Deep Learning Piston Sensing for Sparse Aperture Systems With Simulated Training Data

Xiafei Ma, Zongliang Xie, Haotong Ma, Yangjie Xu, Dong He, and Ge Ren

***Abstract*—The image-based piston sensing method using the convolutional neural network (CNN) is an advanced technique which has good applicability. However, acquiring a large amount of the training dataset required to train a network is difficult to handle in practice. In this letter, we demonstrate the possibility of using a neural network trained by the simulation dataset to accurately sense pistons directly from experimental images. As a demonstration of the proposed scheme, a single CNN developed by computer-generated images is applied for piston measurement of an experimental setup with three sub-apertures. This is particularly helpful for the sparse aperture system with more sub-apertures. We believe that the study in this letter will contribute to the applications of the CNN-based technique for piston sensing.**

***Index Terms*—Sparse aperture systems, piston sensing, deep learning.**

I. Introduction

COMPARED to the traditional monolithic-mirror telescope, the sparse aperture imaging system has many advantages, such as lighter weight, lower cost, and simpler manufacturing technique. However, some challenges come along. Piston sensing is a specific as well as the most key issue in the field of sparse aperture imaging. Presence of piston errors brings about deterioration of telescope images. Therefore, the advanced piston sensing technique will be of great impetus for the realization of high-resolution imaging.

After years of development, great amounts of effort have been devoted to piston sensing. Modified Shack-Hartmann sensors [1], [2], dispersed fringe sensors [3], [4], and pyramid sensors [5] requiring additional optics have been proposed and applied to many practical systems. Phase diversity is an image-based wavefront sensing technique with much simpler optics but a more complex algorithm, which has been applied for co-phasing of multi-aperture systems [6]–[8]. The piston sensing method using the convolutional neural network (CNN) is a newly developed image-based technique, which has been studied a lot recently.

The idea of using a neural network to detect pistons was first proposed in 1990 [9]. With recent development of deep learning in the recognition of objects, the co-phasing technique using CNN has received more attention. Guerra-Ramos et al. constructed two networks to predict the piston step values and the ambiguity range, respectively. In their simulation, an ample capture range of $\pm 11\lambda$ and an accuracy of about $\pm 0.0087\lambda$ were achieved with point sources [10]. Later, multiple CNNs were used to sense pistons with extended objects [11], [12]. Another research demonstrated that accurate piston sensing from broadband point spread functions can be achieved with only one CNN, and results of a two-aperture concept experiment were presented [13]. Then it was also demonstrated by simulation that pistons could be extracted by a single CNN from broadband images of extended objects [14]. Once trained, the CNNs would be a great tool for tackling the issues of piston sensing. However, current researches mainly remain at the simulation stage since acquiring a large amount of training dataset seems challenging in practice. Actually, it is difficult to obtain training samples in many research areas, and this drawback has been discussed as well [15]–[17].

The key aspect of the deep learning method is to learn features with multiple levels of representation. In this feature learning process, a large amount of training data is usually required to determine the most appropriate features for accurate prediction. In particular, hundreds of thousands training images may be needed for accurate piston sensing for systems with numerous sub-apertures. However, it is not easy to collect so many correct images in practice. First, it is needed to pre-calibrate the system so that using the prior loaded pistons as the outputs for training are available. Second, in this image collecting process, unstable factors can make the training data unusable. For example, unexpected errors caused by jitter and delay in the imaging platform during piston is being loaded can result in an inconsistency between the intensity distribution and label value. Consequently, the method fails to approximate the non-linear model precisely and a trained network with poor performance may be obtained. Thus, we need to simplify the data acquisition process in order to improve the effectiveness and the practicability of this method using CNN.

In this work, we demonstrate that a network trained by using simulation data can be applied to piston sensing of experimental systems. The measurement principle is described in Section II. In Section III a simulation model of sparse aperture imaging

Manuscript received 9 June 2022; revised 13 July 2022; accepted 25 July 2022. Date of publication 28 July 2022; date of current version 8 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grants 62005289 and 62175243, in part by the State Key Laboratory of Pulsed Power Laser Technology under Grant SKL2018KF05, in part by the Excellent Youth Foundation of Sichuan Scientific Committee under Grant 2019JDJQ0012, and in part by Youth Innovation Promotion Association, CAS, under Grants 2018411 and 2020372. *(Corresponding authors: Zongliang Xie; Haotong Ma.)*

The authors are with the Key Laboratory of Optical Engineering, Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China (e-mail: maxiafei3860@163.com; zongliang.xie@yahoo.com; mahaotong@163.com; xyj1289467369@163.com; hedong@ioe.ac.cn; renge@ioe.ac.cn).

