Improving Astronomy Image Quality Through Real-time Wavefront Estimation

David Thomas[1], Joshua Meyers[2], Steven M. Kahn[1]
[1]Stanford University
[2]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 throughout their operation, modern telescopes deploy active optics systems, which sense aberrations in the wavefront and correct them in real-time. For narrow field systems, it is possible to correct wavefront aberrations due to both the atmosphere and the optics. For wide-field systems, however, the atmospheric contributions cannot be corrected since they vary considerably over the field of view. Nevertheless, contributions due to variations in the optics can be corrected if the components of the individual elements can be discerned. 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 [16, 15, 6, 35, 31, 29]. 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 atmospheric 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 [10, 1]. The large scientific community behind the Rubin Observatory has developed a mature suite of simulation codes [25, 19, 5] which we used to train and test our model in realistic scenarios. The unlimited supply of simulated observations allows 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 four 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 both above and below focus. The stars that fall on these sensors produce large ring-like *donut images* due to the annular shape of the Rubin primary mirror. The goal of our algorithm is to constrain optics aberrations attributed to the entrance pupil from the intensity patterns in all the donut images in an 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 that 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.

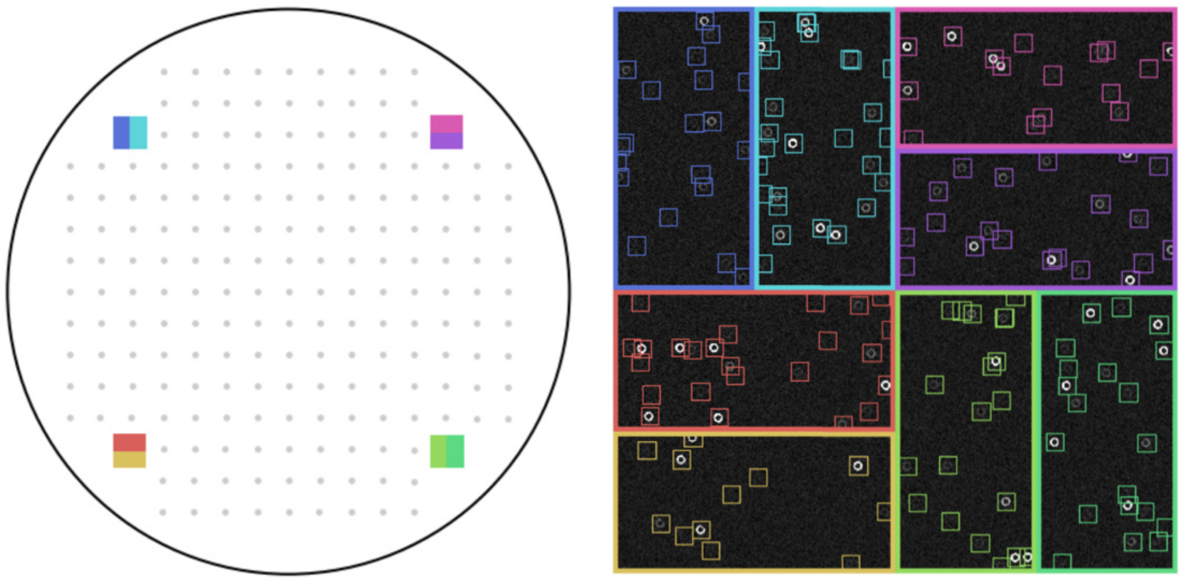

Figure 1. **The Rubin Observatory focal plane and wavefront sensor images.** *Left:* the Rubin Observatory focal plane. The eight solid boxes show the positions of the eight wavefront sensor half-chips and the gray dots show the centers of the remaining 189 science sensors. *Right:* example wavefront 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.

2) We demonstrate that a convolutional neural network can make reasonable estimates of the local wavefront from donut images.

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 point spread function full width at half maximum (PSF FWHM) by 27%, and increases the Strehl ratio by a factor of 6.

