source and any extended scenes at the same time, while the training process is performed without any simulated or real extended scenes. Rather than directly recovering phase information from image plane intensities, we first extract a special feature in the frequency domain that is independent of the original objects but only determined by phase aberrations (a pair of phase diversity images is needed in this process). Then the deep long short-term memory (LSTM) network (a variant of recurrent neural network) is introduced to establish the accurate non-linear mapping between the extracted feature image and phase aberrations. Simulations and an experiment are performed to demonstrate the effectiveness and accuracy of the proposed approach. Some other discussions are further presented for demonstrating the superior non-linear fitting capacity of deep LSTM compared to Resnet 18 (a variant of convolutional neural network) specifically for the problem encountered in this paper. The effect of the incoherency of light on the accuracy of the recovered wavefront phase is also quantitatively discussed. This work will contribute to the application of deep learning to image-based wavefront sensing and high-resolution image reconstruction.

1. Introduction

The resolution of an incoherent imaging system is often limited by phase aberrations. Phase aberrations can arise from a variety of sources including atmospheric turbulence and mirror misalignments and figure errors due to dynamic thermal or gravitational variations. Knowledge of phase aberrations affords either their correction by using adaptive optics or active optics, or post detection deblurring of the imagery [1].

Image-based wavefront sensing is a method of measuring the wavefront phase distribution in the pupil plane using focal plane images [2]. It has several important advantages over the conventional wavefront sensors, such as low requirement for optical hardware that could also be subject to misalignments and no special need for calibration. Therefore, this method is particularly suitable for wavefront sensing in space telescopes [3,4]. Other applications of it include measurement of a laser beam [5] and biological microscopy imaging [6].

It is of great significance that the image-based wavefront sensing approach is not restricted by the objects being imaged and cannot only be applicable to point source but also applicable to different extended scenes at the same time. A point object is not available in many imaging scenarios. Even for astronomical applications, the assumption of a point object involves some risk owing to the abundance of binary stars. If the optical system is expected to be looking at resolved targets, one needs to evaluate the performance with extended scenes, which can be substantially different from that with a point target.

Machine learning with artificial neural networks (ANNs), including deep learning using convolutional neural networks (CNNs), has been introduced to the area of computational

#366468 https://doi.org/10.1364/OE.27.026102
Journal © 2019 Received 30 Apr 2019; revised 4 Aug 2019; accepted 5 Aug 2019; published 29 Aug 2019

imaging [7,8] and image-based wavefront sensing [9–12]. ANNs are input-output information processors composed of parallel layers of elements or neurons which are capable of elementary arithmetic. ANNs can be used to fit the complicated input-output mappings between wavefront phase and intensity images. Image-based wavefront sensing method based on machine learning has several important advantages compared to traditional G-S phase retrieval [13,14] or phase diversity methods [15–17], such as a high efficiency (needless of the time-consuming iteration process) and robustness (completely free from the stagnation problem). Therefore, image-based wavefront sensing using machine learning has a great prospect for development and wide application.

However, the application of the current machine learning wavefront sensing methods to the case of an unknown object is still restricted to some extent. On one hand, some of the current machine learning approaches are only applicable to point sources [9–11]. On the other hand, while some recent researches can be applicable to extended scenes, it seems that the type of the extended scenes is still restricted (for example, selected handwritten numbers taken from the EMNIST database) [12]; Besides, this method needs to generate an extremely large number of different extended objects for training of the network (about 1,000,000), which is not only very time-consuming and hard to implement, but also will pose a great challenge to the storage and computing capacity of the computer (extremely expensive GPU are usually need [7,12]). These facts pose a tremendous obstacle to the promotion of image-based wavefront sensing using machine learning. The underlying reason is that in the current methods the intensities of extended images in the space domain are directly taken as the input of the network. When both of the objects and phase aberrations of the system are completely random and unknown, the applicability and practicability of the current deep learning methods for image-based wavefront sensing are limited.

In this paper, rather than directly recovering phase aberrations from intensities of focal images in the space domain, we first extract a special feature in the frequency domain which is related to phase aberrations but independent of the original objects. Meanwhile, a pair of focal plane images with a known defocus diversity between them is needed. Then the deep long short-term memory (LSTM) network is introduced as the non-linear fitting tool to establish the accurate mapping between the extracted feature image and phase aberrations. LSTM network is a variant of recurrent neural network (RNN), which not only have memory but also can solve the vanishing gradient problem in training due to long term dependencies [18,19]. LSTM networks can recognize and take advantage of the inherent relations between the intensities in the feature image with theirs elaborate memory units. Simulations and an experiment are performed to demonstrate the effectiveness and accuracy of the proposed approach. Some other discussions are also presented to demonstrate the higher fitting accuracy and computation efficiency of the deep LSTM neural network compared to Resnet 18 (a variant of CNN) particularly for the case encountered in this work. The influence of the incoherency of light on the accuracy of the recovered wavefront phase is also quantitatively discussed.

This paper is organized as follows. In Section 2, we introduce a special feature in the frequency domain which is independent of the object being imaged. Then we continue to propose an object-independent wavefront sensing approach using deep LSTM networks in Section 3. Simulations and an experiment are performed to demonstrate the effectiveness of the proposed approach in Section 4. Some other discussions on the proposed approach is presented in Section 5. In Section 6, we conclude the paper.

2. Object-independent feature extraction

We will only consider two images separated by a certain defocus distance. Let us suppose that the object is illuminated with non-coherent quasi-monochromatic light, and the imaging system is a linear shift-invariant system. The intensity distribution of the image plane $\mathbf{i}$ can be modeled by the following equation,

$$\mathbf{i} = \mathbf{o} * \mathbf{s}, \tag{1}$$

where $*$ denotes the convolution operation, $\mathbf{o}$ is the object to be found, $\mathbf{s}$ is the point spread function (PSF) of the optical system. We can see that the information of the object (which is unknown) is included in this equation, which will pose a great challenge for us to establish the accurate non-linear mapping between wavefront phase aberrations and focal-plane images.

To solve this problem, we will establish an equation which is independent of the objects. According to Fourier optics principle, in the frequency domain, Eq. (1) can be rewritten as

$$\mathbf{I} = \mathbf{O} \cdot \mathbf{S}, \tag{2}$$

where $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{S}$ are Fourier transforms of $\mathbf{i}$, $\mathbf{o}$, and $\mathbf{s}$, respectively. $\mathbf{S}$ is optical transfer function. When two images obtained at two different focal planes are available, according to Eq. (2), we can have that

$$\frac{\mathbf{I}_a}{\mathbf{I}_b} = \frac{\mathbf{S}_a}{\mathbf{S}_b}, \tag{3}$$

where the sub-scripts a and b represent that the related variables corresponding to images at two different focal planes. We can see that this equation no longer contains the information of the extended object.