Digital Object Identifier 10.1109/JPHOT.2022.3194509

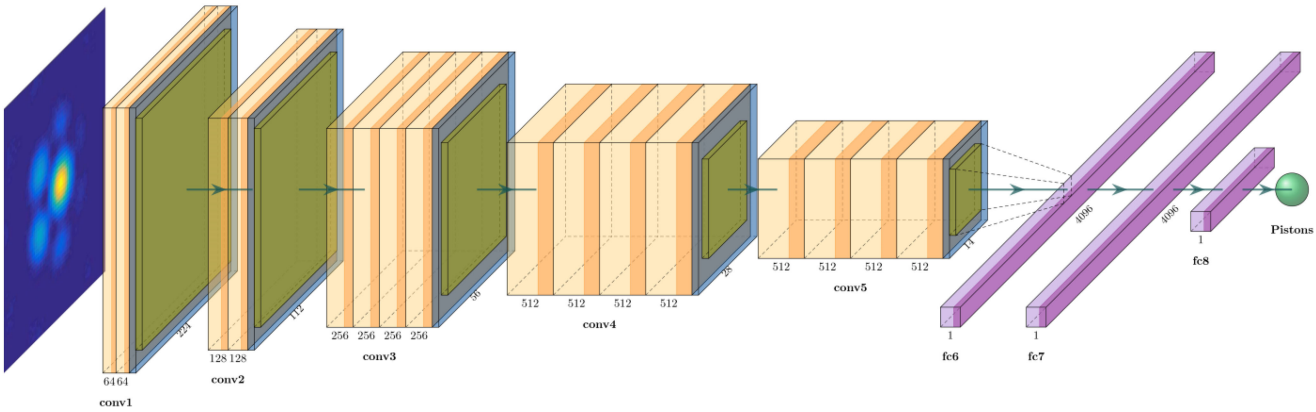

Fig. 1. Schematic CNN architecture. The CNN is composed of five stacked convolution units and three fully-connected layers. Each stacked convolution unit is followed by a BN layer and a max-pooling layer.

system is built to generate training set. On this basis, a CNN is constructed and trained by the computer-generated images, so as to fit the mapping relation between the point spread functions and the corresponding pistons. Section IV reports the piston sensing performance of the method for real images. Fine co-phasing of three-aperture imaging platform is achieved. Finally, concluding thoughts and relevant discussions are offered in Section V.

II. Method Description

According to Fourier optics imaging principle, the focal point spread function (PSF) with monochromatic illumination can be expressed as:

$$h(\mathbf{x}) = |\mathrm{FT}(P(\mathbf{u}))|^2. \tag{1}$$

where $P(\mathbf{u})$ is the generalized pupil function, FT$(\cdot)$ and $\mathbf{u}$ represent Fourier transform and a 2D vector in the pupil plane, respectively.

Under the premise that tip-tilt as well as higher-order aberrations has been corrected, the generalized pupil function of a sparse aperture system can be written as:

$$P(\mathbf{u}) = p(\mathbf{u}-\mathbf{u}_1) + \sum_{n=2}^{N} p(\mathbf{u}-\mathbf{u}_n)\exp\left(\frac{2\pi i}{\lambda} OPD_n\right). \tag{2}$$

where the first term is regarded as the reference sub-aperture, N and $\mathbf{u}_n$ represent the number of all the sub-apertures and the center vector of the nth sub-aperture, respectively. OPD_n is the relative piston between the nth sub-aperture and the reference.

Here we aim to detect the phases of pistons from monochromatic PSFs, which are limited within [0, 2π). When the system has pistons over ($-\lambda/2$, $\lambda/2$), multiple wavelengths [18] combined with multiple CNNs are needed to overcome the 2π ambiguity and extend the capture range.

