

MACHINE LEARNING BASED WAVEFRONT SENSING FOR
THE RUBIN OBSERVATORY

A DISSERTATION
SUBMITTED TO THE INSTITUTE OF COMPUTATIONAL AND
MATHEMATICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

David Thomas
December 2021

Contents

| | |
|--|-----------|
| Abstract | 1 |
| Acknowledgements | 2 |
| | |
| I Machine Learning Based Wavefront Sensing | 8 |
| 1 A New Paradigm For Wide-Field Wavefront Sensing | 9 |
| 1.1 Challenges with Previous Approaches | 9 |
| 1.2 The Double Zernike Polynomials | 14 |
| 1.3 Neural Network Based Phase Estimation | 17 |
| 1.4 Wavefront Estimation Framework | 18 |
| 1.4.1 Estimating Local Wavefronts | 19 |
| 1.4.2 Interpolating the Optics Wavefront | 20 |
| 2 The Rubin Observatory Active Optics System | 22 |
| 2.1 Rubin Telescope | 22 |
| 2.2 Active Control | 24 |
| 2.3 Wavefront Sensors and Donut Images | 26 |
| 3 Simulating High Fidelity Rubin Donut Images | 28 |
| 3.1 Simulating A Single Donut Image | 29 |
| 3.2 Donut Dataset | 34 |
| 3.3 Full Visit Dataset | 34 |

| | |
|---|-----------|
| 4 Training a Convolutional Neural Network To Predict The Local Wavefront | 39 |
| 4.1 Network Architecture | 40 |
| 4.2 Training, Validation, and Testing | 41 |
| 4.3 Neural Network Enabled Insights | 48 |
| 4.3.1 Saliency Maps | 48 |
| 4.3.2 Memorizing Versus Learning | 49 |
| 4.3.3 Adversarial Tradeoffs | 51 |
| 5 Interpolating The Global Wavefront with Least Squares | 53 |
| 5.1 The Interpolation Geometry | 53 |
| 5.2 Evaluating Different Interpolation Strategies | 56 |
| 5.3 Combined Runtime | 60 |
| 5.4 Image Quality Results | 60 |
| 6 Error Covariance Fitting and Optimal Telescope Control | 63 |
| 6.1 Error Covariance Fitting | 63 |
| 6.2 Optimal Control | 66 |
| 7 Conclusions | 72 |
| Bibliography | 75 |

List of Tables

| | | |
|-----|--|----|
| 2.1 | The 50 Telescope Degrees of Freedom | 26 |
| 3.1 | Range of Simulated Control Parameters | 33 |
| 5.1 | Select-Reduce-Fit Results | 57 |
| 5.2 | Improvements To The Optics PSF FWHM And Strehl Ratio | 60 |
| 6.1 | Observation To Perturbation Conversions | 70 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Wavefront and Wavefront Aberrations | 10 |
| 1.2 | Zernike polynomials | 16 |
| 1.3 | Aberration Power in Focal Plane Zernikes | 17 |
| 1.4 | Two Step Approach To Wavefront Sensing | 19 |
| 2.1 | Rubin Optics | 23 |
| 2.2 | Rubin M1M3 Hexapod and Pneumatic Actuators | 25 |
| 2.3 | Rubin M2 Hexapod and Mechanical Actuators | 25 |
| 2.4 | From Focal Plane, To Wavefront Sensor, To Donut Image | 27 |
| 2.5 | Donut Image Intuition | 27 |
| 3.1 | Simuated Donut and Local Wavefront | 29 |
| 3.2 | Survey Footprint | 30 |
| 3.3 | Example Donuts from Donut Dataset | 35 |
| 3.4 | Example Full Visit Sample | 37 |
| 3.5 | Fraction of Blended Sources | 38 |
| 4.1 | Network Architecture | 40 |
| 4.2 | Training Curve | 42 |
| 4.3 | Train Versus Test MSE Loss | 43 |
| 4.4 | Panel of Donut Results | 44 |
| 4.5 | Wavefront Coefficient Error Covariance and Correlation | 45 |
| 4.6 | Panel of Potential Sources of Error | 47 |
| 4.7 | Neural Network Attention | 49 |

| | | |
|-----|---|----|
| 4.8 | Neural Network Label-Shifts | 50 |
| 4.9 | Adversarial Attack | 51 |
| 5.1 | Local Wavefront Predictions Across Four Wavefront Sensors | 54 |
| 5.2 | Fitting Global $\beta_{41}, \beta_{42}, \beta_{43}$ From Local α_4 Predictions | 55 |
| 5.3 | Relative Residuals In Different Scenarios | 59 |
| 5.4 | Illustrative Hubble Telescope Images After Correction From Our Framework | 61 |
| 6.1 | Error Covariance Of Local Wavefront Estimates | 65 |
| 6.2 | Error Covariance Of Global Wavefront Estimates | 67 |
| 6.3 | Control Simulator Algorithm | 68 |
| 6.4 | Controller Results | 71 |

Abstract

We present a new framework for wavefront sensing for wide-field telescopes. 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. In this thesis, we develop simulated observations and images from the upcoming Rubin Observatory to refine and assess this framework. Much of this work is also summarized in [1] and [2], although here we describe it in more detail.

Our unique contributions are as follows. We isolated wavefront sensing problem and decomposed it into two sub-problems. We created benchmark datasets for both sub-problems. We demonstrated a convolutional neural network can predict local wavefront from donut images. We showed that the global wavefront can be interpolated with least squares from the local wavefront estimates. Finally, we developed a wavefront control simulation environment and used it to assess three canonical control strategies.

The algorithm has great practical properties - it is transparent, robust, low latency, and high bandwidth - and achieves stunning performance. In a realistic Rubin 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.

Acknowledgements

The idea to pursue graduate school chrystalized when I was working as a quant for Teza Technologies, a small, academic high frequency trading firm. I witnessed first hand how machine learning and low latency technologies were completely up-ending securities trading. Broad shouldered Wall Street suites were being replaced by hackers and math nerds. As machine learning and automation became increasingly powerful tools for our firm, I began to question whether these tools could spur similar advances in physics and cosmology.

In 2016, I started the Institute for Computational and Mathematical Engineering (ICME) MS program and began working with Philip (Phil) Marshal on high dimensional cosmological inference. In the summer of 2017, Phil introduced me to my eventual advisor, Steven (Steve) Kahn, as someone who could potentially support me for a PhD. I remember our first meeting well. I presented my results. Steve sat there, completely silent. His stern expression gave no indications of his thoughts ... was he even paying attention? After I finished presenting, he asked extremely insightful questions that demonstrated he had been paying careful attention. Indeed, Steve is a great listener. I was also impressed by his creativity and interest in working on problems off the beaten path. Shortly after our first encounter we teamed up to work on star trails, and Steve officially became my PhD advisor.

Over the past four years we focused primarily on subsecond photometry and wave-front sensing. In both cases, after experimenting with and attempting to extend existing approaches, we were able to develop superior approaches based on machine learning. While these solutions seem very natural in hindsight, it is worth reminding any future PhD students reading this, that we proceeded through many many failed

attempts at tackling these problems, received a lot of skepticism around leveraging deep learning initially, before we were finally able to make it work. As Edison says “I just found 2,000 ways not to make a lightbulb; I only needed to find one way to make it work.” In this thesis, we focus on our work on wavefront sensing. We point readers interested in our work on star trail photometry to our published and forthcoming papers.

I also would like to thank Patricia Burchat and Aaron Roodman for their guidance and support. As a student from ICME, there were times where I did not feel welcome by the broader physics community. However, Professors Burchat and Roodman always were extremely welcoming and kind. Professor Burchat has been supportive throughout my PhD and provided key feedback to my research efforts and this thesis. Professor Roodman, who developed a previous generation wavefront sensing system, was also an extremely instrumental mentor. My failed attempts to extend his Fourier transform and optimization based method gave me a clearer sense of the key challenges for Rubin and ultimately allowed me to develop a new approach to tackle them. Both of these Professors also asked great questions during my oral defense.

Professor Stephen Boyd was another crucial member of my defense committee. I enjoyed working with Professor Boyd to map the covariance estimation problem for the global wavefront into a convex optimization problem, and on optimal control. This collaboration was a great example of inter-disciplinary research. While Professor Boyd no longer teaches his infamous CME 364A course, Convex Optimization, his recorded 2008 lectures on YouTube have over one million views. These are largely credited as popularizing the subject, and helped to make this my favorite course at Stanford. I believe it provides an unrivalled foundation for machine learning and statistical learning, and encourage any readers to take it - or at least watch Boyd’s lectures.

I would also like to thank my collaborators. Joshua Meyers has been key a mentor and go-to expert on all things optics. His Batoid Python package significantly lowered the hurdles to doing raytracing and really enabled this work. I have also tried to pick up some Joshism’s, such as his clean code and notebooks, tendency to use widgets, and clear presentations paced with incredible simulations.

Federica Bianco has also been a key collaborator on the star trails work. Federica's enthusiasm for science, which ranges from astrophysics to remote sensing to climate change to extraterrestrial technosignatures - is extremely inspiring. She was always willing to meet our projects where they were, and patiently brainstorm and work through the obstacles that come with trying to break new ground.

I am also extremely grateful to my collaborators at the National Optical Astronomy Observatory in Tucson. Steve helped arrange for me to move out there for two quarters in my third year to collaborate with the team and learn the Rubin (then LSST) software stack and wavefront sensing building blocks. The background required for wavefront sensing is not taught in a traditional physics program. In fact, I have still not come across a good *single* reference. Instead one's knowledge must be gleaned from many disparate sources and through experience. Thus, I am very appreciative of the time Bo Xin and Te-Wei Tsai spent with me to impart this wisdom. I also appreciated collaborating with Sandrine Thomas and Chuck Claver on various challenges of the baseline approach the team was developing for Rubin.

The ICME also played a big role in my development. An ICME graduate can employ more mathematical rigor than most engineers and scientists, and also has strong coding skills. I can't think of better preparation to contribute to innovation in the 21st century. The Directors Margot Gerritsen and Gianluca Laccarino did a great job improving the program throughout my tenure. The former administrative assistant Indira Choudhury was also the spoke that made the institute run, and students graduate. I am forever grateful for her help in this endeavor.

It has been an incredible privilege to attend Stanford University. The school has top faculty, a beautiful campus, and ample support and extracurricular activities. It is also in the heart of Silicon Valley and can serve as a stepping stone into entrepreneurship. Palo Alto is also a great starting point to explore the broader San Francisco Bay Area. Nowhere else can one surf Pacifica in the morning and meet with venture capitalists on Sand Hill Road in the evening.

I am also very appreciative of Stanford's PhD minor programs. In addition to my PhD in Computational and Mathematical Engineering I was also able to complete a PhD Minor in Physics and PhD Minor in Computer Science. This background has

prepared me well to contribute to interfaces between fields where much opportunity remains. I am the first student at Stanford to complete these degrees.

The past three summers I was able to augment my PhD with three formative experiences. In the summer of 2019, I mentored Stanford undergraduate Emily Li. Emily was a bright and enthusiastic young researcher. She absorbed the optics for wave-front sensing like a sponge. This experience taught me how to clearly communicate expectations, share the joy of science, and teach foundational skills. The experience encouraged her to pursue a career in physics. Nothing is more rewarding than seeing one's mentees succeed.

In summer 2020, I interned as a Quantitative Researcher in the Global Quantitative Strategies business at Citadel Investment Group. It was great to see how state of the art quantitative hedge funds operate and compare and contrast the methods used here with those used during my time in high frequency trading. While I cannot go into detail about my work there due to a strict non-disclosure agreement, I can say that there is a lot of value in being fluent in the mathematics behind statistical learning models as well as being a strong programmer and knowing how to accelerate and improve these models and their training on modern compute infrastructure. Again, the ICME combination of mathematical rigor and computational engineering was superb preparation. This experience was also a great reminder of how challenging it is to beat the market. Even with a team of hundreds of highly accomplished PhD researchers, state of the art software and resources, and the accumulation of millions of signals and research findings ... it is difficult to break even. I suspect people who have not experienced this world first hand, and the repeated failures that accompany this ambitious goal, with billions of dollars on the line, cannot truly appreciate how efficient markets really are. Retail trading and investing is pure speculation - and a losing game.

In summer 2021, I interned as a Machine Learning Engineer on the Instagram Ads Ranking and Delivery team at Facebook. I really enjoyed researching and developing large scale recommender systems. These systems, used in everything from Google's search engine to prioritizing content on Facebook's home feed, are some of the most

critical components of large tech businesses and the modern internet. The computational resources and datasets available at companies like Facebook are so superior to those in academia that to perform cutting edge work in this field almost necessitates one to work for a large tech company. Some of my colleagues were former faculty members who moved from academia to research roles at Facebook in order to get around this discrepancy and continue producing impactful research. It seems like it will be increasingly difficult for academia to retain top talent in these fields: large scale recommender systems and machine learning more broadly.

My family has always been incredibly supportive. My father David *Evan* Thomas served in the United States Navy for 25 years. I am incredibly proud of his commitment and dedication to his country. When I was growing up, we would stay with relatives over the holidays so he could work over his vacations to earn enough money to send my sister to a private high school. That was a good investment, because today my sister Danielle Thomas has become a very successful orthopedic surgeon. She is a tough cookie - her favorite tool is the bone saw - and has no problem exceling in a field that is over 90 percent male. She also is a voracious reader. She read over one hundred books in each of the last two years. My mother Valerie Thomas was my original teacher during infancy and has encouraged my pursuits ever since. During my PhD, she would regularly visit the bay area and would send me care packages. She has become an incredible painter. Her paintings decorate my walls; guests tend to be surprised that the artist is, in fact, my mother.

Finally, I would like to give the utmost thanks to my wife Madeleine Scott. We met as undergraduates at MIT and started dating in 2012. She briefly worked at a biotech startup before pursing both an MD and a PhD at Stanford through the eight year, fully funded, and *extremely* competitive Medical Scientist Training Program. She is an incredibly kind person that feels so much for other people that she cannot watch dramatic movies. She loves helping others solve problems and will be an incredible physician. She is also the most talented young scientist that I know. This past year she has had two papers accepted to the revered journal Nature. I am very fortunate to be her collaborator in life.

My wife's parents Frances Donovan and Gary Scott have been incredibly supportive throughout our relationship and especially during the Covid-19 pandemic. During Covid we have stayed with them in Berkeley for months at a time. Their active lives, copious reading, fascinating conversations, and respect for each other set a great example. While, Gary, a Physics PhD who now works as a molecular biology researcher, enjoys shooting down dark matter, lambda-CDM, and other tennets of modern cosmology - I cannot thank him and Frances enough for letting me marry their daughter.

Recently, we welcomed a new member into our family. Kyver Blaze Thomas was born at 3:48 pm on July 24th at the Lucile Packard Children's Hospital. We are absolutely smitten new parents. Kyver is already good-natured like his mom, and inquisitive like his dad. Perhaps one day he will read this thesis.

Part I

Machine Learning Based Wavefront Sensing

Chapter 1

A New Paradigm For Wide-Field Wavefront Sensing

What is the most important theme in computer science? [Class responds.]
No. It is *abstraction*.

John Ousterhout in CS 140: Operating Systems, 2020

1.1 Challenges with Previous Approaches

Keeping telescopes in focus is an art as old as the instruments themselves. Their widespread use spurred the development of theories of optics and aberrations. In 1678, the famed Dutch physicist Christian Huygens proposed that every point that interacts with light may be regarded as the source of a new spherical wave. In 1818, the French physicist Augustin-Jean Fresnel incorporated interference into this model. The resulting Huygens-Fresnel principle is still a widely taught model for wave propagation and provides the physical intuition behind reflection, refraction, and diffraction. The paradigm also introduced the notion of the wavefront.

A wavefront is a two dimensional surface over which the phase of the wave is constant. We also overload this diction to refer to aberrations, or two dimensional

spatial offsets from a reference wavefront. Figure 1.1 shows example aberrations to planar and spherical reference wavefronts.

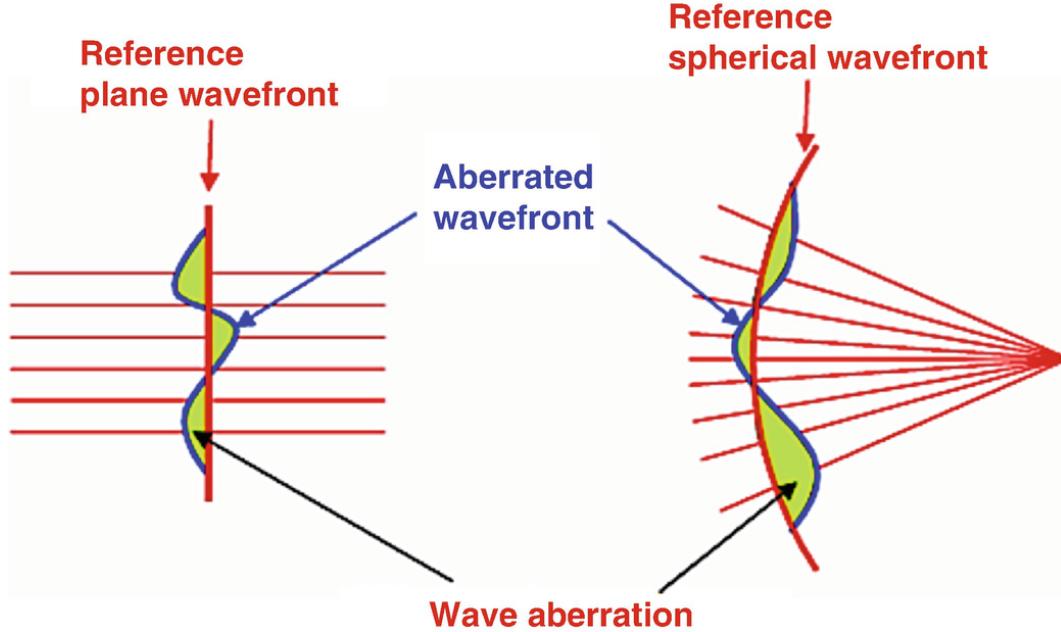


Figure 1.1: Example aberated planar and spherical wavefronts. The horizontal red lines show the direction of propagation. The vertical red line shows the reference wavefront and the blue line shows an example aberrated wavefront. Source: [3].

Wavefronts are useful because the wavefront aberrations in different planes, and through time, largely characterize the image quality of an optic. In the case of ground based telescopes, there are three primary contributions to image quality: the atmosphere, the telescope, and the camera. The field of *adaptive optics* is concerned with controlling deformable mirrors to the optical path to correct for aberrations induced by atmospheric turbulence on 10-100Hz frequencies. However, for wide field telescopes with fields of view on the order of a degree, a common corrections is not possible for the entire field of view. In this context we are primarily concerned with *active optics* which strives to correct the aberrations due to the telescope. New methods of wavefront sensing for active optics is the focus of the first part of this thesis.

One way to characterize the optics of a telescope is through path length differences.

For every position in the image plane, we can measure the path differences to points in the entrance pupil. An alternative way to mathematically express this is through an aberrated wavefront defined at the entrance pupil for each position in the image plane. Then the imaging properties can be computed with Fourier Optics. We refer readers who would like to understand this relation further to Goodman's classic Fourier Optics text [4].

There are two thematic approaches to wide field wavefront sensing: zonal and modal. In the zonal approach, the entrance pupil is partitioned into an array of subapertures. In each of these zones, the wavefront is characterized by its optical path length, local gradient, or local curvature. The wavefront measurement improves with more partitions. The Shack-Hartmann sensor and shearing interferometer are the two most common zonal approaches.

The modal approach treats the wavefront aberration, at a given point in the image plane, as a sum of low order polynomials defined over the entrance pupil. This wavefront measurement improves with higher order polynomial measurements. Modal approaches typically rely on strategically defocused wavefront sensors. The Rubin Observatory focal plane was designed with a modal approach in mind as it does not require lenselets or tweaks to the beam. As described in chapter 2, the Rubin focal plane contains four corner wavefront sensors specifically for this purpose.