Equation (3) can further be rewritten as

$$\frac{\mathcal{F}\{\mathbf{i}_a\}}{\mathcal{F}\{\mathbf{i}_b\}} = \frac{\mathbf{P}_a(\boldsymbol{\psi}) \otimes \mathbf{P}_a(\boldsymbol{\psi})}{\mathbf{P}_b(\boldsymbol{\psi}) \otimes \mathbf{P}_b(\boldsymbol{\psi})}, \tag{4}$$

where $\mathcal{F}\{\cdot\}$ is the Fourier transform operation, $\otimes$ is the auto-correlation operation, $\mathbf{P}_a$ and $\mathbf{P}_b$ are the complex pupil functions at the pupil plane corresponding to the two focal planes, respectively. $\mathbf{P}_a$ and $\mathbf{P}_b$ are functions of the wavefront phase aberrations, $\boldsymbol{\psi}$, i.e.

$$\mathbf{P}_a(\boldsymbol{\psi}) = \mathbf{p}\exp\{j\boldsymbol{\psi}\}, \tag{5}$$

$$\mathbf{P}_b(\boldsymbol{\psi}) = \mathbf{p}\exp\{j(\boldsymbol{\psi} + \boldsymbol{\Delta\psi})\}, \tag{6}$$

where $\mathbf{p}$ represents the binary aperture function with values of 1 inside the pupil and 0 outside, $j \equiv \sqrt{-1}$, and $\boldsymbol{\Delta\psi}$ is a known defocus diversity between the two focal planes.

Equations (4-6) establish an analytic mapping between the extended scene images ($\mathbf{i}_a$ and $\mathbf{i}_b$) at two focal planes and the wavefont phase of the optical system ($\boldsymbol{\psi}$), which does not include the information of the object ($\mathbf{o}$). Here we can define a feature image as

$$\mathbf{f} = \frac{\mathcal{F}\{\mathbf{i}_a\}}{\mathcal{F}\{\mathbf{i}_b\}}, \tag{7}$$

from which we can recover the wavefront phase, $\boldsymbol{\psi}$. Note that here $\mathbf{f}$ is a complex image and in practice we use another feature, i.e.

$$\mathbf{f}_0 = \left| \mathcal{F}^{-1}\left\{ \frac{\mathcal{F}\{\mathbf{i}_a\}}{\mathcal{F}\{\mathbf{i}_b\}} \right\} \right|, \tag{8}$$

where $\mathcal{F}^{-1}\{\cdot\}$ means inverse Fourier transform operation. The reason for us to select this feature is that $\mathbf{f}_0$ is an intensity image (similar with PSF) and the energy of this feature image is mainly concentrated at the central region. To suppress the effects of noise and decrease the complexity of the network, we can reduce the size of the actual feature image used to recover wavefront aberrations (the actual feature image for deep learning is only a portion of $\mathbf{f}_0$ near the central region). Illustration of the extracted feature for different objects in the presence of the same phase aberrations are shown in Fig. 1, where we can see that the feature image is independent of the object.

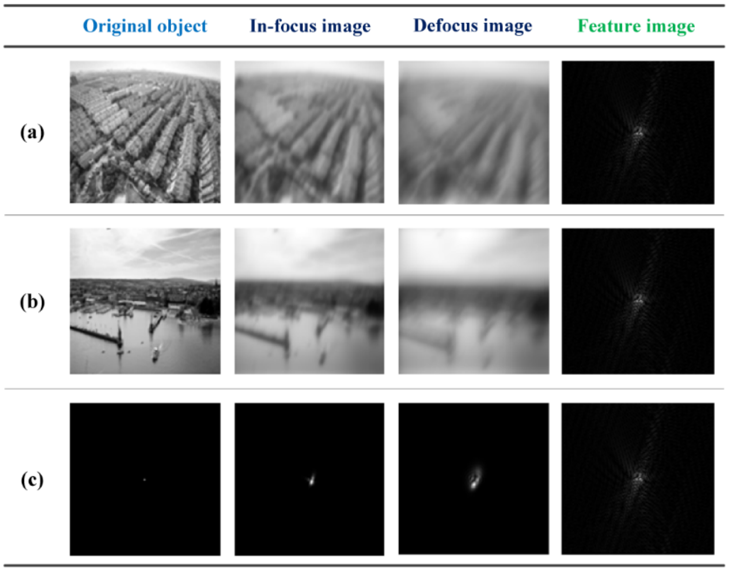

Fig. 1. Illustration of the extracted features for different objects in the presence of the same phase aberrations. (a) and (b) show two extended scenes, the corresponding in-focus and defocus images, as well as the feature images obtained with Eq. (7), respectively. (c) shows a point source, the corresponding in-focus and defocus images as well as the extracted feature image. We can clearly see that the feature images extracted from different objects are the same.

3. Object-independent wavefront sensing approach using deep LSTMs

3.1. Introduction of LSTM network

LSTM network is a variant of recurrent neural network (RNN). A RNN is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit temporal dynamic behavior for a time sequence. Given an input sequence $\mathbf{x} = (x_1, ..., x_T)$, a RNN computes the hidden vector sequence $\mathbf{h} = (h_1, ..., h_T)$ and output vector sequence $\mathbf{o} = (o_1, ..., o_T)$ by iterating the following equations from $t = 1$ to T:

$$\begin{aligned} h_t &= \mathcal{H}\left(W_{xh}x_t + W_{hh}h_{t-1} + b_h\right), \\ o_t &= W_{ho}h_t + b_o, \end{aligned} \tag{9}$$

where the W terms denote weight matrixes, b terms denote bias vectors, and $\mathcal{H}$ is the hidden layer activation function. For basic RNNs $\mathcal{H}$ is an elementwise application of a sigmoid function. Unfold of the basic RNN is shown in Fig. 2. This chain-like nature reveals that RNNs are intimately related to sequences and lists.

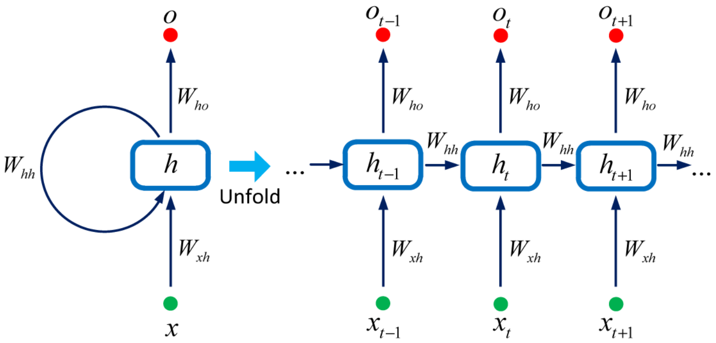

Fig. 2. Diagram of the unfold basic RNN.

In contrast to basic RNNs which use a single tanh layer to compute the hidden state, LSTM networks use purpose-built memory modules to compute the hidden state and store information. The LSTM module is shown in Fig. 3. In each LSTM module, $\mathcal{H}$ is implemented by the following composite function:

$$\begin{aligned}
F_t &= \sigma\left(W_{xF}x_t + W_{hF}h_{t-1} + b_F\right),\\
I_t &= \sigma\left(W_{xI}x_t + W_{hI}h_{t-1} + b_I\right),\\
O_t &= \sigma\left(W_{xO}x_t + W_{hO}h_{t-1} + b_O\right),\\
G_t &= \tan\text{h}\left(W_{xG}x_t + W_{hG}h_{t-1} + b_G\right),\\
c_t &= c_{t-1} \otimes F_t + I_t \otimes G_t,\\
h_t &= \tan\text{h}\left(c_t\right) \otimes O_t,
\end{aligned} \tag{10}$$

where σ is the logistic sigmoid function, tanh is the hyperbolic tangent function, $\otimes$ is an element multiplication, and F, I, O and G are intermediate variables. LSTM cells can well decide internally what to keep in (and what to erase from) memory. LSTM not only can use their memory modules to process sequences of inputs and recognize patterns in them, but also can solve the vanishing gradient problem in training due to long term dependencies. For this reason, we can conveniently construct and train a very deep LSTM network. The intensities of pixels are not independent for a certain image pattern and they have inherent relations. Deep LSTM networks can take full advantage of these relations when the images are decomposed into a series of patches and treated as a sequence.