The CNN applied to the piston sensing, as illustrated in Fig. 1, is a slightly modified VGG-19 architecture with an additional batch normalization (BN) right before each pooling layer, which can yield a speedup in training [19], [20]. Convolution kernel is the essential part of the CNN, which plays a role of feature extractor. We adopt max-pooling and Adam with a mini-batch size of 64 [21]. The ReLU function is applied as the activation function. Differing from that for the general image classification task, the network ends with a regression layer. The schematic illustration of CNN-based piston sensing is presented as Fig. 2. In the training process, the loss function, which is defined as the root mean square error (RMSE) between the outputs and the expected values, can be calculated by

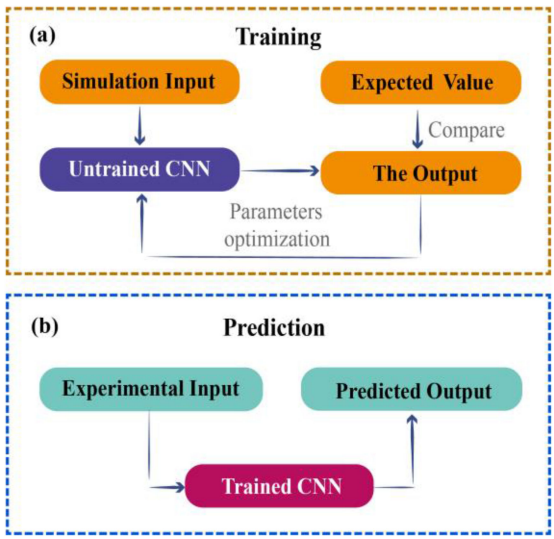

Fig. 2. (a) The training process. (b) The prediction process.

$$L = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(p_k - \hat{p}_k)^2}. \tag{3}$$

where N is number of sub-apertures, and p_k, $\hat{p}_k$ are the ground-truth piston and the estimated output of an input image labeled k, respectively. Then the network parameters are optimized by the gradient descent algorithm after each forward propagation. Iterations are implemented until the loss function is minimized so that the outputs are as close as possible to the expected values. In the prediction process, the trained network is capable of

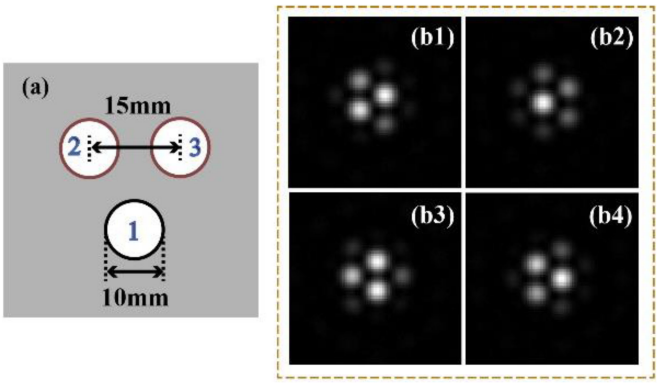

Fig. 3. (a) Configuration of the imaging system. (b1)–(b4) Examples of simulated training images corresponding to different pistons.

sensing the pistons from experimental images directly, which needs only one iteration.

III. Simulation Results

We validate the feasibility of the proposed scheme based on a concept-proof experimental system with three units distributed as shown in Fig. 3(a), in which three sub-apertures with the diameter of 10 mm are arranged according to a regular triangle, and the center distance between two arbitrary units is 15 mm. An unresolved monochrome point of the 632 nm wavelength is imaged by the system with the 400 mm focal length and the 1.67 μm pixel pitch. Then we need to build a simulation model to closely match the experiment setup. Inevitably, Experimental parameters slightly deviate from the ground truth. By comparing the correlation coefficients between the experimental images and simulated images, we choose the simulation parameters as 420 mm focal length and 9.5 mm diameter. In the simulation, sub-aperture 1 is regarded as the reference, and 5000 random piston phase errors ranging from 0 to 2π rad are introduced in sub-aperture 2 and sub-aperture 3. Besides, zero mean and 0.01 variance Gaussian noise is introduced to simulate the practical noisy environment. Then corresponding simulation images are generated, of which several examples are presented as Fig. 3(b1)–(b4).

