



Deep residual learning for low-order wavefront sensing in high-contrast imaging systems

GREGORY ALLAN,^{1,6,*}  IKSUNG KANG,^{2,6}  EWAN S. DOUGLAS,³  GEORGE BARBASTATHIS,^{4,5} AND KERRI CAHOY¹

¹*Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

²*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

³*Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ 85721, USA*

⁴*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

⁵*Singapore-MIT Alliance for Research and Technology (SMART) Centre, 1 Create Way, Singapore 138602, Singapore*

⁶*These authors contributed equally to this work and should be considered co-first authors*

*gregallan@mit.edu

Abstract: Sensing and correction of low-order wavefront aberrations is critical for high-contrast astronomical imaging. State of the art coronagraph systems typically use image-based sensing methods that exploit the rejected on-axis light, such as Lyot-based low order wavefront sensors (LLOWFS); these methods rely on linear least-squares fitting to recover Zernike basis coefficients from intensity data. However, the dynamic range of linear recovery is limited. We propose the use of deep neural networks with residual learning techniques for non-linear wavefront sensing. The deep residual learning approach extends the usable range of the LLOWFS sensor by more than an order of magnitude compared to the conventional methods, and can improve closed-loop control of systems with large initial wavefront error. We demonstrate that the deep learning approach performs well even in low-photon regimes common to coronagraphic imaging of exoplanets.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

The search for planets orbiting other stars uses a variety of techniques for detection and observation. One such technique is the direct imaging of exoplanets, in which a coronagraph is used to attenuate the light from a host star so as not to overwhelm its much dimmer companion planets. The coronagraph can employ masks altering amplitude, phase, or both to modify the optical signal. However, if there are wavefront aberrations due to imperfections in the telescope optics, telescope pointing errors, or atmospheric effects, more starlight will leak through to the image plane. In the case of Earth-like exoplanets around Sun-like stars, the contrast between star and planet can exceed a factor of 10^{10} [1,2]. Under such conditions, even small aberrations will overwhelm the signal from the planet. This sensing regime is known as High-Contrast Imaging (HCI).

The errors at the input of the coronagraph are in both phase and intensity. Optical coating uniformity makes amplitude errors relatively small and well-suited to image plane sensing and correction [1]. We will focus on large amplitude phase errors in this work. The 2D field of phase errors is commonly decomposed into a linear combination of orthogonal circular basis functions called *Zernike polynomials* [3]. To enable HCI in the presence of errors, an adaptive optics (AO) system is used to detect and correct these aberrations. The AO system commonly consists of one or more wavefront sensors (WFS) to measure the aberrations, and one or more deformable mirrors (DM), which adds complementary amplitude or phase delays to the optical field to cancel the errors. In a modal control approach, the wavefront sensor data are processed

to find the coefficients of a linear combination of Zernike polynomials corresponding to the wavefront, which are to be used to induce the proper response in the actuator [4].

Many WFS technologies require that some of the light is split from the main optical system and directed to the sensor, but this has several drawbacks. First, removing some light degrades the system's signal-to-noise ratio (SNR). Second, the light traveling to the WFS takes a different path than the light entering the coronagraph; as a result, residual non-common-path errors are present, and the aberrations are not fully corrected. One sensing modality that does not suffer from these effects is the Lyot-based low order wavefront sensor (LLOWFS) [5]. This sensor only uses the starlight already being rejected by the coronagraph mask for sensing, so it does not degrade the SNR of the planet imaging system. It is called low-order because its purpose is to relatively quickly measure a finite number of low spatial order Zernike polynomials.

A schematic of a phase mask coronagraph with a LLOWFS is shown in Fig. 1. The Lyot stop is an opaque, concentric mask that rejects the starlight that is diffracted away from the center of the optical field by the coronagraph mask. In a system with a LLOWFS, the Lyot stop is mirrored so that the rejected light is reflected and then focused onto a detector. A small defocus is used to break sign ambiguities in the aberration terms.

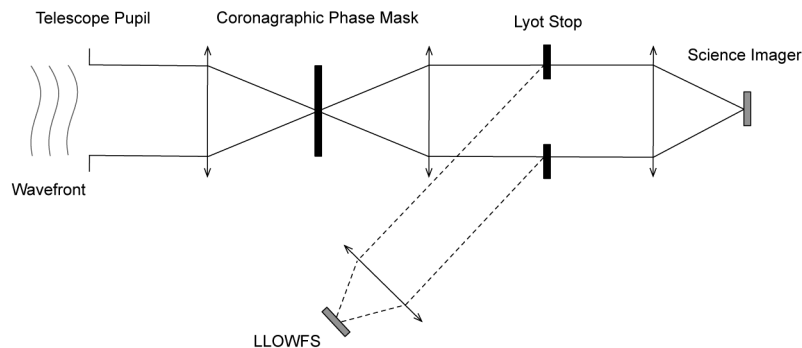


Fig. 1. Schematic of a focal-plane mask coronagraph with a LLOWFS. An aberrated wavefront is incident on the left, is focused to a coronagraph mask which diffracts on-axis light out of the field of view and into a ring which is blocked by the Lyot Stop. The wavefront control mechanism, e.g. deformable mirror(s) — not shown here for clarity — would be located in the optical train ahead of the coronagraph instrument [6].

The sensor response is trained by input of a known phase error consisting of a single Zernike polynomial, and the pattern is recorded. This is repeated for many Zernike polynomials, creating a basis of sensor response modes corresponding to each.

In LLOWFS operation, a signal with an unknown phase error enters the system and creates a pattern on the detector. Least-squares regression using singular-value decomposition is used to find the best linear combination of the previously-observed modes. Assuming linearity, the coefficients of these modes correspond to the coefficients of the Zernike polynomials in the input signal.

A phase mask coronagraph using a four-quadrant phase mask (FQPM) is the representative example examined in this work. The FQPM is a focal plane phase mask where two non-adjacent quadrants have a phase delay of π radians relative to the other two, creating a region of destructive interference where the four quadrants meet [7]. We choose the FQPM because of its high throughput, small inner working angle, and ease of modelling. The FQPM has seen use on several ground-based coronagraphic instruments [8,9].

The method was also tested with a simulated scalar vortex mask [10], but performance was poor at high aberration amplitudes due to degeneracy between the spherical and defocus modes. This phenomenon will be addressed in future work.

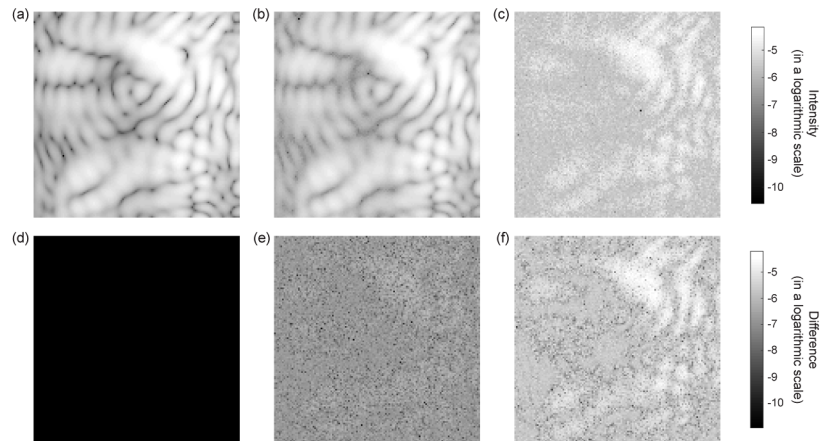


Fig. 2. (a-c) Examples of intensity patterns on LLOWFS with different levels of noise, defined in Table 1. As the noise level increases, details in raw intensity are more obfuscated. (d-f) Difference between a noisy intensity and its corresponding noiseless one. (a,d) Noiseless, (b,e) low noise, (c,f) high noise. (e) Under the low noise condition, the difference mainly represents Gaussian noise, but (f) under the high noise condition, the affect of Poisson statistics becomes dominant, and the noise level is dependent on the level of the raw intensity in (c).

