

# Improving Astronomy Image Quality Through Real-time Wavefront Estimation

David Thomas<sup>1</sup>, Joshua Meyers<sup>2</sup>, Steven M. Kahn<sup>1</sup>

<sup>1</sup>Stanford University

<sup>2</sup>Lawrence Livermore National Laboratory

## Abstract

*We present a new framework for detecting telescope optics aberrations in real-time. The framework divides the problem into two subproblems that are highly amenable to machine learning and optimization. The first involves making local wavefront estimates with a convolutional neural network. The second involves interpolating the optics wavefront from all the local estimates by minimizing a convex loss function. We test our framework with simulations of the Vera Rubin Observatory. In a realistic mini-survey, the algorithm reduces the total magnitude of the optics wavefront by 66%, the optics PSF FWHM by 27%, and increases the Strehl ratio by a factor of 6. The resulting sharper images have the potential to boost the scientific payload for astrophysics and cosmology.*

## 1. Introduction

The signal to noise ratio of most astronomical analyses critically depends on image quality. To maintain optimal image quality during operation, wide-field telescopes deploy active optics systems, which sense aberrations in the optics and correct them in realtime. The primary challenge for these systems is distinguishing the correctable aberrations due to the optics from the dominant and intractable aberrations due to atmospheric turbulence. Here we present a new machine learning framework that is capable of extracting the optics aberrations and improving image quality.

While the immediate application is improving image quality in present and future ground-based telescopes, there are also emerging use cases in space. The simultaneous demands for higher quality images and lighter payloads from space telescopes make large foldable mirrors attractive. These large, lightweight mirrors are more susceptible to environmental disturbances and would benefit from active optics control. Prototypes are already being explored [18, 17, 7, 33, 31, 28]. In this work we demonstrate that our method is capable of improving image quality in the challenging ground-based environment. We suspect its performance will improve in the space environment where at-

mospheric turbulence is absent.

The upcoming ground-based Vera Rubin Observatory (Rubin) has a 3.5 degree field of view and high dimensional optical model that make it the ideal stress test for our framework [12, 1]. The large scientific community behind the Rubin Observatory has developed a mature suite of simulation codes [25, 21, 6] which we used to train and test our model in realistic scenarios. The unlimited supply of virtual telescopes and observations allowed us to assess our method in a more comprehensive range of conditions than would be possible with a real instrument.

The input to our model comes from 4 curvature wavefront sensors in the corner of the Rubin focal plane, shown in Figure 1. Each of these sensors is split into two half-chips which are purposefully offset out of focus. The stars that fall on these sensors produce large ring-like “donut” images. The goal of our algorithm is to constrain optics aberrations attributed to the entrance pupil from the intensity patterns in all the donut images throughout the observation, which can number into the thousands. One of the key challenges is interpolating the optics wavefront across the entire focal plane from donut images in four concentrated regions which collectively cover less than 2% of the total focal plane.

The key breakthrough in our work is the realization that the wavefront sensing problem can be divided into two subproblems that are highly amenable to machine learning and optimization. The first problem is to estimate the local wavefronts, characterized by 18 Zernike coefficients, from individual donut images. The second problem is to interpolate the global optics wavefront, characterized by 56 double Zernike coefficients, from all the local estimates by minimizing a simple convex loss function. The main contributions of this work are:

- 1) We present a new mathematical framework for extracting the optics wavefront across the field of view.
- 2) We demonstrate that a convolutional neural network can make reasonable estimates of the local wavefront from donut images.

