Active Optical Control with Machine Learning: A Proof of Concept for the Vera C. Rubin Observatory

Jun E. Yin*
Harvard
jyin@g.harvard.edu

Daniel J. Eisenstein
Harvard
deisenstein@cfa.harvard.edu

Douglas P. Finkbeiner
Harvard
dfinkbeiner@cfa.harvard.edu

Christopher W. Stubbs
Harvard
stubbs@g.harvard.edu

Yue Wang
MIT
yuewangx@mit.edu

Abstract

The Active Optics System of the Vera C. Rubin Observatory (Rubin) uses information provided by four wavefront sensors to determine deviations between the reconstructed wavefront and the ideal wavefront. The observed deviations are used to adjust the control parameters of the optical system to maintain image quality across the 3.5 degree field of view. In this paper, we use deep learning methods to extract the control parameters directly from the images captured by the wavefront sensors. Our neural net model uses anti-aliasing pooling to boost performance, and a domain-specific loss function to aid learning and generalization. The accuracy of the control parameters derived from our model exceeds Rubin requirements even in the presence of full-moon background levels and mis-centering of reference stars. Although the training process is time consuming, model evaluation requires only a few milliseconds. This low latency should allow for the correction of the optical configuration during the readout and slew interval between successive exposures.

1 Problem Formulation

The Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory (Rubin) [1] will be deepest optical/IR survey ever to cover the majority of the sky. It will provide new insights into the mysteries of dark matter and dark energy, as well as transient phenomena, solar system objects, and Milky Way structure. The telescope's optical system actively compensates for flexure via 50 control parameter, as illustrated in Fig. 1, that adjust the mirror and camera positions (using 2 hexapods with 6 actuators each), and actively deforms the primary, secondary, and tertiary mirrors, using 156 and 72 actuators respectively.

The light arriving from a distant star at the top of the atmosphere may be thought of as a plane wave. Spatial variations in the index of refraction cause the wavefront to distort as it passes through the atmosphere. After the wavefront enters the telescope, the optical system introduces additional distortions. A typical value for the point-spread function (PSF) due to atmospheric effects is estimated to be 0.7 arcsec. The distortions from the optical system are approximately stable over an exposure, and must be determined and corrected from one exposure to the next.

When substantially out of focus, a star produces an annulus corresponding to the part of the primary mirror unobscured by the secondary. These annuli are colloquially known as "donut" images. The donuts reveal shifts in focus by their overall size, but contain far more information than that. The

*corresponding author

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

inner and outer boundaries of each donut deviate from their undistorted shape (in the ideal case, a circle), and the surface brightness of the donut is non-uniform. These perturbations carry sufficient information to enable extraction of the control parameters needed to correct the distortion.

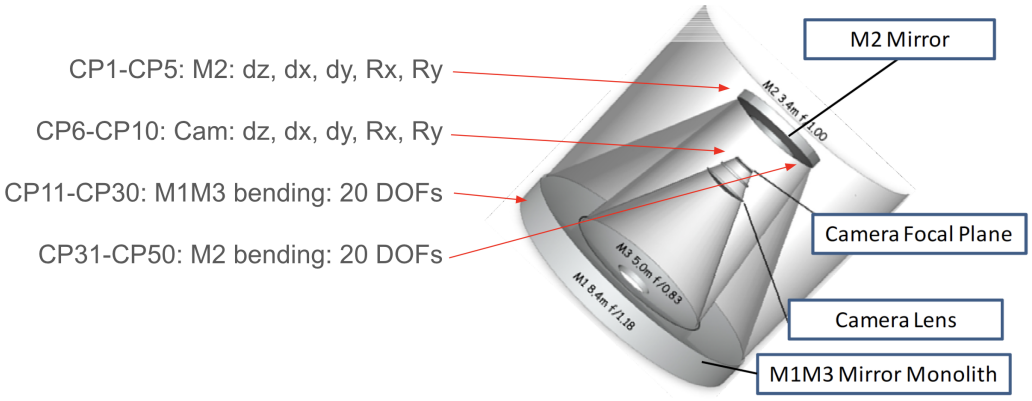

Figure 1: Schematic of the optical system of Vera C. Rubin Observatory. Light is reflected by the M1 mirror, then by the M2 mirror, and finally by the M3 mirror on the way to the camera installed underneath M2. The elements controlled by the 50 control parameters are indicated. This Figure is taken from Refs. [1] and [2].