Finally, we emphasize that while this work focuses on 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 [3]. 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 [4]. Others expanded this concept to predict more wavefront components [11], incorporate temporal history [18, 20], compare reconstruction methods [8], and better characterize atmospheric turbulence [32].

In the past decade, convolutional neural networks (CNNs) [17] have re-emerged and spurred dramatic advances in computer vision [13, 30, 26, 9]. This has created new possibilities for wavefront sensing in astronomy. In [22], the authors created a CNN that could estimate the wavefront from a single PSF image. They used these estimates as initial starting points in a gradient-based optimization and showed this was superior to using random samples. [21] showed wavefront sensing performance could be improved by introducing a preconditioner to broaden the PSF and create more intensity structure for the neural network to exploit. This brings up interesting new design possibilities for wavefront sensors. While conventional Lyot-based low order wavefront sensing methods have a limited dynamic range due to their linear recovery, [2] showed that a CNN can extend the aberration range over which the wavefront can be estimated by an order of magnitude.

Previous work on machine learning based wavefront sensing focuses on sensing the full wavefront aberration. Here we focus on sensing the optics wavefront, across the field of view, in the midst of the dominant atmospheric contribution. This problem presents new challenges, such as how to best aggregate intensity information from throughout the field of view to suppress the spatially correlated error due to the turbulence contribution.

Xin et. al. [33] designed an iterative algorithm that extends conventional curvature sensing [24] to estimate the Rubin optics wavefront at four field positions. While this method has been shown to be in good agreement with the estimates from the Dark Energy Camera active optics system [34], there are some noteworthy limitations. First, each iteration of this algorithm involves complicated image transformations based on the wavefront estimate from the previous iteration, and solving a PDE. There are no guarantees that each step is improving the wavefront estimate, or that the full algorithm will converge. Second, the path-dependent nature of the iterations makes it difficult to characterize the error and benchmark performance. Third, it takes around 10 seconds to process a single donut. Here we present a transparent new approach that leverages the power of machine learning and optimization. Our approach is comprised of only two steps, is easy to characterize, and can process each donut image in 6 *milliseconds*.

3. Wavefront Estimation Framework

The optics wavefront W_{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 [14] to represent the optics wavefront,

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

where β_{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 β_{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

$$\beta = f((D_1, x_1, y_1, z_1), \ldots, (D_n, x_n, y_n, z_n)) \tag{2}$$

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 Fraunhoffer diffraction integral,

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

where $\mathcal{F}$ is the Fourier transform, $P(u, v)$ is the pupil mask, and λ 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

$$w_{\text{tot}}(u, v) = \sum_{j} \alpha_{ij} Z_j(u, v) \tag{4}$$

Convolutional neural networks (CNNs) are particularly well suited for processing images and learning nonlinear mappings. We develop a CNN φ to solve the inverse problem of estimating α_{ij} for $j = 1 \ \ldots m$ 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 β. The total local wavefront at position x_i, y_i is related to the optics wavefront via

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

where ϵ 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 $i = 1, \ldots, m$ we have

$$\alpha e_i = \mathcal{Z} \beta e_i + \epsilon \tag{6}$$

where e_i is the ith unit vector. Then combining the α from the previous subproblem, and computing the corresponding $\mathcal{Z}$, allows us to solve for β,

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

where ℓ is a convex loss function. Algorithm 1 shows the psuedocode.

Algorithm 1: estimates the optics wavefront from donut images.

```
given image I ∈ ℝ^{N×N}
initialize local wavefront estimate α ∈ ℝ^{n×m}
initialize global Zernike basis Z ∈ ℝ^{n×k}
for donut i in 1...n do
    D_i = Crop(I, x_i, y_i)
    α[i, :] = φ(D_i, x_i, y_i, z_i)
    for zernike j in 1...k do
        Z[i, j] = Z_j(x_i, y_i)
    end
end
initialize optics wavefront β ∈ ℝ^{k×m}
for local Zernike i in 1...m do
    β[:, i] = argmin_{β[:,i]} {ℓ(α[:, i], Zβ[:, i])}
end
return β
```