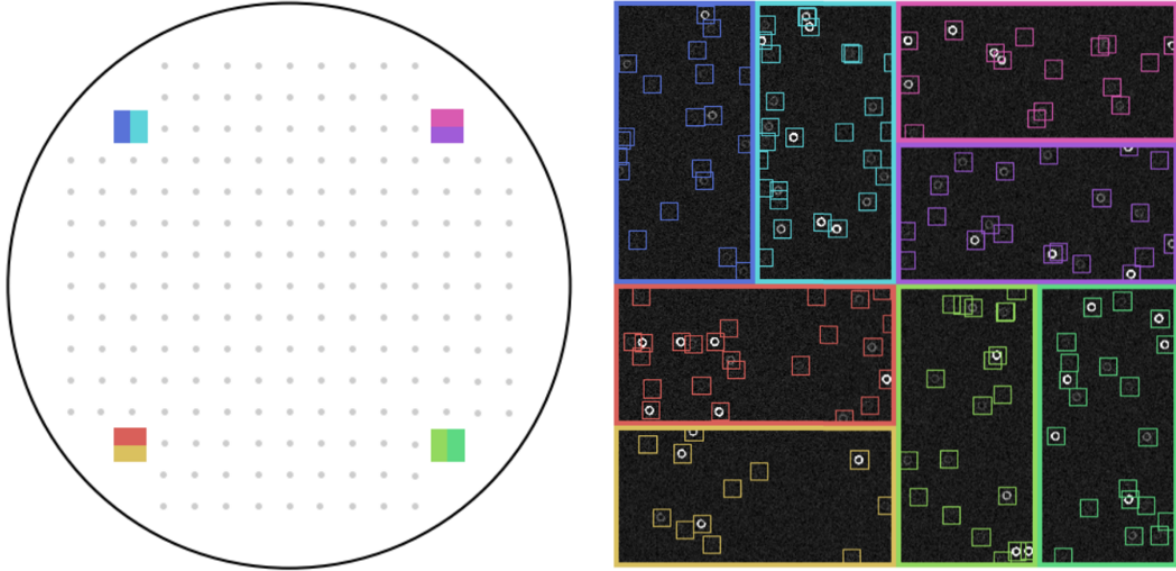


Figure 1. **The Rubin Observatory focal plane and wavefront sensor images.** *Left:* the Rubin Observatory focal plane. The eight solid boxes show the positions of the eight wavefront sensor half-chips and the gray dots show the centers of the remaining 189 science sensors. *Right:* example wavefront sensor images from each of the eight half-chips. The boundary color matches with the region they correspond to on the focal plane. The colored boxes show the donuts on each half-chip.

- 3) We show that fitting the global wavefront from a multitude of local wavefront estimates can suppress the atmospheric contribution.
- 4) We run our framework on a realistic mini-survey where it reduces the total magnitude of the optics wavefront by 66%, the optics PSF FWHM by 27%, and increases the Strehl by a factor of 6.

Finally, we emphasize that while this work focuses on the context of the Rubin Observatory, our framework extends to all wide-field telescopes with curvature wavefront sensors, and potentially to future space telescopes.

## 2. Related Work

The potential for neural networks to learn the non-linear mapping between intensity patterns and aberrations in the pupil plane was first recognized in 1990 [4]. Shortly afterwards this potential was realized as neural networks were deployed to detect turbulence induced distortion on the Multiple Mirror Telescope [27] and to detect aberrations in the primary mirror of the Hubble Space Telescope [5]. Others expanded this concept to predict more wavefront components [13], incorporate temporal history [20, 22], compare reconstruction methods [9], and better characterize atmospheric turbulence [32].

In the past decade, convolutional neural networks (CNN) [19] have re-emerged and spurred dramatic advances in computer vision [15, 29, 26, 11]. This has opened up new

possibilities for wavefront sensing in astronomy. CNNs can complement conventional iterative estimation approaches by providing good starting estimates [24]. They can even solve the full estimation problem in various scenarios [23, 10, 30]. Combining these networks with CMOS sensors opens up the possibility that guide stars could be used for adaptive optics on 2-4m class telescopes [3]. They also have the potential to sense and correct low-order wavefront aberrations for high-contrast astronomical imaging [2].

Most previous work focuses on sensing the dominant source of wavefront error, atmospheric turbulence or simple distortions, on short timescales, typically 10s of milliseconds. In this work, we study the harder problem of sensing subdominant sources of wavefront error, small alignment and mirror surface perturbations, from images integrated over an entire 15 second exposure. This presents new challenges such as how to best aggregate intensity information from throughout the field of view to suppress the dominant and spatially correlated atmospheric turbulence contribution.

## 3. Wavefront Estimation Framework

The optics wavefront  $W_{\text{opt}}$  is a function of two separate planes: the pupil plane parameterized by  $(u, v)$  and the focal plane parameterized by  $(x, y)$ . We use the double Zernike polynomial basis [16] to represent the optics

wavefront,

$$W_{\text{opt}}(u, v, x, y) = \sum_{i=1}^k \sum_{j=1}^m \beta_{ij} Z_i(u, v) Z_j(x, y) \quad (1)$$

where  $\beta_{ij}$  are the coefficients,  $Z_i$  are annular Zernike polynomials over the pupil, and  $Z_j$  are circular Zernike polynomials over the focal plane. The goal of wavefront sensing is to estimate these coefficients  $\beta_{ij}$  from the  $n$  donut images  $D_i$  positioned across the wavefront sensors (see Figure 1). Let the position of donut  $i$  be  $x_i, y_i$  and the defocus offset of the corresponding sensor be  $z_i$ . The wavefront sensing problem is to find  $f$  such that

$$\beta = f((D_1, x_1, y_1, z_1), \dots, (D_n, x_n, y_n, z_n)) \quad (2)$$

We break this into two subproblems.

### 3.1. Estimating Local Wavefronts

In the first subproblem, we estimate the total local wavefront  $w_{\text{tot}}(u, v)$  from donut  $D_i$  at position  $x_i, y_i, z_i$ . The intensity in the donut image is related to the total local wavefront by the Fraunhofer diffraction integral,

$$D \propto |\mathcal{F}\{P(u, v) \exp(2\pi i w_{\text{tot}}(u, v)/\lambda)\}|^2 \quad (3)$$

where  $\mathcal{F}$  is the Fourier transform,  $P(u, v)$  is the pupil mask, and  $\lambda$  is the wavelength. We represent the local wavefront in a basis of annular Zernike polynomials over the pupil, such that the total local wavefront for donut  $i$  at position  $x_i, y_i$  is

$$w_{\text{tot}}(u, v) = \sum_j \alpha_{ij} Z_j(u, v) \quad (4)$$

Convolutional neural networks (CNNs) are particularly well suited for processing images and learning nonlinear mappings. We develop a CNN  $\varphi$  to solve the inverse problem of estimating  $\alpha_i$  from  $(D_i, x_i, y_i, z_i)$ . In Section 4 we describe the implementation of this model in detail.