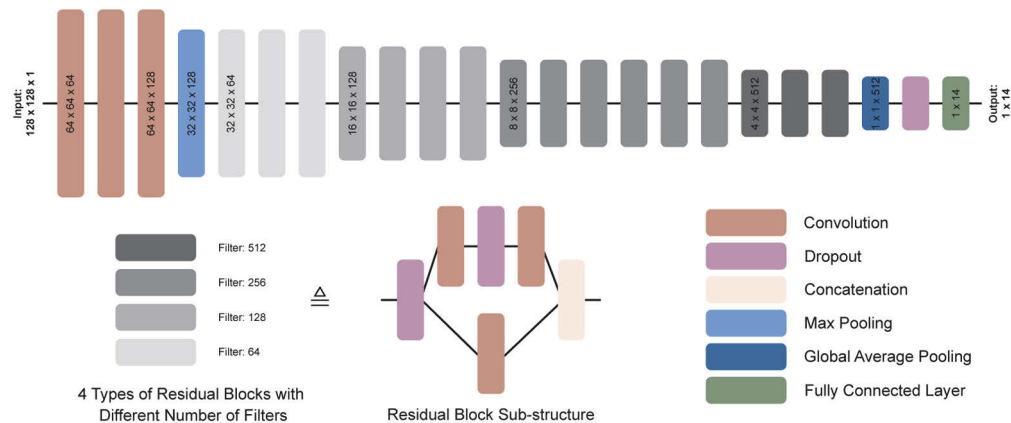


Fig. 3. Proposed DNN architecture. Coronagraph intensity patterns are fed into the network at the furthest left, and predicted Zernike expansion coefficients are given from the furthest right. Different from the adaptation of this network to classification tasks, the last fully connected layer uses a linear activation function, and thus has continuous real values as an output. The proposed architecture consists of 21 M and 15 k trainable and non-trainable parameters, respectively.

The difficulty in any wavefront measurement is that a fixed detector can only measure the intensity of the complex optical field, so phase information is lost. The operation of the coronagraph on the optical field and the conversion of complex field to intensity are nonlinear operations. This means that the response of the LLOWFS can be considered linear only within a small range of wavefront displacement. Demonstrations of LLOWFSs with vortex mask coronagraphs have shown that the response of the LLOWFS to single Zernike polynomial terms is linear if their amplitudes are on the order of ± 100 nm RMS in H-band (centered on 1.65 μm), and becomes non-monotonic for displacements of ± 150 nm or larger. Linearity degrades further in the presence of multiple simultaneous Zernike terms [6].

The limited dynamic range of the LLOWFS means that its use is limited to cases where large low-order aberrations have already been compensated. In ground-based systems a primary wavefront sensing and control (WFSC) system with its own coarse WFS usually accounts for these errors, but a separate optical path means that non-common-path aberration is present and must be accounted for using additional feedback from the instrument and continuous monitoring of control system optical gain [11]. A separate coarse WFS also requires a beamsplitter to pick off light, reducing optical throughput. We propose that by using a nonlinear regression method, the usable range of the LLOWFS may be extended, reducing reliance on additional high-dynamic-range wavefront sensors, and sidestepping accuracy and throughput challenges.

The single-WFS approach is even more applicable to space-based and balloon-based systems, where the WFSC is already more tightly coupled with the coronagraph instrument. Current coronagraph designs, such as the PICTURE-C coronagraph [12,13], employ two-stage wavefront sensors, a high-dynamic range WFS, e.g., a Shack Hartmann Sensor, and a high-precision WFS, e.g. the LLOWFS or a Zernike wavefront sensor [14]. The elimination of the separate coarse WFS would enable the reallocation of limited SWaP (Size, Weight, and Power) resources to other subsystems, such as scientific instruments.

The non-linear regression of Zernike coefficients from the LLOWFS focal-plane intensity is a form of image-based wavefront sensing. Existing general methods for image-based wavefront sensing include iterative and parametric techniques, which suffer from excessive computation time and stagnation at local minima [15]. In high-contrast imaging specifically, the use of image-plane information from the science camera itself is sometimes used to correct for aberrations, eliminating susceptibility to NCPA. [16] is a parametric phase-diversity-based method that recovers Zernike coefficients from two coronagraphic images with a known aberration difference, requiring actuation of an upstream device. It requires long computation times to iteratively propagate a nonlinear optical model. [17,18] use probes of upstream actuators to estimate the EM field at the image plane and directly control for speckle. The latter methods do not recover Zernikes and are only suitable for use with slowly-varying speckles caused by residual WFE in space telescopes.

Neural Networks, and in particular Deep Neural Networks (DNNs) are the subject of much recent interest, as they are prevalently used for image classification [19–22], nonlinear regression [23–26], and image segmentation [27–30] tasks with a comparable or even better performance than the conventional methods. Especially for application to regression tasks, they represent extremely powerful nonlinear function estimators. Some network designs have been shown to give good results in computational phase retrieval from raw intensity measurements [23], even at low-photon counts [31–33]. More recently, the regression of Zernike coefficients has been demonstrated, as is appropriate for a modal wavefront control approach. [34] demonstrates direct retrieval of coefficients from image intensities in a simple imaging system, and [24] shows the validity of the technique and in other optical systems. [15] and [35] achieve ease of training and superior wavefront sensing accuracy by pre-processing the intensity using feature extraction methods, and by leveraging phase diversity. However, requiring multiple images with a known defocus offset is a barrier to the application of these techniques in a closed-loop wavefront control

system, as the defocus actuation time would reduce loop bandwidth. Furthermore, the specific application of the LLOWFS optical system is yet to be addressed.

In this paper, we describe an optical model to simulate the raw LLOWFS intensity corresponding to wavefront aberrations in the pupil plane (Section 2.1). A DNN design is proposed and trained to perform the regression of Zernike coefficients from LLOWFS intensity measurements in Section 2.2. In this section, a training scheme with noisy inputs is also described. We show that the technique allows retrieval of Zernike terms across a wide dynamic range of up to $\sim 1.0 \lambda$, in the presence of simultaneous Zernike terms up to fourth-order, by conducting extensive quantitative analysis on reconstruction accuracy and crosstalk among Zernike terms in Section 3. Concluding thoughts are in Section 4.

2. Method

We train a deep neural network (DNN) to directly retrieve Zernike expansion coefficients from the LLOWFS intensity measurement to solve this complex, nonlinear inversion problem. In particular, we work with datasets with wavefront errors (WFE) of $\sim 1.0 \lambda$. The direct retrieval of Zernike coefficients from the intensity measurement with this extended range of WFEs in a coronagraph environment has not yet been demonstrated. Such retrieval is useful for initial alignment of high-contrast systems, particularly on deployable telescopes, but the linear region of conventional LLOWFS methods in a coronagraph environment is limited, *i.e.* $\sim 0.1 \lambda$ [6]. We show that the simulated case in this work extends roughly an order of magnitude beyond the conventional dynamic range of LLOWFS.

