

Mitigating the nonlinearities in a pyramid wavefront sensor

Finn Archinuk^{a,b}, Rehan Hafeez^{b,c}, Sébastien Fabbro^{b,a,b},
Hossen Teimoorinia^{a,b} and Jean-Pierre Véran^{a,b,*}

^aNational Research Council Canada, Victoria, British Columbia, Canada

^bUniversity of Victoria, Victoria, British Columbia, Canada

^cUniversity of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT. For natural guide star adaptive optics (AO) systems, pyramid wavefront sensors (PWFSs) can provide a significant increase in sensitivity over the traditional Shack–Hartmann but at the cost of a reduced linear range. When using a linear reconstructor, nonlinearities result in wavefront estimation errors, which can have a significant impact on the image quality delivered by the AO system. We simulate a wavefront passing through a PWFS under varying observing conditions to explore the possibility of using a nonlinear machine learning model to estimate wavefront errors and compare with a linear reconstruction. We find significant potential improvements in delivered image quality even with computationally simple models, underscoring the need for further investigation of this approach.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JATIS.9.4.049005](https://doi.org/10.1117/1.JATIS.9.4.049005)]

Keywords: adaptive optics; machine learning; wavefront reconstruction; astronomy

Paper 23052G received May 2, 2023; revised Oct. 16, 2023; accepted Nov. 30, 2023; published Dec. 28, 2023.

1 Introduction

When observing the sky at visible or infrared wavelengths with a large ground-based telescope, atmospheric turbulence causes random distortions in the incoming light (i.e., wavefront errors) that significantly reduce the resolution and contrast of the recorded images. These effects can be corrected in real-time with an adaptive optics (AO) system, and most present and future optical telescopes include one or several AO systems. Although AO systems cannot provide a perfect correction, the residual wavefront errors are small enough so that the delivered image quality becomes limited by the diffraction of the telescope, a fundamental limit. This results in higher resolution, and higher contrast astronomical images unveiling finer details of the structures in the universe. AO also dramatically increases the sensitivity of the observations, significantly reducing the required exposure time to reach a given signal-to-noise ratio on scientific targets, therefore allowing more targets to be observed each night.

A basic AO system, also called single-conjugate AO (SCAO) system, consists of three major components: a deformable mirror (DM), which corrects the wavefront distortions using actuators pushing and pulling a reflective surface; a wavefront sensor (WFS), which measures the residual wavefront errors on a bright guide star; and a real-time controller, which estimates the wavefront from the WFS measurements and updates the DM commands so that the residual wavefront errors are minimized. An AO system, therefore, is a closed-loop feedback system that needs to be operated at typical frame rates of ~ 1 kHz in order to keep up with changes in atmospheric turbulence. This means that every millisecond, a set of WFS measurements is obtained, and the

*Address all correspondence to Jean-Pierre Véran, Jean-Pierre.Veran@nrc-cnrc.gc.ca

DM shape is updated. Beyond SCAO, more sophisticated AO systems involving several WFSs and possibly several DMs have been developed in order to increase the size of the corrected field. However, SCAO systems are very relevant, especially when working on scientific targets close to a bright star, which can be used as a guide star for the WFS. This is the case for the so-called extreme AO systems, such as the Gemini Planet Imager (GPI),¹ which aim at producing very high-contrast images, in which faint stellar companions, such as exoplanets, can be found.

The performance of the AO system is almost always limited by the WFS's ability to accurately measure the instantaneous residual wavefront from a limited number of photons, making the WFS an absolutely critical component. The traditional Shack–Hartmann WFS has been widely used in existing AO systems because it provides a high-level of linearity in reconstructing the wavefront from the WFS. The reconstructor is then implemented as a simple matrix–vector multiplication, an easily parallelizable process that can be executed with very low latency on modern computers.² However, different WFSs such as the pyramid WFS³ (PWFS) have been introduced recently because they are more sensitive, providing a more accurate measurement for a given light level, or conversely, providing the same accuracy for a lower light level (i.e., a fainter guide star). This increase in sensitivity, however, comes with a loss of linearity, which creates additional errors in the wavefront reconstruction process, when a linear reconstructor is used.⁴ With a PWFS, the trade-off between increased sensitivity and loss of linearity can be adjusted by modulating the image of the guide star around the tip of the pyramid during light integration on the WFS detector and by adjusting the modulation radius.⁵ However, even when this trade-off is optimized, the nonlinearity errors can be quite significant. For example, NFIRAOS,⁶ the first light AO system for the future Thirty Meter Telescope, has a PWFS for natural guide star observations, and the PWFS nonlinearity effects account for 64 nm of RMS wavefront error, out of a total budget of 156 nm RMS on a magnitude 8 natural guide star.⁷ Using the so-called Marechal approximation,⁸ RMS wavefront errors can be directly translated into the Strehl ratio, a quantitative metric for image quality. 156 nm RMS corresponds to a Strehl ratio of 70.2% at a wavelength of 1.65 μm (H-Band), down from 74.6% with no nonlinearity errors.

In this paper, we propose to evaluate how one can mitigate nonlinearities in a PWFS by implementing a nonlinear wavefront reconstructor derived using deep learning. We limit ourselves to a traditional AO setup where each WFS measurement is processed for wavefront reconstruction independently, therefore focusing on spatial effects as opposed to temporal effects. We refer to our previous work⁹ for an attempt to use machine learning in order to predict atmospheric turbulence by taking advantage of short-range temporal correlations. Specifically, a convolutional neural network (CNN) as a reconstructor is evaluated as a substitute for a linear system. The CNN is built through standard machine learning methods by training on simulated wavefront maps representative of typical wavefronts measured by AO systems. We show that the CNN is able to capture nonlinearities in the measurement process and provide a reconstruction accuracy that is significantly better than the linear reconstructor.

2 Related Work

Modern observatories are increasingly employing PWFSs, including Keck,¹⁰ the Large Binocular Telescope,¹¹ the Large Magellan Telescope,¹² and the Thirty Meter Telescope¹³ currently under construction. The PWFS tends to replace the more traditional Shack–Hartmann WFS in the Natural Guide Star AO systems because of its increased sensitivity, which enables improved AO correction and/or larger sky coverage.

The sensitivity of the PWFS (i.e., optical gain) changes depending on the level of correction provided by the AO system (better AO correction = lower residuals = higher sensitivity), which in turn depends on observing conditions. If left uncompensated, these changes in sensitivity would cause the AO loop gain to fluctuate, potentially leading to under-performance (gain too low) or even instabilities (gain too high), as well as errors in compensating for noncommon path aberrations (aberrations seen by the science channel but not seen by the WFS channel or vice versa).¹⁴ This problem has been mitigated by estimating the optical gain in real time as conditions change and applying its inverse as part of the wavefront reconstruction process.¹⁵ It was then recognized that each correction mode had its own optical gain,¹⁶ and that adjusting the modal optical gains

can not only account for changes in observing conditions, but also if the adjustment can be performed often enough, mitigate nonlinearities as well.^{17,18} Weinberger et al.¹⁹ have even proposed to use neural networks to estimate the optical gains. However, all these methods assume that the modal optical gains are independent, which is only an approximation, especially in very nonlinear conditions, such as in the case of an unmodulated PWFS. Very recently, this idea was extended with the SIMPC method,²⁰ which computes the entire wavefront reconstruction matrix around the expected level of wavefront residual, in effect linearizing the wavefront reconstruction problem around typical AO residual levels, as opposed to no residuals in the traditional method. How to implement this approach in practice, however, remains a topic for research, as the reconstruction matrix would still have to be updated when observing conditions, and therefore AO residuals, change. In this paper, we explore the alternative approach of giving up on the linear reconstructor altogether and replacing it with a nonlinear reconstructor implemented in the form of a CNN.