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

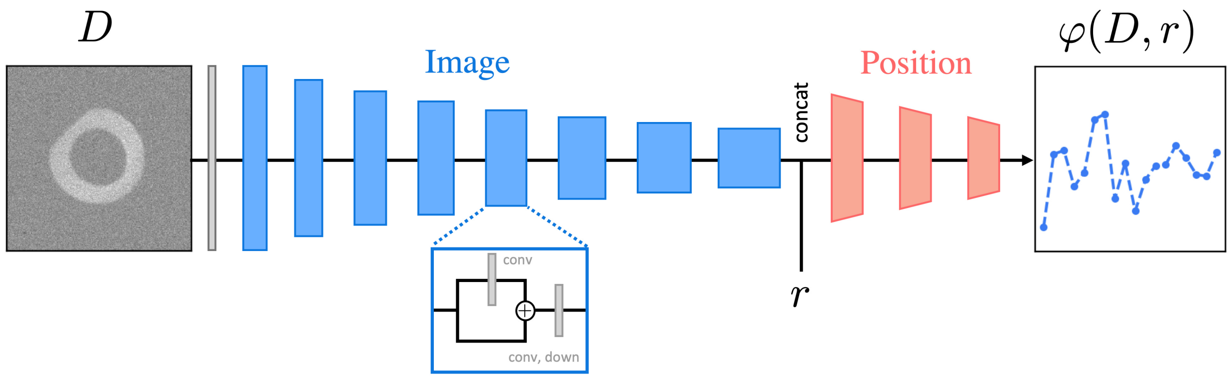

Figure 2. **Network architecture.** The layers of the image component, which reduces the donut image into a 1024 length vector, are in blue and the layers of the position component, which concatenates the position and makes the wavefront estimate, are in red.

are able to suppress this error by roughly a factor of $1/\sqrt{n}$ where n is the number of donuts used.

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 truncate the basis at Z_{21}, a convention set by [33], as the higher order terms have very small coefficients in practice. 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 α, 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 four wavefront sensor images, or up to thousands of donut images. The raytracing necessary to simulate even a single input sample is computationally expensive. In our first subproblem however, the input is only a single donut image. This reduces the computation required to produce a training sample by three orders of magnitude and 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

4.1.1 Donut Training, Validation, and Test Sets

This dataset is used to train the neural network to estimate the local wavefront. Each sample consists of a 256×256 pixel Rubin donut image (see Figure 3), the donut position, and a true local wavefront label. 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 [5]. For each of these observations we queried the Gaia DR2 catalog for sources that would fall on the wavefront sensors [7]. 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 training, validation, and test sets are comprised of 498,071 stars and 100,028 blends, 220 stars and 36 blends, and 1,708 stars and 340 blends respectively.

The simulations start by drawing photons from a blackbody 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 [23]. We use the Batoid Python raytracing package to generate Rubin telescope instances and trace the photons into the detector [19]. We randomly perturb 50 different degrees of freedom for each telescope instance. 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.

4.1.2 Mini-Survey Test Set

This dataset is used for testing the full framework. Each sample corresponds to an observation as opposed to an individual donut. Each sample consists of all the donut images and positions in the observation, plus the full optics wavefront. We used 497 Rubin observations, each containing hundreds to thousands of simulated donuts.

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 same manner as the donuts dataset described above. For each observation, we used the batoid framework to compute the optics wavefront double Zernike coefficients for the perturbed telescope instance.

4.2. Architecture and Training

The input to the neural network is a 256×256 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 consists 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 consists 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.

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 the donut training set with the Adam optimizer [12] and a batch-size of 64. Every 200 batches, we evaluate the MSE of the model on the donut validation set, and keep the best model. After the model is finalized, we evaluate it on the donut test set. The training and test set MSEs are 4.5 ± 3.2 and 4.4 ± 3.5 thousandths of waves on stars respectively, and 9.5 ± 20.0 and 9.6 ± 22.0 thousandths of waves on blends respectively. The performance on the training and test sets is almost identical, which suggests our model is not overfitting.