2.1. Optical model

The LLOWFS system shown in Fig. 1 was numerically simulated using the Fraunhofer approximation. The coronagraph phase mask was implemented as a FQPM, and the LLOWFS detector was placed at a defocus of 2.4 waves. Optical simulations were implemented in a Python 3.7 environment using the Physical Optics Propagation in PYthon (POPPY) package [36]. To allow for fast computation, this simulation was performed assuming monochromatic light at a nominal wavelength of 635 nm, so any chromatic effects of the FQPM and other optical elements are not captured.

Aberration wavefronts were modeled as linear combinations of Zernike polynomials

$$W(\rho, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n c_{n,m} Z_n^m(\rho, \phi) = \sum_{k=1}^{\infty} c_k Z_k(\rho, \phi), \quad (1)$$

where $\sum_{n=0}^{\infty} \sum_{m=-n}^n c_{n,m} < \infty$. A Zernike aberration term is denoted as $Z_n^m(\rho, \phi)$, where n corresponds to the radial order of the term and m the angular meridional frequency. We can alternately number Zernike terms sequentially following the Noll Zernike expansion convention [37]. Following this definition of the Zernike basis, $W(\rho, \phi)$, the aberration (or wavefront error) map in polar coordinates, is made up of a linear combination of these Zernike terms with coefficients c_k . All Z_n^m are normalized to 1 nm RMS wavefront error. We define the overall root mean square (RMS) wavefront error (WFE) for a set of Zernike coefficients according to

$$E^{\text{RMS}} = \sqrt{\overline{(W(\rho, \phi))^2}} = \sqrt{\overline{\sum_{n=0}^{\infty} \sum_{m=-n}^n c_{n,m}^2}} = \sqrt{\overline{\sum_{k=1}^{\infty} c_k^2}}. \quad (2)$$

Here, the overline notation indicates a mean over the polar coordinate variables, *i.e.* ρ, ϕ .

Training and validation data for the neural network was generated using the described optical model, with input wavefront errors constructed from linear combinations of random Zernike coefficients with Noll indices between $k = 2$ (Tip) and $k = 15$ (Oblique Quadrafoil). Since this

method is intended to be applied to systems with various sources of wavefront error (atmospheric, optical misalignment, mirror figure, etc.) the coefficients were drawn from a generic distribution. Specifically, this distribution has overall wavefront error E^{RMS} uniformly distributed between 0 and 1.5 waves such that a contribution from each Zernike term is equally likely. This can be considered an artificially pessimistic case for high-order terms, since low-order terms typically dominate optical systems [38].

Intensity measurements without any noise in a Lyot-based coronagraph with a four quadrant phase mask were generated by simulation. Since the intensity measurements for imaging exoplanets are usually photon-starved, we further added noise according to statistics expected for a CCD (charge-coupled device) or EM-CCD (electron-multiplying CCD) sensor under different conditions as shown in Table 1. The choices of 1000 and 1 average photons per pixel correspond roughly to H-band magnitudes of $m_{\text{H}} = 4.6$ and $m_{\text{H}} = 12.0$, respectively, for an 8 m telescope and a frame rate of 170 Hz, not accounting for system throughput. The low-photon case is dimmer than host stars usually considered suitable candidates for direct imaging [39,40]. We demonstrate that our method of using a DNN to directly retrieve Zernike expansion coefficients works not only in a noiseless, ideal case but also in these noisy cases.

Table 1. Simulated noise levels for noisy intensity patterns.

	Low Noise	High Noise
Average photon counts per pixel (e^-)	1000	1
Readout noise (e^-)	10	10
Electron Multiplying (EM) gain	1	50

2.2. Network architecture and training

The choice of DNN architecture was based upon an understanding of methods commonly used for image processing and regression tasks. Specifically, we first examined various Convolutional Neural Network (CNN) architectures that are known to show good performance on ImageNet classification tasks, such as ResNet [20], Inception [19,22], and DenseNet [21]. Among these architectures, ResNet, in particular, uses a simple alternate branching path, *i.e.* identity mapping, to transfer the previously learned features to the next layer [20], which otherwise are easily lost after several layers. Both a vanilla ConvNet without residual pathways and with the pathways were tried; we have found adding residual pathways to a convolutional neural network architecture helped to preserve and effectively pass on useful information from inputs in the form of encoded low-dimensional nonlinear features. Consequently, compared to several other candidates, ResNet largely minimizes the computational time and consumed memory for training without compromising the quality of reconstruction. This ease of training enables an increase in the practical depth of CNN architectures, ultimately improving its performance.

We modified ResNet with additional convolutional layers at the input, allowing the network to extract sophisticated features out of the two-dimensional image inputs. Dropout layers were also added to further prevent overfitting when training the network [41], which greatly improved the performance of the neural network design. We tuned hyper-parameters to optimize network performance for this application, and added a fully connected layer at the output with a linear activation function, which sets the dimension to that of a 14-element vector of predicted Zernike expansion coefficients. The network design is shown in Fig. 3.

Training of the network was first performed with noiseless intensity patterns. All intensities were expressed in a logarithmic scale before being fed into the neural network. True Zernike coefficients were provided to the network in units of waves. As a training and validation loss function, we chose the root-mean-square error (RMSE) between predicted and true Zernike expansion coefficients. We use a gradient-based stochastic optimization algorithm called *Adam*

[42]. The patterns were randomly grouped into large mini-batches by the stochastic gradient procedure. All the patterns in each large mini-batch collectively determine the gradients for updating weights of the deep neural network [43], which effectively reduces the variance of stochastic gradient updates [44]. The initial learning rate of *Adam* is set 0.002. The learning rate is reduced by half whenever the validation loss stops decreasing and maintains its plateau for 10 epochs. The lower bound of the learning rate is set to be 10^{-7} . The training process uses 2×10^5 training intensity patterns and 5×10^4 validation patterns at each epoch for over 500 epochs.

After obtaining trained weights for the noiseless data, we re-trained the neural network with noisy intensity patterns, initializing with the weights from the previous training. In this training, we set an initial learning rate of 0.001 for the optimizer, and the learning rate was reduced by half after the validation loss plateaued for 5 epochs. Other training parameters remained the same. This retraining can be seen as a form of transfer learning between noiseless and noisy domains. As shown in [45], a deep neural network can learn the underlying physical relationship between intensity and phase using noiseless data. We can transfer this knowledge to the domain of noisy intensity patterns in the form of the previously-trained weights, and subsequently fine-tune them with domain specific knowledge using additional training. This works because many of the features learned in the early layers of a neural network are general enough to be applicable to very different datasets [46]. Starting with the weights learned from noiseless intensity patterns, it is sufficient to train for 250 epochs with noisy patterns.

The Keras 2.2.4 [47] environment with Tensorflow backend [48] was used for training and running the neural network. Training was performed on an Intel Xeon Gold 6248 CPU running at 2.50 GHz, 384 GB of RAM, and a Tesla V100 GPU with 32 GB of VRAM.

After training, the trained weights were used to generate Zernike coefficient estimates for an independent set of 5×10^4 intensity patterns. This test dataset was generated with Zernike coefficients distributed identically to the training and validation data. The results in Section 3. are based on the test dataset estimates. The testing process was performed on a CPU to demonstrate the practicality of implementing this method on systems where no GPU is present. The CPU used was an Intel Core i7-4750HQ CPU running at 2.0 GHz with 8 GB of RAM, and each estimate took 23 ms.