With the training data input, the forward propagation of the data and the feedback of deviation are implemented iteratively. As shown in Fig. 4, after 12000 iterations, the training program is done. Then, other 100 computer-generated images are used to test the performance of the trained network. The average RMSE between the outputs of the trained network and the expected values over simulated testing images is 0.016λ. Fig. 5 shows the distributions of RMSEs over the training set and testing set, from which we can see that all the RMSEs of the pistons are within 0.09λ, and more than 90 percent are within 0.03λ. The simulation results indicate that fine co-phasing with high sensing accuracy is achieved. Next, this network trained by simulated data will be applied to real images to evaluate the feasibility of the approach.

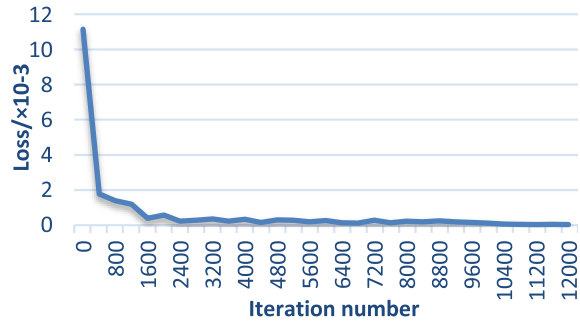

Fig. 4. Evolution of the loss function during training.

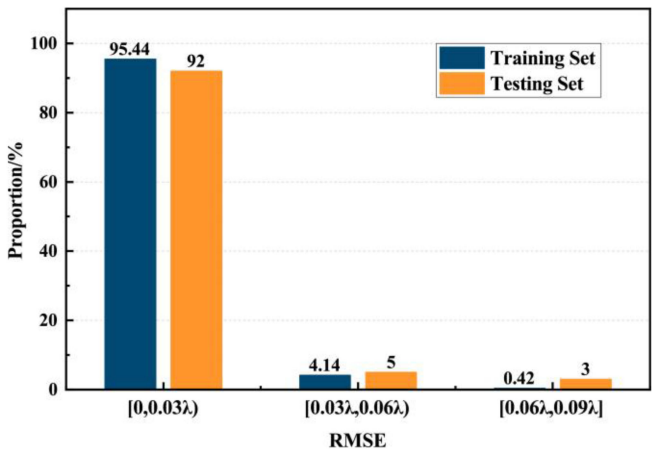

Fig. 5. Histogram of the RMSEs between the predicted values and the ground truth values over simulated training set and testing set.

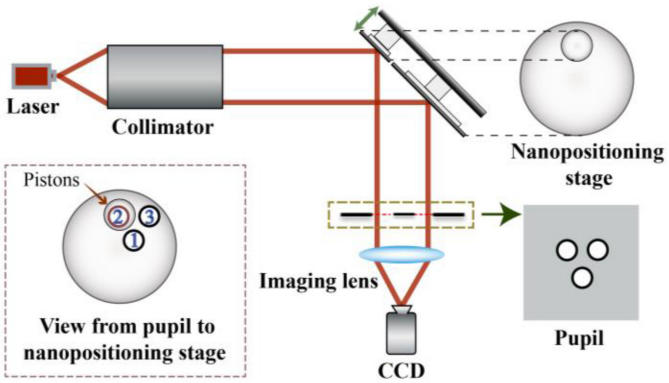

Fig. 6. Schematic diagram of the experimental setup. The nano-positioning stage can bring pistons in the light path of sub-aperture 2 and corresponding experimental images are captured by CCD.

IV. Experimental Results

The concept-proof experimental setup is shown schematically in Fig. 6. First, a beam of laser of wavelength 632 nm is expanded to parallel light by a collimator. Reflected by a nano-positioning stage which is placed at a 45-degree angle, the collimated light through the three-pupil mask is focused by an imaging lens and captured by a CCD of 1.67 um pixel pitch. The nano-positioning