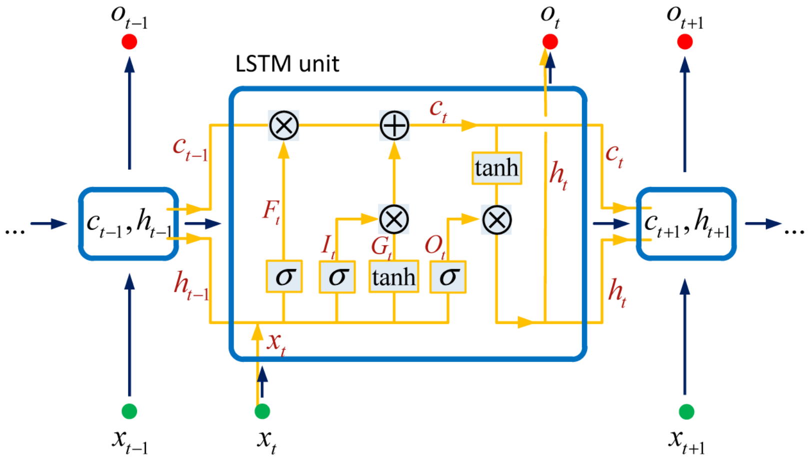

Fig. 3. Diagram of the LSTM unit.

3.2. Object-independent wavefront sensing approach using deep LSTMs

The object-independent wavefront sensing approach using deep learning is illustrated in Fig. 4. Note that LSTM networks are mainly used to process sequences and a feature image cannot directly be taken as the input. Therefore, we first decompose the feature image into a series of patches which can be regarded as a sequence (each patch corresponds to one element in the sequence). The output of the deep LSTM networks is still a sequence. This sequence is then taken as a vector, which serves as the input of a fully-connected layer and at last we can obtain a set of aberration coefficients.

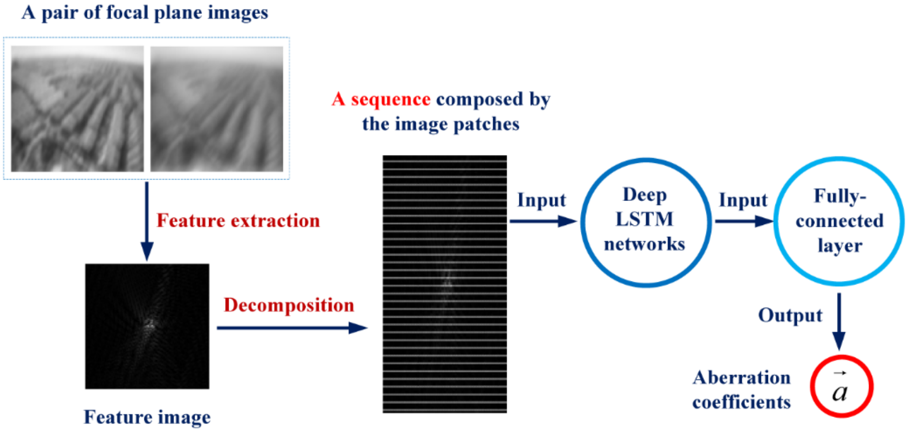

Fig. 4. Sketch map of the object-independent wavefront sensing approach using deep LSTM networks. The feature image extracted from a pair of focal plane images is first decomposed into a series of patches (each patch includes one row of the feature image). Then these patches compose a sequence, which serves as the input of the a deep LSTM network. The output sequence of the deep LSTM network is further taken as a vector, which serves as the input of a fully-connected layer, before a set of aberration coefficients can be obtained.

The application procedure of the object-independent wavefront sensing approach using deep LSTM networks is shown in Fig. 5. Specifically, for certain aberration coefficient ranges, a large number sets of aberration coefficients are randomly generated, and we can calculate the corresponding PSFs at the in-focus and defocus image plane. For each set of PSFs, we can extract a feature image, which is then decomposed into an image patch sequence. The generated aberration coefficients and the image patch sequences compose the output data set and input data set, respectively. The deep LSTM network can then be trained using these data sets. After the deep LSTM network is well trained, it can be applied to those image patch sequences which are obtained from a pair of real scenes even the object being imaged is unknown. Meanwhile, certain magnitude of noise is added to the simulated PSFs to simulate the real noisy condition.

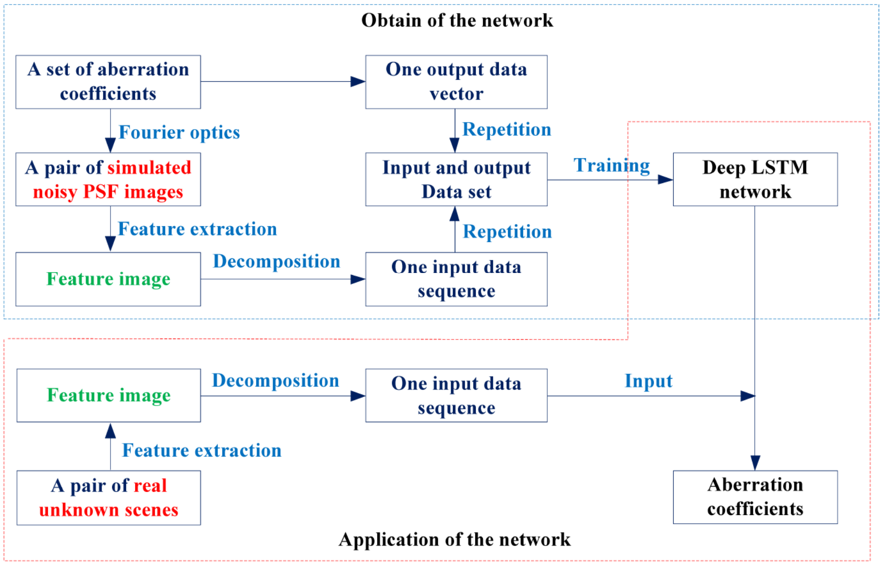

Fig. 5. Application procedure of the object-independent image-based wavefront sensing approach using deep LSTM network. We can see that while the network is trained using simulated PSFs, it can be applicable to unknown extended scenes at the same time.

We can recognize that the deep LSTM network can be trained without any simulated or real extended scenes. In fact, the feature images used to train the deep LSTM network are extracted from the simulated PSF images which are generated according to Fourier optics. On the other hand, after well trained the deep LSTM network can be applied to the feature images which are extracted from the real extended scenes and recover the wavefront phase aberrations. The underlying reason is that the feature image is independent of the object.

4. Simulations and experiment

In this section, simulations and an experiment will be performed to demonstrate the effectiveness of the proposed approach. Meanwhile, the optical system parameters used in the simulation are the same as that in the experiment, and the trained deep LSTM network in the simulation will be directly applied to the images obtained in the experiment.