Neural network approaches have been proposed in order to improve AO correction, but most of this work has been focused on trying to predict the atmospheric turbulence in order to reduce the lag error inherent to all AO systems.^{9,21,22} Machine learning techniques have recently been recognized for their potential to mitigate nonlinearities intrinsic to WFSs.²³ Some advances have been made to apply these techniques to the Shack–Hartmann WFS²⁴ and to design better WFSs.²⁵ A recent paper provides an overview on machine learning applications for wavefront sensing, but only identifies a single application to the PWFS, in ophthalmology,²⁶ where the operating conditions are quite different from those in astronomy. So far, machine learning techniques have been used for highly nonlinear wavefront sensing problems. For example, Orban de Xivry et al.²⁷ looked at measuring noncommon path aberrations directly in the science focal plane (i.e., focal plane WFSing). Landman and Haffert²⁸ demonstrated that CNNs can be used to aid the reconstruction of wavefronts measured with a WFS purposefully made nonlinear in order to increase its sensitivity. Although the task and constraints of their problems differ from what we focus on, their use of deep convolutional architectures demonstrates the adoption of machine learning methods for real-world nonlinear problems.

Concurrently, advancements in image-based machine learning have surged. Transformer architectures have been adapted for images,^{29,30} however, they still tend to require more parameters and data to train with, as a result they are too computationally expensive for millisecond rate real-time implementation. Residual networks (ResNets)³¹ form the backbone of many image-based models. Their efficacy is due to residual connections, which allow more layers to be trained. The advances found by vision transformers have been analyzed and applied to ResNets, which led to the development of ConvNeXt,³² which further builds upon ResNets by parallelizing computational pathways. These promising advancements in computer vision hold potential for our case. However, achieving submillisecond inference times imposes constraints on our choice of architectures, limiting flexibility, as explored in Sec. 5.

3 Data Source

In this study, we utilize simulated random wavefront maps that are generated using the power spectrum method. This method entails taking the Fourier transform of the square root of the desired power spectrum, with the addition of a random phase at each frequency f . Most AO systems have a closed-loop architecture, which means that the WFS does not see the full atmospheric turbulence but instead measures the correction residuals. Accordingly, we have simulated wavefront maps intended to represent typical AO residual wavefronts, which are sensed by the PWFS. Their power spectrum follows a f^{-2} power law, which is shallower than the Kolmogorov $f^{-11/3}$ power law representing the uncorrected atmospheric turbulence. An f^{-2} power law is typical for AO corrected wavefronts for spatial frequencies within the correction range of the DM.³³ However, in order to include more diversity in our training set, as well as to test the robustness of the model, we have also generated wavefront maps with $f^{-1.8}$ and $f^{-2.2}$ power laws. When generating data with power law f^{-p} , we refer to p as the f -value.

The wavefront maps are generated for a Gemini-like $D = 8$ m telescope on a square grid of 176×176 pixels, which provides a sampling of 22 pixels per meter when projected on the primary mirror of the telescope. The pupil is circular with a central obscuration of 1.0 m.

For this pupil, the Karhunen–Loève (KL) modes of the Kolmogorov turbulence have been computed, and each wavefront map is mathematically projected onto the first 1603 KL modes. The resulting 1603 coefficients are the “true” modal coefficients, which, if applied to a modal DM, would minimize the residuals. Of course, in a real system, the true coefficients are not available and must be estimated from the WFS measurements. The goal then is to minimize the error between the estimated coefficients and the true coefficients in order to maximize the delivered image quality.

The PWFS is simulated using the physical optics PWFS module included in PASSATA.³⁴ The PWFS is set up so that it mimics the GPI 2.0 PWFS, which uses an EMCCD220 with 60 pixels in the diameter of each of the four pupil images.³⁵ The PWFS is modeled to be sensitive to a 300-nm wide band centered on $\lambda = 750$ nm. The flux available to the PWFS is derived from the magnitude value of the target, assuming an A0 star and affects the photon noise applied to the PWFS images. Noiseless simulations corresponding to the case of a very bright guide star are also performed for reference purposes. Each PWFS image is captured on a 140×140 pixel grid. All the measurements are obtained with a $3\lambda/D$ modulation, which is typical for such systems.³⁶ The software determines the correct masks to extract the four pupil images, from which 5640 X and Y slopes can be computed. A linear reconstructor is used to obtain an estimate of the 1603 modal coefficients from the slope vector for each wavefront map. The linear reconstruction is obtained via the singular value decomposition inversion of a modal interaction matrix, which contains the measured slopes of each KL mode. This interaction matrix and its inverse are obtained directly from the PASSATA software. The software also ensures that the modes, when presented for the interaction matrix acquisition, have the proper amplitude to stay within the linear range of the PWFS. Figure 1 outlines this simulation process graphically.

The metric to be minimized is the quadratic norm—i.e., the root sum square (RSS)—of the residual coefficients, which is the difference between the 1603 reconstructed and the true modal coefficients. Since the KL modes are orthonormal over the circular pupil function, minimizing the RSS of the residual coefficients maximizes the Strehl ratio—i.e., the optical quality—of the delivered image. The RSS of the residual coefficients is directly related to the root mean square (RMS) of the residual wavefront error. However, the latter also includes the wavefront errors of a higher order than the first 1603 KL modes, which cannot be corrected by the system (fitting error).

It is well known that when the RMS wavefront error is low, the linear reconstructor performs well, but as the RMS increases, the measurement becomes less linear and the reconstruction error increases.⁴ Current PWFS-based AO system accepts this trade-off.^{10–13} With the limitations of a linear reconstructor stated, we use the linear model as a baseline for quality.

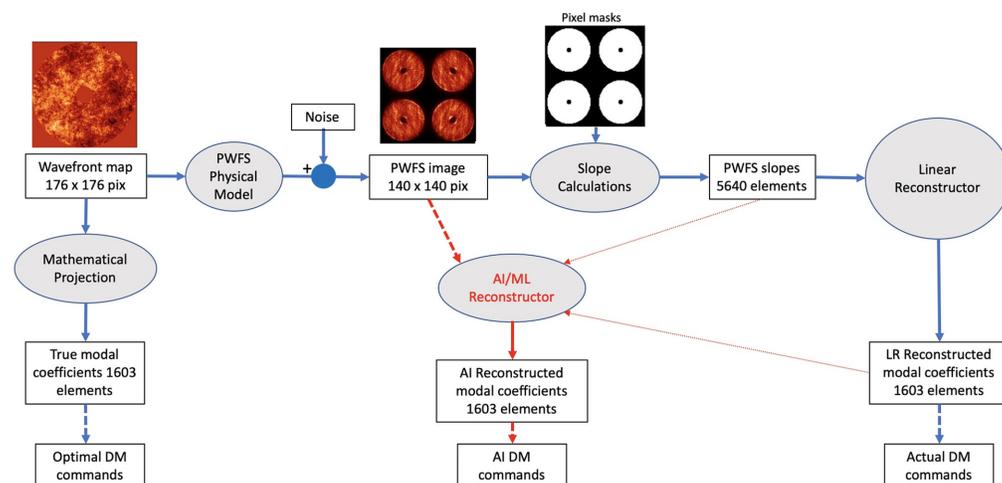


Fig. 1 Outline of the simulation pipeline. Thin red-dotted lines are possible inputs to the ML model that we did not pursue in this work. The thick red lines show the inputs and outputs of the proposed ML model.