The default control parameters are set to correct for the as-built parameters of the telescope structure and mirror imperfections. In practice, changes in environmental conditions and gravity vector will cause changes in the shape and position of the components of the optical system. In this work we are considering the 50 ΔCP values, i.e. the deviations from the default control parameters. The goal of this paper is to create a machine learning model that improves Active Optics System by providing fast, accurate updates to the control parameters.

2 Method

2.1 Data

The training and test images were generated using the wavefront analysis code developed by Roodman et al. [3] for the Dark Energy Camera. Our data set includes 37,000 sets of images (30,000 training, 7,000 test), and each set consists of 8 wavefront ("donut") images, as illustrated in Fig. 2.

Each donut image is generated by simluated 10^8 counts (photoelectrons), resulting in a SNR of about 600 per pixel. We treat these images as having infinite SNR and then add noise as follows. We model the variance of each pixel as the sum of background counts and image counts for each (rescaled) donut, and then add to each pixel a number drawn from a mean-zero Gaussian with that variance. Typical sky background in r band for a 15s exposure is between 800 (no moon) and 5000 (full moon) counts. The sky background for training images is drawn from a uniform distribution between these two values.

2.2 Training Algorithm

We use ResNet18 [4] as the backbone network. The model predicts 50 physical parameters for each set of 8 images, as illustrated in Fig. 3. We investigate the impact of 4 enhancements: anti-aliasing pooling, self-attention, scaled L2 loss, and addition of a PSF term to the loss function. We do the same experiments on data without noise and random shift, with noise but without random shift, and with noise and random shift.

In each case the model is trained for 250 epochs with batch size of 64. We train the NN with Adam [5] with initial learning rate 0.01, and divide the learning rate by 10 at epoch 75, 150 and 200. The

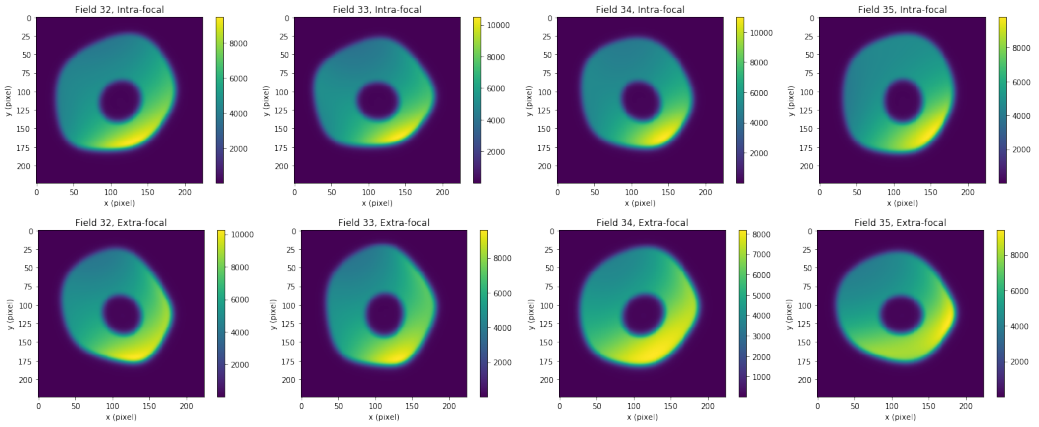

Figure 2: Examples of simulated images from 8 wavefront sensors. The color scale in each image shows the number of electrons per pixel. A perfectly corrected wavefront would be rotationally symmetric, with a uniform distribution of electrons in the donut.