### 3.2. Interpolating the Optics Wavefront

In the second subproblem, we aggregate the local estimates from the first subproblem to constrain  $\beta$ . The total local wavefront at position  $x_i, y_i$  is related to the optics wavefront via

$$w_{\text{tot}}(u, v) = W_{\text{opt}}(u, v|x_i, y_i) + \epsilon(u, v|x_i, y_i) \quad (5)$$

where  $\epsilon$  represents the atmospheric turbulence contribution to the wavefront. Let  $\mathcal{Z}$  be defined such that  $\mathcal{Z}_{ij} = Z_j(x_i, y_i)$ . Then for  $i = 1, \dots, m$  we have

$$\alpha e_i = \mathcal{Z} \beta e_i + \epsilon \quad (6)$$

where  $e_i$  is the  $i$ th unit vector. Then combining the  $\alpha$  from the previous subproblem, and computing the corresponding  $\mathcal{Z}$ , allows us to solve for  $\beta$ ,

$$\beta = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^m \ell(\alpha e_i, \mathcal{Z} \beta e_i) \right\} \quad (7)$$

where  $\ell$  is a convex loss function. The full pseudocode is given in Algorithm 1.

---

**Algorithm 1:** estimates the optics wavefront from donuts images.

---

```

given image  $I \in \mathbb{R}^{N \times N}$ 
initialize local wavefront estimate  $\alpha \in \mathbb{R}^{n \times m}$ 
initialize global Zernike basis  $\mathcal{Z} \in \mathbb{R}^{n \times k}$ 
for donut  $i$  in  $1 \dots n$  do
     $D_i = \text{Crop}(I, x_i, y_i)$ 
     $\alpha[i, :] = \varphi(D_i, x_i, y_i, z_i)$ 
    for zernike  $j$  in  $1 \dots k$  do
         $\mathcal{Z}[i, j] = Z_j(x_i, y_i)$ 
    end
end
initialize optics wavefront  $\beta \in \mathbb{R}^{k \times m}$ 
for local Zernike  $i$  in  $1 \dots m$  do
     $\beta[:, i] = \operatorname{argmin}_{\beta[:, i]} \{ \ell(\alpha[:, i], \mathcal{Z} \beta[:, i]) \}$ 
end
return  $\beta$ 

```

---

The dominant source of error is the atmospheric turbulence contribution to the wavefront. This error is correlated on scales of arcminutes. By processing donuts with reasonable separation and between different wavefront sensors we are able to suppress this error by roughly a factor of  $1/\sqrt{n}$ .

There are two parameters of our algorithm that must be set based on the telescope: the number of Zernike coefficients to use for the pupil  $m$ , and the number of Zernike coefficients to use for the focal plane  $k$ . For the Rubin Observatory we use Zernikes  $Z_4$  through  $Z_{21}$  for the pupil plane. The first three coefficients do not impact image quality, so we exclude them. We use  $Z_1$  through  $Z_3$  for the focal plane. Our simulations show that 90% of the optics wavefront is contained in this truncated basis.

There are two benefits to dividing the wavefront estimation problem into these two subproblems that are worth highlighting. The first is the useful intermediate data products. The local wavefront coefficients  $\alpha$ , which are estimated in the first subproblem, are physically meaningful. Telescope operators can track them during operations and gain further insight into the system. This adds an additional layer of transparency and robustness.

The second benefit is that it makes deep learning approaches feasible. Deep neural networks must be trained on

large datasets to avoid overfitting. The input to the original problem is 4 wavefront sensor images, or up to thousands of donut images. The raytracing necessary to simulate even a single input sample is extremely computationally expensive. In our first subproblem however, the input is only a single donut image. The computation required to simulate a training sample is reduced by three orders of magnitude. This makes it possible to generate simulated datasets that are sufficient for training deep neural networks. In the next section, we highlight the power of these models.

## 4. Experiments and Analysis

### 4.1. Datasets

**Individual Donuts:** comprised of 600,147 simulated  $256 \times 256$  pixel Rubin Observatory donut images (see Figure 3). Each donut image has a corresponding position and true local wavefront label. This dataset is used to train the neural network to estimate the local wavefront.

The sources are chosen to be as realistic as possible. We started by drawing 5,000 r-band observations from a simulated Rubin Observatory observing schedule [6]. For each of these observations we queried the Gaia DR2 catalog for sources that would fall on the wavefront sensors [8]. Then we sampled 200 stars, with replacement, to simulate from each observation. We simulated an additional 100,147 blends - donut images with multiple stars overlapping - so that the network could learn to handle these complicated cases as well.

The simulations start by drawing photons from a black-body distribution based on the star temperature and magnitude from the catalog. We then propagate these through the atmosphere with the help of the GalSim Python package [25]. We use frozen phase screens to represent low spatial frequency turbulence and apply a randomly drawn *second kick* to account for high frequency turbulence. We use the Batoid Python raytracing package to generate Rubin telescope instances and trace the photons into the detector [21]. We randomly perturb 50 different degrees of freedom for each telescope instance in a manner that is consistent with what the telescope is expected to face in operations. These random perturbations are drawn from distributions that represent what we expect the Rubin telescope will face in operations. Finally, we use the GalSim to model the sensor readout. We incorporate custom functions throughout this pipeline to account for additional physical effects such as: chromatic seeing, differential chromatic refraction, charge diffusion in the sensors, bad pixels, and astrometric errors.