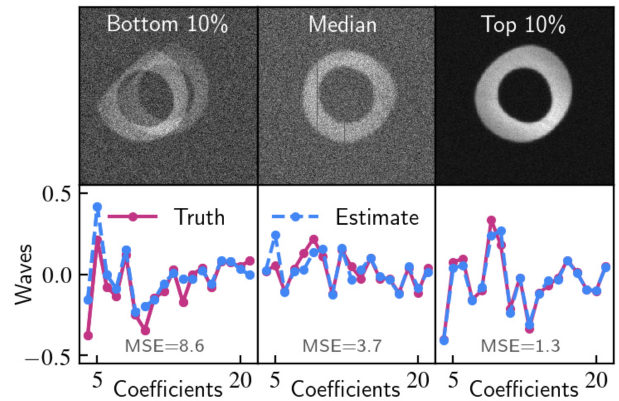

Figure 3. **Three donut images and their wavefront estimates**. *Top:* three donut images drawn from the bottom 10%, median, and top 10% of the MSE distribution respectively. *Bottom:* the three corresponding true and estimated local wavefronts. The y-axis goes from -0.5 waves to 0.5 waves.

4.3. Local Wavefront Results

The wavefront estimates for three representative samples are shown in Figure 3. The wavefront estimates are very close to the truth, even for test samples with MSE in the bottom decile. Going from the bottom decile to the top decile leads to a reduction in the MSE by 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 vignetting 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

$$\nabla_D ||\varphi(D,r)||_2 \tag{8}$$

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

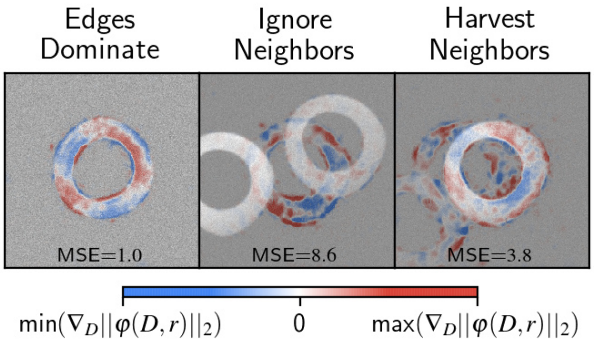

Figure 4. **Saliency maps.** Three input donut images with a color overlay showing where the network is focusing.

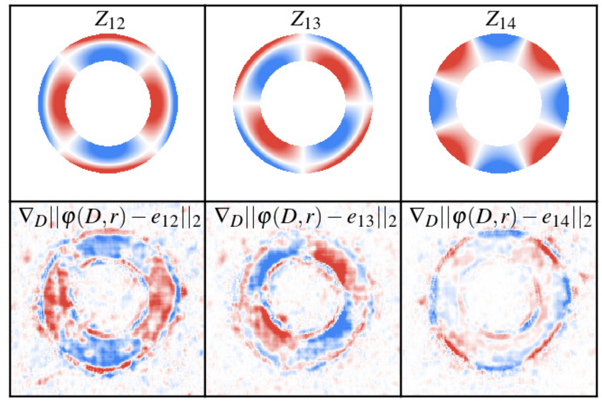

Figure 5. **Network learns Zernike patterns.** *Top:* the Zernike polynomials Z_{12}, Z_{13}, Z_{14}. *Bottom:* the intensity changes needed to move the estimate towards the corresponding Zernike coefficient.

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

$$\nabla_D||\varphi(D,r) - e_i||_2 \tag{9}$$

where e_i is the ith 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 is 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. Full 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 describe the interpolation in 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, which is typically skipped, reduces these estimates across a wavefront sensor. The fit step fits the local coefficients to a global Zernike basis based on the provided loss function.

We examine 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 analyze two variations in the reduce step. Either we make no changes to estimates and effectively skip this step, or we take the median of the estimates on each chip. In the median case, we would then fit against the four points corresponding to the four sensors in the fit step.