3. Results

3.1. Reconstruction Accuracy

For our primary performance metric we select the overall wavefront residual, which can be calculated as the root-sum-square of all Zernike error terms and defined as:

$$\epsilon^{\text{RMS}} = \sqrt{\left(W_{\text{true}}(\rho, \phi) - W_{\text{estimate}}(\rho, \phi)\right)^2} = \sqrt{\sum_{k=1}^{\infty} \left(c_k^{\text{true}} - C_k^{\text{estimate}}\right)^2}. \quad (3)$$

We also define a normalized version of this metric, defined as the ratio of the RMS of the wavefront residual to the RMS of the true aberration as defined in Eq. (2):

$$\text{Normalized error} = \frac{\epsilon^{\text{RMS}}}{E^{\text{RMS}}}. \quad (4)$$

Median values of both metrics are shown in Fig. 4. We can see that the normalized error linearly decreases on a logarithmic scale as RMS WFE increases. It stays below unity only where overall RMS WFE is larger than $\sim 10^{-2}$ waves. Where the overall RMS WFE values are small, even a small error introduced may be amplified or lead to the artifacts. Thus, in this region the normalized error may not be an effective metric to assess the performance of DNN as a nonlinear regressor.

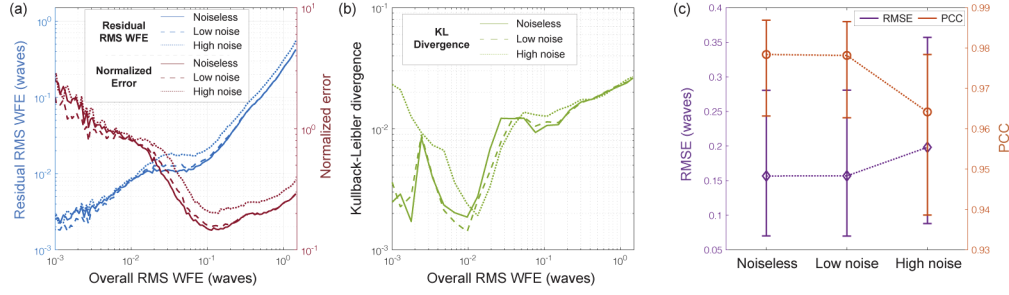


Fig. 4. (a) For all noise levels, the residual RMS WFE monotonically increases as the overall RMS WFE does, which is not surprising, but the normalized error decreases until RMS WFE reaches $\sim 10^{-2}$ waves. (b) Kullback-Leibler (KL) divergence between reconstructions and ground truth was computed within each bin across different noise conditions to address the issue, whose trend is more similar to that of the residual RMS WFE. (c) Overall residual RMS WFE and Pearson correlation coefficient (PCC) were computed over different noise levels. Error bars for each noise level are drawn on top of the medians, denoting 25th and 75th quantiles. In (a-c), low noise case shows similar performance to that of noiseless case, and the performance decreases for higher noise. For (a) and (b), lines are the median of the results.

To address this issue, Kullback-Leibler (KL) divergence was computed along the given overall RMS WFE range. KL divergence is used to measure the difference between two different distributions, defined as:

$$D_{\text{KL}}(p, q) = \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right). \quad (5)$$

Since it is not ratiometric but is the weighted sum of the distributions, it stays comparatively small in the region below $\sim 10^{-2}$ waves compared with the case of the normalized error. In the region larger than $\sim 10^{-2}$ waves, it is to be expected that it increases, as the overall RMS WFE does as well. Taking both metrics into account leads to a better assessment of the performance of the DNN.

Figure 4(c) shows the robustness of the neural network to noise in terms of two quantitative metrics: overall residual RMS WFE and Pearson correlation coefficient (PCC), where PCC shows a linear correlation between two different variables. In Fig. 4(c), PCC was calculated between the residual RMS WFE and RMS WFE as functions of WFE. Both overall residual RMS WFE, defined in Eq. (3), and PCC do not largely deviate from the results of the ideal, noiseless case even in the extremely low photon regime. The strength of using the neural network for this particular regression problem lies in its ability to retrieve each Zernike expansion coefficient out of the intensity patterns with high accuracy despite the presence of noise.

Figure 5 shows the relationships between the true value of each Zernike coefficient and the residual after subtracting the predicted value of that coefficient. Each axis corresponds to a variant of Eqs. (2) and (3) respectively when clamped at the m^{th} Zernike coefficient, defined as:

$$E_{m, \text{clamped}}^{\text{RMS}} = \sqrt{\delta_{mk} \sum_{k=1}^{\infty} (c_k^{\text{true}})^2} = |c_m^{\text{true}}|, \quad (6)$$

$$\epsilon_{m, \text{clamped}}^{\text{RMS}} = \sqrt{\delta_{mk} \sum_{k=1}^{\infty} (c_k^{\text{true}} - c_k^{\text{estimate}})^2} = |c_m^{\text{true}} - c_m^{\text{estimate}}|, \quad (7)$$

which correspond to the absolute value and the residual of the m -th Zernike term, respectively. Figure 5 also includes the normalized version of these values defined similarly to Eq. (4).

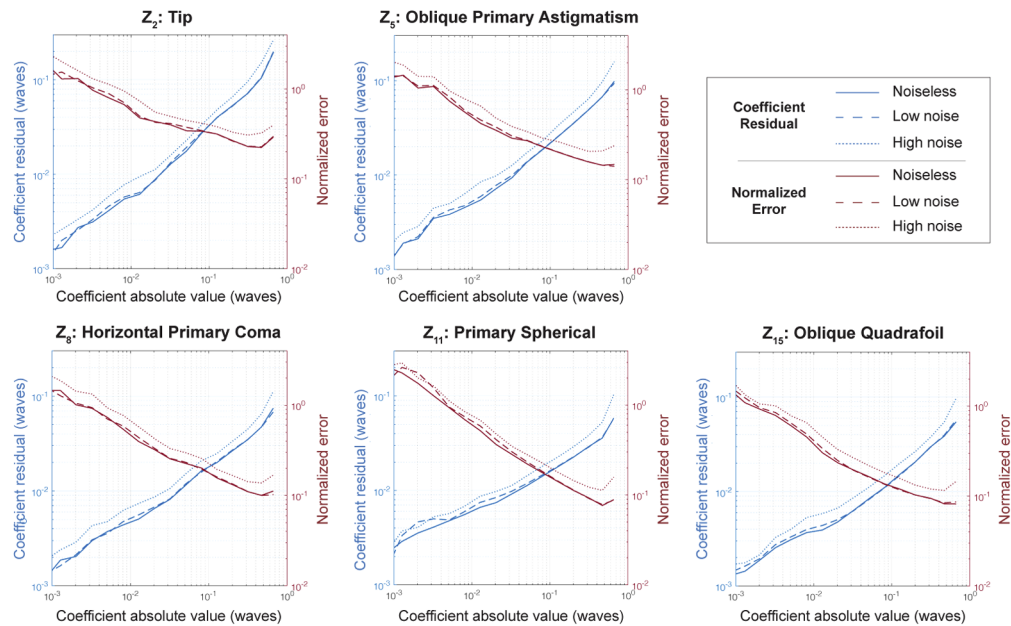


Fig. 5. Relationship between residual RMS WFE and RMS WFE of each Zernike term across different noise levels. Results of other Zernike terms can be found in Fig. 8 in Appendix A.