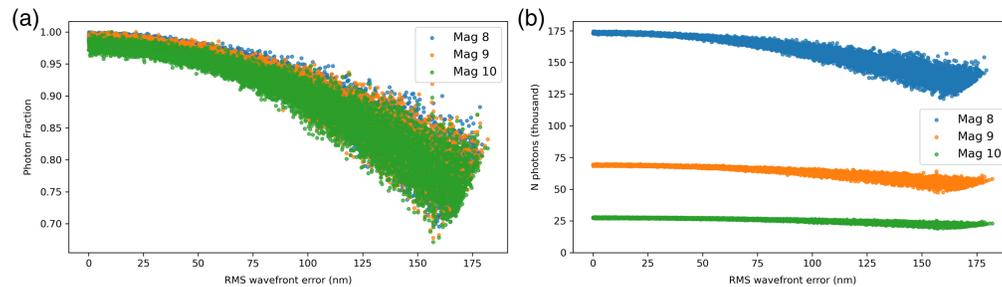


Fig. 2 Flux in the geometric pupil images as a function of the wavefront amplitude for an f -value of $p = 2$ and various magnitudes. (a) Expressed as a fraction of the total number of photons hitting the pupil. This is normalized to the sample with the maximum number of photons for each magnitude. (b) Expressed in absolute number of photons.

Data is generated in groups of 10,000 frames characterized by the magnitude of their guide star and the f -value. Generated wavefront maps are intentionally designed to be statistically independent in order to maximize the information for training the CNN. The models we evaluate are trained on magnitude 8 or 9 simulated sources and tested on magnitude 8, 9, and 10 data.

In order to obtain datasets representing a variety of conditions, we scale each wavefront map to have an RMS wavefront error between 0 and 200 nm. This spans a reasonable range of residual wavefront amplitudes, which vary depending on observing conditions and parameters of the AO system under study. For each frame, the RSS value of the 1603 coefficients will be slightly below the RMS wavefront error because of the fitting error, as discussed previously.

Finally, we explore the fraction of the flux actually contained in the four pupil images used to compute the slopes. Because of diffraction, some photons land outside the geometric pupil images. Figure 2 shows that as the wavefront amplitude increases, a larger fraction of the flux is diffracted outside the geometric pupil images. Diffraction effects result in the loss of up to 30% of photons not reaching the geometric pupil images for frames with the highest RMS values. These photons are unused in a traditional linear reconstructor where slopes are first computed from pixels within the geometric pupil images, but we suspect that they might be key to modeling the nonlinearities of the PWFS.

4 Data Formatting and Preprocessing

The two potential inputs for our model are CCD frames or slopes. Slopes are a reduced representation of the wavefront, which is beneficial for simplifying the CNN model by reducing the number of input features. However, there is information loss in this reduction process, confirmed by our own experiments, which showed better accuracy when inputs from the original CCD frames are used. This could be attributed, at least in part, to the flux beyond the boundaries of the geometric pupil images, which is disregarded during the slope computations, as previously discussed.

The initial processing step involves converting the WFS frame into a reduced-intensity image by normalizing it with the total count and then subtracting the normalized WFS image corresponding to a flat wavefront:

$$\Delta I(\phi) = \frac{I(\phi)}{\Sigma I(\phi)} - \frac{I(\phi = 0)}{\Sigma I(\phi = 0)},$$

where ϕ is the wavefront map, $\Delta I(\phi)$ is the reduced intensity image, $I(\phi)/\Sigma I(\phi)$ is the normalized CCD frame, and $I(\phi = 0)/\Sigma I(\phi = 0)$ is the normalized WFS image for a flat wavefront. The utilization of reduced intensity images offers the advantage of robustness to variations in total illumination, as well as having a flat wavefront image as the reference with zero intensity. Using reduced intensities instead of computing slopes has become a standard practice as the former contains more information than the latter for the wavefront reconstruction process.¹⁷ Of course, for real nonsimulated images, standard image preprocessing, including flat field and background removal, would also have to be performed.

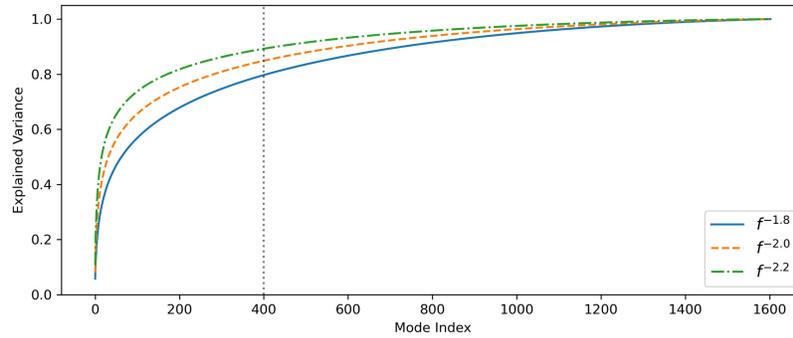


Fig. 3 Energy of cumulative modes. Each line represents 10,000 frames with the specified f -value.

After creating the reduced intensity image, the floating point values of the frame require scaling. Raw values are close to zero, and neural networks perform best when data points utilize the space between ± 1 . We tried a variety of methods to return these values to a useful range: framewise scaling, where the frame has a mean of zero and standard deviation of one; pixelwise scaling, where each pixel location in the training data is scaled to have a mean of zero and standard deviation of one; and fixed scaling, where the whole frame is multiplied by a fixed constant.

Framewise and fixed scaling are the most obvious scaling methods, relying only on the information within the frame. Fixed scaling maintains a direct connection to the reduced intensity image by upscaling the resulting frame by a constant factor. Our experiments suggest that this scaling method was the most robust for changes in magnitude for the set of sources we evaluated. An empirical value of 1000 works well and could be considered as a hyperparameter when further optimizing this model or for adapting this model to a PWFS with a different number of illuminated pixels.

Figure 3 observes the fraction of total modal RMS wavefront error given the number of modes considered. This fraction depends on the f -value, and the three f -values we evaluate are plotted. As expected, low-order modes carry more energy when a steeper (higher f -value) power-law is used. In our work, we focus on the reconstruction of the first 400 modes, which, as shown in the figure, correspond to at least 75% of the total wavefront RMS.

5 Neural Network Architecture

Computer vision research has mostly been focused on deep neural network architectures over the last 10 years. One of the most adopted deep architectures for supervised learning tasks has been the CNN, which exploits translational symmetry by applying successive learned filters in a hierarchical manner, that is, CNNs allow us to efficiently find spatial patterns across the input. We performed significant architecture and hyperparameter tuning, with the greatest impacts on wavefront reconstruction quality being informed by the number and size of filters, and the number of convolutional layers. Here we present a CNN architecture that performs well across the datasets we evaluated.

A visual representation of our CNN architecture is found in Fig. 4. The model uses three convolutional blocks, each with 16 (5×5) filters. A fourth convolutional layer projects these to

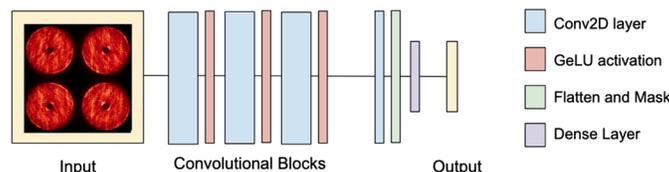


Fig. 4 Our proposed architecture has three convolutional blocks, followed by flattening and modal reconstruction. The input is a pupil image sample following the reduced intensity preprocessing and the output are the first 400 modes.

a single channel. We implemented a custom layer to apply the pupil mask to remove the background pixels from the CCD frame. The reconstructed modal coefficients—the output of the model—are a linear combination of the remaining pupil pixels without a bias coefficient. The custom layer leverages AO-specific knowledge about the task to remove a considerable number of model parameters (≈ 1 million) without significant impact on the model quality. Removing the bias term from the final fully connected layer has a minimal impact on the model quality, but this choice is informed by the fact that the modes from the reduced intensity input should be centered at zero. This architecture has ~ 5.1 million trainable parameters.