We explore three different fitting strategies. The ℓ_1, or absolute loss, is convex and can be found with an iterative optimization algorithm. The ℓ_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 ℓ_h is similar to the ℓ_2 for samples with small error but scales like ℓ_1 for large error. Thus it is similar to ℓ_2 but less sensitive to outliers.

The results of these variations, applied to the local wavefront estimates from the neural network, applied to the mini-survey test set, are shown in Table 1. We compare the true optics wavefront and the residual optics wavefront,

Select	Median Reduce	Fit	% Samples Improved	Relative Residual
Stars and Blends		ℓ_1	99.6	0.48 ± 0.13
		ℓ_2	99.8	0.49 ± 0.12
		ℓ_h	100.0	0.48 ± 0.12
	✓	ℓ_1	97.8	0.67 ± 0.14
	✓	ℓ_2	100.0	0.46 ± 0.12
Stars		ℓ_1	99.8	0.44 ± 0.11
		ℓ_2	100.0	0.43 ± 0.10
		ℓ_h	100.0	0.43 ± 0.10
	✓	ℓ_1	97.2	0.64 ± 0.14
	✓	ℓ_2	99.8	0.41 ± 0.11
Brightest Stars		ℓ_1	99.6	0.37 ± 0.13
		ℓ_2	100.0	0.34 ± 0.12
		ℓ_h	100.0	0.34 ± 0.12
	✓	ℓ_1	97.2	0.60 ± 0.16
	✓	ℓ_2	100.0	0.35 ± 0.12
Labels		ℓ_1	100.0	0.13 ± 0.05
		ℓ_2	100.0	0.06 ± 0.02
		ℓ_h	100.0	0.08 ± 0.04

Table 1. **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.

where the residual is the full true optics wavefront minus the full estimated optics wavefront. The residual wavefront is smaller than the original wavefront for all the samples in the majority of select-reduce-fit variations. The consistency of the improvements makes our method an attractive candidate for deployment.

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 wavefront sensing in crowded fields where almost all of the donuts are blended.

We can also draw conclusions about the variations. Taking the median and fitting with the ℓ_1 norm appears to discard too much information. We also see that the benefit of using median with the ℓ_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 ℓ_h norm also seems to do comparatively well on stars and blends, but loses this advantage on

State	Position	FWHM	Strehl
Before	Center	0.288 ± 0.034	0.093 ± 0.39
After	Center	0.211 ± 0.005	0.555 ± 0.207
Before	Corner	0.314 ± 0.045	0.074 ± 0.32
After	Corner	0.215 ± 0.009	0.400 ± 0.184

Table 2. **Improvement in the optics PSF FWHM and Strehl ratio.** We measure the optics PSF FWHM and Strehl ratio on the original optics wavefront from the observation (Before), and the residual wavefront resulting from subtracting our wavefront estimate from the original optics wavefront (After). The Center position is at the center of the Rubin focal plane; the Corner position is at the center of the $R00$ wavefront sensor in the corner of the focal plane.

the more selective brightest stars selection. We conclude using the ℓ_2 norm, with no median reduce, to fit the brightest stars, is the best variation.

The combined implementation is very fast. The neural network processes donuts in 6 milliseconds on a 2.4 GHz Intel Xeon CPU with a single Nvidia V100 GPU. Our algorithm can process a full observation, both the local estimates and interpolation, in under 300 milliseconds. This is around two orders of magnitude smaller than the typical telescope exposure time, which suggests latency will not prevent its adoption.

The next step is to take this model and measure the repercussions of subtracting its estimate from the true wavefront on both the PSF FWHM and Strehl ratio [28]. We compute these by calculating the local wavefronts at the center and one 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 2.

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 provides an illustrative example of how the improvements to the optics PSF from our method can improve image quality. We apply the Rubin optics PSF, before and after subtracting the optics wavefront estimated by our framework, to six classic Hubble Telescope images in the absence of other significant PSF contributions.