There are two previously described approaches to wavefront sensing that are relevant for Rubin. The first, by Roodman et. al. [5], is the algorithm used for the Dark Energy Camera on the Blanco Telescope [6]. It is based on the Fraunhofer diffraction integral:

$$I(x, y) \propto |\mathcal{F}\{P(\rho, \theta \exp(2\pi i W(\rho, \theta)/\lambda))\}|^2 \quad (1.1)$$

where I is intensity; \mathcal{F} is the two-dimensional Fourier transform; P is the pupil function; W is the wavefront; λ is the wavelength; and x, y and ρ, θ parameterize the image and pupil planes respectively. The wavefront is modeled as a sum of

orthonormal Zernike polynomials:

$$W(\rho, \theta) = \sum_i a_i Z_i(\rho, \theta) \quad (1.2)$$

Starting from a set of telescope control parameters, the wavefront coefficients can be computed. Then the Fraunhofer diffraction integral can be used to generate a corresponding intensity image. The difference between the donut images and the images from the forward model,

$$\mathcal{L} = \sum_{\text{donut } d} \|I_{d,\text{true}} - I_{d,\text{forward}}\|_2^2 \quad (1.3)$$

can be used as a loss function. Then we can estimate the telescope parameters by using an optimizer to find the parameters that minimize this loss function. This works well when there are a few non-degenerate parameters and the problem remains approximately convex.

While this approach has been successfully applied to the Blanco Telescope, there are two impediments to applying it to the Rubin Observatory. The first is that Rubin has much higher dimensionality and complete degeneracy between some of the parameters of interest. The Blanco Telescope is a prime focus telescope with a single mirror; the Rubin Observatory is a modified Paul-Baker telescope with three mirrors. The Blanco active optics system controls 8 parameters; the Rubin active optics system strives to control 50 parameters. The Rubin control parameter optimization surface is highly nonconvex. In addition to these theoretical hurdles, the fast beam of the Rubin optics also necessitates much larger Fourier transforms pushing the runtime of the algorithm beyond the 15 seconds we are targeting (see next chapter for more details).

The second approach is based on the transport of intensity equation (TIE):

$$\frac{\partial I}{\partial z} = -(\nabla I \cdot \nabla W + I \nabla^2 W) \quad (1.4)$$

where z is the optical axis. It builds on the curvature wavefront sensing technique developed by Francois, Claude, and Nicolas Roddier [7]. Their key insight was that

$\frac{\partial I}{\partial z}$ can be approximated by subtracting two donut images on different sides of focus. Then W in the TIE equation can be solved for in Fourier space or with a Zernike polynomial series expansion. The large central obscuration, fast f -number, off-axis distortion and vignetting, and different field positions of intra and extra-focal donuts are four challenges to using curvature sensing for Rubin. Xin et. al. [8] proposed solutions and extended curvature sensing to estimate the Rubin optics wavefront at four field positions. Below we give a high-level summary of the operations this algorithm performs, on a single wavefront sensor, to estimate the wavefront at the center of the corresponding wavefront sensor:

1. Crop N pairs of intra-focal and extra-focal donuts
 $\{(D_{\text{intra},1}, D_{\text{extra},1}), \dots, (D_{\text{intra},N}, D_{\text{extra},N})\}.$
2. Make initial wavefront estimates $W_0 = 0$, $W_1 = \text{guess}.$
3. Set $i = 1.$
4. While $\|W_i - W_{i-1}\|_2^2 < \text{tolerance}$ repeat,
 - (a) Apply nonlinear transformation f which migrates donut image from its field position to the center of the wavefront sensor, based on current wavefront estimate W_i . We have $\{(D'_{\text{intra},1}, D'_{\text{extra},1}), \dots, (D'_{\text{intra},N}, D'_{\text{extra},N})\}$ where $D'_{*,k} = f(D_{*,k}, W_i).$
 - (b) Apply correct g for intensity and vignetting differences between the pairs of donuts. We have $\{(D''_{\text{intra},1}, D''_{\text{extra},1}), \dots, (D''_{\text{intra},N}, D''_{\text{extra},N})\}$ where $D''_{*,k} = g(D'_{*,k}, W_i).$
 - (c) Set $(\partial_z I)_k \propto D''_{\text{intra},k} - D''_{\text{extra},k}.$
 - (d) Solve TIE for W_{i+1} with $(\partial_z I)_k$ by series expansion.
 - (e) Set $i = i + 1.$
5. Return W_i as wavefront for the center of the wavefront sensor.

Afer this process the algorithm uses the wavefront estimates at the center of the four corner wavefront sensors to constrain 50 telescope control parameters (these parameters are described in the next chapter).

This approach has been developed by a dedicated team over the last decade. Parts of the algorithm, such as step 4 (d) are mature and have been validated on real data [9]. However, despite significant effort, the code for the full algorithm on Rubin remains incomplete. Thus it is difficult to benchmark or compare to alternatives.

The primary challenge on theoretical grounds is the convergence. There are no convergence gaurentees. The wavefront estimate may not improve between the iterations in step 4. Some tests suggest that the algorithm typically converges for small perturbations to the nominal optics configuration. However, for large perturbation the transformations f and g will have more error. A different strategy might be required.

The algorithm is also fairly computationally expensive. It takes approximately 15 seconds to perform 10 iterations of step 4 on 10 pairs of donuts. However, a typical wavefront sensor image contains close to a thousand donut images. A lot of potentially useful donut information is ignored.

There are also a few challenges on practical grounds. The functions f and g used in step 4 depend on the wavefront estimate and are fairly complicated. This makes the algorithm challenging to interpret and debug. Also, since f and g depend on the wavefront estimate, it is difficult to characterize the error in the full algorithm, because each step depends on the preceeding steps.

After spending a year supporting efforts to implement this algorithm, I realized the prudence in developing an alternative approach. I sought to develop an algorithm that was simple, transparent, robust, low-latency, and high-throughput. The next two sections develop two key paradigms behind the algorithm.

1.2 The Double Zernike Polynomials

The Zernike polynomials are a sequence of polynomials $\{Z_i\}$ that are widely used to characterize wavefronts in optics [10]. They are defined over a unit disk, or annulus,

and normalized such that

$$\int_{R_{\text{inner}}}^{R_{\text{outer}}} \int_0^{2\pi} Z_i(\rho, \theta) Z_j(\rho, \theta) d\rho d\theta = \delta_{ij} \quad (1.5)$$

Figure 1.2 shows the first 21 Zernike polynomials. For characterizing a wavefront in the pupil plane we follow the convention of using $Z_4 - Z_{21}$ by Xin et. al. [8]. The first three polynomials are ignored because while they effect the position of the image, they do not impact the image quality. The polynomials beyond 21 are ignored because they are only weakly excited by typical Rubin perturbations.

For small telescopes, this basis can be used to characterize the wavefront in the entrance pupil. For wide-field telescopes like the Rubin Observatory, the wavefront changes non-trivially across the field of view. Thus the wavefront is a four dimensional function of both the entrance pupil and focal plane. One way to model this would be to have another polynomial basis $\{P_i\}$ for the focal plane x, y . Then we could write the full wavefront as $W(x, y, \rho, \theta) = \sum_i \sum_j \beta_{ij} P_i(x, y) Z_j(\rho, \theta)$. The coefficients β are indexed by the image plane index i and pupil plane index j . The double Zernike polynomials are the $P_i(x, y) Z_j(\rho, \theta)$ where the field polynomials are circular Zernike polynomials $P_i = Z_i$ [11]. Thus β_{ij} is the mode of circular polynomial Z_i in the focal plane, and annular polynomial Z_j in the image plane.

In practice we also truncate the focal plane index. This leaves us with a finite set of coefficients to represent the full wavefront across the pupil and focal planes. We found that for the Rubin Observatory, most of the aberration power is concentrated in the first three focal plane polynomials. We simulated 500 perturbed telescope states and numerically calculated the double Zernike coefficients out to index 36. Then we studied the fraction of the aberrated wavefront contained in the first k focal plane Zernikes. We found that the first three polynomials capture 90% of the wavefront. These first three polynomials are a constant offset, a tip-plane, and a tilt-plane - collectively they define a plane. We suspect a similar pattern holds for other wide field telescopes. It suggests that the 54 double Zernike coefficients β_{ij} where $1 \leq i \leq 3$ and $4 \leq j \leq 24$ offer a good characterization of the wavefront.

This separation of the wavefront into pupil and focal plane components is key

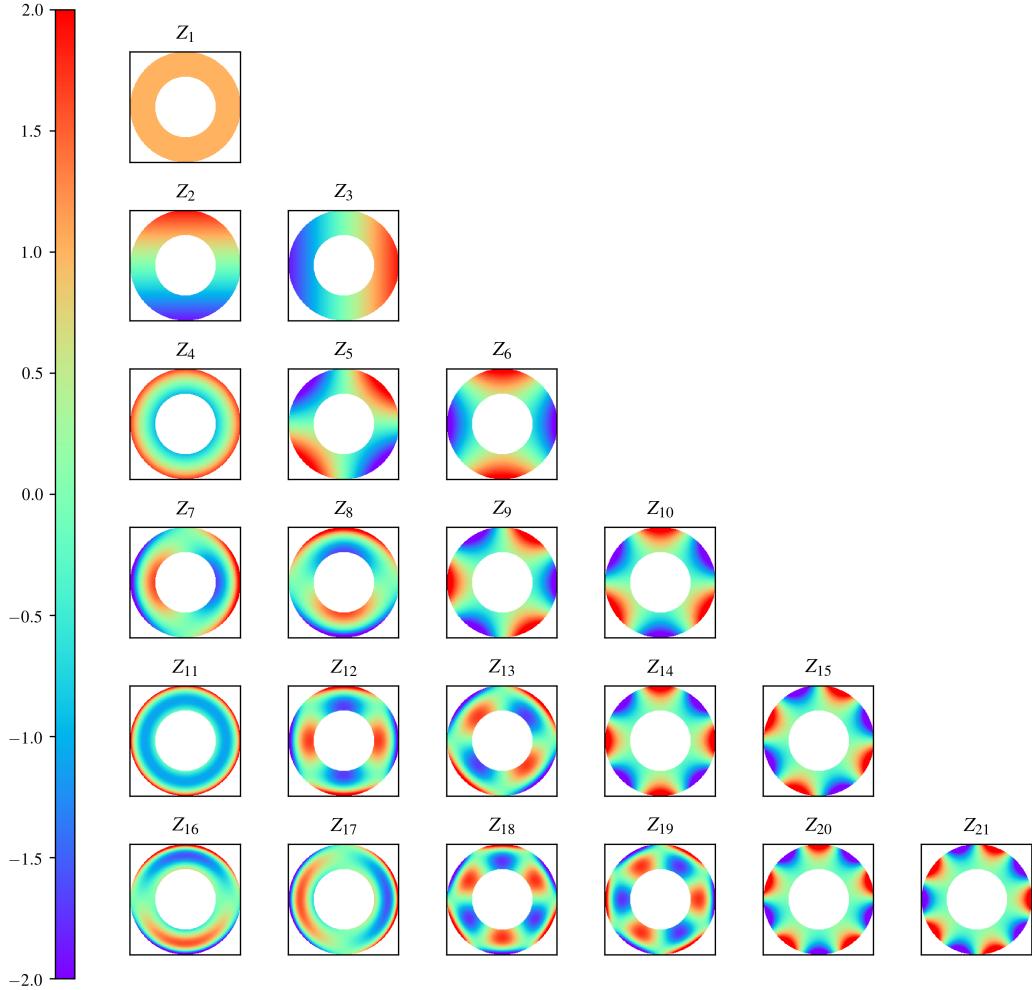


Figure 1.2: The Zernike polynomials Z_1 through Z_{24} with the Noll index scheme. The obscuration of the annular domain is consistent with the obscuration in the Rubin optics.

to this work. Our key insight was as follows. At any point in the image plane, we can use the image to constrain the pupil wavefront at that position - we deem this the *local* wavefront. Then we can interpolate the focal, or *global*, wavefront from all the local estimates. In section 1.4, we describe techniques that are very well suited for each of these subproblems. Before then, we summarize recent progress in neural network based phase estimation.

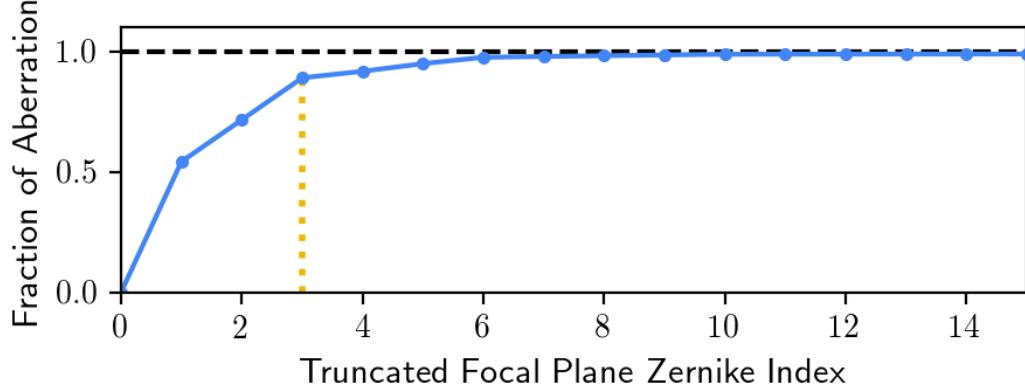


Figure 1.3: The fraction of the aberration that is present in the first k focal plane Zernike polynomials, where k is the truncation index on the x-axis.

1.3 Neural Network Based Phase Estimation

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 [12]. Shortly afterwards, this potential was realized as neural networks were deployed to detect turbulence induced distortion on the Multiple Mirror Telescope [13] and to detect aberrations in the primary mirror of the Hubble Space Telescope [14]. Others expanded this concept to predict more wavefront components [15], incorporate temporal history [16, 17], compare reconstruction methods [18], and better characterize atmospheric turbulence [19].

In the past decade, convolutional neural networks (CNNs) [20] have re-emerged and spurred dramatic advances in computer vision [21, 22, 23, 24]. This has created new possibilities for wavefront sensing in astronomy. In [25], 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. [26] 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, [27] 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. Our approach is comprised of only two steps, is easy to characterize, and can process each donut image in 6 *milliseconds*.

1.4 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 [11] to represent the optics wavefront,

$$W_{\text{opt}}(u, v, x, y) = \sum_{i=1}^k \sum_{j=1}^m \beta_{ij} Z_i(u, v) Z_j(x, y) \quad (1.6)$$

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 2.5). 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

$$\beta = f((D_1, x_1, y_1, z_1), \dots, (D_n, x_n, y_n, z_n)) \quad (1.7)$$

We break this into two subproblems.

1. Estimate the local wavefront with a CNN at each donut position.

2. Interpolate the global wavefront from the local wavefront estimates across the focal plane.

These steps are outlined in Figure 1.4.

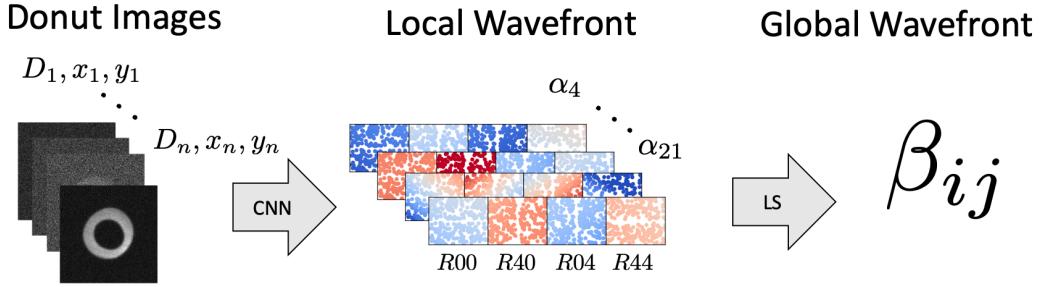


Figure 1.4: In the first step, a convolutional neural network (CNN) processes each donut crop on the four wavefront sensors. In the second step, from each local wavefront coefficient we interpolate three global wavefront coefficients with least squares optimization (LS).

1.4.1 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 (equation 1.1). 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

$$w_{\text{tot}}(u, v) = \sum_j \alpha_{ij} Z_j(u, v) \quad (1.8)$$

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 $j = 1 \dots m$ from (D_i, x_i, y_i, z_i) . In chapter 4 we describe the implementation of this model in detail.

1.4.2 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

$$w_{\text{tot}}(u, v) = W_{\text{opt}}(u, v|x_i, y_i) + \epsilon(u, v|x_i, y_i) \quad (1.9)$$

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 $i = 1, \dots, m$ we have

$$\alpha e_i = \mathcal{Z}\beta e_i + \epsilon \quad (1.10)$$

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

$$\beta = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^m \ell(\alpha e_i, \mathcal{Z}\beta e_i) \right\} \quad (1.11)$$

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

Algorithm 1: estimates the optics wavefront from donut 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 Sernike  $j$  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 $1/\sqrt{n}$ 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 [8], 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 chapter 4, we highlight the power of these models.

Chapter 2

The Rubin Observatory Active Optics System

Some people don't like change, but you need to embrace change if the alternative is disaster.

Elon Musk

2.1 Rubin Telescope

The Vera C. Rubin Observatory is a 8.4-m wide-field telescope designed to measure the structure and evolution of the universe. The roots of the observatory date back twenty years ago to a 2001 proposal for The Dark Matter Telescope [28]. The original scientific motivation was to produce premier weak lensing catalogs. Since then, the facility has gone through two name changes (it was the Large Synoptic Survey Telescope for many years) and the purview has expanded to include cataloging the solar system, mapping the milky way, and exploring the transient and variable sky [29]. Site construction began on Cerro Pachón, Chile in 2015 and is expected to be completed in 2023 [30, 31].

Rubin will have a $319 \text{ meter}^2 \text{degrees}^2$ étendue, ten times larger than any previous or currently planned telescope. The large étendue is achieved by a novel three mirror

design with a very fast $f/1.234$ beam. The design is optimized to attain seeing limited image quality across the 9.6 deg^2 field of view and passband spanning $320 - 1050 \text{ nm}$. Figure 2.1 shows the light path through the optics. The incident light first strikes the annular primary mirror M1, which has an outer diameter of 8.4 m and an effective filled aperture of 6.4 m . Next the light visits the 3.4 m convex secondary mirror M2 and the 5.0 m concave tertiary mirror M3. Then the light passes through three refractive camera lenses and onto the 64 cm focal plane.

The primary and tertiary mirrors, M1 and M3, share an inner and outer radius respectively. This was chosen so that they could be fabricated from a single monolithic blank using spin-cast borosilicated technology. We refer to this collective mirror as M1M3.