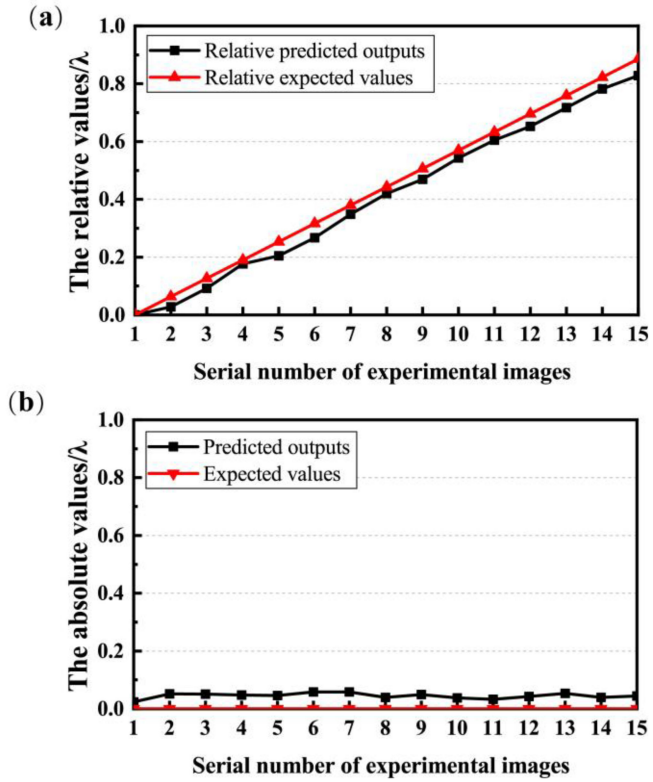

Fig. 7. (a) The relative values of results of sub-aperture 2 on experimental images. (b) The absolute values of results of sub-aperture 3 on experimental images.

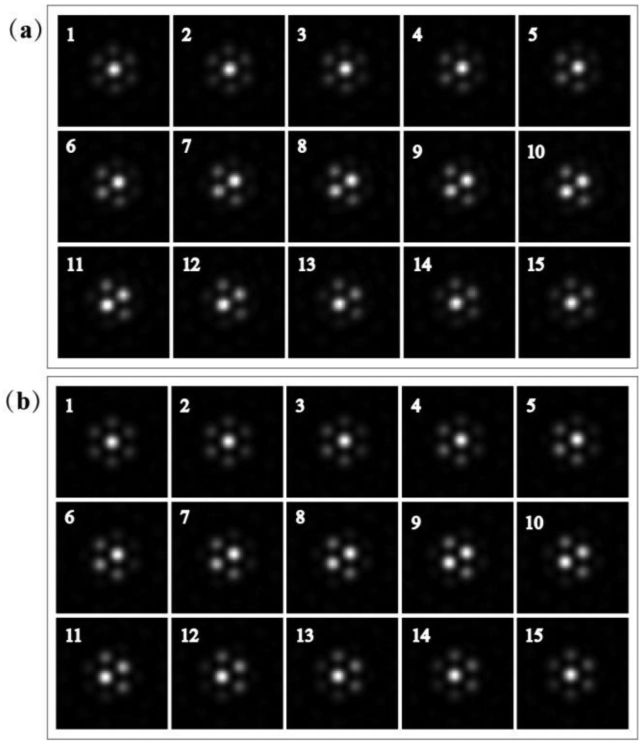

Fig. 8. Comparison of (a) experimental images and (b) computer-generated images with 420 mm focal length and 9.5 mm diameter.

stage is composed of a large fixed reference mirror with 150 mm diameter and a sub-mirror with 30 mm diameter.

In order to avoid the inconsistence of labels and better test the actual piston sensing performance of the simulation-trained CNN, the experiment setup is first calibrated by using the chromatic phase diversity [22]. In image collection process, the light passed through sub-aperture 3 as well as the reference sub-aperture is reflected by the same fixed mirror. Meanwhile, the sub-mirror embedded in the fixed mirror is driven to move along the normal axis, thus introducing pistons in the light path of sub-aperture 2. In the range of 0 to 560 nm pistons, 15 images are captured at interval of 40 nm. These 15 experimental images as shown in Fig. 8(a), the intensities of which are normalized, are input to the trained network, and then pistons on sub-aperture 2 and sub-aperture 3 are output. In order to release effects of the calibration errors, we evaluated the accuracy by calculating the relative values of the predicted results of sub-aperture 2, which are plotted in Fig. 7(a). As there actually no optical path difference between sub-aperture 3 and the reference sub-aperture, we used the predicted results of sub-aperture 3 as evaluation metrics, which are plotted in Fig. 7(b). It can be seen that the predicted outputs of the network trained by using the simulation data on experimental images are in good consistence with the expected values. High sensing accuracy with an average piston error of 0.033λ for sub-aperture 2 and 0.045λ for sub-aperture 3 are obtained.