Up to $\sim 0.5 \times 10^{-2}$ waves of Zernike coefficient value, the normalized error is larger than 1.0, meaning that the reconstruction using DNN does not do well for small signals; the conventional linear method [6] shows a better performance in this region. However, the deep residual learning method has its strength when the coefficient value increases, larger than $\sim 10^{-2}$ waves; the normalized error linearly decreases on a logarithmic scale. Deep residual learning effectively increases the applicable range of wavefront errors for closed-loop compensation up to $\sim 1.0 \lambda$. We note that the poor performance of this method at small wavefront displacements means it cannot be used on its own to achieve the low-aberration regime necessary for high-contrast imaging. For detection of an Earth-like planet around a sun-like star at 10 parsecs, tip-tilt terms must be controlled to better than 10^{-3} waves [49]. Therefore, the deep residual learning method is most suited to use in initial optical alignment before switching to the more accurate but low-dynamic range linear method.

We assess the performance of the DNN in the presence of only a single aberration term, which allows a more direct comparison to the linear method of [6]. A specialized test dataset was generated, in which each Zernike mode is taken in isolation and swept from -1.5 to 1.5 waves RMS. It was tested with the DNN which was trained with noiseless patterns with linear combination of Zernike coefficients. Figure 6 shows the responses of the trained DNN to four of these modes (please refer to Fig. 9 in Appendix A for analysis.). Included for comparison are response curves generated using the linear regression method described in [5].

We observe that the responses of Z_5 and higher terms, except Z_{11} , are highly linear with a slope of around 1, even out to 1.5 waves RMS. The responses to Z_2 and Z_3 remain largely linear but with a lesser slope, while terms Z_4 and Z_{11} exhibit a highly nonlinear response. These results stand in contrast to the linear method, whose response becomes non-monotonic for aberrations of more than about 0.2 waves RMS [49].

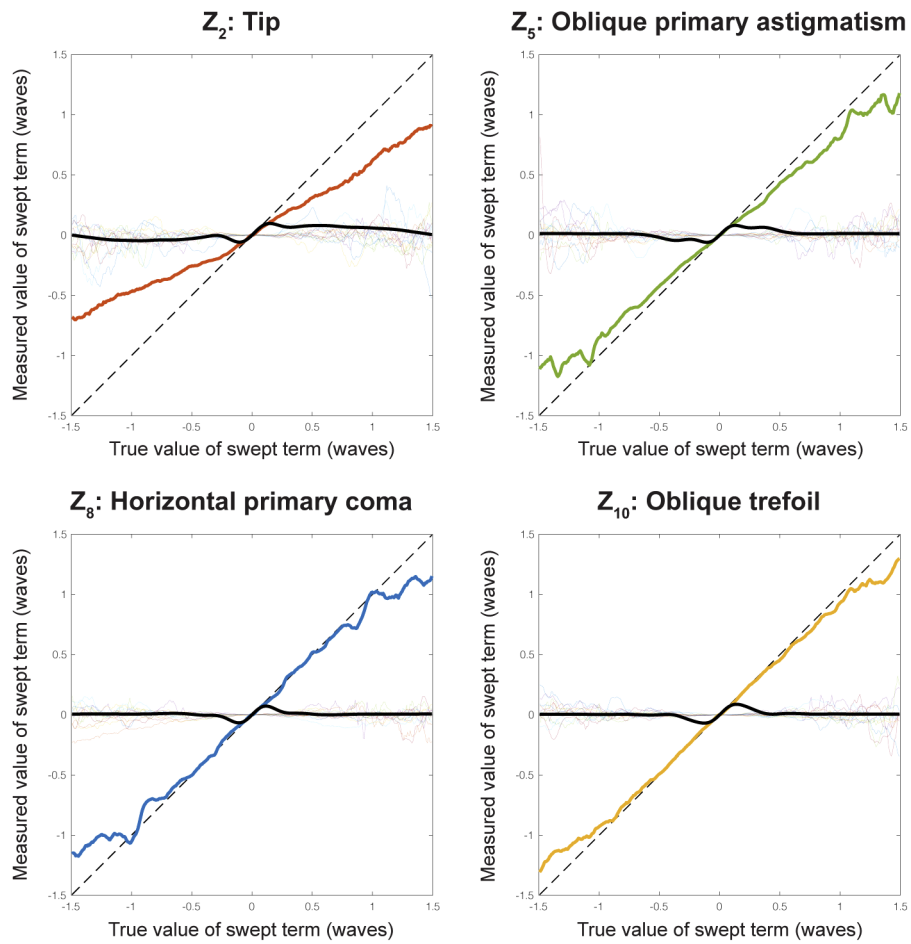


Fig. 6. Response of the DNN-based sensor (colored lines) and a linear method (black bold line) to individual Zernike terms. In each figure, a single aberration mode is applied in an amount varied between -1.5 and 1.5 waves. The measured value of each coefficient in response to these aberrations is shown. Clearly, the DNN-based sensor has a more extensive dynamic range with good linearity than the linear method. The responses for other modes are included in Fig. 9 in Appendix A.

3.2. Crosstalk

Another measure of the performance of the wavefront estimate is the degree to which the value of one zernike coefficient affects the estimate of the other coefficients. Ideally, as Zernike terms form an orthogonal basis, if the neural network were to correctly learn the basis, there should be no term-to-term interaction. However, Fig. 7 displays the correlation between the error in the mean reconstruction for pairs of Zernike terms, showing that some degree of crosstalk exists among Zernike terms. Each element in Fig. 7(a) through Fig. 7(c) can be represented by the equation

$$\rho_{pq} = \rho(\epsilon_p^{\text{RMS}}, \epsilon_q^{\text{RMS}}), \quad p, q = 2, 3, \dots, 15, \quad (8)$$

$$\text{where } \rho(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}$$

and ϵ_p^{RMS} and ϵ_q^{RMS} are the residual RMS WFEs of p and q th Zernike term, respectively, according to the Noll Zernike expansion convention. $\rho(x, y)$ is the Pearson correlation coefficient (PCC) that is essentially the normalized cross-covariance between x and y .

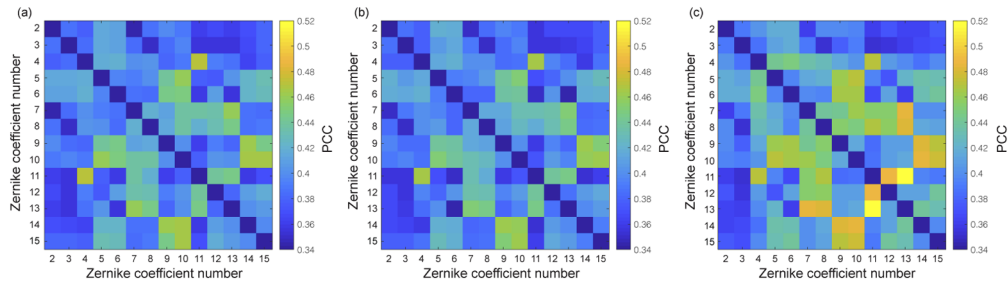


Fig. 7. (a-c) Pearson correlation coefficients (PCC) between residual RMS WFE of one Zernike term and that of another Zernike term when the intensity patterns are (a) noiseless, (b) low noise, and (c) high noise. Diagonals are left blank. Crosstalk among Zernike terms increases as images are more affected with noise.