The local wavefront labels are calculated with Batoid. For each perturbed telescope instance, a grid of rays are traced from the entrance pupil through the corresponding field position to the exit pupil. Then Zernike polynomials are fit to the optical path differences between the rays.

These coefficients are the entries of the labels.

**Full Observations:** consists of 497 Rubin observations, each containing hundreds to thousands of simulated donuts. This dataset is used for testing the full framework, which ingests all the donuts in a given observation and produces an optics wavefront estimate.

All the donuts in an observation are simulated with the same atmosphere and sky background. The observations are drawn from Rubin Observatory scheduler simulations and the sources correspond to Gaia queries. Each star in the observation is simulated in the manner described above. In addition to the positions and local wavefront labels for each donut, there is also an optics wavefront for the entire observation. We use the batoid package to compute the relevant double Zernike coefficients from each telescope instance.

### 4.2. Architecture and Training

The input to the neural network is a  $256 \times 256$  pixel donut image and position  $r = (x, y, z)$ . The network, shown in Figure 2, has two components: an image component and a position component. The image component reduces the donut image to a 1024 dimensional vector. The position component combines this vector with the position input and estimates the 18 local wavefront coefficients.

The image component is comprised of eight repeated convolution blocks which decrease the tensor height and width and increase the depth, all by a factor of two. The convolution block has a convolution skip connection followed by the downsampling convolution. The position component is comprised of three linear layers which each reduce the dimensionality of the tensor. All convolution and linear layers are followed by ReLU and batchnorm layers, except the final linear layer.

Sources	Count		MSE	
	Train	Test	Train	Test
Stars	498,071	1,708	$4.5 \pm 3.2$	$4.4 \pm 3.5$
Blends	100,028	340	$9.5 \pm 20.0$	$9.6 \pm 22.0$

Table 1. Train and test results on stars and blends. The mean-squared-error (MSE) is in units of thousandths of waves.

We use the mean-squared-error (MSE) between the estimated and true wavefront coefficient as the loss function. We train the model for 8 epochs over 598,099 training samples with the Adam optimizer [14]. Then we evaluate the model on a held out test set of 2,048 samples. The MSE on the training and test sets is almost identical (see Table 1), which suggests it is not overfitting.

### 4.3. Local Wavefront Results

The wavefront estimates for a few representative samples are shown in Figure 3. The wavefront estimates are very close to the truth, even for test samples with MSE in



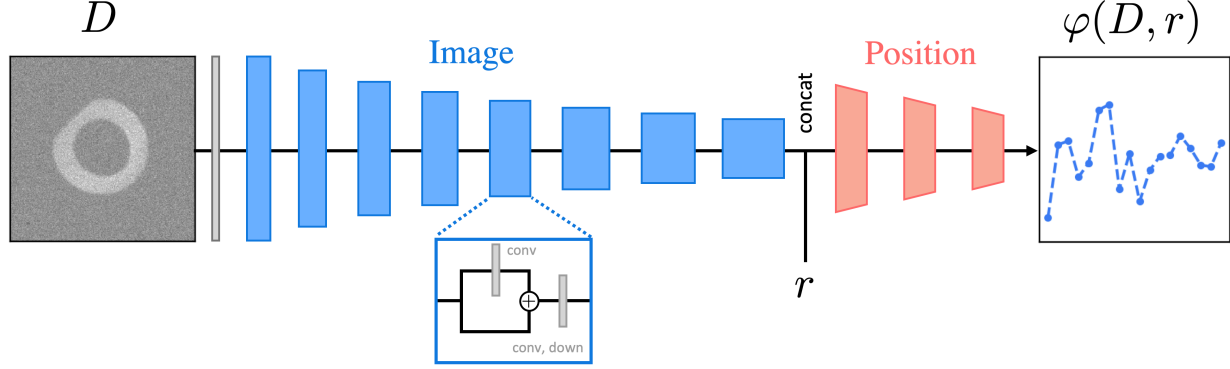


Figure 2. **Network architecture.** The layers of the image component, which reduces the donut image into a 1024 length vector, are in blue and the layers of the position component, which concatenates the position and makes the wavefront estimate, are in red.

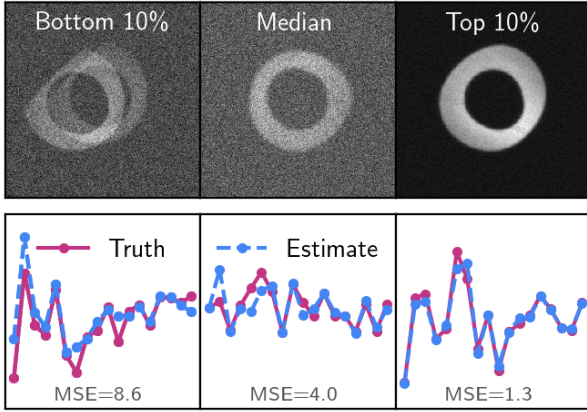


Figure 3. **Three donut images and their wavefront estimates.** *Top:* three donut images drawn from the bottom 10%, median, and top 10% of the MSE distribution respectively. *Bottom:* the three corresponding true and estimated local wavefronts. The y-axis goes from -0.5 waves to 0.5 waves.

the bottom decile. Going from the bottom decile to the top decile leads to a reduction in the MSE of almost an order of magnitude. In Subsection 4.4, we explore different ways to take advantage of this large discrepancy when interpolating the full optics wavefront.