4.1. Simulations

For a certain set of system parameters (11.5mm aperture size, 150mm focal length, 632.8nm wavelength, 5.5um pixel size, and 1mm defocus distance), we first train a suitable deep LSTM network according to the procedure shown in Fig. 5. We generate a number of (10000) aberration coefficient sets and the corresponding PSF images for training. The aberration coefficients considered here are simply 2nd~9th Fringe Zernike coefficients corresponding to tip-tilt, focus, astigmatism, coma and spherical aberration, which are randomly generated within the range of [-0.5λ, 0.5λ]. The size of the selected PSF image is 64×64 and the actual size of the extracted feature image for training is 32×32. In other words, the number of the input sequence is 32 and each sequence includes 32 elements. The number of layers of the deep LSTM network is 128. A learning algorithm called Adam was used for optimizing the network with an initial learning ratio of 0.0001, a batch size of 10, and a number of epochs of 40. The dropout ratio of the dropout layer was 0.2. Meanwhile, a proper level of Gaussian noise (50dB) has been introduced to the generated PSF images to simulate the practical noisy condition in the training process. Besides, an error in defocus distance ([-0.1mm,0.1mm]) has also been taken into consideration. The codes are implemented in Python and Keras and were executed on a computer with an Intel Core 6700 CPU running at 3.4 GHz, with 16 GB of

RAM, and no GPU is used here. The training time is about 4 hours. Compared to other researches on image-based wavefront sensing with deep learning, the requirement on hardware in our approach is much lower (in [12] the codes are executed on a computer with an Intel Xeon 6134 CPU running at 3.2 GHz, with 192 GB of RAM, and an NVIDIA Tesla V100 GPU with 16 GB of VRAM). Therefore, this deep learning approach is much easier to implement and more convenient for popularization and application.

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. Figure 6(a) shows the error distribution for the training set which is particularly used to train the deep LSTM network. Figure 6(b) shows the error distribution for the test set which is only used to test the performance of the trained network. We can see that the trained deep LSTM network can accurately establish the non-linear mapping between the extracted feature images and the phase aberrations. Specifically, the mean of the absolute errors between the targets and outputs of the network are 1.51e-4 waves for training set and 1.67e-4 waves for the test set, respectively.

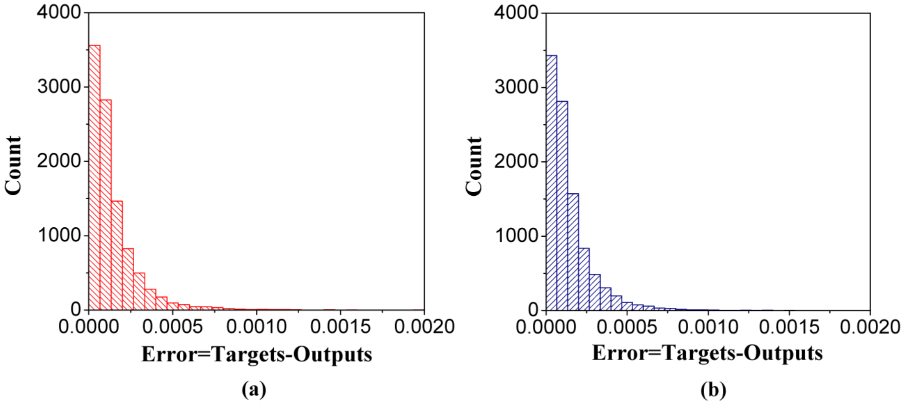

Fig. 6. Distributions of the error between the targets and the actual outputs of the deep LSTM network for the training set (a) and test set (b). We can see that deep LSTM network can accurately fit the non-linear mapping between the extracted feature image and the phase aberrations.

To further demonstrate the accuracy of the proposed approach, the trained deep neural network is applied to several simulated aberrated extended scenes for wavefront sensing and image reconstruction. Specifically, three sets of aberration coefficients are randomly introduced and three sets of aberrated extended scenes including the in-focus and defocus scenes (with a noise level of 50dB and a random error in the defocus distance of [-0.1mm, 0.1mm]) are obtained correspondingly according to Fourier optics [Eq. (1)]. Three feature images are then extracted and three sets of aberration coefficients are recovered after the feature images are decomposed into sequences and taken as input of the trained deep LSTM network. The comparisons between the introduced aberration coefficients and the recovered aberration coefficients are shown in Table 1. The mean error between the introduced aberration coefficients and the aberration coefficients recovered from simulated aberrated extended scenes is 2.3e-3 waves, which indicates a high accuracy of wavefront reconstruction from unknown extended scenes. Note that while the tip-tilt terms are considered in the generation of the PSFs for training the deep LSTM network, they are not included in the outputs, since they have no influence on imaging quality.

Table 1. The comparisons between the introduced aberration coefficients (A) and the aberration coefficients recovered from simulated aberrated extended scenes (B)

		C4	C5	C6	C7	C8	C9
1	A	−0.2152	0.4419	−0.3327	−0.3365	0.4790	−0.2915
	B	−0.2129	0.4387	−0.3318	−0.3413	0.4683	−0.3023
2	A	−0.4872	0.3704	−0.5773	−0.1084	−0.4520	−0.3175
	B	−0.4947	0.3799	−0.5704	−0.1049	−0.4638	−0.3205
3	A	−0.0724	−0.2104	0.5957	−0.5947	0.1701	−0.0171
	B	−0.0705	−0.2088	0.5889	−0.6051	0.1667	−0.0133

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

The recovered aberration coefficients can be used to reconstruct the original extended scenes through deconvolution operation. The results of image reconstruction from the three sets of aberrated images are shown in Fig. 7. By comparing the reconstructed extended scenes with original ones as well as the aberrated ones, we can recognize that the resolution of the reconstructed image is greatly improved, which is comparable to the original extended images. This fact can demonstrate the accuracy of the recovered aberration coefficients indirectly. The extracted feature images corresponding to different phase aberrations are also shown in Fig. 7.

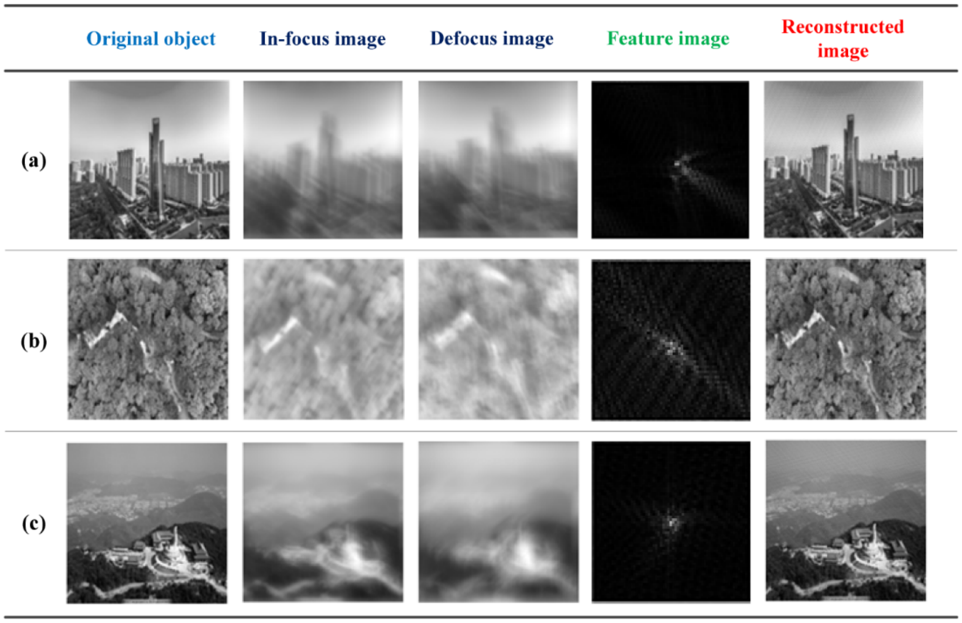

Fig. 7. Simulation results of image reconstruction with the recovered aberration coefficients. We can see that the resolution of the reconstructed images can be effectively improved, which demonstrates the accuracies of the recovered aberration coefficients from the side.