As for the ideal, noiseless case, a term-to-term crosstalk is noticeable between Z_4 and Z_{11} , representing Defocus and Primary Spherical, respectively. Since both of the terms are perfectly circularly symmetric but for their radial orders, the neural network fails perceive these two terms as perfectly orthogonal. Additional inter-term crosstalk can be noticed between Zernike terms of Foil but with different radial orders, e.g. Z_9 , Z_{10} and Z_{14} , Z_{15} . Based on the observations, we conclude that the neural network inadvertently introduces spurious crosstalk between Zernike terms with the same or adjacent angular meridional frequencies.

As the noise level increases, as in Fig. 2, the features in the intensity patterns become more unclear, and the underlying orthogonality among Zernike terms is obfuscated, which thus contributes to the increase in the degree of term-to-term crosstalk, such as in Fig. 7(c).

4. Conclusion

We have shown that wavefront sensing using DNNs can be applied to the LLOWFS optical system. A DNN incorporating convolutional layers and ResNet concepts is able to directly retrieve Zernike coefficients from LLOWFS intensity patterns. The sensor response is monotonic even for wavefront error up to 1.5 waves. Compared to the traditional linear technique, the DNN is less exact for very small aberrations, but provides greatly extended dynamic range, and can operate in the presence of multiple large aberration terms even in a low-photon regime. In fact, the reconstruction error as a percentage of actual aberration amplitude is shown to decrease as the aberration grows, provided it is within the range of the training data. The reconstruction error is

less than 20% for aberrations of 0.02λ or more RMS. For systems where wavefront aberrations are expected to exceed the range of conventional LLOWFS use, this DNN-based technique could take the place of additional hardware, such as a SHWFS for coarse sensing.

There remain practical challenges in the application of these methods to a physical telescope system. An experimental demonstration is necessary to ensure that the reconstruction is robust to sources of noise and systematic error that may not be captured in these simulations. A particular challenge is the generation DNN training data on a real optical system. The challenges of response matrix generation for a conventional LLOWFS (e.g. SHWFS calibration and optical system stability) are magnified here due to the large number of examples required. It would be preferable to leverage a physical model of the coronagraphic system, though reliance on purely simulated training examples will be subject to systematic error due to differences between the model and physical system. Indeed, there have been several attempts on training neural networks with synthetic data based on a computer model and testing with experimental acquisitions in various fields, e.g. in optical tomographic reconstruction [50,51], and deep-sea acoustic source imaging [52]. Note that in the former case, the performance of a network trained on entirely simulated data was found to be sufficient. The latter work leverages transfer learning, as discussed in Section 2.2. In this case, the network is first trained on simulated data and then re-trained on a small amount of real-world data. Future work will assess these approaches in a laboratory environment. Specifically, the sensitivity of the reconstruction to realistic noise sources and to errors in model parameters, and the applicability of transfer learning methods. Future study will also assess the use of other machine learning techniques in wavefront sensing for coronagraphs, e.g. recurrent neural networks [35], deep belief networks [26], or and feature-based approaches [15].

The DNN technique is very flexible, and can potentially be applied to many different optical systems. Other common wavefront sensor geometries, such as the Pyramid and Zernike WFS, also become nonlinear at high displacements and could potentially be augmented by application of DNNs. Future work should be undertaken to explore the suitability of this method for phasing of segmented aperture telescopes, which are planned for future space systems like LUVVOIR and HABEX [53].

Appendix A. Supplementary figures

In this section, we provide supplementary figures to Figs. 5 and 6 to provide further understanding on other results than those presented in the main text.

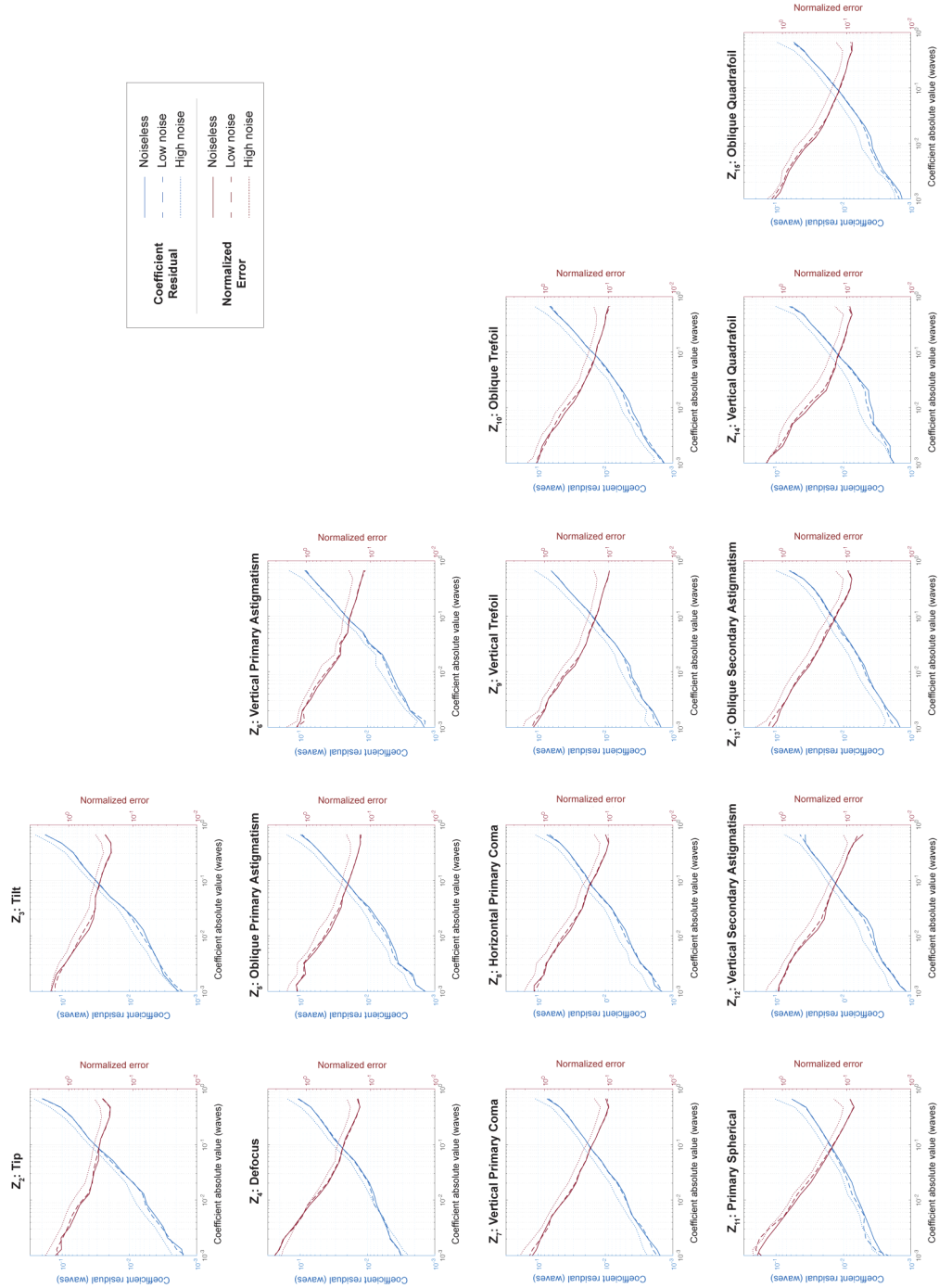


Fig. 8. Supplement to Fig. 5. This figure displays the residual and normalized error of each Zernike term over the absolute value of the coefficient.