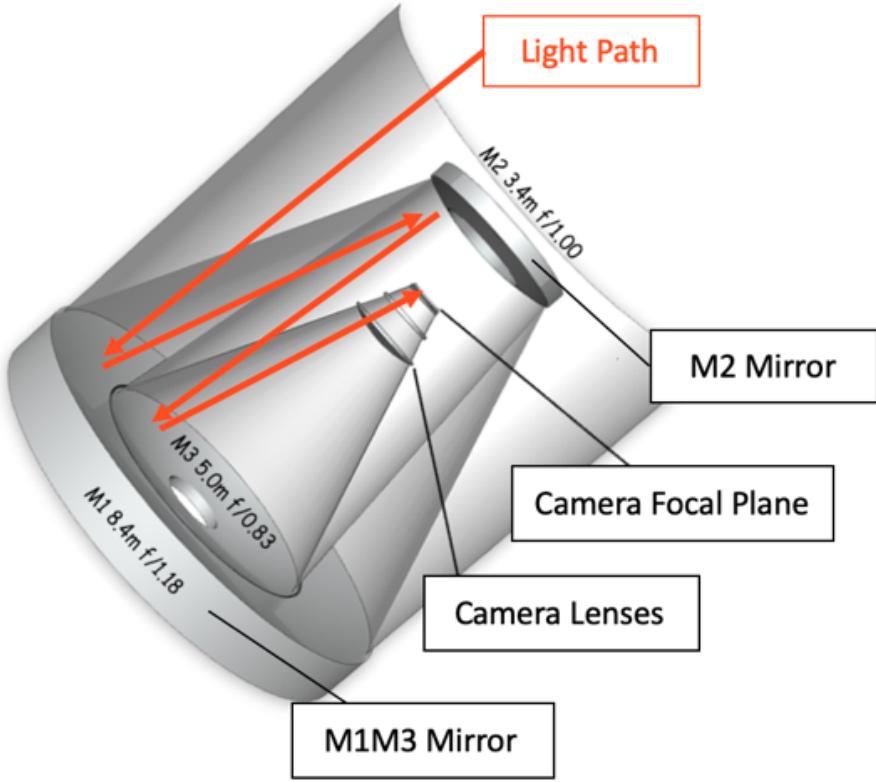


Figure 2.1: The optic elements of the Rubin telescope. The light path through the optics is shown in orange.

2.2 Active Control

There are two telescope properties that incentivize the creation of telescopes with large primary mirrors. First, the étendue, or light collecting power of a telescope is proportional to the area of the primary mirror. Second, the diffraction limited seeing is inversely proportional to the diameter of the primary mirror. Historically these primary mirrors were thick in order to maintain their nominal surface figure. However, the maximum size possible by these manufacturing processes is 5-6 meters of diameter.

In the late 1980s active optics technology was developed to enable the construction of a generation of telescopes with 8 meter primary mirrors. The mirrors for telescopes with this technology are thin and malleable. For Rubin, the mirrors are made out of borosilicate, which also quickly equilibrates with the ambient temperature. Active optics works by actively adjusting degrees of freedom of the telescope, including the mirror surface, based on a feedback system. The main sources of deviation to correct for are mechanical stresses stemming from the dynamics, wind, and gravity and temperature gradients in the mirror surfaces.

Figure 2.2 shows the active components of the combined primary and tertiary mirrors M1M3. The position in the x,y, and z directions is controlled by six large actuator hardpoints that form a hexapod. The mirror surface is controlled by 156 pneumatic actuators. The distribution of the actuators and their axes is explained in detail in [32], although these details are not critical for this thesis. Given that even the large hardpoint actuators are pneumatic it is apt to say that the mirror sits on air.

Figure 2.3 shows the active components of the aspherical secondary mirror M2. The position in the x,y, and z directions is controlled by six large tangential electromechanical actuators. The mirror surface is controlled by 72 axial electromechanical actuators. The details of the configuration, metrology, and testing are described further in [33]. Again, these details are not strictly necessary for what follows.

While it is important to understand the underlying physical interfaces for controlling the telescope, we use a key abstraction to reduce the dimensionality of the

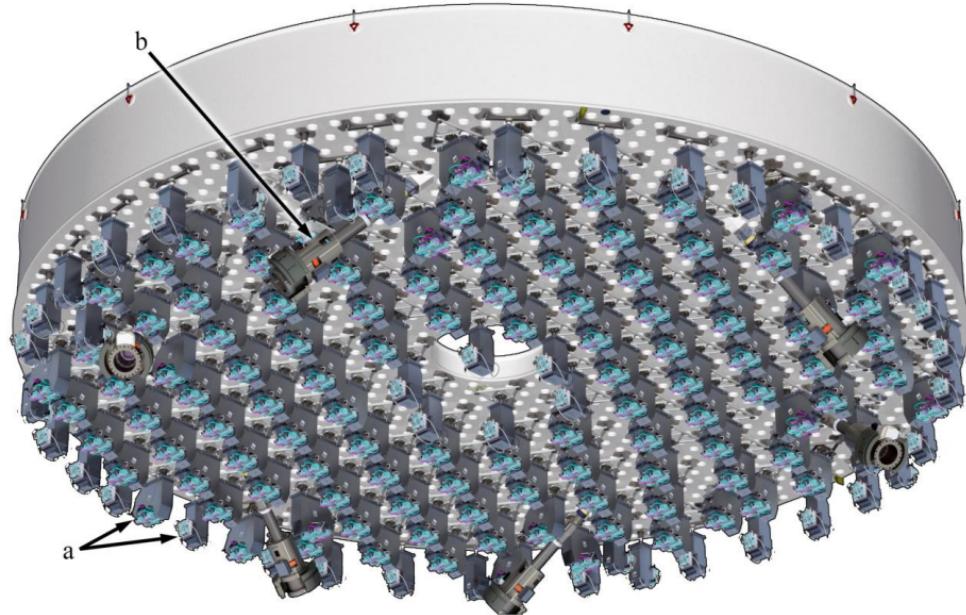


Figure 2.2: View of the bottom face of the M1M3 mirror showing: (a) pneumatic figure control actuators and (b) hard point hexapod. Source: [32].

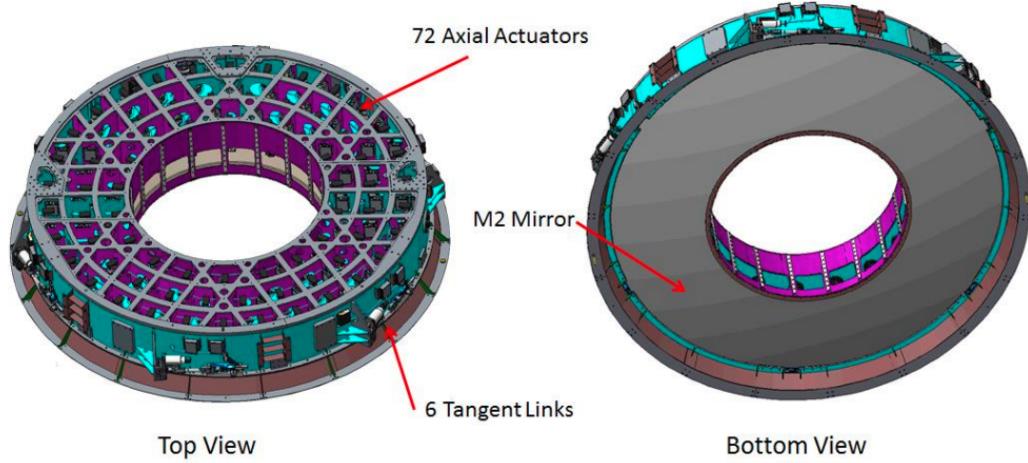


Figure 2.3: Top and bottom views of the secondary mirror M2 showing the mirror cell, mirror, axial actuators, and tangential hard points. Source: [33].

problem. Instead of modeling the state of the mirror surfaces with hundreds of predicted actuator deviations, we use low order bending modes. These are the modes

that are most excited by the aforementioned sources of perturbations. Not only does this allow us to reduce the dimensionality by an order of magnitude, it also better captures the deviations we care about. We use finite element simulations to determine which modes have the most power and how many modes are required to maintain Rubin image quality specifications.

At the time this work was initiated, all the bending modes had not been confirmed. Thus we decided to parameterize deviations to the mirror surface with the first 20 annular Zernike polynomials. We do not believe this choice has significant impact on the analysis. Table 2.1 shows the 50 degrees of freedom - 5 for each hexapod and 20 for each mirror surface. It is worth emphasizing that this telescope's predecessor, the Dark Energy Camera on the Blanco Telescope, only controls 8 degrees of freedom.

Table 2.1: The 50 telescope optics control parameters. Notation: dx is a translation in the x-axis and rx is a rotation along the x-axis.

| | |
|-------------------|--------------------------|
| Camera hardpoints | dx, dy, dz, rx, ry |
| M2 hardpoints | dx, dy, dz, rx, ry |
| M1M3 figure | 20 surface bending modes |
| M2 figure | 20 surface bending modes |

2.3 Wavefront Sensors and Donut Images

In the last section we discussed the degrees of freedom we are trying to predict and control. In this section we describe the feedback system that makes this possible.

The Rubin focal plane is comprised of 189 science sensors, 4 wavefront sensors, and 8 guidestar sensors [34]. The wavefront sensors, in the four corners of the focal plane, are used for active optics. Each $4k \times 4k$ pixel sensor consists of two $2k \times 4k$ pixel half-chips. One half-chip is 1.5 mm intra-focal; the other is 1.5 mm extra-focal. The stars on these images are called *donuts* because of their annular, donut-like shape. Figure 2.5 shows the layout of the focal plane and simulated donut images.

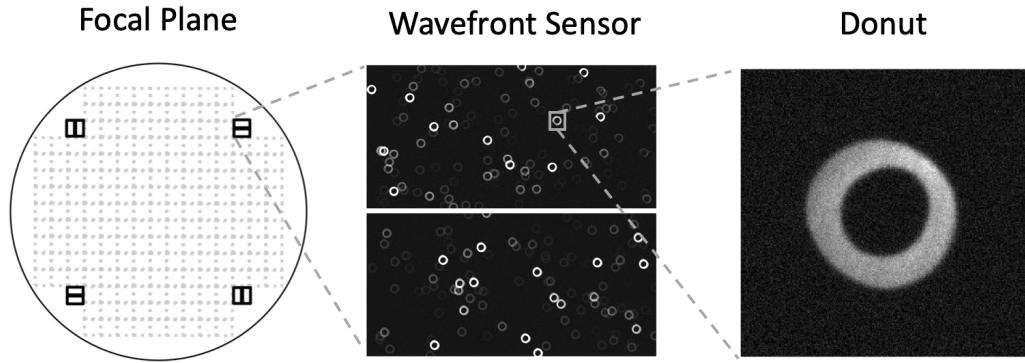


Figure 2.4: Left: the Rubin focal plane; science sensors are in gray; wavefront sensors are in black. Middle: One wavefront sensor split into intra-focal and extra-focal half-chips. Right: a single 256×256 pixel crop of a donut.

The annular shape of the donut image stems from the outline of the M1 aperture. Figure 2.5 highlights this intuition. The intra-focal and extra-focal sensors cut off the Rubin beam before it has converged. The intensity pattern in these larger images provide more information about the location of perturbations that are accounted for across the entrance pupil.

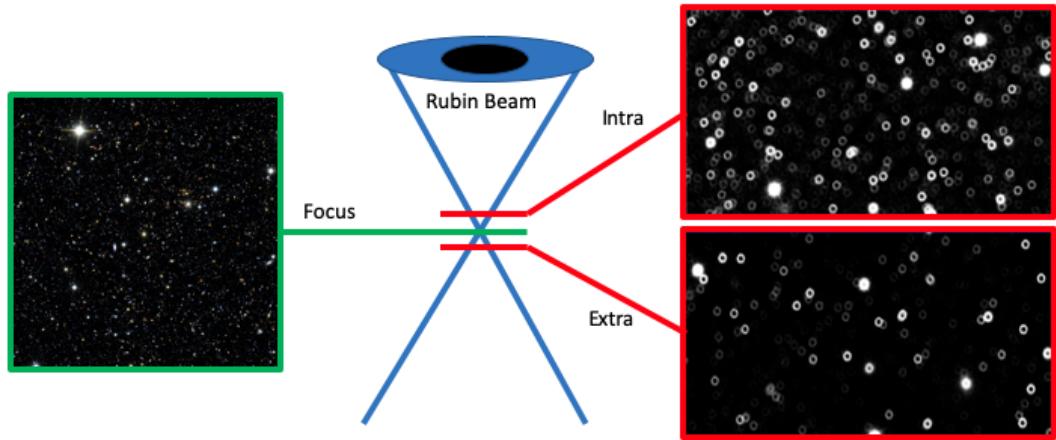


Figure 2.5: The images formed from placing sensors at different positions along the optical path. The optical path from the pupil to focus is shown in blue. The image produced at focus is shown in green. The intra-focal and extra-focal images are shown in red. The donut shape stems from the annular entrance pupil.

Chapter 3

Simulating High Fidelity Rubin Donut Images

Perhaps one day we will have machines that can cope with approximate task descriptions, but in the meantime, we have to be very prissy about how we tell computers to do things.

Richard Feynman

Simulation is a cornerstone of this work. The Rubin Observatory is not complete, which in some sense necessitates using simulated data. However, even if the observatory was complete and we could collect real data, it would still be challenging to adequately sample the 50 dimensional state space of the telescope. In real life, to configure the Rubin telescope in a perturbed state takes on the order of 10 seconds. In code, initializing a telescope instance in a specific configuration takes single digit microseconds. In real life, there is only one telescope, whereas using the Sherlock cluster at Stanford University, we can initialize the telescope in up to roughly one thousand parallel processes. In short, we exploit simulation to create richer datasets to train and assess our framework.

In the first section, we describe the details of simulating a single donut. In the following sections, we describe the two datasets used in this work.

3.1 Simulating A Single Donut Image

We use high fidelity Rubin Observatory image simulations to train our model and analyze its performance. By selecting different observational, atmospheric, and telescope parameters for each donut, we produced a simulated training dataset that spans a wide variety of conditions. The output of each simulation i is a 256×256 pixel donut image D_i , and the corresponding true local wavefront $\alpha_i \in \mathbb{R}^{18}$. Figure 3.1 shows an example pair of end products.

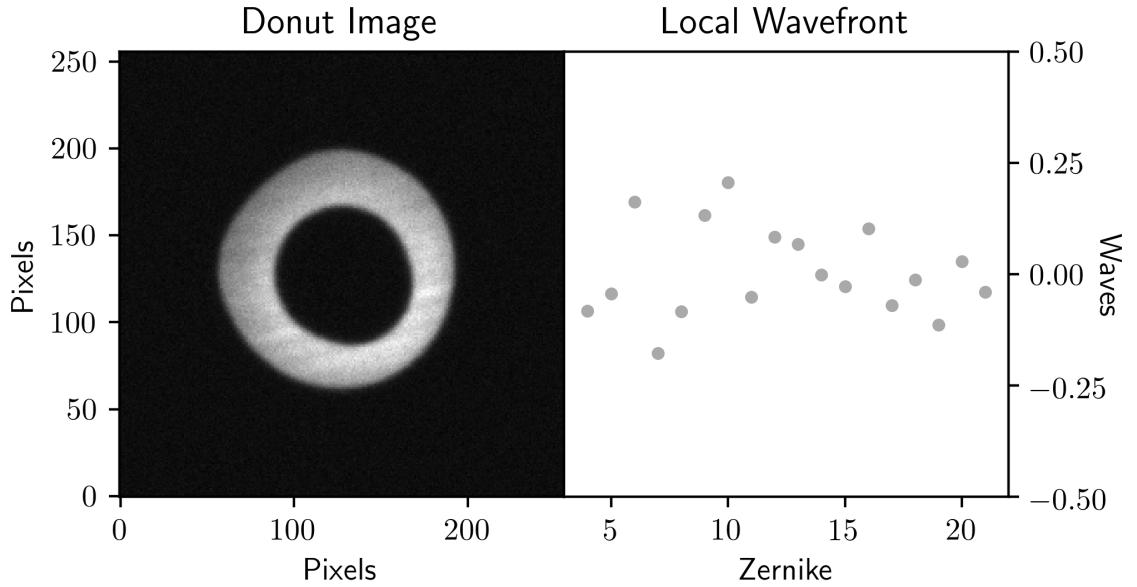


Figure 3.1: *Left:* a simulated 256×256 pixel donut image at the (x, y) field position $(-1.1048^\circ, 1.0868^\circ)$. *Right:* the corresponding local wavefront annular Zernike coefficients ($Z_4 - Z_{21}$ inclusive), measured in units of 622.20 nm waves.

The simulations are comprised of sequential steps that trace a photon's journey from a distant source to a count in a pixel in an image. The first step is to generate photons in the sky corresponding to a source from the observation. We create a baseline survey to draw observations from. The survey is comprised of 5497 randomly drawn r-band visits from the Operations Simulator (OpSim) [35]. Figure 3.2 shows the distribution of the observations across the sky. This survey samples a wide range

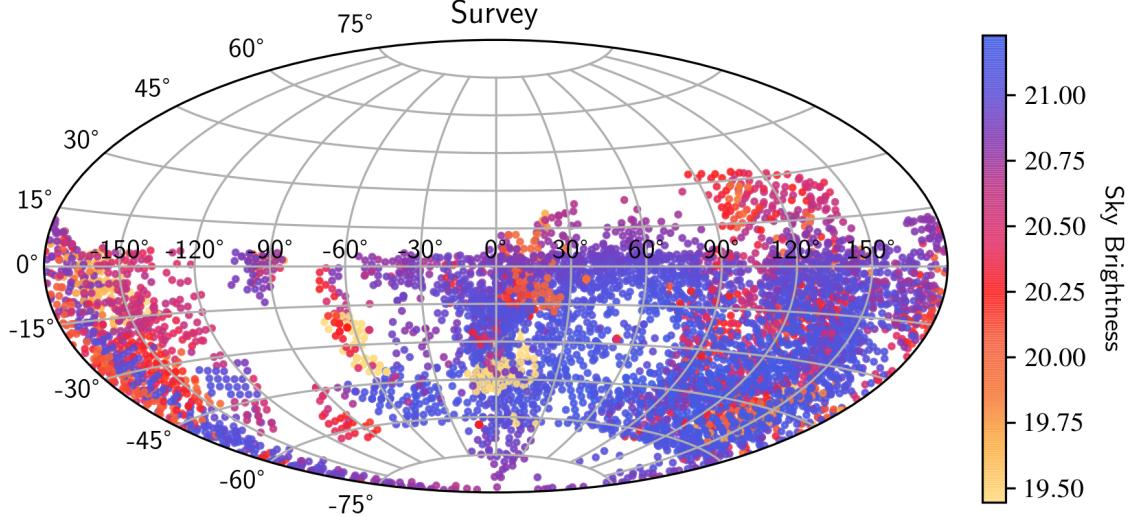


Figure 3.2: The right ascension and declination of the observations we use for our simulations. Each point corresponds to an observation; the color shows the variation in sky brightness.

of sky brightnesses, atmospheric seeing, and sky positions.

For each of these visits, we determine the fields of view covered by the four corner wavefront sensors. Then we query the Gaia DR2 catalog for sources within these regions and with G magnitudes less than 18 [36]. The mean number of sources per catalog is 1101 with a standard deviation of 1720. The largest catalog contains 12684 sources. Once we have the catalog, we convert the Gaia G , G_{bp} , and G_{rp} magnitudes to the SDSS r magnitude based on Equation 3.1 from the Gaia data release documentation [37].

$$x = G_{bp} - G_{rp} \quad (3.1)$$

$$r = G + 0.12879 - 0.24662 \cdot x + 0.027464 \cdot x^2 + 0.049465 \cdot x^3 \quad (3.2)$$

We clip SDSS r magnitudes below 14 to save simulation time. The simulation starts by selecting a source and generating the corresponding photons. The number of

photons is determined by the SDSS magnitude and the wavelengths are drawn from a black body distribution with spectral radiance,

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \quad (3.3)$$

based on the temperature of the source. After the photons are generated, the next stage of simulation propagates them through the atmosphere.

The primary contribution from the atmosphere is atmospheric turbulence. We use the GalSim Python package to split the atmospheric turbulence power spectrum into low and high spatial frequencies and simulate their contribution separately [38]. The low spatial frequencies modulate the centroid of the source photons. The high spatial frequencies scatter the photons around this centroid, sometimes called the *second kick* [39]. The phase screens are initialized with altitudes and velocities from atmospheric tomography profiles [40]. The Fried parameter for the atmosphere is determined by the airmass and seeing from the OpSim observation and by the central wavelength of the *r* bandpass, 622.20 nm.