Optimizing the architecture for real-time inference would require considerable work and is outside the scope of this analysis. Larger models have been demonstrated to run at 6 ms on NVIDIA V100 GPUs³⁷ suggesting our model is likely to be able to be optimized to the required latency (≤ 1 ms) on more modern GPUs.

5.1 Model Training

Models were trained with 17,100 simulated frames and stopped when the error of the validation set of 900 frames stopped decreasing. Models were trained on either magnitude 8 or 9 data. We used the AdamW³⁸ optimizer with a learning rate of 8×10^{-5} and weight decay of 1×10^{-5} . Training used a batch size of 256. Model quality was assessed against the ground-truth modes using mean squared error as a metric. Our intention is to optimize the entire set of reconstructed modal coefficients instead of specific modes; therefore, we did not adjust the scaling of the modal coefficients during training. We let larger magnitudes of the lower modes act as output weighting so that the lower modes (especially tip, tilt, and defocus) were proportionally more important for the model to optimize for. By separating reconstruction quality by total RMS wavefront error and guide star magnitude, we gain a better understanding of the reconstruction limitations of the models. From these metrics, we were able to develop model architectures and regularization hyperparameters iteratively. We observed our model performing poorly in the low RMS region relative to the linear reconstructor and attempted sample weighting so that low RMS frames were more impactful when calculating the gradient to update model parameters. Sample weighting followed the exponential decay equation: ab^x , where a was set to an initial value of 1000, b was the exponential decay of 0.975, and x was equal to the RMS of the sample.

Figure 5 shows the effects of sample weighting on otherwise identical models. Results are the RSS of the reconstructed modal coefficients versus ground truth and have been separated by

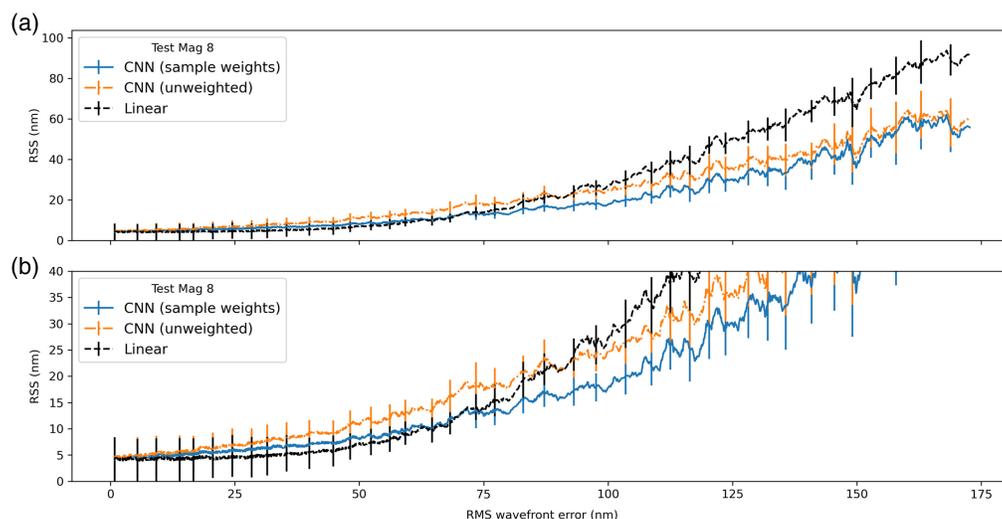


Fig. 5 Weighting samples by RMS value improves model performance. (a) The overall change and (b) a zoomed-in view of the lower (<125 nm) RMS cases. Samples were weighted based on their RMS, with low RMS samples being higher weighted. Moving averages are plotted to help visualize the trend, and error bars are the standard deviation of the window.

RMS of the input wavefront error. Figure 5(a) shows improvement at the higher RMS wavefront error levels, and Fig. 5(b) zooms into the lower RMS wavefront error cases. We expected sample weighting to cause a loss in reconstruction quality for the higher RMS region; however, it acted as a regularization term and improved the reconstruction quality over all RMS values. The reason this may be acting as a regularizer is that by forcing the model to focus on the linear region, simpler solutions were imposed across the model.

6 Results

6.1 Wavefront Reconstruction

Figure 6 shows sample reconstruction residuals, comparing the traditional linear reconstruction method and the CNN approach. We have limited the wavefront to the first 400 KL modes and have looked at wavefront maps of different amplitudes. We see that for the high amplitude wavefront map (134 nm RMS), the residual provided by the CNN is significantly smaller than the residual provided by the linear reconstructor (69 versus 85 nm RMS), whereas, for the low-amplitude wavefront map, the linear reconstructor produces slightly lower residuals (8 versus 10 nm RMS).

6.2 Error by Mode

Our architecture reconstructs the first 400 modes, which are evaluated against the true modes for the different magnitude test sets. By separating reconstruction quality as a function of mode index, Fig. 7 shows that relative to the linear approximation, our model makes the best reconstructions for the lowest order modes, with diminishing returns as the mode index increases. Each test session consists of 2000 frames. The main figures show reconstruction error relative to the true coefficient. The inset figures show the improvement in RMS wavefront error relative to the linear reconstruction error, which more clearly shows that our CNN model outperforms the linear reconstructor for lower order modes. We see that for low-order modes, reduction in RMS reconstruction error of up to 40% is possible, which is quite significant.

6.3 Overall Reconstruction Quality

In the results presented in the previous section, modes are examined for the entire range of 0 to 200 nm RMS of wavefront error. Here we investigate the overall wavefront reconstruction

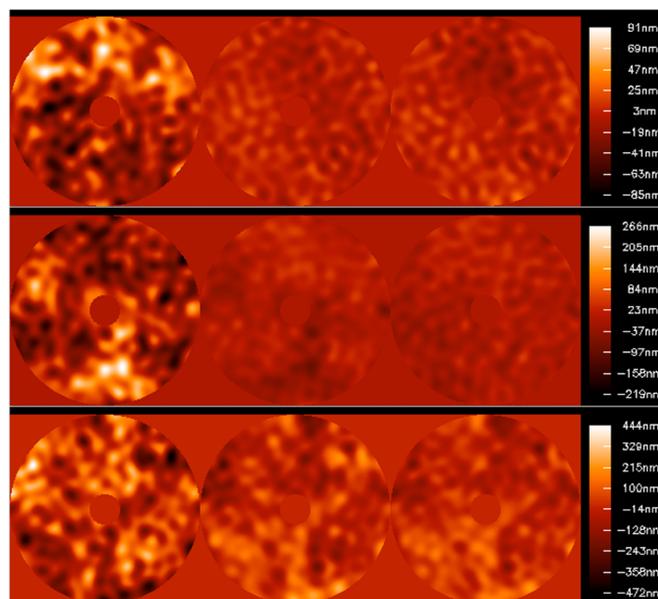


Fig. 6 Wavefront reconstruction residuals for different levels of incoming wavefront. Left: incoming wavefront map, 400 KL modes (30/80/134 nm RMS). Middle: residual wavefront error map after linear reconstruction (8/25/85 nm RMS). Right: residual wavefront error map after CNN reconstruction (10/19/69 nm RMS).