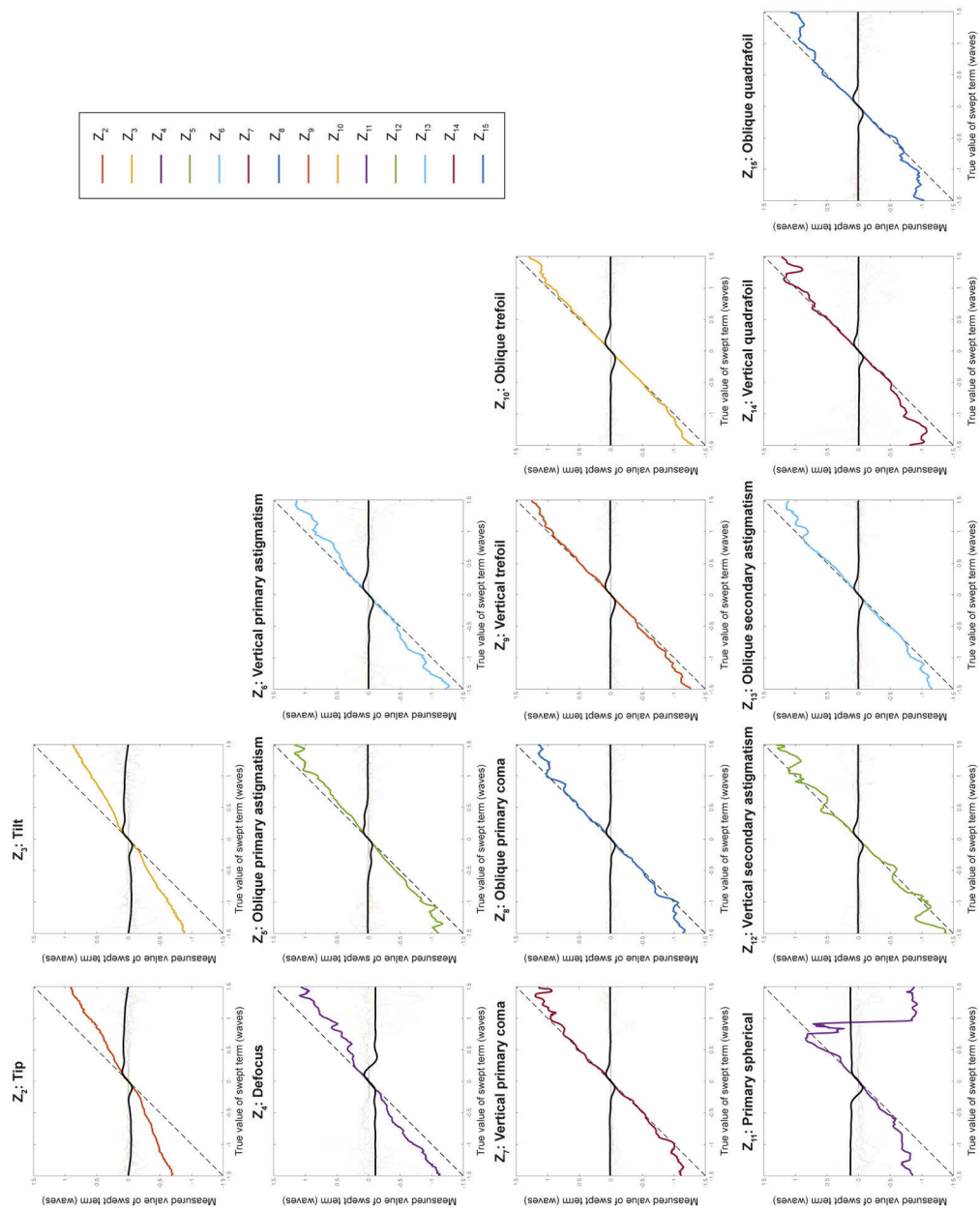


Fig. 9. Supplement to Fig. 6. This figure displays the individual response of each Zernike term out of the trained neural network (colored lines) and the linear method (black bold line) with simultaneous Zernike terms.

Funding

Korea Foundation for Advanced Studies; Jet Propulsion Laboratory (1640749); Defense Advanced Research Projects Agency (AMA-19-0015, W31P4Q-16-C0089); Intelligence Advanced Research Projects Activity (FA8650-17-C-9113).

Acknowledgments

Thanks to Alexander Knoedler, Ondrej Čierny, Changyeob Baek and Clara Park for their contributions to preliminary proof-of-concept studies as part of MIT coursework; to Leonid Podgorelyuk for helpful discussions, and to Mo Deng for some constructive comments. G. Allan acknowledges JPL/WFIRST and DARPA for partial support, and I. Kang acknowledges partial support from the KFAS (Korea Foundation for Advanced Studies) scholarship and the Intelligence Advanced Projects Activity. The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing resources (HPC, database, consultation) that have contributed to the research results reported within this paper. This research made use of POPPY, an open-source optical propagation Python package originally developed for the James Webb Space Telescope project [36,52].