We see significant degradation in the accuracy of estimates on donut images with three or more overlapping neighbors. We suspect this may be due to the distribution of our training set, where the frequency of blended donuts with  $n$  neighbors decreases exponentially with  $n$ . We also see that brighter sky backgrounds, significant vetting near the far corners of the sensors, and the atmospheric seeing are all weakly correlated with decreasing accuracy. This is in line with expectations.

Along with studying the output of the network, we also

probe some of its intrinsic behavior. This serves as a sanity check and can give us humans clues as to where further improvements may lie. In order to uncover what donut features the network is paying most attention to, we take gradients of the norm of the estimates with respect to the pixel intensity values, or

$$\nabla_D ||\varphi(D, r)||_2 \quad (8)$$

The entries of this gradient with the largest magnitude are the pixels that the estimate is most sensitive to.

Two notable trends emerge from this analysis (shown in Figure 4). The first trend is that the estimates are most sensitive to the edges of the donuts. This raises the question of whether only having the shape of the boundary, perhaps extracted with a simple thresholding algorithm, is sufficient to achieve comparable accuracy. The second finding is on the blended donuts. The network exhibits bimodal behavior. Either it ignores the overlapping regions of the target donut, or it uses information from all the donuts. It seems as if the network is making a binary decision as to which way to go. It would be interesting to learn what factors contribute to this decision and see whether these insights would be relevant to other de-blending problems in astronomy.

In a similar vein, we also took a single donut, and calculated the intensity changes that would move the estimate towards a specific zernike coefficient. We take the gradient of the norm of the estimate minus the target zernike coefficient with respect to the pixel intensity values, or

$$\nabla_D ||\varphi(D, r) - e_i||_2 \quad (9)$$

where  $e_i$  is the  $i$ th unit vector. The results for three example coefficients are shown in Figure 5. The similarity between the intensity changes and the Zernike polynomials is striking, especially given that the network has not been explicitly trained to learn these patterns. It demonstrates that the network is engaging in higher order learning, where it

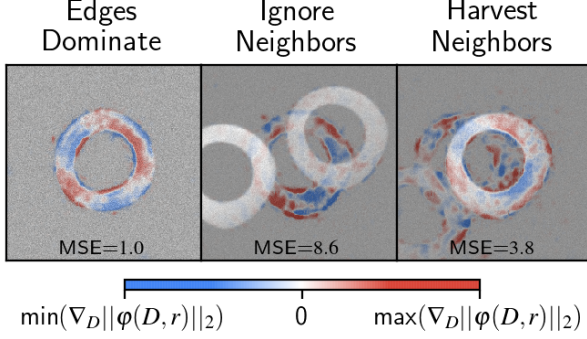


Figure 4. **Saliency maps.** Three input donut images with a color overlay showing where the network is focusing.

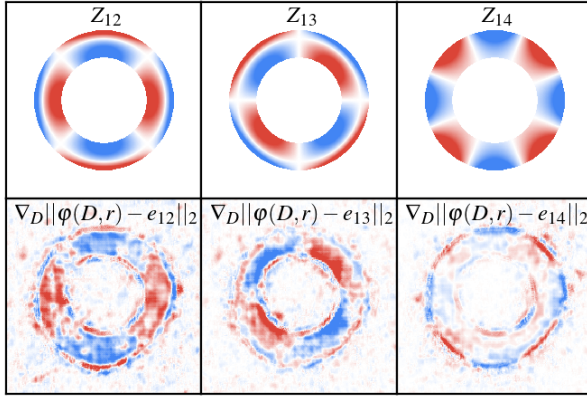


Figure 5. **Network learns Zernike patterns.** *Top:* the Zernike polynomials  $Z_{12}, Z_{13}, Z_{14}$ . *Bottom:* the intensity changes that would most rapidly move the estimate to the corresponding Zernike coefficient.

goes beyond memorizing to learning general features of the problem space.

We showed a neural network is capable of making accurate local wavefront estimates for the first stage of our algorithm. It is the responsibility of the second stage to aggregate all these estimates and estimate the full optics wavefront.