In addition to the turbulence contribution, we also incorporate chromatic seeing and differential chromatic refraction [41]. Both of these are functions of the wavelengths of individual photons. We model chromatic seeing by scaling the turbulence induced perturbations based on wavelength. We model chromatic refraction by refracting the photons along the meridian based on the airmass. At the conclusion of this step, there is a bundle of photons at the entrance pupil of the telescope.

Next, we trace photons from the entrance pupil of the primary mirror to the plane of the detector. The batoid Python package was used to construct model Rubin telescope instances and perform the ray tracing [42]. We constructed telescope instances that have 50 degrees of freedom that are analogous to those in Table 2.1. For simplicity, we used 20 Zernike polynomials ($Z_4 - Z_{23}$) to represent the surface bending modes. We skip the first three Zernike modes because these surface perturbations would be degenerate with the hardpoint control parameters.

In order to determine a reasonable range for the degrees of freedom, we first computed the single parameter excitation levels, shown in Table 3.1, that induce

the global wavefront to have root-mean-square error of 0.3 wavelengths, or *waves*. We set the amplitude of a given parameter by drawing from a normal distribution that is centered on the center of the range and has a standard deviation of half the range multiplied by $\sqrt{5/50}$. This means that the total error across all the control parameters will be roughly 5 times the error that on an individual parameter would make the global wavefront 0.3 waves (where a wave is the central wavelength of the r bandpass, or 622.20 nm).

Once the photons are at the detector, we simulate sensor effects. We use GalSim to simulate the aggregation of photons into silicon pixels. The silicon model includes electron diffusion to account for the nonlinear brighter-fatter effect [43]. We also use GalSim to introduce shot noise from the sky background and simulate the camera readout. Then we zero both individual pixels and columns following sensor statistics taken in the lab. At this point, we have a 256×256 pixel digital image of counts like the one shown in Figure 3.1.

For each donut image, we also simulate the corresponding local wavefront. We use the Batoid package to compute the wavefront at the donut's field position x, y . The local wavefront α is represented by the 18 annular Zernike coefficients which have the largest impact on image quality ($Z_4 - Z_{21}$ inclusive). These coefficients serve as the labels for the neural network described in the next Chapter.

In addition to donuts from individual stars, we also simulated blended donuts. These arise when the donut images from two or more stars overlap. In these simulations, the photons from all the overlapping stars are aggregated on the sensor, then the sensor effects are simulated once. In crowded fields, the majority of the donuts will be blends. This data helps us assess this scenario and the density limits of the fields where we can successfully apply our algorithm.

In the next two Sections we describe the datasets used to train the neural network and to test the full framework, respectively.

Table 3.1: The range in each degree of freedom where the root-mean-square of the global wavefront is below 0.3 waves (one wave is 622.20 nm). These ranges are used to create realistically deformed telescope instances in our simulations. Notation: dx is a translation in the x-axis; rx is a rotation along the x-axis; Z_4 is the fourth annular Zernike polynomial over the annulus of the corresponding surface.

| Parameter | | Min | Max | Parameter | | Min | Max |
|--------------------|--|-------|------|------------------|--|-------|------|
| Camera dx (um) | | -580 | 580 | M2 dx (um) | | -118 | 118 |
| Camera dy (um) | | -580 | 580 | M2 dy (um) | | -118 | 118 |
| Camera dz (um) | | -10.7 | 5.39 | M2 dz (um) | | -11.1 | 11.1 |
| Camera rx (urad) | | -42.0 | 42.0 | M2 rx (urad) | | -17.6 | 17.6 |
| Camera ry (urad) | | -42.0 | 42.0 | M2 ry (urad) | | -17.6 | 17.6 |
| M1M3 Z_4 (nm) | | -111 | 56.4 | M2 Z_4 (nm) | | -59.1 | 117 |
| M1M3 Z_5 (nm) | | -42.4 | 42.4 | M2 Z_5 (nm) | | -68.9 | 68.9 |
| M1M3 Z_6 (nm) | | -42.4 | 42.4 | M2 Z_6 (nm) | | -68.9 | 68.9 |
| M1M3 Z_7 (nm) | | -65.0 | 65.0 | M2 Z_7 (nm) | | -76.4 | 76.4 |
| M1M3 Z_8 (nm) | | -65.0 | 65.0 | M2 Z_8 (nm) | | -76.4 | 76.4 |
| M1M3 Z_9 (nm) | | -45.9 | 45.9 | M2 Z_9 (nm) | | -70.6 | 70.6 |
| M1M3 Z_{10} (nm) | | -45.9 | 45.9 | M2 Z_{10} (nm) | | -70.6 | 70.6 |
| M1M3 Z_{11} (nm) | | -130 | 66.2 | M2 Z_{11} (nm) | | -62.8 | 70.4 |
| M1M3 Z_{12} (nm) | | -51.5 | 51.5 | M2 Z_{12} (nm) | | -68.3 | 68.3 |
| M1M3 Z_{13} (nm) | | -51.5 | 51.5 | M2 Z_{13} (nm) | | -68.3 | 68.3 |
| M1M3 Z_{14} (nm) | | -48.1 | 48.1 | M2 Z_{14} (nm) | | -73.6 | 73.6 |
| M1M3 Z_{15} (nm) | | -48.1 | 48.1 | M2 Z_{15} (nm) | | -73.6 | 73.6 |
| M1M3 Z_{16} (nm) | | -73.6 | 73.6 | M2 Z_{16} (nm) | | -68.0 | 68.0 |
| M1M3 Z_{17} (nm) | | -73.6 | 73.6 | M2 Z_{17} (nm) | | -68.0 | 68.0 |
| M1M3 Z_{18} (nm) | | -49.7 | 49.7 | M2 Z_{18} (nm) | | -64.8 | 64.8 |
| M1M3 Z_{19} (nm) | | -49.7 | 49.7 | M2 Z_{19} (nm) | | -64.8 | 64.8 |
| M1M3 Z_{20} (nm) | | -49.2 | 49.2 | M2 Z_{20} (nm) | | -76.5 | 76.5 |
| M1M3 Z_{21} (nm) | | -49.2 | 49.2 | M2 Z_{21} (nm) | | -76.5 | 76.5 |
| M1M3 Z_{22} (nm) | | -67.7 | 51.5 | M2 Z_{22} (nm) | | -59.0 | 57.2 |
| M1M3 Z_{23} (nm) | | -51.3 | 51.3 | M2 Z_{23} (nm) | | -65.7 | 65.7 |

3.2 Donut Dataset

The donut dataset consists of samples with three components: the simulated donut image D , the focal plane position r - which includes distance along the focal z axis, and the true local wavefront α . The dataset is used to train, validate, and test models for the task,

$$(D, r) \rightarrow \alpha \quad (3.4)$$

Modern neural networks, like the one we develop in Chapter 4, typically have millions of parameters. In order to constrain these parameters and simultaneously avoid overfitting, these models must be trained on copious amounts of data.

For each of the first 5,000 visits from the observing schedule, we simulate 200 individual donut images. The most computationally expensive part of these simulations is generating the atmospheric phase screens. We speed up the simulations by using one atmosphere for every 100 donuts. Telescope instantiation, on the other hand, costs a few microseconds, so we draw a different 50-dimensional perturbation to the nominal telescope state for each donut image. We also simulate 100,147 blends from the same observations.

Following the standard machine learning cross validation paradigm, we break this data up into training, validation, and test segments. 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. Figure 3.3 shows 100 example donuts from the training set.

3.3 Full Visit Dataset

The full problem involves going from n donut images and positions to the 54 coefficients β ,

$$(D_1, r_1), \dots, (D_n, r_n) \rightarrow \alpha_1, \dots, \alpha_n \rightarrow \beta \quad (3.5)$$

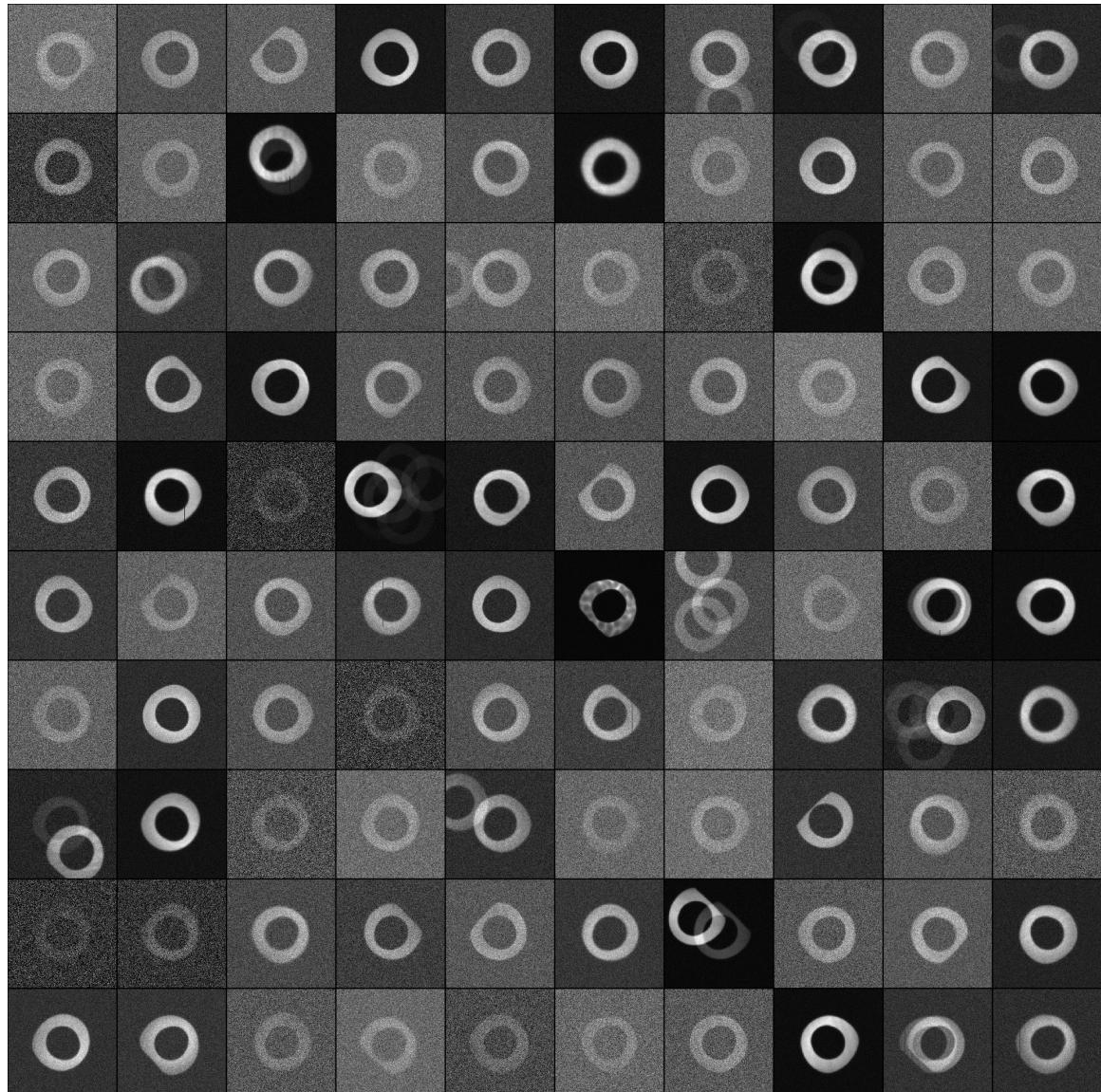


Figure 3.3: Example donuts from the training set. The different degrees of vignetting, missing columns, blending, and turbulence can be seen with the naked eye.

In addition to the donut dataset, we also need a dataset of full visits, where each visit contains many donut images, so that we can assess how well our framework predicts β .

Each sample in this dataset consists of all the donut images and positions in the observation, plus the full optics wavefront β . We used the last 497 observations from the scheduler simulations for the observations. For each visit, we simulated all the stars, and blends, from the Gaia catalog that fall on the corner wavefront sensors. The mean, median, and maximum number of sources in a sample is 787, 222, and 4,998 respectively. Figure 3.4 shows an example sample.

All the donuts within a single observation are simulated with the same atmosphere and sky background. Otherwise, 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.

The largest difference between observations is the density of sources in the field. The number of sources in a sample ranges from 111 to 4998. On average, 32 percent of the sources are blends. However, for the crowded fields this number can be over 90 percent. Figure 3.5 shows how the fraction of sources that are blends increases with the total number of sources in the visit. Thus, while for less crowded fields ample un-blended stars are available, for crowded fields, the wavefront sensing algorithm's ability to process blends is critical.

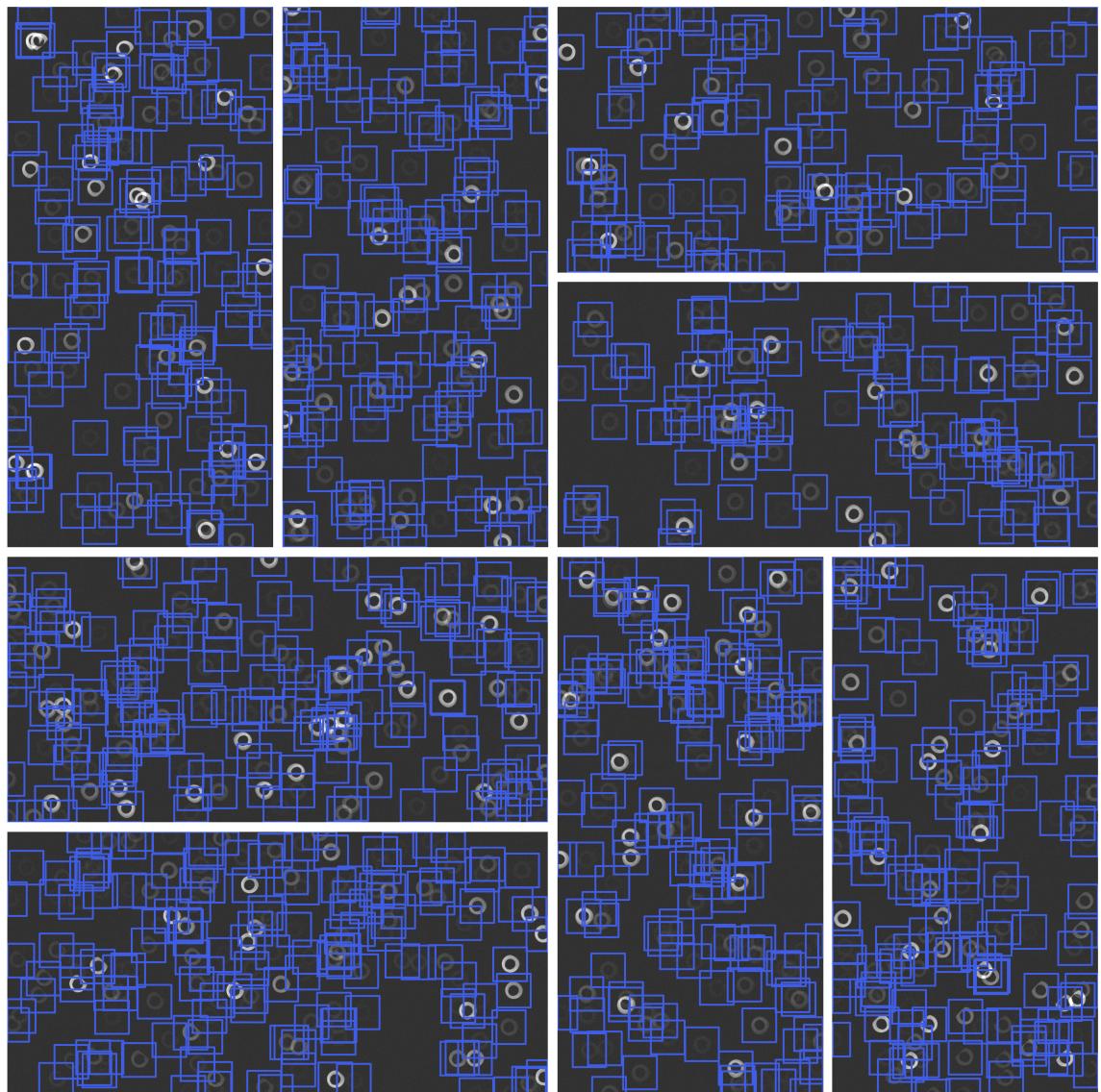


Figure 3.4: The eight wavefront sensor half-chips in an example sample from the full visit dataset. This sample contains 1112 donut images which are highlighted in blue boxes.

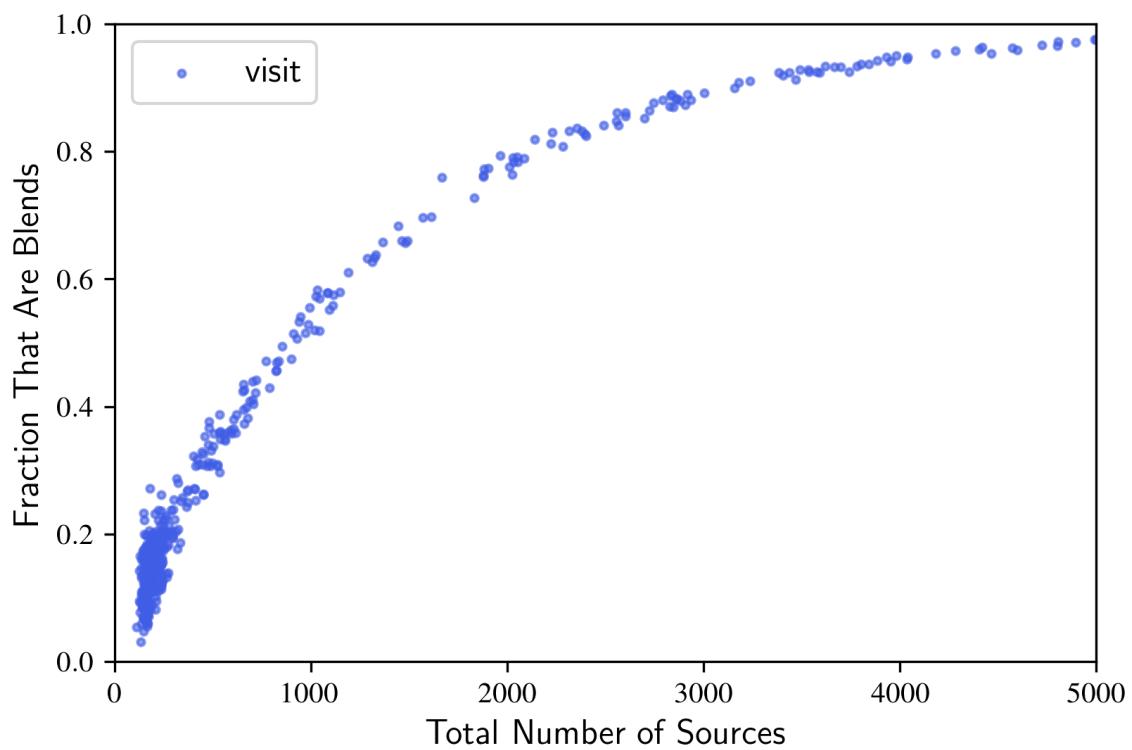


Figure 3.5: The fraction of sources that are blends versus the total number of sources for the 497 visits in the full visit dataset.

Chapter 4

Training a Convolutional Neural Network To Predict The Local Wavefront

Uncrumpling paper balls is what machine learning is about: finding neat representations for complex, highly folded data manifolds.