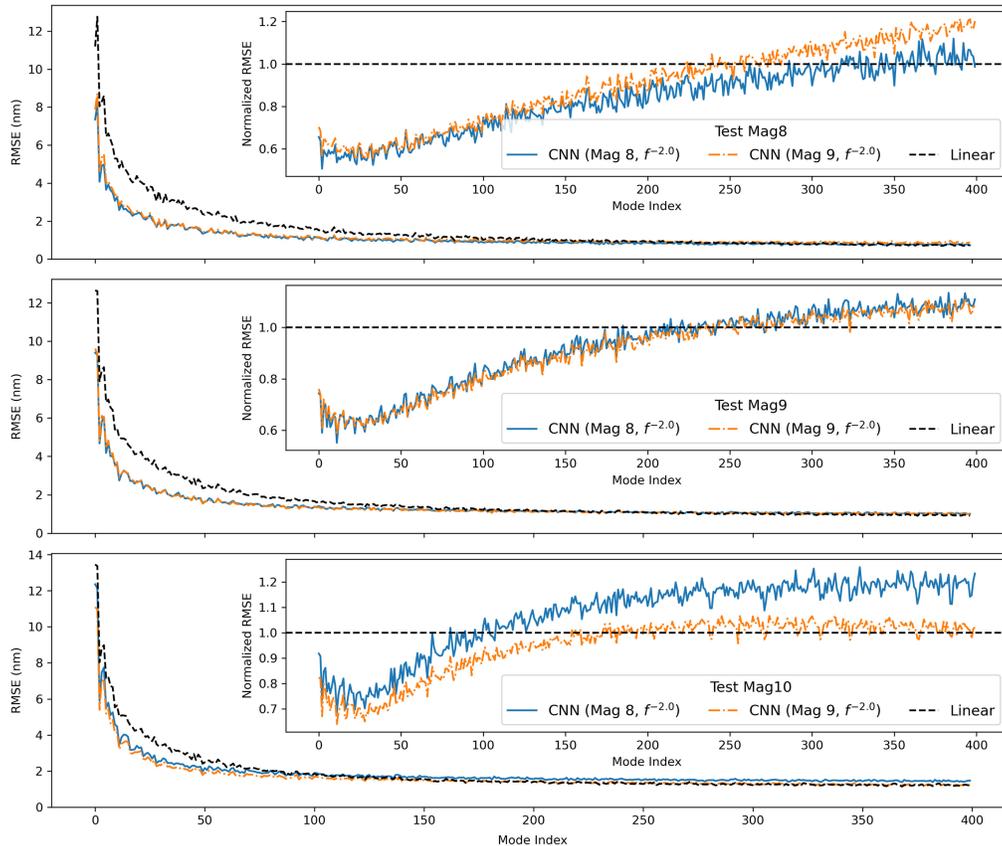


Fig. 7 Error by mode of the proposed model on three guide stars of different magnitudes. The main panels show the CNN outperforms the linear model at lower modes. Inset panels show reconstruction change relative to the linear baseline.

quality for a given RMS. Here the RMS wavefront error is calculated using the RSS of the 1603 true modes, so it does not include the fitting error, which is why the 200 nm RMS value is not reached. Figure 8 shows the moving average trend our model reconstructs as a function of RMS. The CNN model outperforms the linear reconstruction for wavefront RMS values above 75 nm for magnitude eight sources, above 85 nm for magnitude nine sources, and above 120 nm for magnitude 10 sources. These results are expected since higher RMS values correspond to higher amplitudes and, therefore, to higher levels of nonlinearities. For a wavefront error of 150 nm RMS, which is a typical residual for a PWFS-based NGS AO system on a bright star (magnitude 8), such as NFIRAOS (see Sec. 1), the linear method results in a residual RMS wavefront error of 71 ± 9 nm compared to our method, which achieved 44 ± 11 nm. This reduction is about 40% in RMS residual wavefront error. For an AO system with 150 nm RMS of residual wavefront error using a linear reconstructor, the CNN would bring the residual wavefront error to $\sqrt{150^2 - 71^2 + 44^2} = 139$ nm RMS. Using the Marechal approximation in H-band as in Sec. 1, this would bring the Strehl ratio from 72.2% to 75.6%. When observing a point source with AO, the sensitivity, which is inversely proportional to the exposure time needed to reach a given signal-to-noise ratio, is roughly proportional to the square of the Strehl ratio.³⁹ Therefore, the CNN-based reconstructor could provide an increase in sensitivity of about 10%, which is quite significant since this directly translates into a 10% increase in observing efficiency for the telescope. Interestingly, the magnitude 9 CNN model produces reconstructions better than linear reconstructor in the linear region, which is especially visible on the magnitude 9 and 10 data. We expect this is due to the CNN model implicitly applying a smoothing function that reduces the noise propagation. Models in Fig. 8 have a moving average applied with a window size of 21 to highlight the trend. Error bars show one standard deviation.

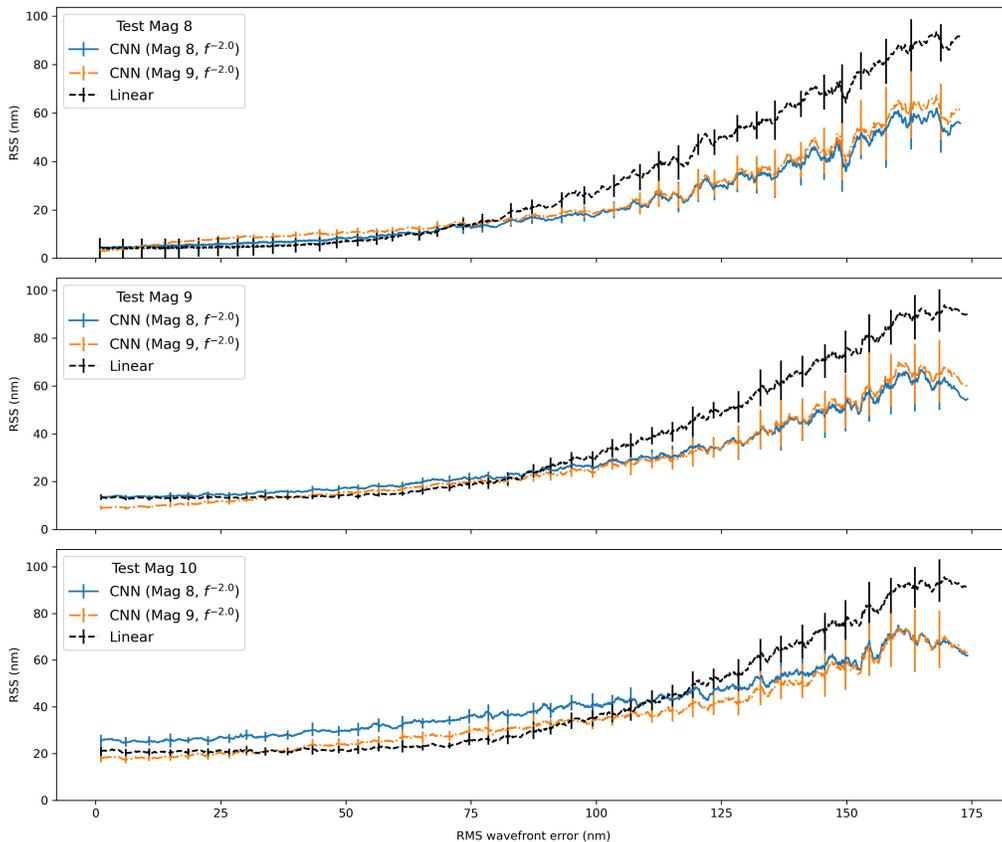


Fig. 8 Wavefront reconstruction quality as a function of RMS for magnitudes 8, 9, and 10. Models were trained on either magnitude 8 or 9 data with an f -value of $p = 2$ power law. Moving averages are plotted to help visualize the trend, with error bars showing one standard deviation for that window.