#### 4.4. Optics Wavefront Results

The second stage uses the local wavefront estimates from donuts in the four wavefront sensor images to interpolate the optics wavefront across the entire focal plane. We break the interpolation down into three steps that are reminiscent of a standard data query: select, reduce, and fit. The select step decides which donuts and corresponding local estimates to use in the interpolation. The reduce step reduces these estimates across a wavefront sensor. The fit step fits the local coefficients to a global Zernike basis based on the

Select	Median Reduce	Fit	% Samples Improved	Relative Residual
Stars and Blends	✓ ✓	$\ell_1$	99.6	$0.48 \pm 0.13$
		$\ell_2$	99.8	$0.49 \pm 0.12$
		$\ell_h$	100.0	$0.48 \pm 0.12$
		$\ell_1$	97.8	$0.67 \pm 0.14$
		$\ell_2$	100.0	$0.46 \pm 0.12$
Stars	✓ ✓	$\ell_1$	99.8	$0.44 \pm 0.11$
		$\ell_2$	100.0	$0.43 \pm 0.10$
		$\ell_h$	100.0	$0.43 \pm 0.10$
		$\ell_1$	97.2	$0.64 \pm 0.14$
		$\ell_2$	99.8	$0.41 \pm 0.11$
Brightest Stars	✓ ✓	$\ell_1$	99.6	$0.37 \pm 0.13$
		$\ell_2$	100.0	$0.34 \pm 0.12$
		$\ell_h$	100.0	$0.34 \pm 0.12$
		$\ell_1$	97.2	$0.60 \pm 0.16$
		$\ell_2$	100.0	$0.35 \pm 0.12$
Labels		$\ell_1$	100.0	$0.13 \pm 0.05$
		$\ell_2$	100.0	$0.06 \pm 0.02$
		$\ell_h$	100.0	$0.08 \pm 0.04$

Table 2. **Optics wavefront results on different select-reduce-fit variations.** Each row contains the results for a different combination of select, reduce, and fit steps. The penultimate column contains the percentage of the number of samples where the residual improved. The final column contains the relative residual: the total magnitude of the residual divided by the total magnitude of the true wavefront. The best variation on neural network estimates is highlighted in blue. The best variation on the true label estimates is highlighted in gold.

provided loss function.

We explore multiple variations in each of these steps to find which combination works best. We explore selecting donuts from all the sources (stars and blends), from only the non-blended stars (stars), the non-blended 10 brightest stars per chip (brightest stars), and using the true labels (labels). The results on the true labels provide a sanity check and bound the performance we can expect to achieve with alternatives.

We also explore two variations in the reduce step. Either we make no changes to estimates, or we take the median of the estimates on each chip. Then in the fit stage we fit against 4 points corresponding to the four sensors.

We explore three different fitting strategies. The  $\ell_1$ , or absolute loss, is convex and can be found with an iterative optimization algorithm. The  $\ell_2$ , or least squares loss, has an analytic solution. This has the added benefit of making error propagation analytic as well. Finally, the Huber loss  $\ell_h$  is similar to the  $\ell_1$  for samples with small error but scales like  $\ell_2$  for large error. Thus it is similar to  $\ell_2$  but less sensitive to outliers. The results of these variations are shown in Table 2.

State	Position	FWHM	Strehl
Original	Center	$0.288 \pm 0.034$	$0.093 \pm 0.39$
Corrected	Center	$0.211 \pm 0.005$	$0.555 \pm 0.207$
Original	Corner	$0.314 \pm 0.045$	$0.074 \pm 0.32$
Corrected	Corner	$0.215 \pm 0.009$	$0.400 \pm 0.184$

Table 3. **The optics PSF FWHM and Strehl ratios before (original) and after (corrected) subtracting out the optics wavefront estimated by our framework.**

The comparisons of interest are between the true optics wavefront and the residual optics wavefront, where the residual is the true wavefront minus our estimated wavefront. The residual wavefront is smaller than the original wavefront for all the samples in the majority of select-reduce-fit variations. This consistency makes it hard to argue against deploying it.

This experiment also taught us that more data is not always better. Ignoring the blended donuts leads to a clear improvement in performance. So does ignoring all but the brightest stars. This suggests that we should prioritize making accurate predictions on the best donuts, perhaps at the expense of making consistent estimates on all the donuts. It also may have consequences for doing wavefront sensing in very crowded fields where there may not be non-blended donuts.

We can also draw conclusions about the variations. Taking the median and fitting with the  $\ell_1$  norm appears to discard too much information. We also see that the benefit of using median with the  $\ell_2$  norm goes away as the select becomes more selective. This is likely because the outliers, which the median reduce suppresses, get filtered and are no longer an issue. The  $\ell_h$  norm also seems to do comparatively well on stars and blends, but loses this advantage on the more selective brightest stars selection. We conclude using the  $\ell_2$  norm, with no median reduce, to fit the brightest stars, is the best variation.

The next step is to take this model and measure the repercussions of subtracting its estimate from the true wavefront on both the point spread function full width half maximum (FWHM) and Strehl ratio. We compute these by calculating the local wavefronts at the center and a corner of the focal plane. Then we take the fourier transform of the resulting pupil plane aberration to get the point spread function. The results for the original and corrected wavefronts are shown in Table 3.

The optics PSF FWHM decreases considerably, especially when compared to the standard deviation of the original. The Strehl ratio increases in an even more extreme manner. Figure 6 shows how the improvements to the optics PSF from our method could improve classic Hubble Telescope images in the absence of other significant PSF contributions.