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. W. A. Traub and B. R. Oppenheimer, "Direct imaging of exoplanets," in *Exoplanets*, S. Seager, ed. (University of Arizona, 2010), pp. 111–156.
2. L. Pueyo, "Direct imaging as a detection technique for exoplanets," in *Handbook of Exoplanets*, H. J. Deeg and J. A. Belmonte, eds. (Springer, 2018), pp. 705–765.
3. F. Zernike, "Diffraction theory of the knife-edge test and its improved form, the phase-contrast method," *Mon. Not. R. Astron. Soc.* **94**(5), 377–384 (1934).
4. J. W. Hardy, *Adaptive optics for astronomical telescopes* (Oxford University, 1998).
5. G. Singh, F. Martinache, P. Baudoz, O. Guyon, T. Matsuo, N. Jovanovic, and C. Clergeon, "Lyot-based low order wavefront sensor for phase-mask coronagraphs: Principle, simulations and laboratory experiments," *Publ. Astron. Soc. Pac.* **126**(940), 586–594 (2014).
6. G. Singh, J. Lozi, O. Guyon, P. Baudoz, N. Jovanovic, F. Martinache, T. Kudo, E. Serabyn, and J. Kuhn, "On-sky demonstration of low-order wavefront sensing and control with focal plane phase mask coronagraphs," *Publ. Astron. Soc. Pac.* **127**(955), 857–869 (2015).
7. D. Rouan, P. Riaud, A. Boccaletti, Y. Clenet, and A. Labeyrie, "The four-quadrant phase-mask coronagraph. i. principle," *Publ. Astron. Soc. Pac.* **112**(777), 1479–1486 (2000).
8. A. Boccaletti, P. Riaud, P. Baudoz, J. Baudrand, D. Rouan, D. Gratadour, and F. Lacombe, "The four-quadrant phase mask coronagraph. iv. first light at the very large telescope," *Publ. Astron. Soc. Pac.* **116**(825), 1061–1071 (2004).
9. P. Huguenaier, E. Serabyn, B. Mennesson, J. K. Wallace, R. O. Gappinger, M. Troy, E. E. Bloemhof, J. Moore, and C. D. Koresko, "Astronomical near-neighbor detection with a four-quadrant phase mask (fqpm) coronagraph," in *Space Telescopes and Instrumentation I: Optical, Infrared, and Millimeter*, vol. 6265 (International Society for Optics and Photonics, 2006), p. 62651G.
10. G. Foo, D. M. Palacios, and G. A. Swartzlander, "Optical vortex coronagraph," *Opt. Lett.* **30**(24), 3308–3310 (2005).
11. S. Esposito, A. Puglisi, E. Pinna, G. Agapito, F. Quirós-Pacheco, J. Véran, and G. Herriot, "On-sky correction of non-common path aberration with the pyramid wavefront sensor," *Astron. Astrophys.* **636**, A88 (2020).
12. C. B. Mendillo, J. Brown, J. Martel, G. A. Howe, K. Hewawasam, S. C. Finn, T. A. Cook, S. Chakrabarti, E. S. Douglas, and D. Mawet et al., "The low-order wavefront sensor for the picture-c mission," in *Techniques and Instrumentation for Detection of Exoplanets VII*, vol. 9605 (International Society for Optics and Photonics, 2015), p. 960519.
13. C. B. Mendillo, K. Hewawasam, G. A. Howe, J. Martel, T. A. Cook, and S. Chakrabarti, "The picture-c exoplanetary direct imaging balloon mission: first flight preparation," in *Techniques and Instrumentation for Detection of Exoplanets IX*, vol. 11117 (International Society for Optics and Photonics, 2019), p. 1111707.
14. P. Janin-Potiron, M. N'Diaye, P. Martinez, A. Vigan, K. Dohlen, and M. Carbillet, "Fine cophasing of segmented aperture telescopes with zelda, a zernike wavefront sensor in the diffraction-limited regime," *Astron. Astrophys.* **603**, A23 (2017).
15. 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).
16. B. Paul, L. Mugnier, J.-F. Sauvage, K. Dohlen, and M. Ferrari, "High-order myopic coronagraphic phase diversity (coffee) for wave-front control in high-contrast imaging systems," *Opt. Express* **21**(26), 31751–31768 (2013).
17. A. E. Riggs, N. J. Kasdin, and T. D. Groff, "Recursive starlight and bias estimation for high-contrast imaging with an extended kalman filter," *J. Astron. Telesc. Instrum. Syst.* **2**(1), 011017 (2016).
18. L. Pogorelyuk and N. J. Kasdin, "Dark hole maintenance and a posteriori intensity estimation in the presence of speckle drift in a high-contrast space coronagraph," *Astrophys. J.* **873**(1), 95 (2019).

19. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2015), pp. 1–9.
20. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2016), pp. 770–778.
21. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2017), pp. 4700–4708.
22. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, (2017).
23. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica* **4**(9), 1117–1125 (2017).
24. 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).
25. A. Kendall and R. Cipolla, "Geometric loss functions for camera pose regression with deep learning," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (2017), pp. 5974–5983.
26. X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amarutunga, "Ensemble deep learning for regression and time series forecasting," in *2014 IEEE symposium on computational intelligence in ensemble learning (CIEL)*, (IEEE, 2014), pp. 1–6.
27. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention (MICCAI)*, (Springer, 2015), pp. 234–241.
28. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
29. P. Moeskops, J. M. Wolterink, B. H. van der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, "Deep learning for multi-task medical image segmentation in multiple modalities," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, (Springer, 2016), pp. 478–486.
30. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, (IEEE, 2016), pp. 565–571.
31. A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Phys. Rev. Lett.* **121**(24), 243902 (2018).
32. M. Deng, S. Li, A. Goy, I. Kang, and G. Barbastathis, "Learning to synthesize: Robust phase retrieval at low photon counts," *Light: Sci. Appl.* **9**(1), 36 (2020).
33. I. Kang, F. Zhang, and G. Barbastathis, "Phase extraction neural network (phenn) with coherent modulation imaging (cmi) for phase retrieval at low photon counts," *Opt. Express* **28**(15), 21578–21600 (2020).
34. S. W. Paine and J. R. Fienup, "Machine learning for improved image-based wavefront sensing," *Opt. Lett.* **43**(6), 1235–1238 (2018).
35. Q. Xin, G. Ju, C. Zhang, and S. Xu, "Object-independent image-based wavefront sensing approach using phase diversity images and deep learning," *Opt. Express* **27**(18), 26102–26119 (2019).
36. M. Perrin, J. Long, E. Douglas, A. Sivaramkrishnan, and C. Slocum, "Poppy: Physical optics propagation in python," *Astrophysics Source Code Library (ASCL)* (2016).
37. R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**(3), 207–211 (1976).
38. D. W. Kim, C. J. Oh, A. Lowman, G. A. Smith, M. Aftab, and J. H. Burge, "Manufacturing of super-polished large aspheric/freeform optics," in *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II*, vol. 9912 (International Society for Optics and Photonics, 2016), p. 99120F.
39. J. R. Males and O. Guyon, "Ground-based adaptive optics coronagraphic performance under closed-loop predictive control," *J. Astron. Telesc. Instrum. Syst.* **4**(01), 1 (2018).
40. E. S. Douglas, J. R. Males, J. Clark, O. Guyon, J. Lumbres, W. Marlow, and K. L. Cahoy, "Laser Guide Star for Large Segmented-aperture Space Telescopes. I. Implications for Terrestrial Exoplanet Detection and Observatory Stability," *Astron. J.* **157**(1), 36 (2019).
41. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
42. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
43. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT, 2016).
44. M. Li, T. Zhang, Y. Chen, and A. J. Smola, "Efficient mini-batch training for stochastic optimization," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2014), pp. 661–670.
45. M. Deng, S. Li, I. Kang, N. X. Fang, and G. Barbastathis, "On the interplay between physical and content priors in deep learning for computational imaging," arXiv preprint arXiv:2004.06355 (2020).
46. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Adv. Neural Inform. Process. Syst. (NIPS)*, (2014), pp. 3320–3328.
47. F. Chollet, "Keras: Deep learning library for theano and tensorflow," URL: <https://keras.io/> (2015).
48. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker,

- V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467 (2016).
49. G. Singh, "Low-order wavefront control and calibration for phase mask coronagraphs," Ph.D. thesis, École Doctorale d'Astronomie & Astrophysique d'Île-de-France (2015).
 50. A. Goy, G. Rughoobur, S. Li, K. Arthur, A. I. Akinwande, and G. Barbastathis, "High-resolution limited-angle phase tomography of dense layered objects using deep neural networks," *Proc. Natl. Acad. Sci.* **116**(40), 19848–19856 (2019).
 51. I. Kang, A. Goy, and G. Barbastathis, "Limited-angle tomographic reconstruction of dense layered objects by dynamical machine learning," <https://arxiv.org/abs/2007.10734> (2020).
 52. W. Wang, H. Ni, L. Su, T. Hu, Q. Ren, P. Gerstoft, and L. Ma, "Deep transfer learning for source ranging: Deep-sea experiment results," *J. Acoust. Soc. Am.* **146**(4), EL317–EL322 (2019).
 53. M. D. Perrin, R. Soummer, E. M. Elliott, M. D. Lallo, and A. Sivaramakrishnan, "Simulating point spread functions for the james webb space telescope with webbpsf," in *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, vol. 8442 (International Society for Optics and Photonics, 2012), p. 84423D.