François Chollet

In the first step, we use a convolutional neural network (CNN) to predict the relationship between donuts and their corresponding local wavefronts. There are three inputs to the network: the 256×256 pixel donut image D , the position r of the sensor. This position $r = (x, y, d)$ is comprised of the (x, y) field position in radians, and the offset d along the z-axis indicating whether the defocus is intra-focal or extra-focal. The output is the vector of Zernike coefficients $\alpha \in \mathbb{R}^{18}$. The sub-problem is to find a function approximator ϕ ,

$$\phi(D, r) \rightarrow \alpha(x, y) \text{ such that } W(u, v|x, y) \approx \sum_{j=4}^{21} \alpha_j(x, y) Z_j(u, v)$$

Predicting the local wavefront $W(u, v|x, y)$ from donuts and their metadata is an intrinsically challenging problem because the optics contribution to the image is

sometimes degenerate with, and typically subdominant to, the atmospheric contribution. Thus, even for very bright donuts with no effective sources of background noise, the model will have a systematic ceiling. In this section, we demonstrate that a CNN can produce accurate estimates despite this challenge.

4.1 Network Architecture

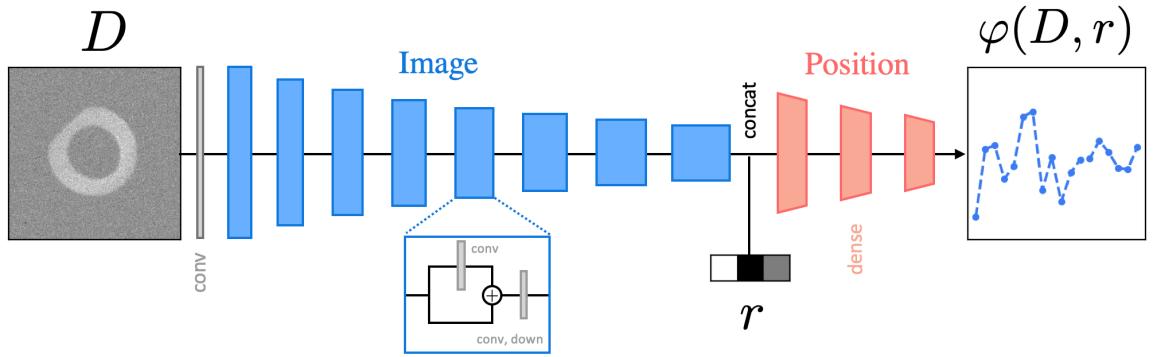


Figure 4.1: The network architecture and example inputs and output. The image network is in blue and the position network is in red.

Modern CNN architectures are governed by an astounding number of design decisions and hyper-parameter options. Here we show that a simple network can make highly accurate wavefront estimates without many additional design or hyper-parameter optimizations. This shows that this step can also serve as a network in other optics algorithms.

The network, shown in Figure 4.1, has two sub-networks: an image network and a position network. We will describe the architecture through the convolutions that are applied and how they change the data tensors. Each 2-dimensional convolution or linear layer, except the final linear layer, is followed by a 2-dimensional batch normalization layer (batchnorm) and a rectified linear activation (ReLU) [20, 44, 45]. The batch normalization protects against covariate-shift and provides weak regularization and the ReLU activation introduces non-linearity.

The image network reduces the donut image to a 1024 dimensional vector. The first component is a convolution that increases the depth of the input from 1 to 8, and retains the 256 by 256 height and width. This is followed by a sequence of 8 Down-Blocks. Each Down-Block has an initial convolution that does not change the tensor dimensions, followed by a convolution that increases the depth by a factor of 2 and reduces the height and width by a factor of 2. There is a skip connection between these two convolutions for better gradient flow [46]. In the 8th and final Down-Block, the depth is not increased. Thus by the end of these steps we have a depth of $\text{depth} = 8 \cdot 2^7 = 1024$ and $\text{height} = \text{width} = 256 \cdot 2^{-8} = 1$. This $1 \times 1 \times 1024$ output is a learned representation of the input image. In total, the image network has 28,326,360 trainable parameters.

The position network combines the $1 \times 1 \times 1024$ output from the image network vector with the 3-dimensional position input $r = (x, y, d)$ and estimates the 18 local wavefront coefficients. It first concatenates and flattens these tensors into a combined 1027-dimensional tensor. Then three linear layers reduce the dimensionality to 266, 69, and 18 respectively (the sequential dimensions are log-linear). After the first linear layer we insert a dropout layer with a 90% keep probability [21] to improve generalization. In total, the position network has 293,801 trainable parameters.

We developed the model in PyTorch [47]. We determined this balance of parameters between sub-networks through testing. In the next Section we describe how the network is trained.

4.2 Training, Validation, and Testing

In addition to the large number of decisions that go into the network architecture, there are also a lot of decisions on the configuration of the training. For the loss function, we use the mean-squared-error (MSE) between the estimated and true wavefront coefficients and found it to be superior to the mean-absolute-error. We evaluate batches of 64 samples at a time and use the Adam optimizer to compute the parameter updates for the backward pass after computing the loss function in the forward pass [48]. We train the model for 8 epochs over the donut dataset. After each epoch

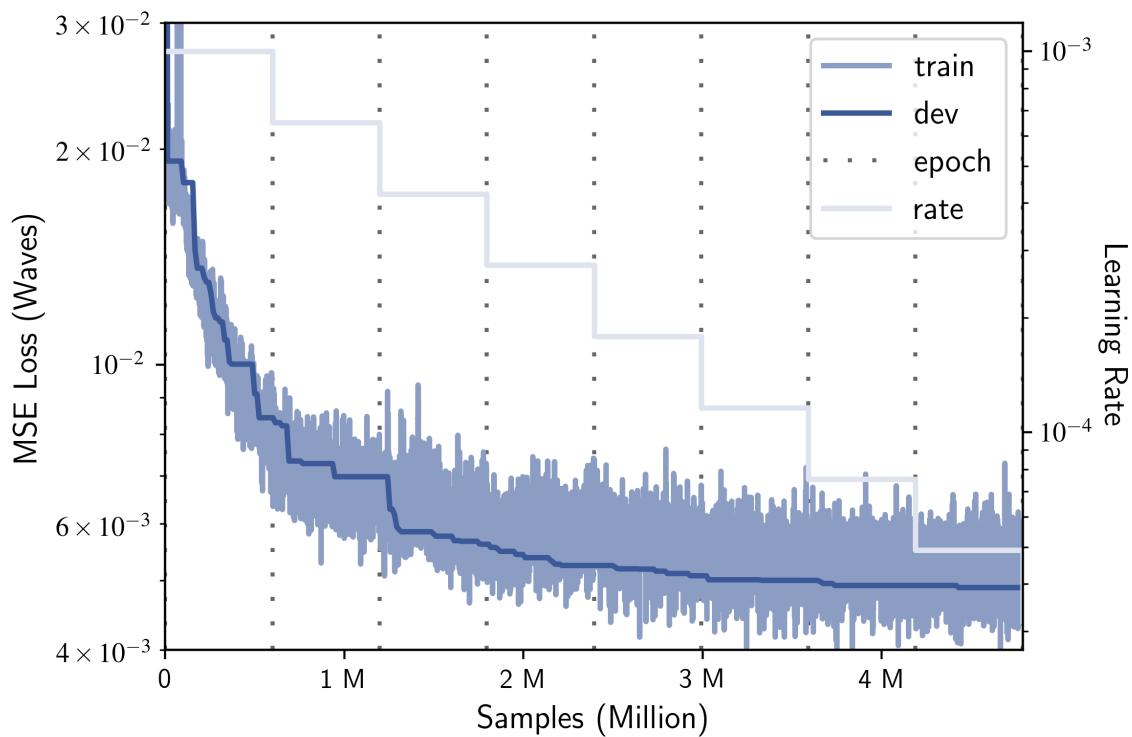


Figure 4.2: The training loss (train), validation loss (dev), and learning rate versus the number of samples used for training the neural network. The MSE loss corresponds to the y-axis on the left; the learning rate corresponds to the y-axis on the right.

we multiply the learning rate by a factor of 0.65. Every 200 batches, we evaluate the MSE of the model on the donut validation set, and keep the best model. Figure 4.2 shows the training and validation (dev) errors, and the learning rate, versus the number of training samples.

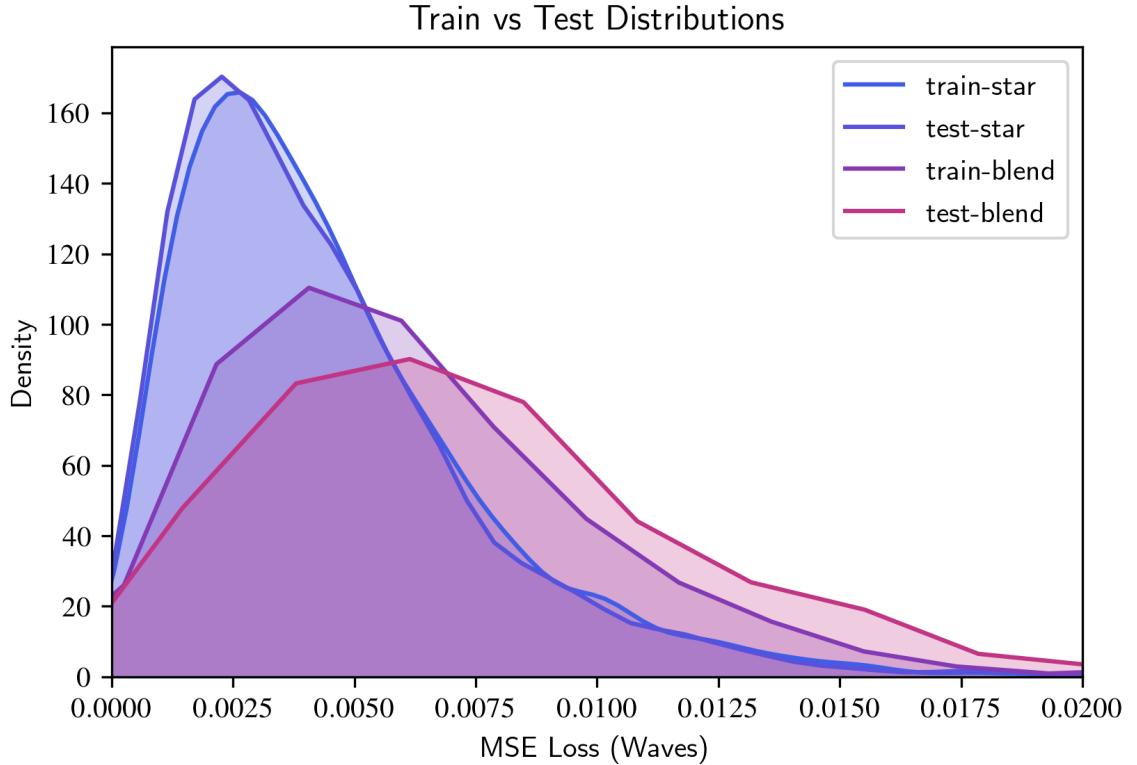


Figure 4.3: The distributions of the training and test MSE loss on stars (donuts with no overlaps) and blends (donuts with overlaps).

We went through multiple iterations with different training configurations to arrive at this final configuration. To assess the models we used the validation set. Once we finalized the training configuration, we used the held out test set for a final assessment of the network. We compared the MSE loss on the training and test sets to check for overfitting. Figure 4.3 shows the comparison. The training and test set MSE losses 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.

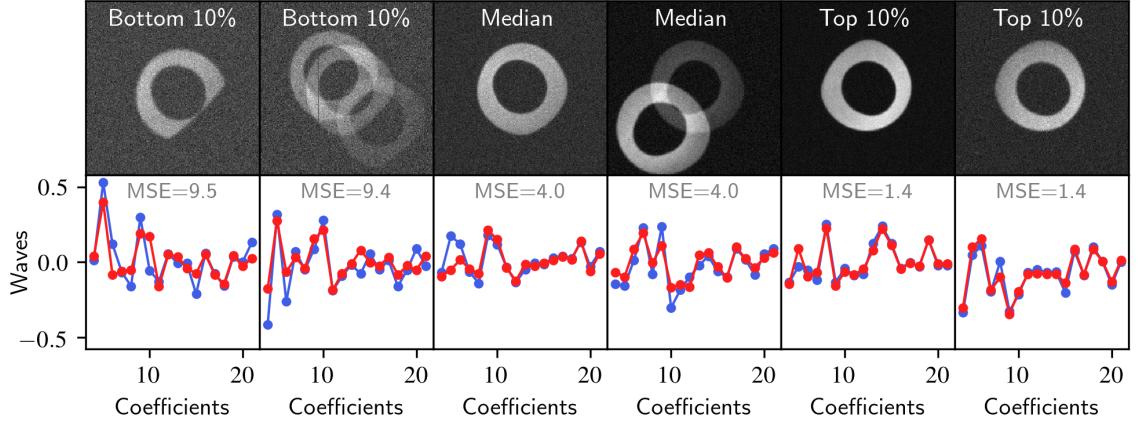


Figure 4.4: Top: six donut images; two from the bottom 10% of the MSE distribution, two from the median, and two from the top 10%. Bottom: the corresponding true (blue) and neural network estimated (red) local wavefronts. The MSEs in the bottom row are in units of thousandths of waves.

We also inspected the performance of the network on many samples. Figure 4.4 shows the performance on six donuts from across the MSE loss distribution. At the bottom decile, the network achieves a MSE of 9.5 thousandths of a wave. While benchmarks do not exist for this task, based on our experience with other approaches and previous generation telescopes, this already seems like a competitive result. The median MSE is 4.0 thousandths of a wave. Here a few differences between the true and estimated coefficients can be seen by eye. At the top decile, the network achieves a MSE of 1.4 thousandths of a wave. The match between the true and estimated coefficients is astounding. It is perhaps more astounding to consider that the network is able to achieve this when the optics contribution that we are estimating is sub-dominant to the atmospheric contribution.

It is also important to examine the correlation between the errors on different wavefront coefficients to see if there are systematic errors and biases. Figure 4.5 shows the covariance and correlation matrices for the error on the different wavefront coefficients. The covariance diagonal matches with the coefficients that are most excited in our donut dataset. For example, the Z_5, Z_6 astigmatism coefficients have the

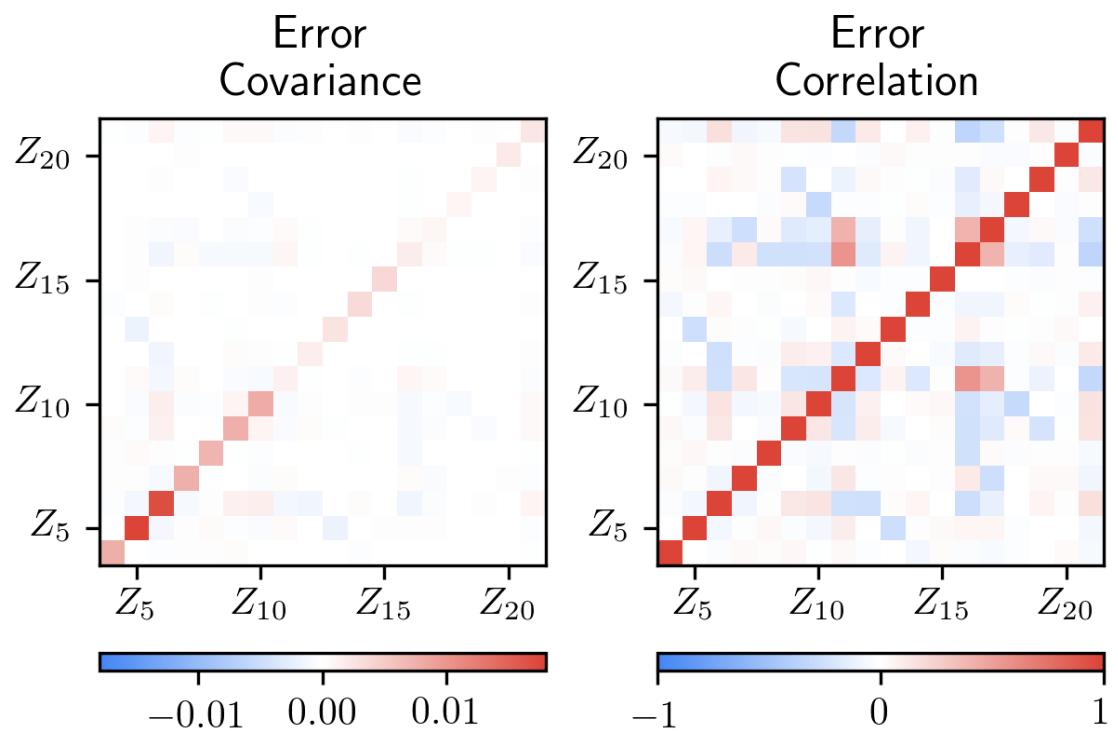


Figure 4.5: The covariance (left) and correlation (right) of the error between different local wavefront coefficients.

largest error, but also have the largest values in the true wavefront. The correlation matrix shows that the errors are fairly independent across coefficients. The largest correlations are 0.57 between Z_{11}, Z_{16} ; 0.41 between Z_{11}, Z_{17} ; 0.40 between Z_{16}, Z_{17} ; and -0.34 between Z_{16}, Z_{21} .

There are a lot of parameters that determine a donut image. Another natural question is whether any of these are correlated with the error. Figure 4.6 shows 8, out of the many, error sources we examined. We found that the error generally increases when the background noise increases or foreground number of photons in the donut decreases, as expected. The ratio of these contributions serves as a proxy for the signal-to-noise ratio.

Interestingly, the seeing does not have much impact, suggesting that the network is really able to differentiate between PSF contributions due to the atmosphere and due to the optics. Before training the network, we had also expected the radial distance, a proxy for the amount of vignetting, to be correlated with the error. However, in practice, it seems like the network is able to utilize the (x, y) coordinates in the position to account for this effect and process vignetted donuts accurately.

The biggest surprise was the impact of chromatic effects. The temperature of the source is reasonably correlated with the error. The fact that there is also some error correlation with airmass suggests some of this stems from differential chromatic refraction and some from the variations in wavelength across Rubin’s r-passband and how this contributes to PSF structure. In future versions of this model, one could imagine training the network to also predict the temperature, or peak wavelength, of the source. This would better encourage it to differentiate the chromatic effects during training.

Finally, there is a clear difference between the stars (sources with no overlaps) and blends (sources with overlaps). The network performs much worse on the blends. This has important implications for which sources, and even fields, to do wavefront sensing in. As the number of overlapping neighbors increases, the error increases. However, it is important to note that there are fewer samples with large numbers of overlapping neighbors so these distributions are less reliable. In Section 5, we explore whether ignoring blends improves the performance of the global wavefront

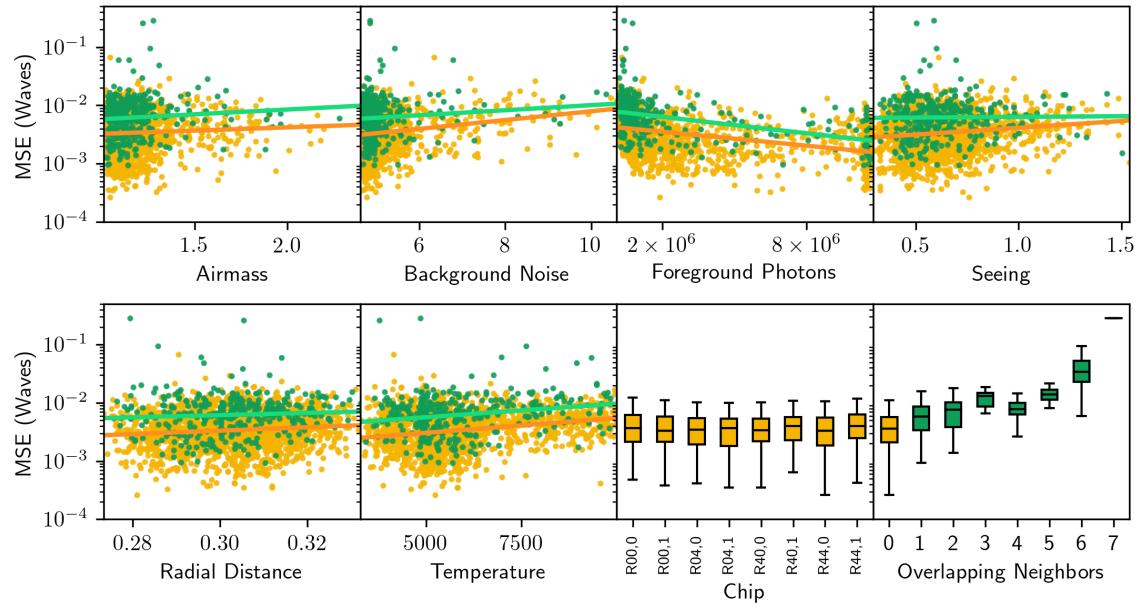


Figure 4.6: The MSE, in waves, versus the feature distributions for 8 different potential sources of error. The stars are highlighted in yellow; the blends are in green. We also show the best fit line for the 6 non-categorical features. Units: Background Noise is the standard deviation of photons in the background, Foreground Photons is the total number of photons in the donut, Seeing is in arcseconds, Radial Distance is in meters, Temperature is in Kelvin.

interpolation.