Generally, the simulation should be set as close to the experiment as possible to assure the accuracy. Whereas, the deviation between the simulation and the experiment is inevitable. As mentioned above, we computed the correlation coefficients between experimental images and computer-generated images with corresponding piston values added, as presented in Fig. 8(b). There shows a more relevant results when we set the simulation parameter values as 420 mm focal length and 9.5 mm diameter. The correlation coefficients can be calculated by using

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)\left(B_{mn} - \overline{B}\right)}{\sqrt{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2 \sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2}}, \quad (4)$$

where A and B represent the experimental images and computer-generated images, respectively. A and B has the same size of $m \times n$. $\overline{A}$ and $\overline{B}$ represent the mean of all pixel values in image A and image B, respectively. We can see that there is still a minor discrepancy between real images and simulated images, especially in images 11–15. Our analysis considers that the main factor may be the cumulative errors produced in the procedure of driving sub-mirror.

To further evaluate the tolerance of the proposed scheme, we generated another two simulated training sets by changing the focal length and the sub-aperture diameter of the simulation model, and then used the retrained networks to implement piston sensing with experimental images. The detecting results are shown in Table I, where there seems no significant deterioration

TABLE I
EXPERIMENTAL PERFORMANCE OF PISTON SENSING USING THE NETWORKS TRAINED BY SIMULATIONS WITH DIFFERENT SYSTEMS

System parameters	f/mm d/mm	420 9.5	400 10	440 10
Average errors/λ	Sub-apertures 2	0.033	0.041	0.048
	Sub-apertures 3	0.045	0.049	0.053

of detection accuracy. The actual sensing accuracy that fine phasing requires is also available.

V. CONCLUSION AND DISCUSSION

In conclusion, it has been demonstrated experimentally that computer-generated training data can be utilized to obtain a well-trained network for piston sensing in real imaging system. The testing results show that the proposed approach has the ability of high-accuracy sensing and generalization. it is expected that elimination of the demand for real training images will further promote the practical application of piston sensing technique.

If unknown aberrations result in obvious differences between the intensity distributions generated by the simulation and experiment, the network trained by the simulation data might be unusable. Two approaches might be available for the problem above: the first is to apply the adaptive optics to the practical system so that the effects of aberrations could be eliminated; the other is to take the intrinsic wavefront of the system into account in the simulation, thus producing images close to the practical ones.

Atmospheric turbulence is problematic, which however can be decomposed into dominant segmented piston and tip-tilt errors by the sparse pupil distribution. If the decomposed errors are detected and corrected rapidly enough, the atmospheric turbulence can be effectively relieved. There are mainly two approaches to overcoming turbulence. On the one hand, we need to improve the sensing speed. At present, it takes 10ms to extract pistons from a single image using the graphics processing unit (GPU) of NVDIA GeForce GTX 1080Ti. The speed can be further improved by optimizing the network structure and upgrading computational resource. On the other hand, it is necessary to study on how to use CNN to sense both the piston and tip-tilt errors.

When using such a single CNN to directly fit the mapping relations between monochromatic PSFs and pistons, we also find that the pupil configurations may affect the effectiveness of the network. Vast numerical simulations are performed using the same CNN structure and the results show that while the network is able to converge for the noncentrosymmetrical configurations with 5 or more sub-apertures, it fails for centrosymmetrical configurations. Further researches are needed to find the theoretical reasons, but at the same time, this issue can be solved with a diaphragm used to break the symmetry [14].

We again note that this image-based piston sensing technique using monochromatic light suffers from 2π ambiguity, which can be solved by the use of multiple wavelengths. Besides, more researches about the proposed approach on extended objects and solutions to overcome the effects of centrosymmetry will be developed.

REFERENCES

[1] G. Chanan et al., "Phasing the mirror segments of the Keck telescopes: The broadband phasing algorithm," *Appl. Opt.*, vol. 37, no. 1, pp. 140–155, 1998.