4.2. Experiment

A simple experiment is then performed to validate the effectiveness of the proposed approach. The sketch map of the experimental setup is shown in Fig. 8. In this figure, the laser light passes through an extended hole (a resolution board) after it passes through a scattering matter, and therefore the extended hole can be seen as an extended light source. Here the scattering matter can be any semitransparent solution (such as milk) which is used to change the parallel laser light to scattered light. The position of the detector is located conjugate to the extended hole. The optical system is composed by only one lens. The aberrations of this simple optical system change with field. Different figures in the resolution board can simulate different objects at different field positions. The aperture size of the stop is 11.5mm, the distance between the lens and the detector is 150mm, the wavelength of the laser is 632.8nm, the pixel size of the detector is 5.5um, and the defocus distance is 1mm. The detector is located on an adjustable stage and therefore we can introduce a known defocus diversity. The parameters of the system are the same as that used in the simulation (in fact, the parameters used in the simulations are determined according to this experimental system).

We can select several isolated figures in the image plane which correspond to different field angles of the optical system and therefore they are blurred by different phase aberrations. These extended scenes can be used to recover the wavefront phase aberrations with the approach proposed in this paper. Then the recovered phase aberrations are further used to perform deconvolution and reconstruct the extended scenes collected from the experimental setup. Several sets of in-focus and defocus extended scenes as well as the reconstructed images are shown in Fig. 9. We can see that the resolutions of the reconstructed images are effectively improved, which demonstrates the accuracies of the recovered phase aberrations from the side.

While the traditional phase diversity algorithm has some disadvantages in efficiency and robustness, we can compare the results of this method with the results of deep LSTM network to further evaluate the accuracy of them quantitatively. In the process of performing phase diversity approach, we can use different optimization methods [20,21] and guarantee that a global optimum is obtained. The comparisons between the results of phase diversity approach and the proposed approach are presented in Table 2. We can see that the results of these two different approaches bear strong similarities with each other. Specifically, the mean absolute error between the results of these two approaches is 8.7e-3 waves. This fact will further demonstrate the accuracies of the recovered phase aberrations from the side.

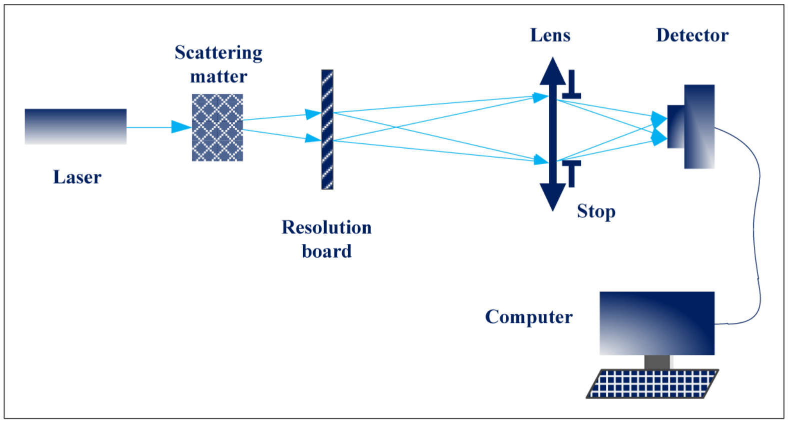

Fig. 8. Sketch map of the experimental setup. Here the scattering matter can be any semitransparent solution which is used to change the parallel laser light into scattered light.

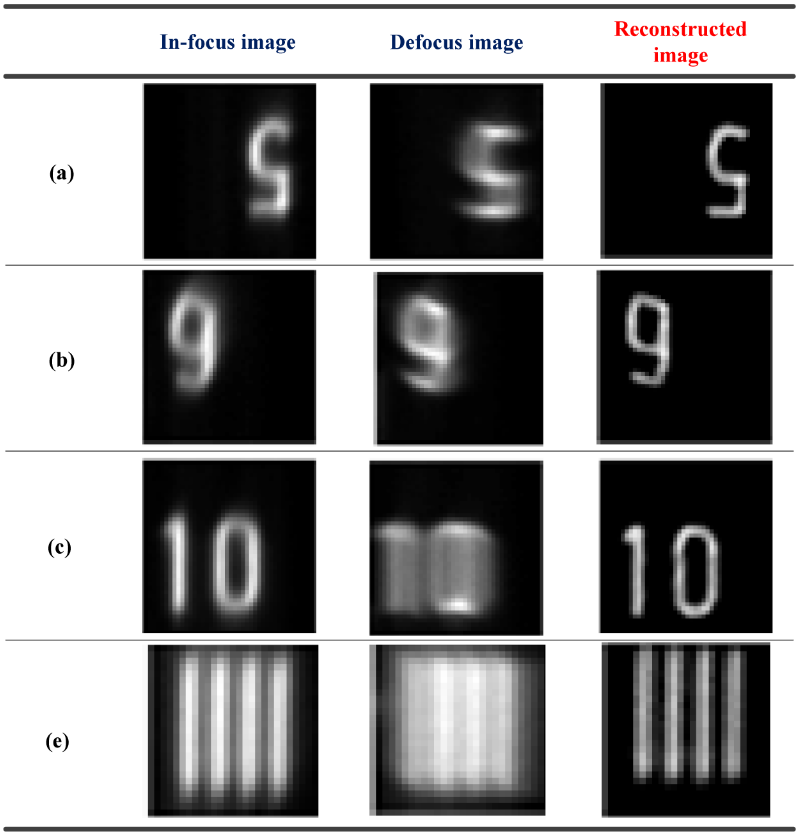

Fig. 9. Experimental results of image reconstruction with the recovered aberrations. We can see that the resolution of the reconstructed images is effectively improved compared to original aberrated images, which demonstrates the accuracy of the recovered aberration coefficients from the side.

Table 2. The comparisons between the aberration coefficients obtained with deep LSTM network (A) and the aberration coefficients recovered using traditional phase diversity algorithm (B) for the 4 pair of images

		C4	C5	C6	C7	C8	C9
1	A	−0.3731	−0.4192	0.0577	−0.1071	0.2688	−0.0382
	B	−0.3808	−0.4212	0.0406	−0.1036	0.2654	−0.0401
2	A	−0.2781	−0.2256	0.2271	−0.1091	−0.1206	−0.0349
	B	−0.2797	−0.2189	0.2217	−0.1071	−0.1170	−0.0332
3	A	−0.2995	−0.5066	0.1244	0.0745	−0.0429	−0.0259
	B	−0.2983	−0.4938	0.1183	0.0801	−0.0690	−0.0281
4	A	−0.3401	−0.1895	0.1066	−0.0939	−0.1874	−0.0245
	B	−0.3381	−0.1929	0.1162	−0.0733	−0.1807	−0.0328

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

While in the experiment we only use simple objects, which seems similar with the case of reference [12]. However, the case presented in our paper has some fundamental differences from the case presented in reference [12]. In this reference, EMNIST database is used to train the network and the trained network is then applied to those samples selected from EMNIST. However, in our manuscript, the network is trained without any simulated or real extended scenes, i.e., the training process does not have any relation with the images obtained with our experimental setup. The trained network in our manuscript not only can apply to these simple objects but can also be applicable to any other objects, for it is independent of the objects being imaged. Besides, since in the simulations we have used complicated extended scenes to demonstrate the effectiveness and accuracy of the proposed approach, here we only need to validate the practicality of the proposed approach in the experiment.