## 5. Conclusion

## References

- [1] P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, S. Bailey, et al. Lsst science book, version 2.0. 2009. 1
- [2] Gregory Allan, Iksung Kang, Ewan S. Douglas, George Barbastathis, and Kerri Cahoy. Deep residual learning for low-order wavefront sensing in high-contrast imaging systems. *Optics Express*, 28(18):26267, Aug. 2020. 2
- [3] Torben Andersen, Mette Owner-Petersen, and Anita Enmark. Neural networks for image-based wavefront sensing for astronomy. *Optics Letters*, 44(18):4618, Sept. 2019. 2
- [4] J. R. P. Angel, P. Wizinowich, M. Lloyd-Hart, and D. Sandler. Adaptive optics for array telescopes using neural-network techniques. *Nature*, 348(6298):221–224, Nov. 1990. 2
- [5] Todd K. Barrett and David G. Sandler. Artificial neural network for the determination of Hubble Space Telescope aberration from stellar images. *Applied Optics*, 32(10):1720–1727, Apr. 1993. 2
- [6] Francisco Delgado and Michael A. Reuter. The LSST Scheduler from design to construction. In *Observatory Operations: Strategies, Processes, and Systems VI*, volume 9910 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 991013, July 2016. 1, 4
- [7] Nicholas Devaney, Fiona Kenny, Alexander V. Goncharov, Matthias Goy, and Claudia Reinlein. Development of a prototype active optics system for future space telescopes. *Applied Optics*, 57(22):E101, Aug. 2018. 1
- [8] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, et al. Gaia Data Release 2. Summary of the contents and survey properties. *Astronomy and Astrophysics*, 616:A1, Aug. 2018. 4
- [9] Hong Guo, Nina Korablinova, Qiushi Ren, and Josef Bille. Wavefront reconstruction with artificial neural networks. *Optics Express*, 14(14):6456–6462, July 2006. 2
- [10] Hongyang Guo, Yangjie Xu, Qing Li, Shengping Du, Dong He, Qiang Wang, and Yongmei Huang. Improved machine learning approach for wavefront sensing. *Sensors*, 19(16), 2019. 2
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2
- [12] Ž. Ivezić, J. A. Tyson, E. Acosta, R. Allsman, S. F. Anderson, J. Andrew, J. R. P. Angel, T. S. Axelrod, J. D. Barr, A. C. Becker, et al. Lsst: from science drivers to reference design and anticipated data products. 2008. 1
- [13] Mark B. Jorgenson and George J. Aitken. Neural network prediction of turbulence-induced wavefront degradations with applications to adaptive optics. In Firooz A. Sadjadi, editor, *Adaptive and Learning Systems*, volume 1706 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 113–121, Aug. 1992. 2
- [14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun,

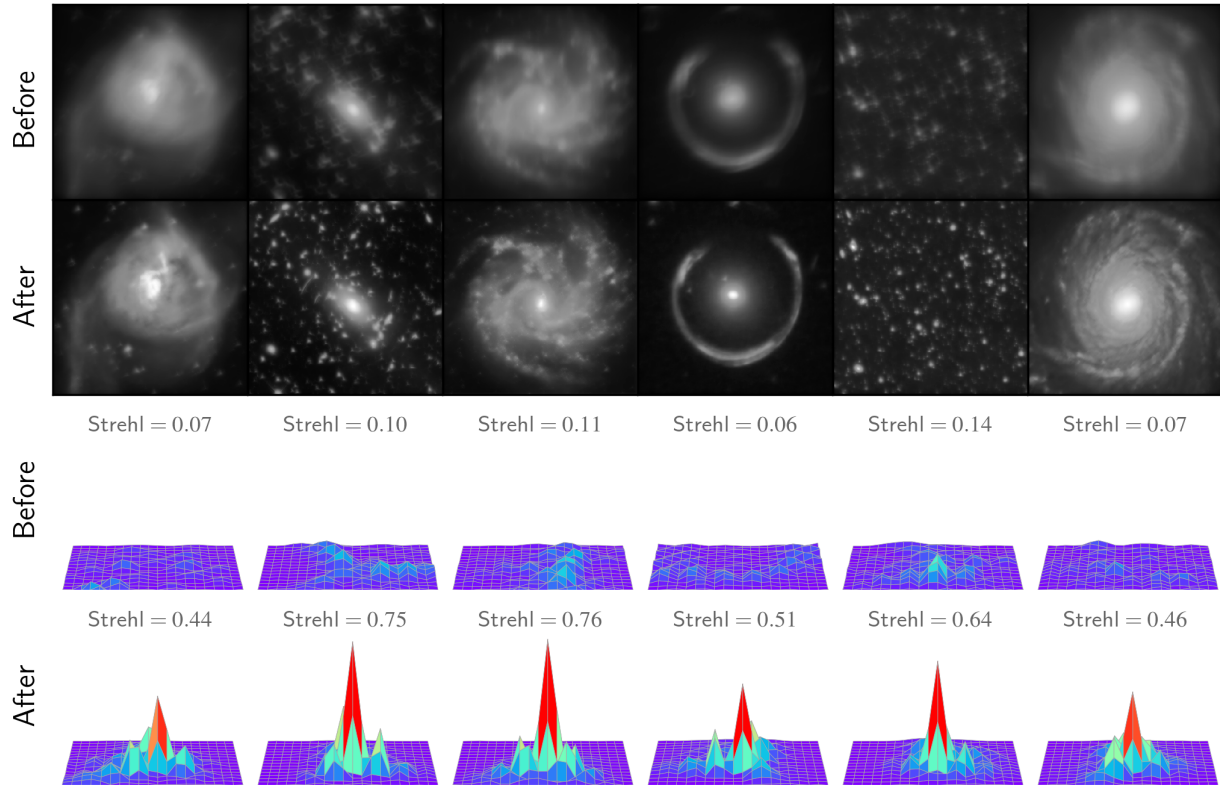


Figure 6. **Hubble Telescope images before (original) and after (corrected) subtracting the optics wavefront estimated by our framework.** The images are treated as if their angular extent was 3 arcseconds. The corresponding optics PSFs are displayed in the rows below.

- editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 4
- [15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 10971105, Red Hook, NY, USA, 2012. Curran Associates Inc. 2
- [16] Ivo W Kwee and Joseph J M Braat. Double zernike expansion of the optical aberration function. *Pure and Applied Optics: Journal of the European Optical Society Part A*, 2(1):21–32, jan 1993. 2
- [17] Marie Laslandes, Claire Hourtoule, Emmanuel Hugot, Marc Ferrari, Christophe Devilliers, Arnaud Liotard, Céline Lopez, and Frédéric Chazallet. Last results of MADRAS, a space active optics demonstrator. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 10564 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1056413, Nov. 2017. 1
- [18] Marie Laslandes, Emmanuel Hugot, Marc Ferrari, Claire Hourtoule, Christian Singer, Christophe Devilliers, Céline Lopez, and Frédéric Chazallet. Mirror actively deformed and regulated for applications in space: design and performance. *Optical Engineering*, 52:091803, Sept. 2013. 1
- [19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2
- [20] M. Lloyd-Hart and P. McGuire. Spatio-temporal prediction for adaptive optics wavefront reconstructors. In *European Southern Observatory Conference and Workshop Proceedings*, volume 54 of *European Southern Observatory Conference and Workshop Proceedings*, page 95, Jan. 1996. 2
- [21] Joshua Meyers. batoid. <https://github.com/jmeyers314/batoid>, 2020. Accessed: 2020-04-07. 1, 4
- [22] Dennis A. Montera, Byron M. Welsh, Michael C. Roggemann, and Dennis W. Ruck. Prediction of wave-front sensor slope measurements with artificial neural networks. *Applied Optics*, 36(3):675–681, Jan. 1997. 2
- [23] Yohei Nishizaki, Matias Valdivia, Ryoichi Horisaki, Katsuhisa Kitaguchi, Mamoru Saito, Jun Tanida, and Esteban Vera. Deep learning wavefront sensing. *Optics Express*, 27(1):240, Jan. 2019. 2
- [24] Scott W. Paine and James R. Fienup. Machine learning for improved image-based wavefront sensing. *Optics Letters*, 43(6):1235, Mar. 2018. 2



- [25] B. T. P. Rowe, M. Jarvis, R. Mandelbaum, G. M. Bernstein, J. Bosch, M. Simet, J. E. Meyers, T. Kacprzak, R. Nakajima, J. Zuntz, H. Miyatake, J. P. Dietrich, R. Armstrong, P. Melchior, and M. S. S. Gill. GALSIM: The modular galaxy image simulation toolkit. *Astronomy and Computing*, 10:121–150, Apr. 2015. [1](#), [4](#)
- [26] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision*, 115(3):211252, Dec. 2015. [2](#)
- [27] D. G. Sandler, T. K. Barrett, D. A. Palmer, R. Q. Fugate, and W. J. Wild. Use of a neural network to control an adaptive optics system for an astronomical telescope. *Nature*, 351(6324):300–302, May 1991. [2](#)
- [28] Damien Sucher, Guillaume Butel, Guillaume Briche, Jean-François Blanc, Arnaud Liotard, Marc Bernot, Mikael Carlván, Aurélien Suau, Nisrine Louh, Lauriane Galtier, Sebastien Guionie, Thierry Viard, Stéphanie Behar-Lafenetre, Fabrice Champandard, Jean-Bernard Ghibaudo, and Vincent Costes. Active optics for space telescopes. In *Astronomical Optics: Design, Manufacture, and Test of Space and Ground Systems II*, volume 11116 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1111611, Sept. 2019. [1](#)
- [29] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13*, page 25532561, Red Hook, NY, USA, 2013. Curran Associates Inc. [2](#)
- [30] David Thomas, Joshua Meyers, and Steven M. Kahn. Wide-field wavefront sensing with convolutional neural networks and ordinary least squares. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 11448 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 114484H, Dec. 2020. [2](#)
- [31] T. Viard, J. F. Blanc, C. Devilliers, F. Champandard, B. Bailly, F. Falzon, J. B. Ghibaudo, D. Sucher, G. Briche, V. Costes, and C. Du Jeu. Active optics for next generation of space observation instruments. In *International Conference on Space Optics & ICSO 2018*, volume 11180 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1118008, July 2019. [1](#)
- [32] S. J. Weddell and R. Y. Webb. Reservoir Computing for Prediction of the Spatially-Variant Point Spread Function. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):624–634, Nov. 2008. [2](#)
- [33] Pingwei Zhou, Dongxu Zhang, Guang Liu, and Changxiang Yan. Development of space active optics for a whiffletree supported mirror. *Applied Optics*, 58(21):5740, July 2019. [1](#)