4.3 Neural Network Enabled Insights

At a high level, neural networks are differentiable and parameterized function approximators. The same properties that make them trainable, make them effective tools for exploring tasks in new ways. Here our network φ has been trained to solve the task $\varphi(D, r) = \alpha$. In the next three subsections, we perform mathematical manipulations on φ to gain insight into the behavior of our network and more general properties of this problem.

4.3.1 Saliency Maps

What aspects of the donut images is the neural network using to make its predictions? Following in the footsteps of [49], we define a metric that shows the pixels the predictions of the network are most sensitive to. We call this *attention* and compute it by taking the magnitude of the gradient of the norm of the local wavefront prediction with respect to *the pixels*. The explicit formula is

$$\text{Attention} = |\nabla_D||\varphi(D, r)||_2|$$

This gradient with respect to pixels is similar to the mechanism behind style transfer [50].

Figure 4.7 shows the attention for six samples. The first pattern that emerges is that the edges dominate, particularly when there is significant vignetting. The network focuses on the outline of a donut, in tandem with its position, to determine the wavefront. This suggests it may be possible to develop a successful network that takes only the outline of the donuts as input. It is not clear how we would have picked up this insight from the other approaches. This is an example of how the neural network can teach us new things about a problem and help us discover new approaches.

The second pattern is that, in the case of blends, the network either utilizes or

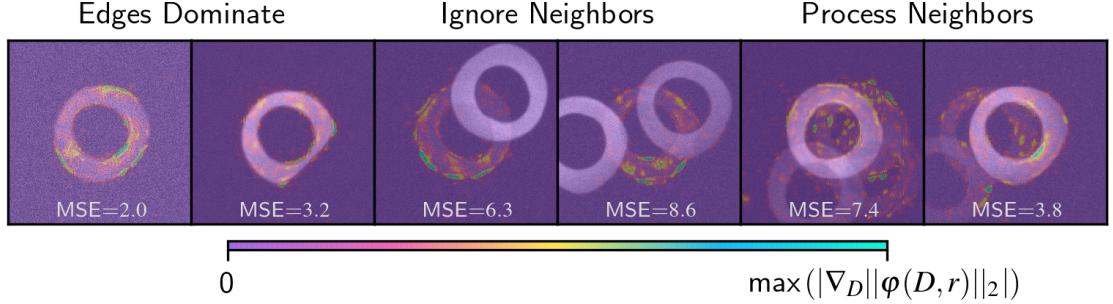


Figure 4.7: The neural network attention values for six samples.

ignore the neighboring donuts. It is curious that this behavior is so bimodal. For example, where in the network is this decision being made? We suspect answering this question could unlock ways to further improve the performance of the network on blends.

4.3.2 Memorizing Versus Learning

A common grievance of opponents to machine learning being applied in physics is that the network is memorizing, rather than truly learning, the relationships between the input and output. Here, we test this. We examine the intensity changes that would most quickly move the predictions on a base donut towards the Zernike of interest, which we deem the label-shift. We take the gradient of the norm of the difference of the network output and the target Zernike unit vector, or explicitly

$$\text{Label-Shift} = \nabla_D ||\varphi(D, r) - Z_i||_2$$

Figure 4.8 shows the results for the Zernike unit vectors Z_4 through Z_{21} under the respective annular Zernike polynomial. 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. For example, the network even learns the 5 cycle periodicity of Z_{20} and Z_{21} . It also learns the complicated radial dependence of Z_{18} . This demonstrates that the network is engaging in higher order learning, where

it is learning general features of the problem space.

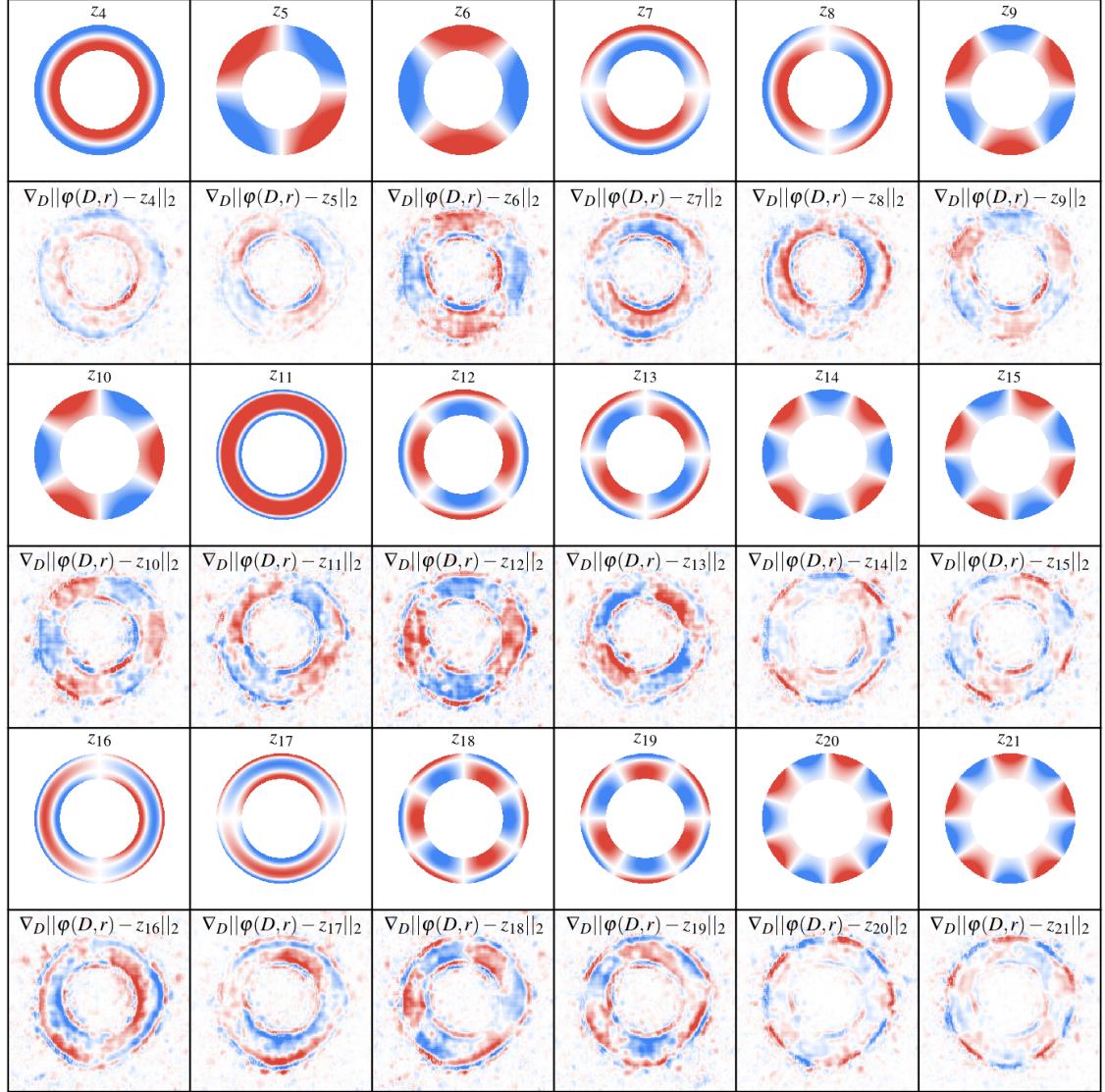


Figure 4.8: The odd rows show the annular Zernike polynomials Z_4 through Z_{21} . The even rows show the neural network label-shifts for the corresponding Zernike unit vectors.

4.3.3 Adversarial Tradeoffs

The input space for our problem, the image and the position, has $256 \times 256 + 3 = 65539$ dimensions. This is an extremely high dimensional space. It is not hard to find subtle changes in this space that are hardly discernable *by eye*, or at a global level, but which can dramatically change the output of the network. In [51], the authors show how to use these adversarial examples to improve training. Here, we examine a few adversarial examples to understand how our network could be hacked.

An attacker has two competing objectives. The first is to find a modified donut image D' that is similar to D ; the second is to find a modified donut image D' such that $\varphi(D', r)$ is close to some target $t' \in \mathbb{R}^{18}$. The Adversarial-Objective is,

$$\text{Adversarial-Objective} = \|\varphi(D', r) - t'\|_2^2 + \|D - D'\|_2^2$$

In order to find such a D' we can minimize this objective function with respect to D' . We use the Adam optimizer to perform gradient descent. Two example results are shown in Figure 4.9. In both cases we only change one coefficient in the target t' .

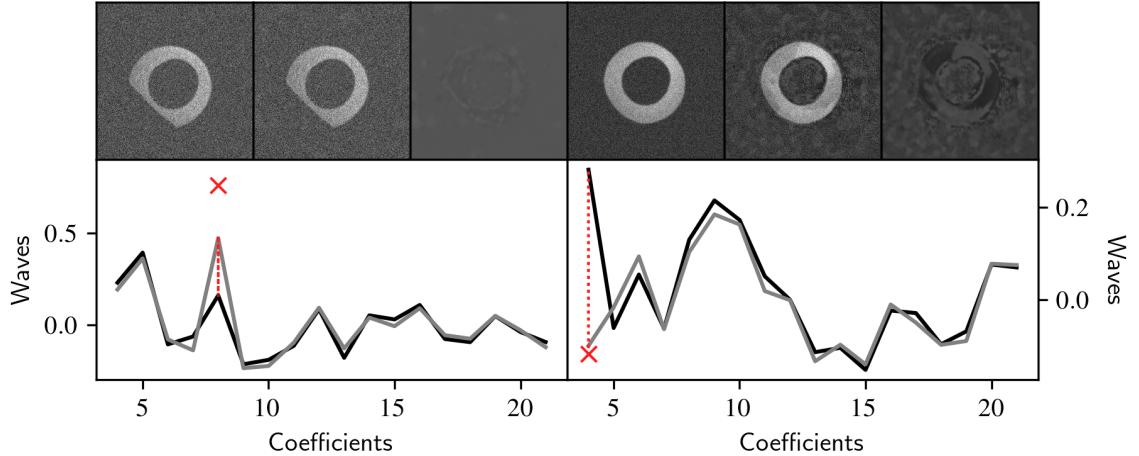


Figure 4.9: Two example adversarial attacks. For each of the attacks we show three donut images, from left to right: the original donut image D , the new donut image D' , and the difference $D' - D$. Below we show the original prediction $\varphi(D, r)$ in black, the target t' as a red “x”, and the new prediction $\varphi(D', r)$ in gray.

In practice, one must scale the ratio of the two terms in the Adversarial-Objective to get good convergence. This ratio determines the relative priority of similarity between D and D' versus $\varphi(D', r)$ and t' . In the example on the left, we prioritized the similarity of the donuts. Here, the two donuts appear the same by eye. However, the coefficient for Z_8 is increased significantly, although not all the way to the target - while keeping most of the other coefficients close to their original values. This highlights the power of this kind of attack.

In the second example, the difference between the donut images is clearly discernible. Here the Z_4 coefficient changes dramatically, and largely hits the target, and even changes signs. This shows how the prediction can be manipulated to the target with sufficient changes to the donut.

In the context of the Rubin Observatory, a closed environment, adverse attacks need not be feared. However, the algorithm we develop here serves a general building block for other wavefront sensing methods. These algorithms are also used for defense purposes. In this context, the sensitivity to adversarial attacks is an important characteristic to be aware of, and potentially protect against.

Chapter 5

Interpolating The Global Wavefront with Least Squares

In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual.

Galileo Galilei

5.1 The Interpolation Geometry

In the second step, we interpolate the global wavefront from the local wavefront predictions from the first step. We interpolate the global wavefront coefficients β_{jk} based on the predictions α from the previous step. More explicitly, we use ordinary least squares (LS) to solve

$$\text{LS}(\alpha(x_1, y_1), \dots, \alpha(x_n, y_n)) \rightarrow \beta_{jk} \text{ such that } W(u, v, x, y) = \sum_{j=4}^{21} \sum_{k=1}^3 \beta_{jk} Z_j(u, v) Z_k(x, y)$$

In the previous step, the 18 predictions $\{\alpha_4(x_i, y_i), \dots, \alpha_{21}(x_i, y_i)\}$ for a source image D at position (x_i, y_i) . In this step, we start by aggregating the predictions for a single coefficient j for all n donuts in the field of view of the observation, or

$\{\alpha_j(x_1, y_1), \dots, \alpha_j(x_n, y_n)\}$. Then the problem reduces to minimizing the distance between the α_j estimates and $\sum_{k=1}^3 \beta_{jk} Z_k(x, y)$ for each index separately. Figure 5.1 shows the true and predicted α_4 values on the four wavefront sensors.

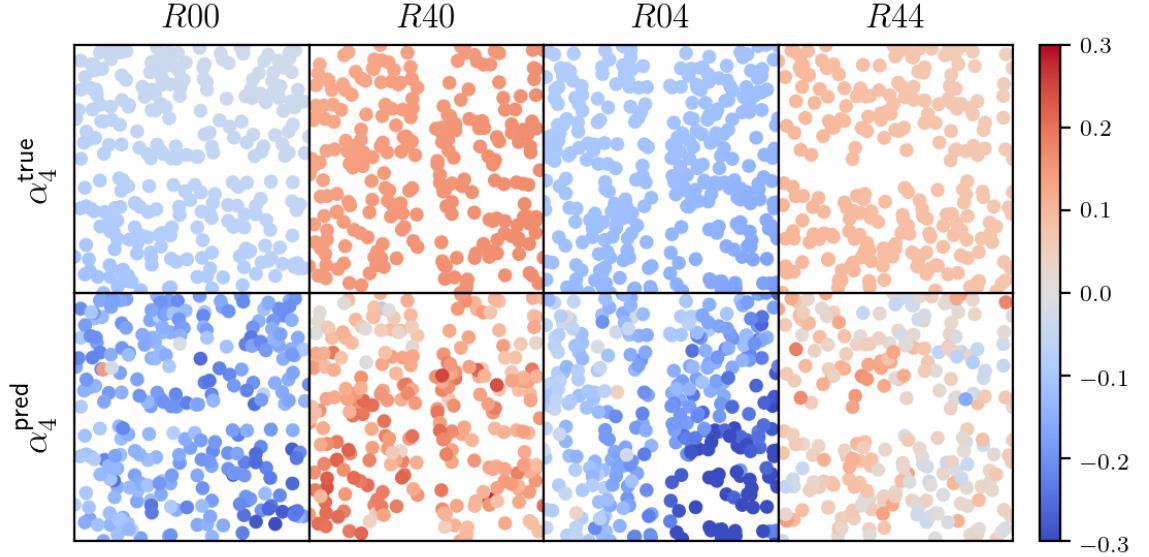


Figure 5.1: The two rows correspond to the true (α_4^{true}) and predicted ($\alpha_4^{\text{predicted}}$) Z_4 coefficients. The four columns correspond to the four corner wavefront sensors. The colormap is in units of waves.

The first three Zernike coefficients establish a plane. Thus we are effectively fitting a plane by the three coefficients along the k index, for each local wavefront Zernike coefficient, or j index. Figure ?? shows how $\beta_{41}, \beta_{42}, \beta_{43}$ are fit from the local α_4 predictions.

The explicit minimization is

$$\min_{\beta} \left\| \begin{bmatrix} \alpha_j(x_1, y_1) \\ \vdots \\ \alpha_j(x_n, y_n) \end{bmatrix} - \begin{bmatrix} Z_1(x_1, y_1) & Z_2(x_1, y_1) & Z_3(x_1, y_1) \\ \vdots & \vdots & \vdots \\ Z_1(x_n, y_n) & Z_2(x_n, y_n) & Z_3(x_n, y_n) \end{bmatrix} \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \\ \beta_{j3} \end{bmatrix} \right\|_2$$

which we solve with ordinary least squares. Letting α_j, Z, β_j be the vectors and matrices from the equation above, we have the analytic solution $\beta_j = (Z^T Z)^{-1} Z^T \alpha_j$.

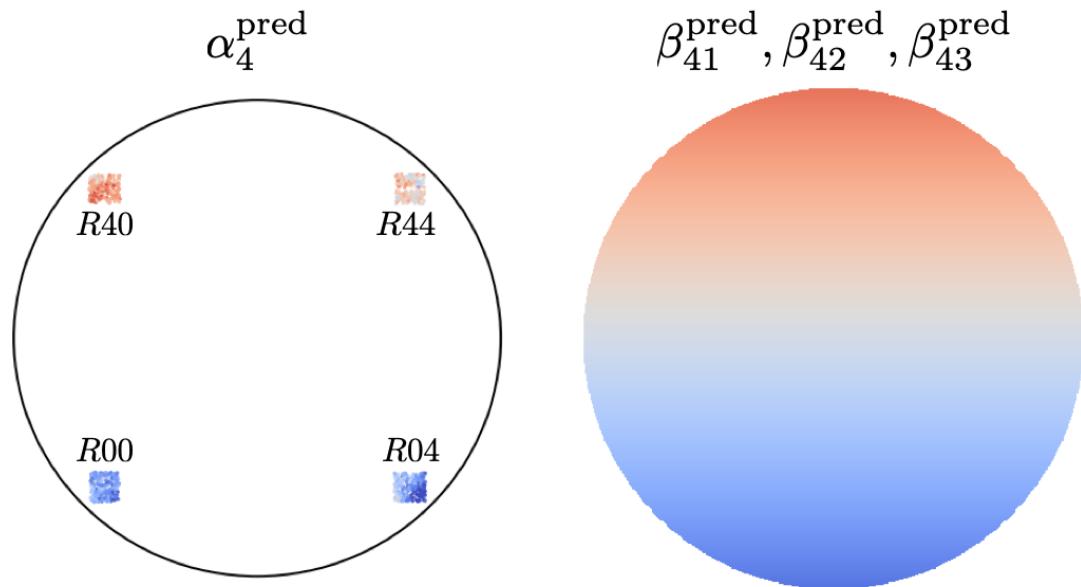


Figure 5.2: *Left:* the predicted coefficients α_4^{pred} from Figure 5.1 shown at their corresponding field locations within the field of view. *Right:* the plane characterized by the three fit double Zernike coefficients $\beta_{41}^{\text{pred}}, \beta_{42}^{\text{pred}}, \beta_{43}^{\text{pred}}$. The two circles represent the field of view of the Rubin Observatory.

After solving this equation for each j , we have β_{jk} for $4 \leq j \leq 21$ and $1 \leq k \leq 3$. In the next section we describe a few variations to this basic framework.

5.2 Evaluating Different Interpolation Strategies

The interpolation is comprised of 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 full visit dataset, are shown in Table 5.1. We compare the true global wavefront and the residual global wavefront, where the residual is the true wavefront minus the interpolated optics wavefront. The residual 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 |
| | \checkmark | ℓ_1 | 97.8 | 0.67 ± 0.14 |
| | | ℓ_2 | 100.0 | 0.46 ± 0.12 |
| | \checkmark | 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 |
| | \checkmark | 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 5.1: 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 red.