5. Other discussions

In this section, we will further present some discussions on the proposed method. On one hand, we will compare the non-linear fitting capacity of deep LSTM network with Resnet 18 (a variant of convolutional neural network) specifically for the problem encountered in this paper. On the other hand, considering that the in practice the incident light is usually not purely coherent light, we will further discuss the influence of the incoherency of light on the accuracy of the proposed method.

5.1. Comparison between deep LSTM network and Resnet 18 as deep learning tool

In this paper, deep LSTM network is selected as the mathematical tool to establish the non-linear mapping between the extracted feature image and the aberration coefficients. In fact, CNNs can also do this work. The main reason for us to select deep LSTM network as the non-linear fitting tool is that for the problem encountered in this paper the non-linear fitting accuracy of the deep LSTM network is much higher than the CNN model used by us and the structure of the deep LSTM network is also more simple. The fitting accuracies of deep LSTM and Resnet 18 [22,23] (a variant of CNN) for the fitting problem of this paper will be compared and discussed in this section.

When the 2nd~9th Fringe Zernike coefficients (corresponding to tip-tilt, focus, astigmatism, coma and spherical aberration) are considered, which are randomly generated within the range of [-0.5λ, 0.5λ], the fitting accuracies of Resnet 18 for the training set and test set are shown Fig. 10. The numbers of the feature images for training and testing of the Resnet 18 are both 10000. Other conditions for training the Resnet 18 are also the same as those for training the deep LSTM (except that the training time for Resnet 18 is much longer). Comparing Fig. 10 with Fig. 6, we can see that the residual fitting error using deep LSTM is much smaller than that using Resnet 18. Specifically, the mean absolute errors between the targets and outputs of the network are 7.12e-3 waves for training set and 1.14e-2 waves for the test set in Fig. 10, respectively.

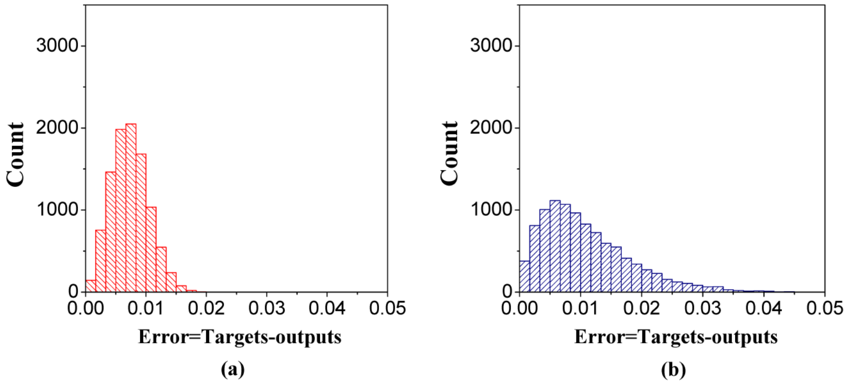

Fig. 10. Distributions of the error between the targets and the actual outputs of the Resnet 18 for the training set (a) and test set (b). Comparing this figure with Fig. 6 we can recognize that the fitting accuracy of Resnet 18 is much less than deep LSTM for the problem encountered in the paper.

When we further increase the number of Fringe Zernike coefficients that are considered (2nd~21th Fringe Zernike coefficients), which are randomly generated within the range of [-0.5λ, 0.5λ], the training results using deep LSTM and those using Resnet 18 are presented in

Fig. 11 and Fig. 12, respectively. We can see that in this case the fitting error of deep LSTM is about two orders of magnitude smaller than the fitting error of Resnet 18 for the problem encountered in the paper. Specifically, the mean absolute errors between the targets and outputs of the deep LSTM are 2.37e-4 for the training set and 2.96e-4 for the test set, respectively. On the other hand, the mean absolute errors between the targets and outputs of the Resnet 18 are 4.93e-2 for the training set and 6.34e-2 for the test set, respectively. Besides, when we further increase the number of Fringe Zernike coefficients that are considered (2nd~37th Fringe Zernike coefficients), the training results for deep LSTM network are shown in Fig. 13. In this case, the mean absolute errors between the targets and outputs of the deep LSTM are 6.71e-4 for the training set and 9.04e-4 for the test set, respectively. We can recognize that while the accuracy actually decreases as the number of aberration coefficients increases, the degradation in accuracy is still not obvious for deep LSTM network.

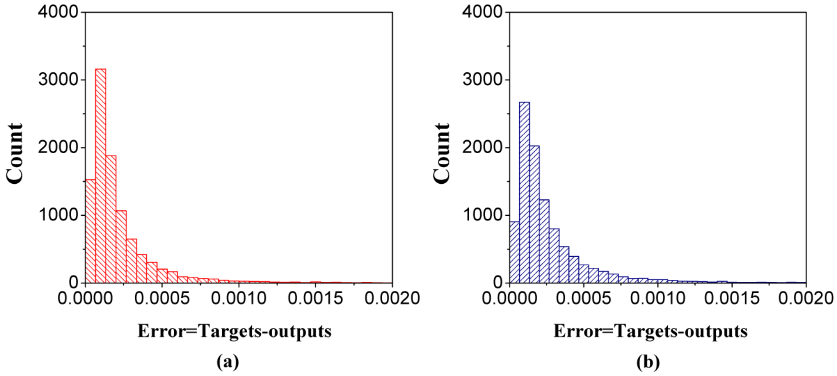

Fig. 11. Distributions of the error between the targets and the actual outputs of the deep LSTM for the training set (a) and test set (b) when 2nd~21th Fringe Zernike coefficients are considered. Comparing this figure with Fig. 6 we can recognize that the fitting accuracy of the deep LSTM is not sensitive to the number of the outputs.

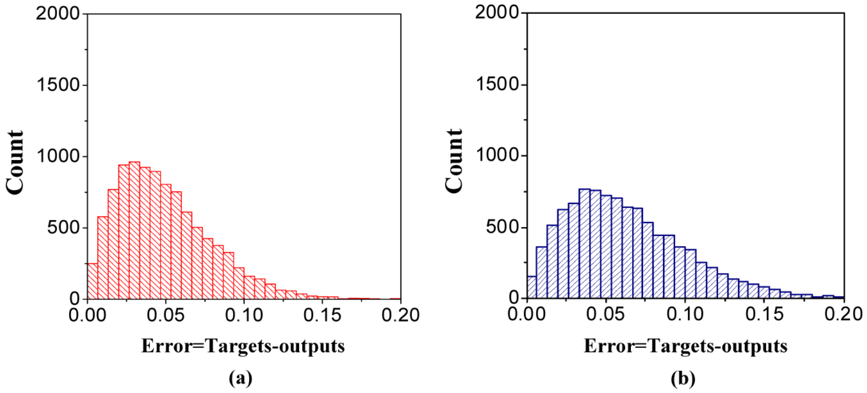

Fig. 12. Distributions of the error between the targets and the actual outputs of the Restnet 18 for the training set (a) and test set (b) when 2nd~21th Fringe Zernike coefficients are considered. Comparing this figure with Fig. 11 we can recognize that the fitting accuracy of Resnet 18 is much less than deep LSTM for the problem encountered in the paper. Besides, the fitting accuracy decreases obviously as the number of the outputs increases.