6.3.1 Single-mode reconstruction

We can inspect single modes across different values of total RMS wavefront error to better understand where the model accumulates errors. Figure 9 displays the absolute errors of modes 0 (tilt), 10, and 100 when trained to reconstruct magnitude 8 samples. The magnitude of the errors decreases with the mode index, as expected. The improvement in wavefront reconstruction is most noticeable at higher RMS and is no better than the linear model when the RMS is below 75 nm. This is again expected since the PWFS has a linear behavior for low RMS wavefront errors.

We confirm the CNN gives more importance to the expected (physical) modes using saliency mapping. A saliency map takes a model output and determines the importance of input features used to calculate it. Multiple methods exist to combine the gradients, and we used Grad-CAM introduced by Selvaraju et al.⁴⁰ and implemented by the Xplique library.⁴¹ Figure 10(a) shows which pixels were important for determining the specified mode. Figure 10(b) provides reference measurements of the corresponding Karhunen–Loève modes.

6.4 Model Robustness

We classify model robustness in two ways; the model must be able to handle changes in atmospheric conditions and adapt to dimmer sources. Throughout the previous section, we have presented results showing how wavefront reconstructions perform when we change the magnitude, and as expected, brighter sources provide more photons to reduce error in modal reconstructions. We now turn to robustness versus the statistical characteristics of the measured wavefronts.

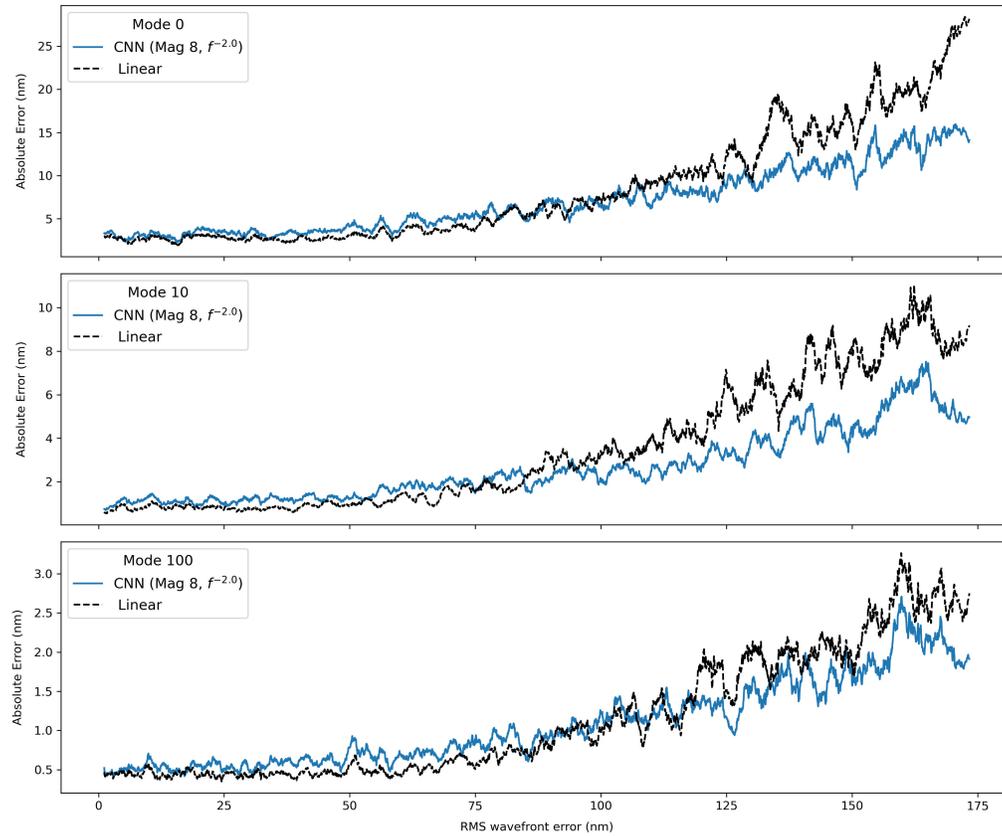


Fig. 9 Selection of three modes showing reconstruction quality by RMS on magnitude 8 samples. Moving averages are plotted to help visualize the trend.

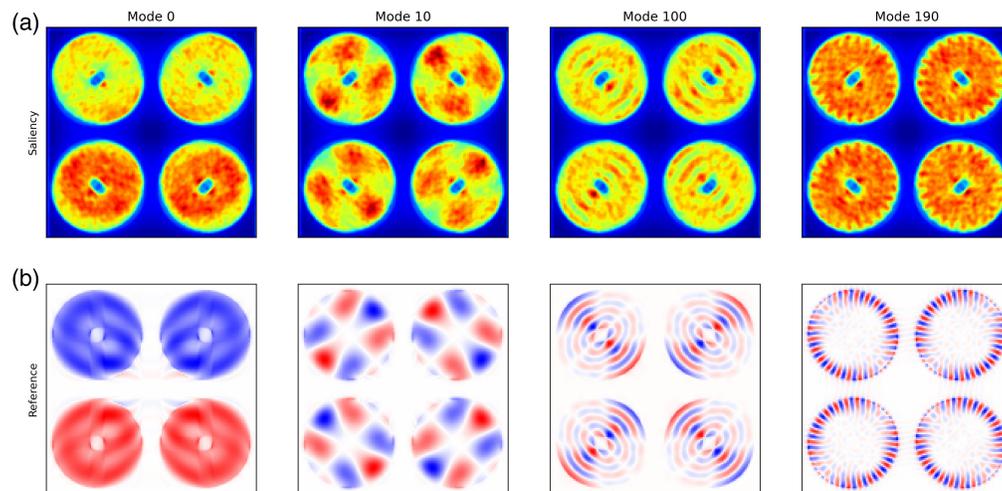


Fig. 10 Selection of modes comparing saliency and idealized Karhunen–Loève modes. (a) The importance of each pixel to the model when determining the modal output. (b) The corresponding Karhunen–Loève mode.

6.4.1 f -value robustness

The f -value defines a power law outlining the decreasing importance of modes. Atmospheric conditions with higher f -values correspond to lower frequency distortions that are captured by lower order modes. It is important that our model is not overfitted to a narrow f -value window, which would limit a real-world application. To test f -value robustness, we prepared two models:

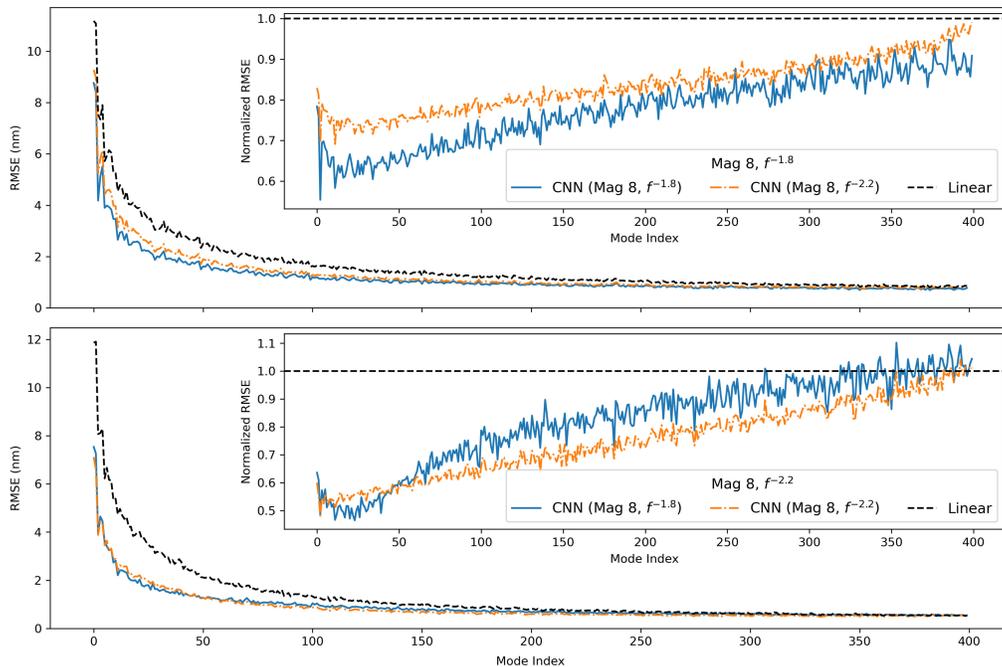


Fig. 11 Two models were trained on different f -values and testing was performed on both sets of data. The main figures show the error against the true modal coefficient. The insert figures show error relative to the linear baseline.

one trained with 17,100 frames of $f^{-2.2}$ and the other trained with 17,100 frames of $f^{-1.8}$. Each model made reconstructions on 2000 samples of both $f^{-2.2}$ and $f^{-1.8}$. The model trained on $f^{-2.2}$ acted as the baseline for the $f^{-2.2}$ testing set, and reconstructions made by the model trained on $f^{-1.8}$ showed how the reconstruction quality drops with a change in f -value. This was repeated for the $f^{-1.8}$ model acting as the baseline. Figure 11 shows the drop in model performance caused by changing the f -value during testing. Both models performed significantly better than the linear model, though not as well as the one trained directly on the corresponding f -value data. Interestingly, the loss in reconstruction quality was anisotropic with the normalized RMSE of both models having different characteristics: $f^{-1.8}$ made better reconstructions for a small window of modes on the $f^{-2.2}$ test set, whereas the $f^{-2.2}$ model performed uniformly worse on the $f^{-1.8}$ test set. These observations suggest the models were learning filters to prioritize modes more common in their training data.

These two models were then evaluated by looking at frame quality as a function of RMS in the same way as in Sec. 6.3. In Fig. 12, we see both models outperforming the linear model for higher RMS frames (above 100 nm). Here we can more clearly see that the model trained on $f^{-1.8}$ atmospheric conditions does not transfer to $f^{-2.2}$ atmospheric conditions as well as the inverse.

7 Conclusion and Future Work

In this work, we demonstrated that a neural network is able to take into account nonlinearities in the measurement process of an AO PWFS and make considerable improvements over a linear reconstructor in delivered image quality. These improvements are most noticeable in the low-order modes, and for RMS errors above 75 nm, which is typical for astronomical AO systems. When measuring a 150 nm RMS wavefront residual, the portion due to reconstruction errors is found to be reduced by about 40%, which results in a gain of more than 3 points in Strehl ratio at H-band (1.65 μ m). When observing a point source, this gain could translate to an increase in observing efficiency of up to 10%, which is quite significant.

We see diminished returns on fainter sources where WFS noise dominates, but still potential improvements are observed in higher (>125 nm) RMS conditions. A result of focusing on the overall reconstruction means that the quality of a mode reconstruction decreases with index,

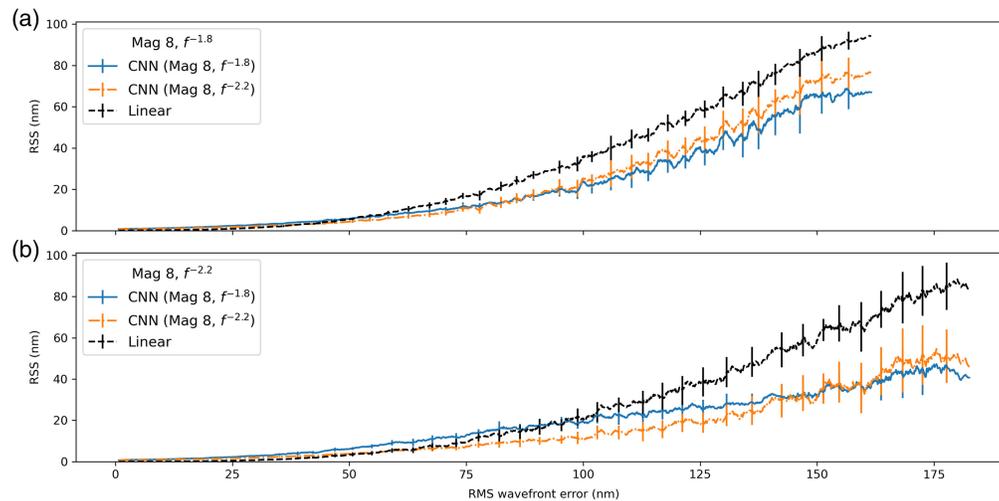


Fig. 12 Error by RMS of two models trained on different f -values. (a) Models reconstructing $f^{-1.8}$ samples and (b) reconstructing $f^{-2.2}$ samples. Moving averages are plotted to help visualize the trend. Error bars are one standard deviation for that window.

which can be explained by our model using convolutional filters to identify lower frequency patterns. Increasing the number of filters—especially at the lower layers—should improve the reconstruction quality, but this was not found in practice. The cause for the difficulty reconstructing higher order modes with our CNN architecture is unclear but may be related to these modes concentrating power toward the edge of the pupils. Why exactly that would be a problem and how to mitigate it is future work.

A major goal of this work is to test the robustness of a fixed model to reconstructions outside of its training range. This occurs in two dimensions: changes in magnitude and changes in f -value. The improvements for low-order mode reconstructions remained for magnitude 10 data when the model was trained on magnitude 8 data. When our model was trained on magnitude 9 data, this robustness was further realized, suggesting that training on more varied samples could further improve the final model. The other axis of robustness, f -value change, was evaluated by training models on different power-law distributions of residual wavefront error. These models were then evaluated on the opposing dataset, where we found they were susceptible to changes in this f -value. This change is not unexpected as the convolutional layers in the model would be more tuned for higher or lower frequency patterns depending on the training set.

This paper is focused on proof-of-concept through numerical simulations, and, at this point, we have not thought in detail on how this approach could be implemented on a real system. For the training, one could simply use simulated wavefront maps and simulated PWFS measurements, making sure that the parameters of the target AO systems are correctly captured in the model. For a more realistic training, based on physical measurements, one could imagine installing a high precision WFS at the output of the AO system, and a wavefront generation device, such as a DM or a spatial light modulator or a phase screen at the input, the latter providing sample wavefront maps that the PWFS could measure while the former would provide the truth measurement. Since training can be performed at a relatively slow speed if necessary, commercial devices could be used. The other challenge is to run the ML-reconstructor in real time, with a latency low enough to track the atmospheric turbulence (typically <1 ms). This implementation would replace the matrix–vector multiplication used to implement the linear reconstructor in conventional AO systems and would probably require a combination of architecture optimization and advanced computation hardware.

Other future work includes training on more varied datasets to improve robustness, and investigating the potential benefits of using different CNN models for different observing conditions.

Code and Data Availability

Code is available upon request.