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 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.

We used this variation to explore how our framework performs in five scenarios. We used the relative residual $|\beta^{\text{res}}| / |\beta^{\text{true}}| = \sum_{jk} |\beta_{jk}^{\text{true}} - \beta_{jk}^{\text{pred}}| / \sum_{jk} |\beta_{jk}^{\text{true}}|$ to measure our algorithm's capability to reproduce the true global wavefront. A value close to zero means the predictions match the true global wavefront almost perfectly; a value close to one means the predictions capture almost none of the true global wavefront. Figure 5.3 shows the results.

In the first scenario, we evaluated the second step of our algorithm on the true local wavefront labels from the first step. The relative residual is 6%, which validates that our simulations are consistent and there are no sources of catastrophic numerical error. In the second scenario, we added uncorrelated Gaussian noise to each local wavefront coefficient. The variance of the noise for each coefficient was equal to the variance of the neural network error on that coefficient in the test data. Despite this noise, the relative residual remained very low at 10%. This shows that our approach is not overly sensitive to uncorrelated noise.

In the third scenario, we fed the local wavefront predictions from the neural network to the second step. These have significant correlated error due to the atmosphere. The relative residual is 34%, extremely promising, with a tight 12% standard deviation. Furthermore, the right tail decays rapidly. Not a single relative residual is greater than one. Our framework improves the global wavefront on every observation in the full visit dataset!

In the fourth scenario we add noise as in the second scenario, effectively doubling the error. The relative residual increases slightly to 38%. This suggests that the network is fairly robust to uncorrelated noise.

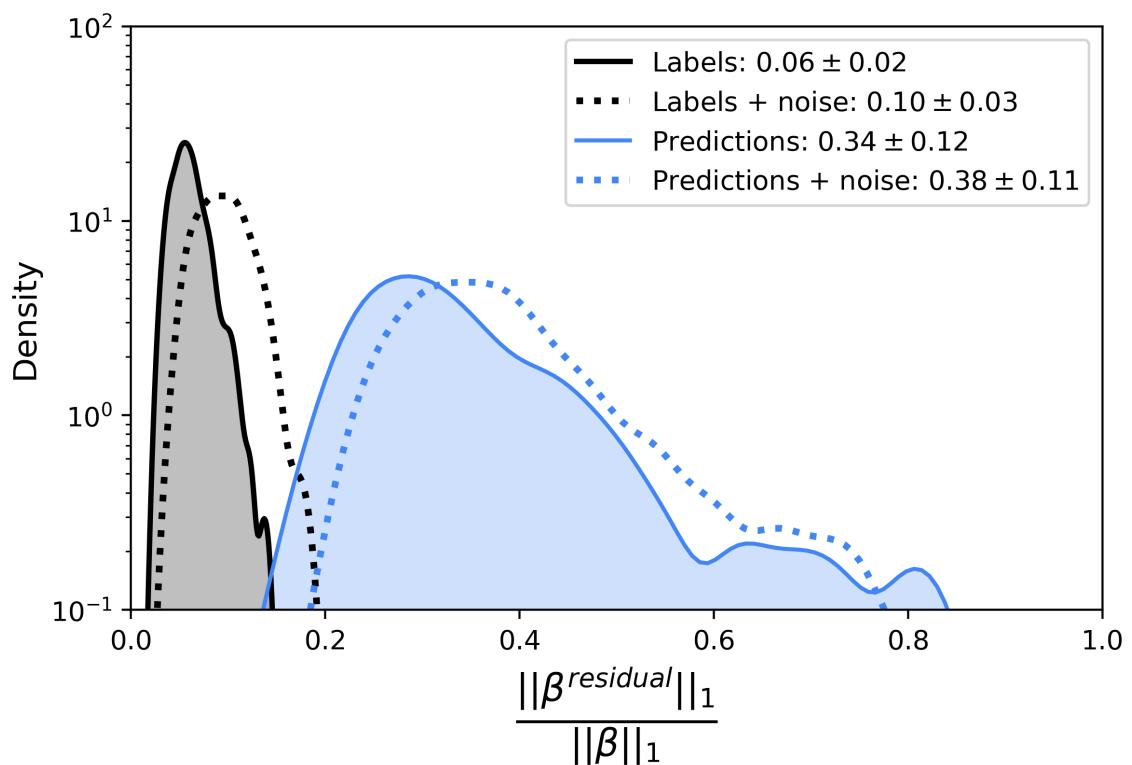


Figure 5.3: The distributions of relative residuals in four different scenarios. The distributions from performing the interpolation over the labels with and without noise are shown in black. The distributions from performing the interpolation over the local wavefront predictions from the neural network with and without noise are shown in blue.

| 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 5.2: 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.

5.3 Combined Runtime

The run-time is another key advantage of our approach. It takes just under 4 seconds to process 811 donut images on a 2.4 GHz Intel Xeon CPU with a single Nvidia V100 GPU. It takes an additional 5 milliseconds to run the least squares optimization. The total processing time per donut image is 5 milliseconds; the total processing time per donut of the Rubin Observatory AOS under development is around 10 seconds[8, 52]. Our scheme can process donuts 2,000 times faster, and is capable of processing 7,800 donuts in a single 39 second Rubin visit. The low latency allows it to process all the donuts in most observations.

5.4 Image Quality Results

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 [53]. 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 5.2.

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 ?? 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.

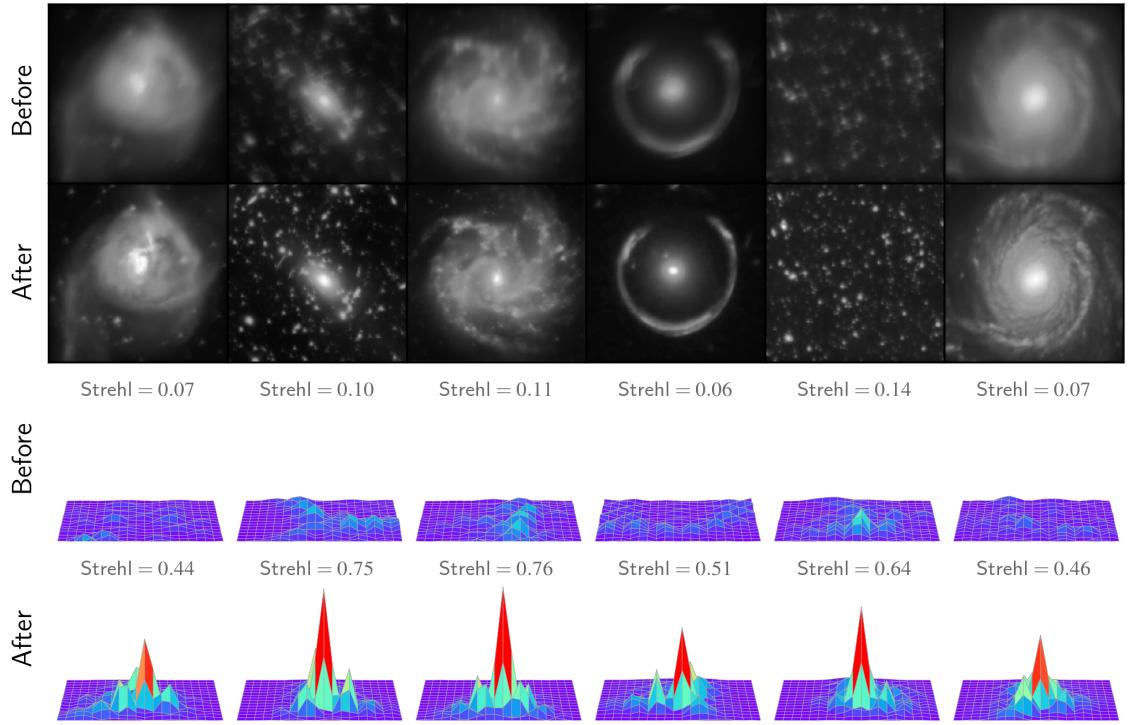


Figure 5.4: An illustration of the effect of wavefront correction using our techniques on image resolution. We use actual Hubble Space Telescope images to show the effect. The Before images are degraded by the wavefront aberrations we have simulated. The After images show their reconstruction after wavefront 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.

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.

Chapter 6

Error Covariance Fitting and Optimal Telescope Control

By failing to prepare, you are preparing to fail.

Benjamin Franklin

In this chapter, we transition from developing our wavefront sensing framework to showing how it can be used to optimally control a telescope like the Rubin Observatory. The connecting bridge is the error covariance matrix of the estimated global wavefront coefficients. In Section 6.1, we show how our framework allows this covariance to be easily computed. Then in Section 6.2, we show how we can use this in control simulations and assess different control strategies.

6.1 Error Covariance Fitting

The key breakthrough is that since the interpolation step of our algorithm has a closed form, the error covariance is a closed form function of the error covariance of our local wavefront predictions from the first step. In order to do this we first fix the positions of the 10 brightest stars on each of the four sensors. For indexing, let (x_1, y_1) be the first position on the first corner sensors, (x_{11}, y_{11}) be the first position

on the second sensor, ... and (x_{31}, y_{31}) be the first position on the fourth sensor. Then we define γ as

$$\gamma = \begin{bmatrix} \alpha_4(x_1, y_1) \\ \dots \\ \alpha_4(x_{10}, y_{10}) \\ \alpha_5(x_1, y_1) \\ \dots \\ \dots \\ \alpha_{21}(x_{10}, y_{10}) \\ \alpha_4(x_{11}, y_{11}) \\ \dots \\ \dots \\ \alpha_4(x_{21}, y_{21}) \\ \dots \\ \dots \\ \alpha_4(x_{31}, y_{31}) \\ \dots \\ \dots \\ \alpha_{21}(x_{40}, y_{40}) \end{bmatrix}$$

Thus $\gamma \in \mathbb{R}^{720}$, where 720 stems from 10 positions, 4 sensors, and 18 coefficients. We also make the empirically confirmed assumption that the individual local wavefront coefficient predictions are zero centered. Then the error covariance between γ_1, γ_2 measures the error covariance of the α_4 predictions across donut locations on the first wavefront sensors; the error covariance between γ_1, γ_{11} measures the error covariance between the α_4 and α_5 predictions at the same donut position; the error covariance between γ_1, γ_{181} measures the error covariance of the α_4 predictions between positions on two different corner wavefront sensors. Thus a lot of different forms of covariance are encoded in the covariance between elements of γ . The full error covariance on the local wavefront predictions is $\Sigma_\gamma = \text{Cov}(\gamma, \gamma)$.

We assume that the error γ is a multi-dimensional gaussian random variable. Then

the covariance from the maximum likelihood estimate is $\Sigma_\gamma = (X - E[X])(X - E[X])^T$ where $X \in \mathbb{R}^{497 \times 720}$ such that $X_i = \gamma^{(i)}$ the local wavefront errors from the i th observation. Figure 6.1 shows the local estimate covariance matrix.

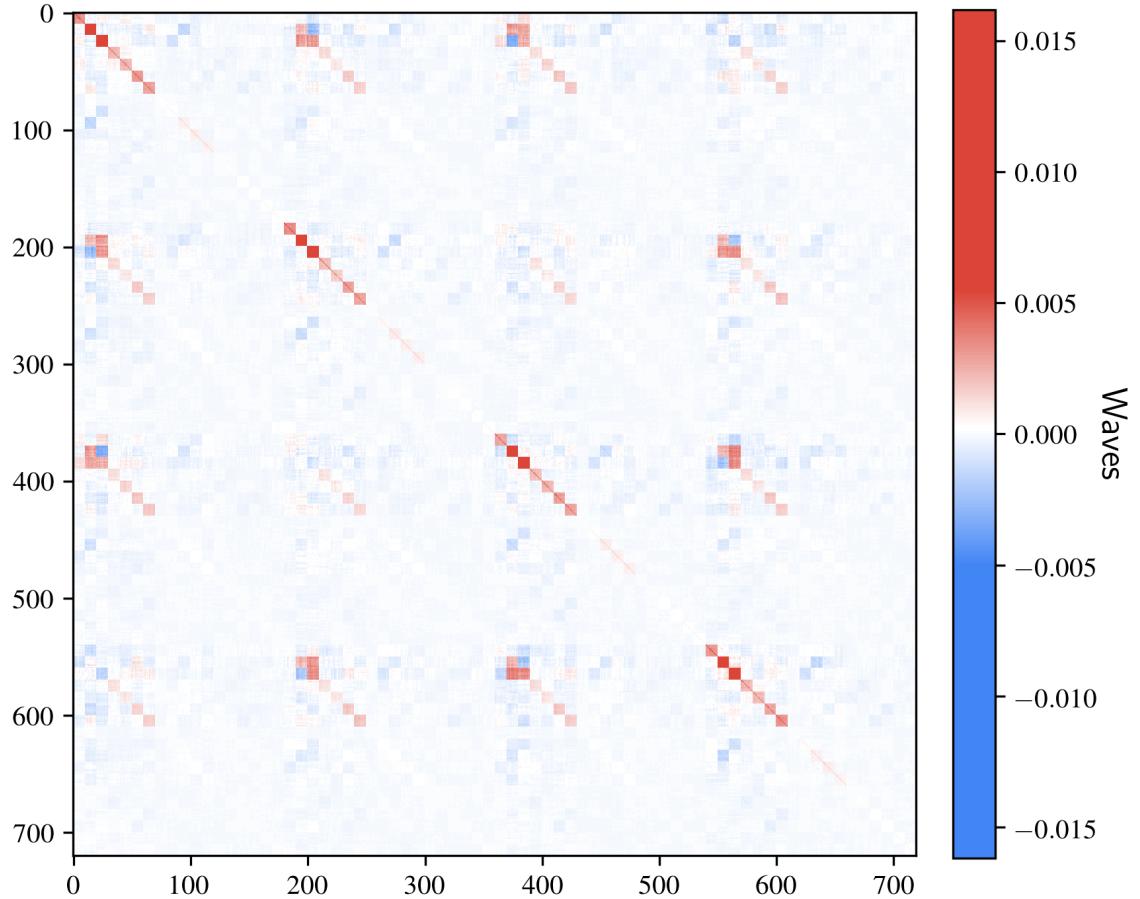


Figure 6.1: The error covariance matrix Σ_γ for the local wavefront error estimates in units of waves.

Recall that if $y = Bx$, then $\text{Cov}(y, y) = BC\text{Cov}(x, x)B^T$. In our case we can define an A , such that $\beta = (A^T A)^{-1} A^T \gamma$. This A is

$$A = \begin{bmatrix} I_{18} \otimes \begin{bmatrix} Z_1(x_1, y_1) & Z_2(x_1, y_1) & Z_3(x_1, y_1) \\ \vdots & \vdots & \vdots \end{bmatrix} \\ I_{18} \otimes \begin{bmatrix} Z_1(x_{10}, y_{10}) & Z_2(x_{10}, y_{10}) & Z_3(x_{10}, y_{10}) \\ Z_1(x_{11}, y_{11}) & Z_2(x_{11}, y_{11}) & Z_3(x_{11}, y_{11}) \\ \vdots & \vdots & \vdots \end{bmatrix} \\ I_{18} \otimes \begin{bmatrix} Z_1(x_{20}, y_{20}) & Z_2(x_{20}, y_{20}) & Z_3(x_{20}, y_{20}) \\ Z_1(x_{21}, y_{21}) & Z_2(x_{21}, y_{21}) & Z_3(x_{21}, y_{21}) \\ \vdots & \vdots & \vdots \end{bmatrix} \\ I_{18} \otimes \begin{bmatrix} Z_1(x_{30}, y_{30}) & Z_2(x_{30}, y_{30}) & Z_3(x_{30}, y_{30}) \\ Z_1(x_{31}, y_{31}) & Z_2(x_{31}, y_{31}) & Z_3(x_{31}, y_{31}) \\ \vdots & \vdots & \vdots \end{bmatrix} \\ I_{18} \otimes \begin{bmatrix} Z_1(x_{40}, y_{40}) & Z_2(x_{40}, y_{40}) & Z_3(x_{40}, y_{40}) \end{bmatrix} \end{bmatrix}$$

where $I_{18} \in \mathbb{R}^{18 \times 18}$ is the identity matrix, \otimes is the kronecker product, and $Z_i(x_j, y_j)$ is the j th circular Zernike polynomial being evaluated at focal plane position x_j, y_j . Then the error covariance matrix for β is $\Sigma_\beta = (A^T A)^{-1} A^T \Sigma_\gamma A (A^T A)^{-1}$. It is shown in Figure 6.2.

The global wavefront covariance Σ_β , which can be computed based on the empirically determined Σ_γ , is a key advantage to our wavefront sensing framework. We can use this statistical error model to characterize the performance of the system through time.

6.2 Optimal Control

We developed a simulator in order to assess different control strategies. The simulator simulates the global wavefront coefficients over discrete time steps. There are three steps at each time step. First, a perturbation is applied. In practice, this would be caused by changes to the wind, gravitational stress, thermal gradients, etc. Second, the global wavefront prediction is made with the error covariance from the previous section. Third a control strategy is applied with this estimate. Figure 6.3 summarizes

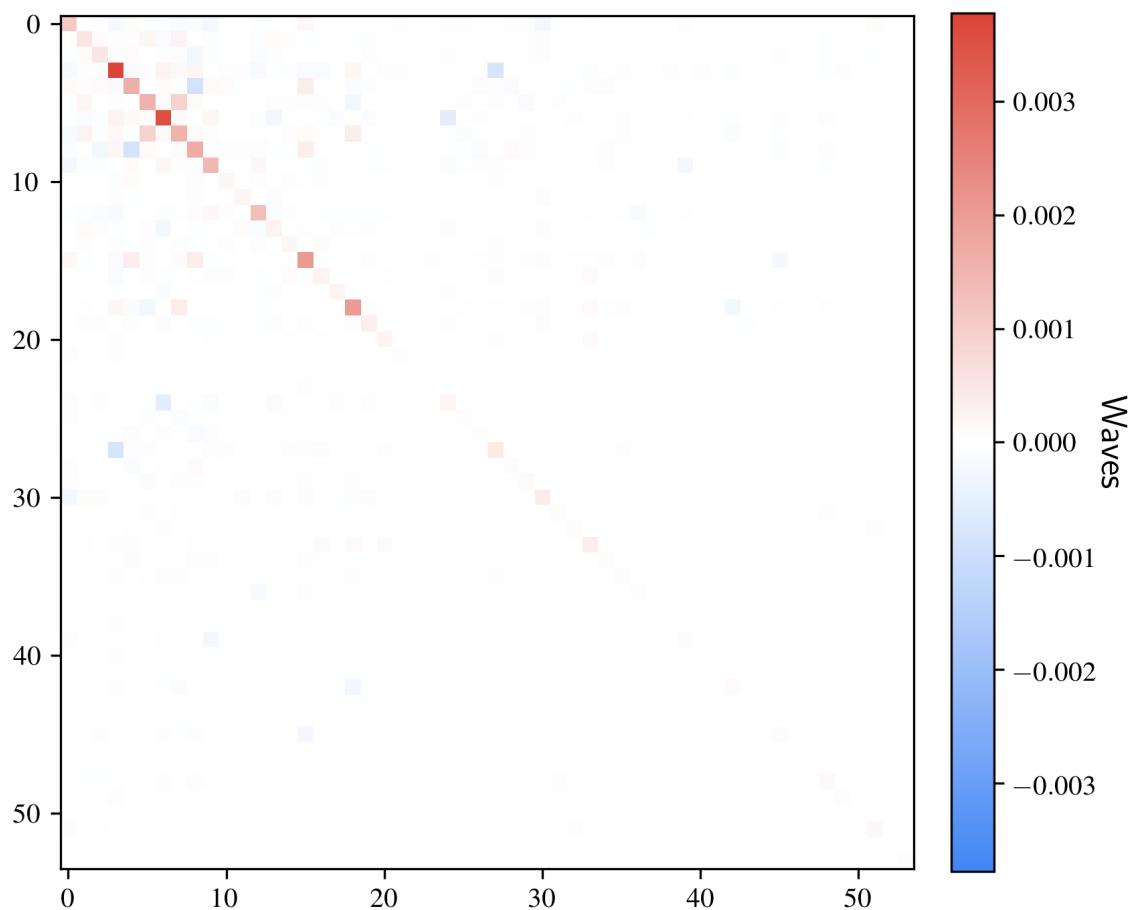


Figure 6.2: The error covariance matrix Σ_β for the global wavefront error estimates in units of waves.

the simulator algorithm.