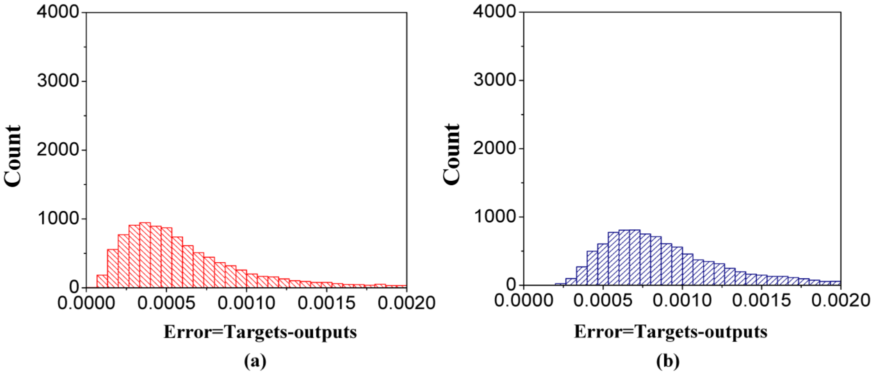

Fig. 13. Distributions of the error between the targets and the actual outputs of the deep LSTM for the training set (a) and test set (b) when 2nd~37th Fringe Zernike coefficients are considered. We can recognize that while the accuracy actually decreases as the number of aberration coefficients increases, the degradation in accuracy is still not obvious for deep LSTM network.

It is possible that if we increase the scale of the data set, the training time, as well as the computational performance of the computer (by using GPU), the non-linear fitting accuracy of the Resnet 18 will increase correspondingly. However, in this case the non-linear fitting accuracy of the deep LSTMs will increase correspondingly. In other words, deep LSTMs have a higher fitting accuracy than Resnet 18 under equal conditions for the problem encountered in the paper. The underlying reason may be that the feature images in this paper include some sparse and scattered points (as shown in Fig. 7) and it may a little hard to further extract features through convolution. CNNs are more suitable for the cases where there are some typical geometrical features (such as edge, texture, line, curve and so on).

Besides, the time it costs to recover aberration coefficients is very short using deep LSTM network. Specifically, recovering 10000 sets of aberration coefficients only costs 3.5 seconds. In other words, recovering one sets of aberration coefficients only costs 0.35 milliseconds. On the other hand, as presented in [12], this value is 9.2 milliseconds for CNNs (in our simulation, this value is tens of milliseconds due to the lower configuration of our computer). The underlying reason is that the convolution process is very time-consuming, especially there are usually hundreds of convolutional layers in those deep CNNs. While it seems that the feature image should be decomposed into a sequence and the elements in this sequence need to be put into the computer one after the other which will cost some time. However, this time is far less than the time it costs for large amount of convolution computation. We also find that when increasing the number of aberration coefficients to be recovered the training time for deep LSTM is nearly unaffected (in the training process, we set the training iterations are same for different number of aberration coefficients). Therefore, deep LSTMs not only have a higher accuracy, but also have a higher efficiency than the CNNs for the problem encountered in this paper. This fact indicates that the proposed deep learning approach is more suitable for wavefront sensing of dynamic phase screens which changes very quickly. Note that in this case a beam splitter is usually needed to obtain a pair of images with a defocus diversity simultaneously.

Therefore, while CNNs have achieved great successes in large-scale image recognition and classification, they are not always the best choices. We should select suitable deep learning tool according to the specific situations and specific issues.

5.2. Influence of the incoherency of incident light on wavefront sensing accuracy

In the sections presented above, we only consider the case of coherent light, i.e., we suppose the spectrum bandwidth of the incident light is infinitely small. However, in practice, the incident light is usually incoherent, even we use optical filter to restrict the bandwidth. In this case, if we still use the deep LSTM network which is trained with the coherent model to recover the wavefront phase, some error will be introduced to the results.

In this subsection, we will quantitatively discuss the incoherency of the incident light on the accuracy of the recovered aberration coefficients under the condition that the deep LSTM network is trained with coherent model. Specifically, we consider 4 cases, as shown in Table 3, which includes two different numbers of aberration coefficients and two different ranges of aberration coefficients. For each case, we first train a deep LSTM network based on the coherent model presented above. Then, for each case, we generate a series of extended scenes with different spectrum bandwidths. Here the bandwidth changes from 5nm to 600nm and the step size is 5nm. For each bandwidth we generate 100 extended scenes to simulate the true incoherent scenes. Meanwhile, for simplicity we suppose that the intensity of light spectrum is uniformly distributed within the bandwidth. The deep LSTM network which is trained with the coherent model is then applied to these extended scenes. Certain image noise and the error in the defocus distance are also considered in this process. The mean absolute error between the aberration coefficients recovered from these 100 extended scenes and those true values is used to evaluate the wavefront sensing error for each bandwidth. These errors for different cases are shown in Fig. 14.

Table 3. Four cases considered in the simulations for demonstrating the influence of the incoherency of incident light when deep LSTM network is trained based on the coherent model

	Number of aberration coefficients	Range of each aberration coefficient
Case 1	$2^{nd}\sim9^{th}$	[-0.5λ, 0.5λ]
Case 2	$2^{nd}\sim9^{th}$	[-1λ, 1λ]
Case 3	$2^{nd}\sim21^{th}$	[-0.5λ, 0.5λ]
Case 4	$2^{nd}\sim21^{th}$	[-1λ, 1λ]

The aberration coefficients are measured in λ (λ = 632.8nm)

The following conclusions can be drawn from Fig. 14:

(1) On one hand, we can easily see that the accuracy of the recovered aberration coefficients decreases as the spectrum bandwidth increases. As the bandwidth increases, the difference between the actual imaging model and the coherent model used to train the network increases, which will lead to decreases in wavefront sensing accuracy.

(2) On the other hand, we can still recognize that the accuracy is nearly unaffected when the spectrum bandwidth is comparatively small (<150nm). In practice, we can put an optical filter before the detector to restrict the bandwidth to guarantee the wavefront sensing accuracy.

(3) Besides, we can also recognize the influence of the incoherency of light on wavefront sensing accuracy is nearly unaffected by the number of aberration coefficients to be recovered. However, it seems that this influence is larger for a larger range of aberration coefficients.