References

1. B. Macintosh et al., “The Gemini Planet Imager: looking back over five years and forward to the future,” *Proc. SPIE* **10703**, 107030K (2018).
2. T. Y. Chew, R. M. Clare, and R. G. Lane, “A comparison of the Shack–Hartmann and pyramid wavefront sensors,” *Opt. Commun.* **268**(2), 189–195 (2006).
3. R. Ragazzoni et al., “Multiple spatial frequencies wavefront sensing,” in *Adapt. Opt. for Extremely Large Telesc. 5—Conf. Proc.*, arXiv:1808.00373 (2017).
4. I. Shatokhina, V. Hutterer, and R. Ramlau, “Review on methods for wavefront reconstruction from pyramid wavefront sensor data,” *J. Astron. Telesc. Instrum. Syst.* **6**(1), 010901 (2020).
5. C. Véronaud, “On the nature of the measurements provided by a pyramid wave-front sensor,” *Opt. Commun.* **233**, 27–38 (2004).
6. J. Crane et al., “NFIRAOS adaptive optics for the Thirty Meter Telescope,” *Proc. SPIE* **10703**, 107033V (2018).
7. K. Hardy et al., “Thirty meter telescope adaptive optics system error budgets and requirements traceability,” in *Proc. AO4ELT5 Conf.* (2017).
8. V. N. Mahajan, “Strehl ratio for primary aberrations in terms of their aberration variance,” *J. Opt. Soc. Am.* **73**, 860 (1983).
9. R. Hafeez et al., “Forecasting wavefront corrections in an adaptive optics system,” *J. Astron. Telesc. Instrum. Syst.* **8**, 029003 (2022).
10. C. Z. Bond et al., “Adaptive optics with an infrared pyramid wavefront sensor at keck,” *J. Astron. Telesc. Instrum. Syst.* **6**, 039003 (2020).
11. S. Esposito et al., “Large binocular telescope adaptive optics system: new achievements and perspectives in adaptive optics,” *Proc. SPIE* **8149**, 814902 (2011).
12. S. Esposito et al., “Wavefront sensor design for the GMT natural guide star AO system,” *Proc. SPIE* **8447**, 84471L (2012).
13. J.-P. Véran et al., “Pyramid versus Shack–Hartmann: trade study results for the NFIRAOS NGS WFS,” in *Adapt. Opt. for Extremely Large Telesc. 4 – Conf. Proc.* (2015).
14. S. Esposito et al., “Non-common path aberration correction with nonlinear WFSs,” in *Adapt. Opt. for Extremely Large Telesc. IV (AO4ELT4)*, p. E36 (2015).
15. S. Esposito et al., “On-sky correction of non-common path aberration with the pyramid wavefront sensor,” *Astron. Astrophys.* **636**, A88 (2020).
16. V. Deo et al., “A modal approach to optical gain compensation for the pyramid wavefront sensor,” *Proc. SPIE* **10703**, 1070320 (2018).
17. V. Chambouleyron et al., “Pyramid wavefront sensor optical gains compensation using a convolutional model,” *Astron. Astrophys.* **644**, A6 (2020).
18. V. Chambouleyron et al., “Focal-plane-assisted pyramid wavefront sensor: enabling frame-by-frame optical gain tracking,” *Astron. Astrophys.* **649**, A70 (2021).
19. C. Weinberger et al., “Design and training of a deep neural network for estimating the optical gain in pyramid wavefront sensors,” in *Imaging and Appl. Opt. Congr. 2022 (3D, AOA, COSI, ISA, pcAOP)*, Optica Publishing Group, p. JF1B.6 (2022).
20. G. Agapito et al., “Non-modulated pyramid wavefront sensor. Use in sensing and correcting atmospheric turbulence,” *Astron. Astrophys.* **677**, A168 (2023).
21. R. Swanson et al., “Closed loop predictive control of adaptive optics systems with convolutional neural networks,” *MNRAS* **503**(2), 2944–2954 (2021).
22. A. P. Wong et al., “Machine learning for wavefront sensing,” *Proc. SPIE* **12185**, 121852I (2022).
23. Y. Nishizaki et al., “Deep learning wavefront sensing,” *Opt. Express* **27**, 240–251 (2019).
24. Y. He et al., “Deep learning wavefront sensing method for Shack–Hartmann sensors with sparse sub-apertures,” *Opt. Express* **29**, 17669–17682 (2021).
25. E. Vera, F. Guzmán, and C. Weinberger, “Boosting the deep learning wavefront sensor for real-time applications,” *Appl. Opt.* **60**, B119–B124 (2021).
26. C. A. Diez, F. Shao, and J. Bille, “Pyramid and Hartmann–Shack wavefront sensor with artificial neural network for adaptive optics,” *J. Mod. Opt.* **55**(4-5), 683–689 (2008).
27. G. Orban de Xivry et al., “Focal plane wavefront sensing using machine learning: performance of convolutional neural networks compared to fundamental limits,” *Mon. Not. R. Astron. Soc.* **505**, 5702–5713 (2021).
28. R. Landman and S. Y. Haffert, “Nonlinear wavefront reconstruction with convolutional neural networks for fourier-based wavefront sensors,” *Opt. Express* **28**, 16644 (2020).
29. A. Dosovitskiy et al., “An image is worth 16 × 16 words: transformers for image recognition at scale,” in *9th Int. Conf. Learn. Represent., ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, OpenReview.net (2021).
30. Z. Liu et al., “Swin transformer: hierarchical vision transformer using shifted windows,” in *IEEE/CVF Int. Conf. Comput. Vision (ICCV)*, pp. 9992–10002 (2021).

31. K. He et al., “Deep residual learning for image recognition,” in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 770–778 (2015).
32. Z. Liu et al., “A convnet for the 2020s,” in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 11966–11976 (2022).
33. L. Jolissaint, “Synthetic modeling of astronomical closed loop adaptive optics,” *J. Eur. Opt. Soc.* **5**, 10055 (2010).
34. G. Agapito, A. Puglisi, and S. Esposito, “PASSATA: object oriented numerical simulation software for adaptive optics,” *Proc. SPIE* **9909**, 99097E (2016).
35. J. Fitzsimmons et al., “GPI 2.0: design of the pyramid wave front sensor upgrade for GPI,” *Proc. SPIE* **11448**, 114486J (2020).
36. A. Madurowicz et al., “GPI 2.0: optimizing reconstructor performance in simulations and preliminary contrast estimates,” *Proc. SPIE* **11448**, 114482H (2020).
37. D. D. Thomas, J. E. Meyers, and S. M. Kahn, “Improving astronomy image quality through real-time wavefront estimation,” in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. Workshops (CVPRW)*, pp. 2076–2085 (2021).
38. I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *7th Int. Conf. Learn. Represent., ICLR 2019*, New Orleans, LA, May 6-9, 2019, OpenReview.net (2019).
39. P. Hickson, “Atmospheric and adaptive optics,” *Astron. Astrophys. Rev.* **22**, 76 (2014).
40. R. R. Selvaraju et al., “Grad-cam: visual explanations from deep networks via gradient-based localization,” *Int. J. Comput. Vision* **128**, 336–359 (2017).
41. T. Fel et al., “Xplique: a deep learning explainability toolbox,” in *Workshop Explainable Artif. Intell. for Comput. Vision (CVPR)* (2022).

Finn Archinuk is a master’s student at the University of Victoria of the Computational Biology Research and Analytics Lab. He received his BS degree in microbiology from the University of Victoria. He has a varied publication resume including secondary metabolism in poplar and galaxy quantization. He briefly worked as a machine learning researcher at the National Resource Council Canada, where this work originates.

Biographies of the other authors are not available.