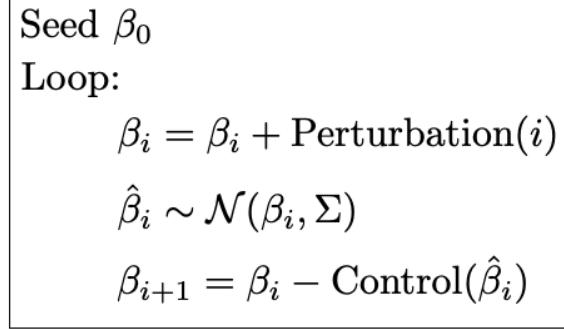


Figure 6.3: The control simulator algorithm. The sequence of perturbations, error covariance matrix, and control strategy are passed by the client.

We used this simulator to test four different control strategies. The first is the null control strategy which does not make any changes to the control parameters, or

$$\text{Control-Null}(\hat{\beta}_i) = \hat{\beta}_i$$

This is useful for null tests.

The second is a gain controller. It makes the update

$$\text{Control-Gain}(\hat{\beta}_i) = g \cdot \hat{\beta}_i$$

where g is the gain, which is typically less than 1. When $g = 1$, this reduces to the null controller.

The third is a proportional-integral-derivative (PID) controller. It extends the gain controller with two additional additive terms. One term accumulates the estimates, the other uses the difference between the current and previous estimates as a proxy for a derivative. It makes the update

$$\text{Control-PID}(\hat{\beta}_i) = c_p \cdot \hat{\beta}_i + c_i \sum_{j=1}^i \hat{\beta}_j + c_d \cdot (\hat{\beta}_i - \hat{\beta}_{i+1})$$

where c_p, c_i, c_d are the configurable model parameters. When $c_i = c_d = 0$ it reduces

to the gain controller.

The fourth is a linear-quadratic-gaussian (LQG) controller. In control theory, the separation principle says that the problem of designing an optimal controller for a stochastic system can be decomposed into (i) an optimal estimator of the state and (ii) an optimal deterministic controller. LQG adheres to this decomposition. A Kalman filter (or linear-quadratic state estimator) is used for state estimation and a linear-quadratic regulator is used for control. For our discrete time problem, the update is

$$\begin{aligned}
\text{Control-LQC}(\hat{\beta}_i) &= K_i \cdot x_i \\
x_{i+1} &= x_i - u_i + L_i(\hat{\beta}_i - x_i + u_i) \\
x_0 &= \hat{\beta}_0 \\
L_i &= P_i(P_i + \Sigma_\beta + V)^{-1} \\
P_{i+1} &= P_i - P_i(P_i + \Sigma_\beta + V)^{-1}P_i + V \\
P_0 &= \Sigma_\beta + V \\
K &= (S_h + R)^{-1}S \\
S_{i+1} &= S_i - S_i(S + R)^{-1}S + Q \\
S_0 &= Q
\end{aligned}$$

where Q, R, V and the horizon h are configurable parameters.

In order to make the sequence of perturbations in the simulator as realistic as possible we developed code to convert contiguous observations from OpSim into sequences of perturbations. We use two types of perturbations: Jump and Box.

Jump perturbations are applied once at a specific time step. They are parameterized by a time step and scale. The 56-dimensional global wavefront perturbation is drawn from a zero-mean multivariate normal distribution. The covariance is Σ_β times the scale.

Box perturbations are parameterized by a step, duration, and scale. It begins at

| Observation | Perturbation |
|-----------------------|---|
| First 50 Observations | Box(start=0, duration=50, scale=2) |
| Last 50 Observations | Box(start=end-50, duration=50, scale=2) |
| Slew > 5 | Jump(step=step, scale=5) |
| Slew > 20 | Jump(step=step, scale=100) |
| Filter Change | Jump(step=step, scale=20) |

Table 6.1: The mapping between catalog observations and their corresponding perturbations for the control simulator.

the step time step and lasts for the duration number of time steps. At each time step, the 56-dimensional global wavefront perturbation is drawn from a zero-mean multivariate normal distribution. The covariance is Σ_β times the scale. It is worth noting that for each timestep within the duration, a different perturbation is drawn and applied. Table 6.1 shows how we convert OpSim sequences of observations into sequences of perturbations.

Each of the control strategies has parameters. We randomly selected 11 nights within the 10 year observing period for Rubin. We convert each of these nights into a sequence of perturbations for the control simulator. Then we fit the control strategy parameters on the first 10 nights, and evaluate them on the held out final night. The penalty combines the norm of the state, which is zero in the nominal state, and the size of the update.

$$\text{Penalty} = \|\beta_{i+1}\|_2^2 + \frac{1}{2} \|\text{Control}(\hat{\beta}_i)\|_2^2$$

For the gain controller, we found a gain of 0.39 was optimal. For the PID controller, we found $c_p = 0.37, c_i = 5.4 \times 10^{-6}, c_d = .15$ was optimal. For the LQG controller, we found $Q = 0.84 \cdot I_{56}, R = 0.88 \cdot I_{56}$ with a fixed horizon $h = 5$ and V equal to the covariance of the state difference between steps with the null controller on the 10 training nights. The results on the final test night are shown in Figure 6.4.

The first thing to note is that the performance of the three non-trivial controllers is very similar. The cumulative penalty for the different methods is 1901.78, 16.22, 15.61, and 16.92 for the null, gain, PID, and LQG controllers respectively. It seems

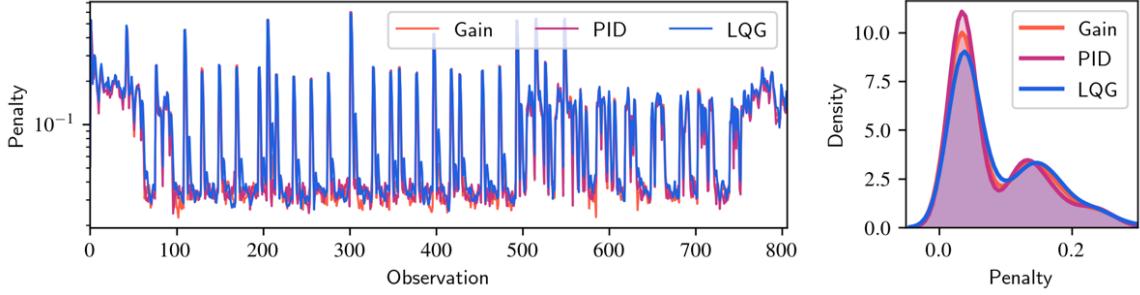


Figure 6.4: Left: the penalty of the different controllers through time on the test night. Right: the distributions of the penalties.

like the PID controller best balances the positives from its sophistication and parameterization with the negatives that can arise from overfitting.

As of yet, there has not been an investigation of the tradeoffs between different control strategies for Rubin. The current plan is to spend time during commissioning to take data and develop a strategy on the mountain. However, each extra day on the mountain not only delays the survey but also runs up the project budget. The framework we have presented here has the potential to save this critical commissioning time by enabling this work on the mountain to be done now, with this framework.

Chapter 7

Conclusions

Don't be afraid to give up the good to go for the great.

John D. Rockefeller

In this dissertation, we gave readers a comprehensive tour of our new two step wavefront sensing framework.

In Chapter 1, we provided an overview of our framework. We started by describing the limitations of existing methods. Then we described the key conceptual breakthrough in our framework, that wavefront sensing can be decomposed into two sub-problems. The first sub-problem is to use a neural network to predict the local wavefront. The second sub-problem is to interpolate the global wavefront, represented by double Zernike polynomials, from the local wavefront estimates. We also demonstrated how 90% of the aberration power is concentrated in the lowest three field Zernike polynomials, which makes the interpolation feasible.

In Chapter 2, we covered the Rubin Observatory and its active optics system. We described the components of the telescope and the reduced 50 degrees of freedom that the active optics system strives to control.

Given that the Rubin Observatory is currently under construction, this work critically depends on high fidelity simulations. In Chapter 3, we described all the relevant physical effects and how they are modelled in our simulations. We also introduced the donut dataset and full visit dataset. We hope that these datasets can continue to

serve as a research resource and benchmark for comparison moving forward.

In Chapter 4, we described how we trained a convolutional neural network to predict the local wavefront. After stepping through the architecture in detail, we explain the training process and assess the network’s performance. We also characterized different potential sources of error. Ultimately our model made state of the art wavefront predictions. We believe this task, and our approach using a neural network to solve it, can serve a valuable component in many other wavefront sensing and optics algorithms.

In Chapter 5, we described how least squares is used to interpolate the global wavefront. We explored many select-reduce-fit combinations and found that using only the brightest stars, without reducing within a wavefront sensor, with an l_2 fit gave the best results. Our algorithm improved the wavefront in all 497 observations in the full visit dataset. It also reduced the total magnitude of the optics global wavefront by 66%, the optics PSF FWHM by 27%, and increased the Strehl ratio by a factor of 6. The illustrative Hubble images provide a sense of how much impact this image quality improvement can have on the downstream science applications.

In Chapter 6, we described our telescope control simulator. We evaluated the performance of three classic optimal control strategies with varying degrees of sophistication. This framework we have built on top of our wavefront sensing algorithm enables research on telescope control strategies to begin now, and potentially save time on the mountain during the commissioning of the Rubin Observatory.

Finally, in addition to achieving incredible performance, it is worth stressing that our algorithm has many practical advantages.

- **Transparency:** 2 steps.
- **Robustness:** 2x noise leads to 11% performance degradation.
- **Low Latency:** 5 milliseconds per donut.
- **High Bandwidth:** 7,800 donuts in 39 seconds.

There is no existing alternative that achieves this combination of properties. Our two step wavefront sensing algorithm has the potential to improve image quality on

multiple current and upcoming wide-field telescopes. We hope physicists continue to be receptive to interdisciplinary collaborations with those from applied mathematics and machine learning backgrounds. Leveraging these techniques has the potential to continue advancing many experiments and investigations in physics and cosmology.

Bibliography

- [1] 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, December 2020.
- [2] David Thomas, Joshua Meyers, and Steven M. Kahn. Improving astronomy image quality through real-time wavefront estimation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2076–2085, 2021.
- [3] Mazen M. Sinjab and Arthur B. Cummings. *Introduction to Wavefront Science*, pages 65–93. Springer International Publishing, Cham, 2018.
- [4] Joseph W Goodman. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, 1, 2005.
- [5] Aaron Roodman, Kevin Reil, and Chris Davis. Wavefront sensing and the active optics system of the dark energy camera. In *Ground-based and Airborne Telescopes V*, volume 9145 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 914516, July 2014.
- [6] B. Flaugher, H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufort, G. M. Bernstein,

- R. A. Bernstein, B. Bigelow, M. Bonati, D. Boprie, D. Brooks, E. J. Buckley-Geer, J. Campa, L. Cardiel-Sas, F. J. Castander, J. Castilla, H. Cease, J. M. Cela-Ruiz, S. Chappa, E. Chi, C. Cooper, L. N. da Costa, E. Dede, G. Derylo, D. L. DePoy, J. de Vicente, P. Doel, A. Drlica-Wagner, J. Eiting, A. E. Elliott, J. Emes, J. Estrada, A. Fausti Neto, D. A. Finley, R. Flores, J. Frieman, D. Gerdes, M. D. Gladders, B. Gregory, G. R. Gutierrez, J. Hao, S. E. Holland, S. Holm, D. Huffman, C. Jackson, D. J. James, M. Jonas, A. Karcher, I. Karliner, S. Kent, R. Kessler, M. Kozlovsky, R. G. Kron, D. Kubik, K. Kuehn, S. Kuhlmann, K. Kuk, O. Lahav, A. Lathrop, J. Lee, M. E. Levi, P. Lewis, T. S. Li, I. Mandrichenko, J. L. Marshall, G. Martinez, K. W. Merritt, R. Miquel, F. Muñoz, E. H. Neilsen, R. C. Nichol, B. Nord, R. Ogando, J. Olsen, N. Palaio, K. Patton, J. Peoples, A. A. Plazas, J. Rauch, K. Reil, J.-P. Rheault, N. A. Roe, H. Rogers, A. Roodman, E. Sanchez, V. Scarpine, R. H. Schindler, R. Schmidt, R. Schmitt, M. Schubnell, K. Schultz, P. Schurter, L. Scott, S. Serrano, T. M. Shaw, R. C. Smith, M. Soares-Santos, A. Stefanik, W. Stuermer, E. Suchyta, A. Sypniewski, G. Tarle, J. Thaler, R. Tighe, C. Tran, D. Tucker, A. R. Walker, G. Wang, M. Watson, C. Weaverdyck, W. Wester, R. Woods, and B. Yanny and. THE DARK ENERGY CAMERA. 150(5):150, oct 2015.
- [7] 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.
- [8] Bo Xin, Chuck Claver, Ming Liang, Srinivasan Chandrasekharan, George Angeli, and Ian Shipsey. Curvature wavefront sensing for the large synoptic survey telescope. *Applied Optics*, 54(30):9045, October 2015.
- [9] 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, pages 1577 – 1586. International Society for Optics and Photonics, SPIE, 2016.

- [10] Robert J. Noll. Zernike polynomials and atmospheric turbulence*. *J. Opt. Soc. Am.*, 66(3):207–211, Mar 1976.
- [11] 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.
- [12] 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, November 1990.
- [13] 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.
- [14] 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, April 1993.
- [15] 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, August 1992.
- [16] 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, January 1996.
- [17] 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, January 1997.

- [18] Hong Guo, Nina Korablinova, Qiushi Ren, and Josef Bille. Wavefront reconstruction with artificial neural networks. *Optics Express*, 14(14):6456–6462, July 2006.
- [19] 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, November 2008.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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, December 2015.
- [24] 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.
- [25] Scott W. Paine and James R. Fienup. Machine learning for improved image-based wavefront sensing. *Optics Letters*, 43(6):1235, March 2018.
- [26] Yohei Nishizaki, Matias Valdivia, Ryoichi Horisaki, Katsuhisa Kitaguchi, Mamoru Saito, Jun Tanida, and Esteban Vera. Deep learning wavefront sensing. *Optics Express*, 27(1):240, January 2019.

- [27] 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, August 2020.
- [28] J. A. Tyson, D. M. Wittman, and J. R. P. Angel. The Dark Matter Telescope. In Tereasa G. Brainerd and Christopher S. Kochanek, editors, *Gravitational Lensing: Recent Progress and Future Go*, volume 237 of *Astronomical Society of the Pacific Conference Series*, page 417, January 2001.
- [29] LSST Science Collaboration, Paul A. Abell, et al. LSST Science Book, Version 2.0. *arXiv e-prints*, page arXiv:0912.0201, December 2009.
- [30] Sandrine J. Thomas, Jeffrey Barr, Shawn Callahan, Andy W. Clements, Felipe Daruich, Juan Fabrega, Patrick Ingraham, William Gressler, Freddy Munoz, Doug Neill, Tiago Ribeiro, Jacques Sebag, Eduardo Serrano, Brian Stalder, Roberto Tighe, Tomislav Vucina, and Bo Xin. Vera C. Rubin Observatory: telescope and site status. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 11445 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 114450I, December 2020.
- [31] Jacques Sebag, Charles F. Claver, Sandrine J. Thomas, Kevin Reil, Jeffrey Barr, Keith Bechtol, Felipe Daruich, Juan Fabrega, Leanne P. Guy, Patrick Ingraham, Victor Krabbendam, Margaux Lopez, Freddy Munoz, Doug Neill, Wil O'Mullane, Michael Reuter, Austin Roberts, Tiago Ribeiro, Eduardo Serrano, Cristian Silva, Brian Stalder, Roberto Tighe, and Bo Xin. Vera C. Rubin Observatory system integration, test, and commissioning: strategy and status. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 11445 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 114451U, December 2020.
- [32] Douglas R. Neill, Gary Muller, Ed Hileman, Joe DeVries, Constanza Araujo, William J. Gressler, Paul J. Lotz, Dave Mills, Jacques Sebag, Sandrine Thomas, Mike Warner, and Oliver Wiecha. Final design of the LSST primary/tertiary

- mirror cell assembly. 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 99060Q, July 2016.
- [33] Douglas R. Neill, Gregory Bogan, Dale Zajac, Constanza Araujo, William J. Gressler, Joe DeVries, Edward A. Hileman, Paul J. Lotz, Dave Mills, Sandrine Thomas, Thomas A. Sebring, Jacques Sebag, Mike Warner, and Oliver Wiecha. LSST secondary mirror system final design. 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 990667, July 2016.
- [34] Scot S. Olivier, Vincent J. Riot, David K. Gilmore, Brian Bauman, Steve Pratuch, Lynn Seppala, John Ku, Martin Nordby, Mike Foss, Pierre Antilogus, Nazario Morgado, Benoit Sassolas, Raffaele Flaminio, and Christophe Michel. LSST camera optics design. In Ian S. McLean, Suzanne K. Ramsay, and Hideki Takami, editors, *Ground-based and Airborne Instrumentation for Astronomy IV*, volume 8446, pages 1943 – 1955. International Society for Optics and Photonics, SPIE, 2012.
- [35] 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.
- [36] 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, August 2018.
- [37] Josep M. Carrasco. Photometric relationships with other photometric systems. https://gea.esac.esa.int/archive/documentation/GDR2/Data_processing/chap_cu5pho/sec_cu5pho_calibr/ssec_cu5pho_PhotTransf.html, 2020. Accessed: 2020-04-07.

- [38] 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, April 2015.
- [39] 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.
- [40] Luc Gilles and Brent L. Ellerbroek. Split atmospheric tomography using laser and natural guide stars. *Journal of the Optical Society of America A*, 25(10):2427, September 2008.
- [41] Joshua E. Meyers and Patricia R. Burchat. Impact of Atmospheric Chromatic Effects on Weak Lensing Measurements. *Astrophysical Journal*, 807(2):182, July 2015.
- [42] Joshua Meyers. batoid. <https://github.com/jmeyers314/batoid>, 2020. Accessed: 2020-04-07.
- [43] C. Lage, A. Bradshaw, and J. A. Tyson. Measurements and simulations of the brighter-fatter effect in CCD sensors. *Journal of Instrumentation*, 12(3):C03091, March 2017.
- [44] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

- [45] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.
- [46] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [47] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [48] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimiza-
tion. 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.
- [49] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *Workshop at International Conference on Learning Representations*, 2014.
- [50] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [51] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and Harnessing Adversarial Examples. *arXiv e-prints*, page arXiv:1412.6572, December 2014.

- [52] Douglas Neill, George Angeli, Chuck Claver, Ed Hileman, Joseph DeVries, Jacques Sebag, and Bo Xin. Overview of the LSST active optics system. In *Modeling, Systems Engineering, and Project Management for Astronomy VI*, volume 9150 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91500G, August 2014.
- [53] Karl Strehl. Aplanatische und fehlerhafte abbildung im fernrohr. *Zeitschrift für Instrumentenkunde*, 15(7):362–370, 1895.

David Thomas

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Professor Steven Kahn) Principal Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Professor Patricia Burchat)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Professor Stephen Boyd)

Approved for the Stanford University Committee on Graduate Studies.