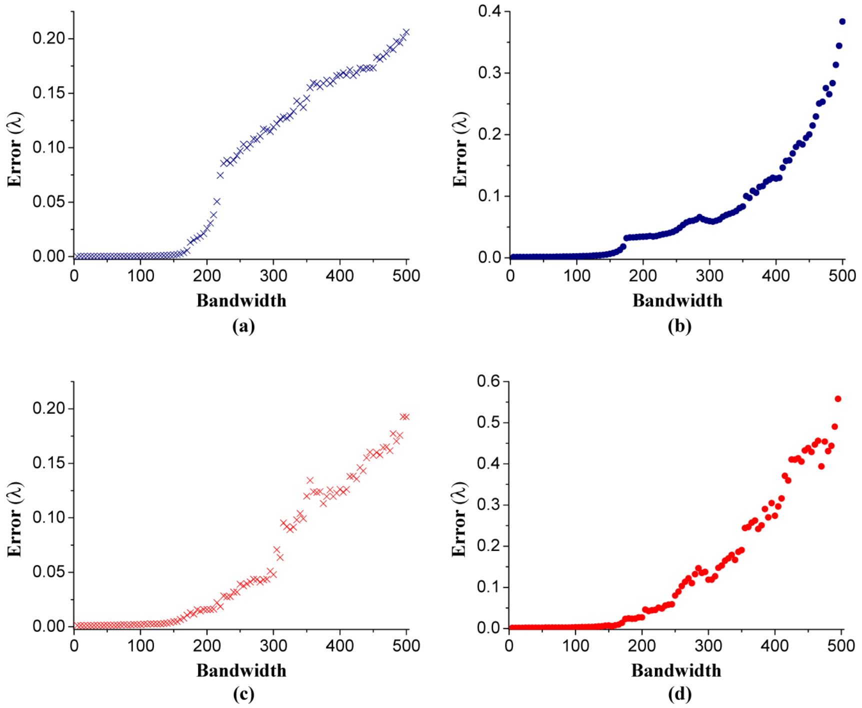

Fig. 14. Influence of the incoherency of light on the accuracy of the deep LSTM network which is trained with the coherent model for 4 cases with different numbers and different ranges of aberration coefficients. (a), (b), (c) and (d) show the mean absolute error of the recovered aberration coefficients which corresponds to Case 1, Case 2, Case 3 and Case 4, respectively. On one hand, we can easily see that the accuracy decreases as the spectrum bandwidth increases. On the other hand, we can still recognize that the accuracy is nearly unaffected when the spectrum bandwidth is comparatively small (<150nm).

6. Conclusion

This paper proposes an image-based wavefront sensing approach using deep learning, where both of the training and application of the neural network are independent of the object. compared to the traditional phase retrieval algorithms, this approach is more robust (free from stagnation problem) and has a far higher efficiency. This work contributes to the wide application of deep learning to image-based wavefront sensing and high-resolution image reconstruction.

One of the core innovations of this paper lies in that we first extract a feature image in the frequency domain which is related to phase aberrations but independent of the original extended objects. Due to this operation, we can train a deep neural network without using any simulated or real extended scenes while this network can be used to recover the aberration coefficients from any extended scenes. In other words, this image-based wavefront sensing approach is totally object-independent. Another key point is that we demonstrate deep LSTM networks can serve as a powerful deep learning tool for image-based wavefront sensing. We also compare the performance of deep LSTM with Resnet 18 for the fitting problem encountered in this paper under the same conditions, and we find that the former has a higher fitting accuracy and a higher computation efficiency. This conclusion indicates that CNNs are not always the best choices for deep learning and we should select suitable deep learning tool according to the specific situations and specific issues.

Since the training process is performed without any simulated or real extended scenes, our approach is much easier to implement and has a far lower requirement on computational performance of the computer. The underlying reason is that the extraction of the object-independent feature image eliminates the interfere of the unknown object on image-based wavefront sensing. On the contrary, the current deep learning methods usually directly recover wavefront phases from the intensities of unknown extended images in the space domain. The training process in these methods usually needs an extremely large number of the extended scenes, a high performance computer (in general expensive GPU is also needed) as well as a very long time for training. Therefore, the deep learning approach proposed in this paper is more convenient for application and generalization.

Note that the proposed approach needs a pair of images with a roughly known phase diversity between them, while it seems that some other existing deep learning methods only need one image [11,12]. However, we should point out that 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, simultaneous collection of multiple intensity images with certain phase diversities are usually needed [24,25]. This is the underlying reason for why we need two images with a phase diversity between them.

Funding

National Key Research and Development Program (2016YFE0205000 and 2016YFF0103603); National Natural Science Foundation of China (NSFC) (61705223, 61805235).

References

1. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**(15), 2758–2769 (1982).
2. R. A. Gonsalves and R. C. Hidlaw, "Wavefront sensing by phase retrieval," Proc. SPIE **207**, 32–39 (1979).
3. 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).
4. 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).
5. N. Védrenne, L. M. Mugnier, V. Michau, M. T. Velluet, and R. Bierent, "Laser beam complex amplitude measurement by phase diversity," Opt. Express **22**(4), 4575–4589 (2014).
6. K. F. Tehrani, Y. Zhang, P. Shen, and P. Kner, "Adaptive optics stochastic optical reconstruction microscopy (AO-STORM) by particle swarm optimization," Biomed. Opt. Express **8**(11), 5087–5097 (2017).
7. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light Sci. Appl. **7**(2), 17141 (2018).
8. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**(9), 1117–1125 (2017).
9. 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).
10. G. Ju, X. Qi, H. Ma, and C. Yan, "Feature-based phase retrieval wavefront sensing approach using machine learning," Opt. Express **26**(24), 31767–31783 (2018).
11. S. W. Paine and J. R. Fienup, "Machine learning for improved image-based wavefront sensing," Opt. Lett. **43**(6), 1235–1238 (2018).
12. Y. Nishizaki, M. Valdivia, R. Horisaki, K. Kitaguchi, M. Saito, J. Tanida, and E. Vera, "Deep learning wavefront sensing," Opt. Express **27**(1), 240–251 (2019).
13. 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).
14. 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), 2200–2216 (1973).
15. R. A. Gonsalves and R. C. Hidlaw, "Wavefront sensing by phase retrieval," Proc. SPIE **207**, 32–39 (1979).
16. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Opt. Eng. **21**(5), 829–832 (1982).
17. 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).
18. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," in *9th International Conference on Artificial Neural Networks: ICANN '99* (1999), pp. 850–855.
19. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," IEEE Trans. Neural Netw. Learn. Syst. **28**(10), 2222–2232 (2017).
20. D. Yue, S. Xu, and H. Nie, "Co-phasing of the segmented mirror and image retrieval based on phase diversity using a modified algorithm," Appl. Opt. **54**(26), 7917–7924 (2015).
21. P. G. Zhang, C. L. Yang, Z. H. Xu, Z. L. Cao, Q. Q. Mu, and L. Xuan, "Hybrid particle swarm global optimization algorithm for phase diversity phase retrieval," Opt. Express **24**(22), 25704–25717 (2016).

22. S. Targ, D. Almeida, and K. Lyman, “Resnet in Resnet: Generalizing Residual Architectures,” arXiv preprint arXiv:1603.08029 (2016).
23. S. Ayyachamy, V. Alex, M. Khened, and G. Krishnamurthi, “Medical image retrieval using Resnet-18 for clinical diagnosis,” Proc. SPIE **10954**, 1095410 (2019).
24. R. L. Kendrick, D. S. Acton, and A. L. Duncan, “Phase-diversity wave-front sensor for imaging systems,” Appl. Opt. **33**(27), 6533–6546 (1994).
25. B. H. Dean and C. W. Bowers, “Diversity selection for phase-diverse phase retrieval,” J. Opt. Soc. Am. A **20**(8), 1490–1504 (2003).