For the Rubin Observatory however, the PSF is dominated by the atmosphere, which as indicated cannot be corrected. For the Rubin observatory the assumed PSF width is of order 0.71 arcseconds, of which 0.65 is contributed by the atmosphere. Therefore, the improvement in the Strehl and image quality is not as dramatic. Nevertheless, this improvement is still important for use on nights with unusually good atmospheric seeing and also for applications that are especially sensitive to the PSF width.

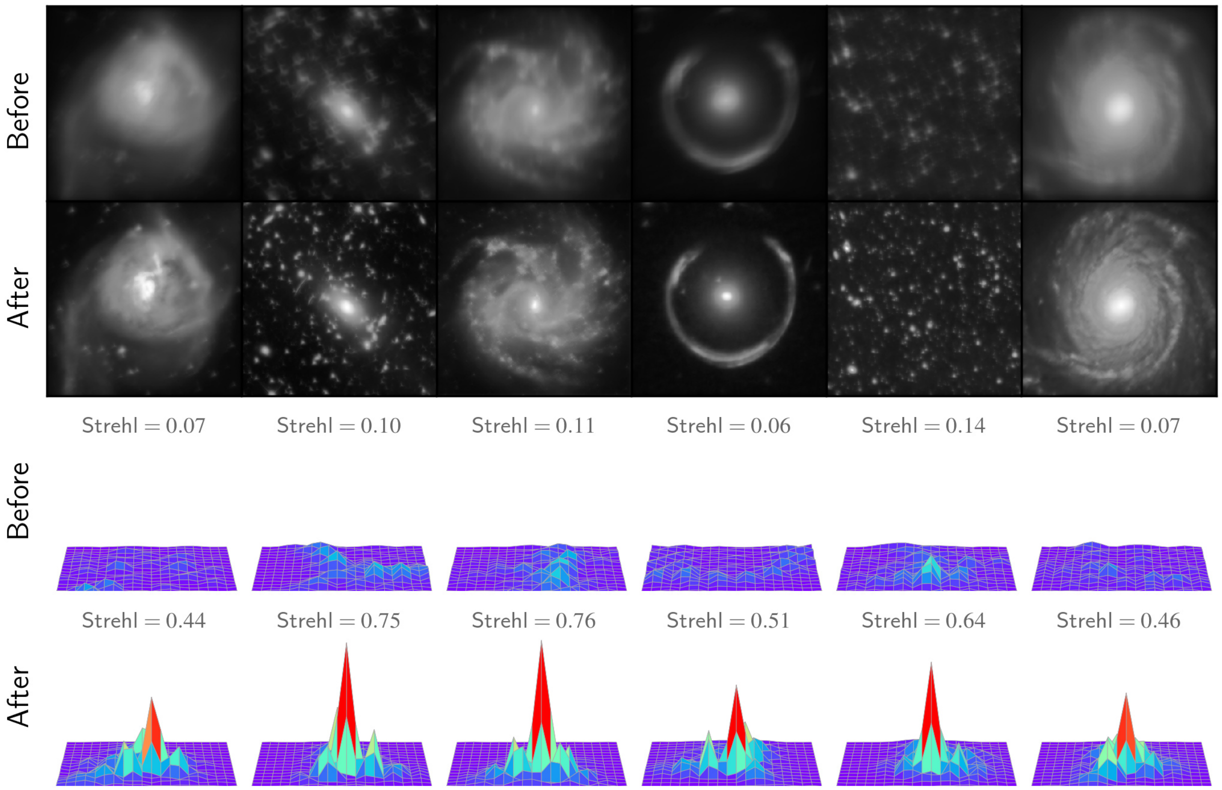

Figure 6. **Hubble Telescope images before and after subtracting the optics wavefront estimated by our framework.** An illustration of the effect of wavefront correction using our techniques on image resolution. We use actual Hubble Space Telescope images to show the effect. The Before images are degraded by the wavefront aberrations we have simulated. The After images show their reconstruction after wavefront estimation and correction using the technique we describe in this paper. On the bottom two rows, we show the effective PSF, both before and after wavefront correction. The images have an angular extent of 3 arcseconds and the PSFs are displayed on a 0.16×0.16 arcsecond grid.