[2] G. Chanan, C. Ohara, and M. Troy, "Phasing the mirror segments of the Keck telescopes II: The narrow-band phasing algorithm," *Appl. Opt.*, vol. 39, no. 25, pp. 4706–4714, 2000.

[3] M. A. van Dam, B. A. McLeod, and A. H. Bouchez, "Dispersed fringe sensor for the Giant Magellan Telescope," *Appl. Opt.*, vol. 55, no. 3, pp. 539–547, 2016.

[4] F. Shi et al., "Segmented mirror coarse phasing with a dispersed fringe sensor: Experiment on NGST's wavefront control testbed," in *Proc. SPIE*, 2003, vol. 4850, pp. 318–328.

[5] S. Esposito, E. Pinna, A. Puglisi, A. Tozzi, and P. Stefanini, "Pyramid sensor for segmented mirror alignment," *Opt. Lett.*, vol. 30, no. 19, pp. 2572–2574, 2005.

[6] R. G. Paxman and J. R. Fienup, "Optical misalignment sensing and image reconstruction using phase diversity," *J. Opt. Soc. Amer. A*, vol. 5, no. 6, pp. 914–923, 1988.

[7] R. L. Kendrick et al., "Wide-field Fizeau imaging telescope: Experimental results," *Appl. Opt.*, vol. 45, no. 18, pp. 4235–4240, 2006.

[8] R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1072–1085, 1992.

[9] J. R. P. Angel, P. Wizinowich, M. Lloyd-Hart, and D. Sandler, "Adaptive optics for array telescopes using neural-network techniques," *Nature*, vol. 348, no. 6298, pp. 221–224, 1990.

[10] G. R. Dailos, D. G. Lara, T. S. Juan, and R. R. M. Jose, "Piston alignment of segmented optical mirrors via convolutional neural networks," *Opt. Lett.*, vol. 43, no. 17, pp. 4264–4267, 2018.

[11] D. Li, H. Xu, D. Wang, and D. Yan, "Large-scale piston error detection technology for segmented optical mirrors via convolutional neural networks," *Opt. Lett.*, vol. 44, no. 5, pp. 1170–1173, 2019.

[12] M. Hui, W. Li, M. Liu, L. Dong, L. Kong, and Y. Zhao, "Object-independent piston diagnosing approach for segmented optical mirrors via deep convolutional neural network," *Appl Opt*, vol. 59, no. 3, pp. 771–778, 2020.

[13] X. Ma, Z. Xie, H. Ma, Y. Xu, G. Ren, and Y. Liu, "Piston sensing of sparse aperture systems with a single broadband image via deep learning," *Opt. Exp.*, vol. 27, no. 11, pp. 16058–16070, 2019.

[14] X. Ma, Z. Xie, H. Ma, Y. Xu, D. He, and G. Ren, "Piston sensing for sparse aperture systems with broadband extended objects via a single convolutional neural network," *Opt. Lasers Eng.*, vol. 128, 2020, Art. no. 106005.

[15] Y. Gao, X. Liu, and J. Xiang, "Fault detection in gears using fault samples enlarged by a combination of numerical simulation and a generative adversarial network," *IEEE/ASME Trans. Mechatronics*, early access, Dec. 29, 2021, doi: 10.1109/TMECH.2021.3132459.

[16] Y. Gao, X. Liu, and J. Xiang, "FEM simulation-based generative adversarial networks to detect bearing faults," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4961–4971, Jul. 2020.

[17] X. Liu, H. Huang, and J. Xiang, "A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine," *Knowl-Based Syst.*, vol. 195, no. 1, 2020, Art. no. 105653.

[18] M. G. Löfdahl and H. Eriksson, "An algorithm for resolving 2π ambiguities in interferometric measurements by use of multiple wavelengths," *Opt. Eng.*, vol. 40, no. 6, pp. 984–990, 2001.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017, *arXiv:1412.6980*.

[22] D. Mourard et al., "Group and phase delay sensing for cophasing large optical arrays," *Monthly Notices Roy. Astronomical Soc.*, vol. 445, no. 2, pp. 2082–2092, 2014.