5. Conclusion

A new perspective can facilitate significant progress on a classic problem. For wide-field telescopes, discerning the aberrations due to the optics from those due to atmospheric turbulence is a critical challenge. In this work, we decomposed this classic wavefront sensing problem into two new subproblems. The first subproblem is to estimate local wavefronts from donut images. The second subproblem is to interpolate the optics wavefront across the focal plane from all of the local estimates in the observation. We achieved notable performance by developing specialized techniques for each subproblem, namely a deep convolutional neural network and a select-reduce-fit interpolation scheme. The combined framework reduced the optics PSF FWHM by 27% and increased the Strehl ratio by a factor of 6 on a simulated Rubin Observatory mini-survey. This subproblem decomposition may be a valuable paradigm for designing future wavefront sensing strategies.

Our specific implementation also has unique practical advantages. First, it is extremely fast. It is capable of processing an entire observation in less than 300 milliseconds. Second, it is easy to monitor. There are no iterations; just two steps. The intermediate local wavefronts passed between steps are a useful physical observable that can help telescope operators supervise the system. Third, it is easy to characterize. Both steps are characterized by well known, physical, and interpretable error metrics, such as the MSE of the wavefront coefficients or PSF FWHM of the full optics wavefront. Fourth, it is robust. It improved the optics PSF in all 497 observations in the mini-survey test set. These properties make it an attractive candidate for current and future active optics systems. We are particularly excited about the prospect of our framework being used on space telescopes, where robustness is at a premium.

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.

[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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] Hong Guo, Nina Korablinova, Qiushi Ren, and Josef Bille. Wavefront reconstruction with artificial neural networks. *Optics Express*, 14(14):6456–6462, July 2006.

[9] 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.

[10] 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.

[11] 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.

[12] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Joshua Meyers. batoid. https://github.com/jmeyers314/batoid, 2020. Accessed: 2020-04-07.

[20] 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.

[21] 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.

[22] Scott W. Paine and James R. Fienup. Machine learning for improved image-based wavefront sensing. *Optics Letters*, 43(6):1235, Mar. 2018.

[23] J. R. Peterson, J. G. Jernigan, S. M. Kahn, A. P. Rasmussen, E. Peng, Z. Ahmad, J. Bankert, C. Chang, C. Claver, D. K. Gilmore, E. Grace, M. Hannel, M. Hodge, S. Lorenz, A. Lupu, A. Meert, S. Nagarajan, N. Todd, A. Winans, and M. Young. Simulation of Astronomical Images from Optical Survey Telescopes Using a Comprehensive Photon Monte Carlo Approach. *Astrophysical Journal, Supplement*, 218(1):14, May 2015.

[24] Francois Roddier, Claude Roddier, and Nicolas Roddier. Curvature Sensing: A New Wavefront Sensing Method. In G. Michael Morris, editor, *Statistical Optics*, volume 0976, pages 203 – 209. International Society for Optics and Photonics, SPIE, 1988.

[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.

[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.

[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.

[28] Karl Strehl. Aplanatische und fehlerhafte abbildung im fernrohr. *Zeitschrift für Instrumentenkunde*, 15(7):362–370, 1895.

[29] Damien Sucher, Guillaume Butel, Guillaume Briche, Jean-François Blanc, Arnaud Liotard, Marc Bernot, Mikaël. Carlavan, 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.

[30] 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.

[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.

[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.

[33] Bo Xin, Chuck Claver, Ming Liang, Srinivasan Chand rasekharan, George Angeli, and Ian Shipsey. Curvature wavefront sensing for the large synoptic survey telescope. *Applied Optics*, 54(30):9045, Oct. 2015.

[34] Bo Xin, Aaron Roodman, George Angeli, Chuck Claver, and Sandrine Thomas. Comparison of LSST and DECam wavefront recovery algorithms. In Helen J. Hall, Roberto Gilmozzi, and Heather K. Marshall, editors, *Ground-based and Airborne Telescopes VI*, volume 9906 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 99064J, July 2016.

[35] 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.