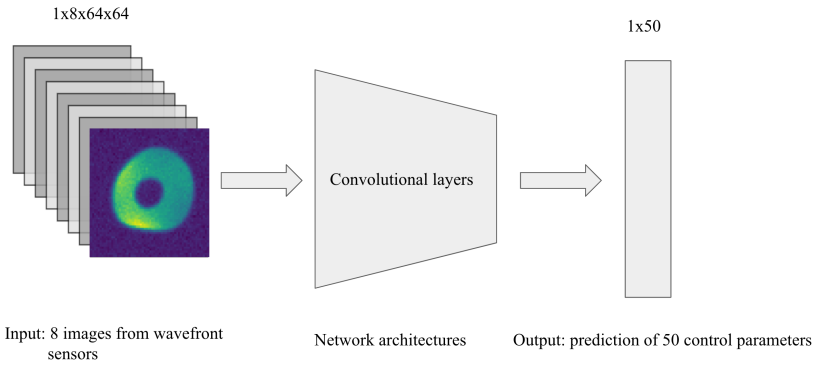

Figure 3: **Schematics of the algorithm**. **Left**: The input to the network is eight channel images from the wavefront sensors, 64×64 pixels each, or 12.8 arcsec for 0.2 arcsec pixels. **Middle**: The convolutional neural network structure is based on ResNet18 with some additional task-specific sub-network architectures. **Right**: The output from the network is the prediction of the values of the 50 control parameters.

training time is order of 6 GPU hours. We use mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) as error metrics.

Anti-aliasing pooling [6] addresses the problem of shift equivariance. For our application we would like translational equivariance,[2] so we employ anti-aliasing pooling. At each convolutional layer we first densely evaluate max pooling on the feature maps, then use a box filter to blur feature maps in each channel, and finally subsample the feature maps using the required stride.

Following [7, 8], we use self-attention modules to denoise feature maps [9]. Self-attention produces a non-local operation. Feature maps refers to both input images and output of the subsequent CNN layers. It calculates the response at a position as a weighted sum of the features at all positions in the input feature map [7]. It helps the network to link information from disparate parts of an image, and in some cases allows a neural network to be less sensitive to noisy inputs.

[2]More precisely, we want translational *invariance* of the control parameters. Other parameters like the (x, y) position of the donut should be *equivariant* (i.e. varying smoothly and proportionally with the input shift), but those are nuisance parameters in our case.

2.3 Loss Function

We use $\mathbf{y}$ to denote the predicted control parameter vector while $\mathbf{y}^*$ is the ground truth. The loss function is formulated as follows,

$$L(\mathbf{y}, \mathbf{y}^*) = \sum_j \alpha_j L_2(y_j, y_j^*) + \beta f(\mathbf{y}^* - \mathbf{y}) \tag{1}$$

where α_j is a scaling factor calculated according to error tolerance of the control parameter and β is a scaling factor to control the relative weight between two terms. We denote $\sum_j \alpha_j L_2(y_j, y_j^*)$ as scaled L_2 loss and $f(\mathbf{y}^* - \mathbf{y})$ as the PSF loss. From our experiments, $\beta = 10$ optimizes the results most but the result is not very sensitive to the choice of β.

The scaled L2 loss function uses each parameter's error tolerance to scale its L2 loss. This is in contrast to the naive L2 loss where the loss of every parameter is weighted equally, regardless of the required tolerance. Because some parameters have units of angle, and some distance, it is not meaningful to combine them without appropriate scaling. The PSF loss function explicitly penalizes a poor PSF (large FWHM).There are many combinations of CPs that produce similar PSFs. We care less about achieving the exact CP values than about finding values that yield a good PSF. Adding a PSF term to the loss function enhances performance substantially.

3 Results and Discussion

From our experiments, we observed that

- using the scaled L2 loss and adding a PSF term to the loss function enhances performance substantially;
- including anti-aliasing pooling, and augmenting the training data to include randomly shifted donuts, the resulting model performance is insensitive to image shift;
- including self-attention modules in the CNN led to modest changes in performance.

The CNN recovers CPs from the donuts within the required tolerance. As shown in Fig. 4 left, the prediction RMSE over error tolerance for all 50 control parameters are under 1. Fig. 4 right shows that when selecting the 10% worst cases based on PSF size, we construct the PSF for typical seeing of ~ 0.65 arcsec, measure its FWHM, and find that the prediction error of the CPs makes only a small contribution to the FWHM. It keeps the AOS contribution to the PSF below 0.079 arcsec, as desired.

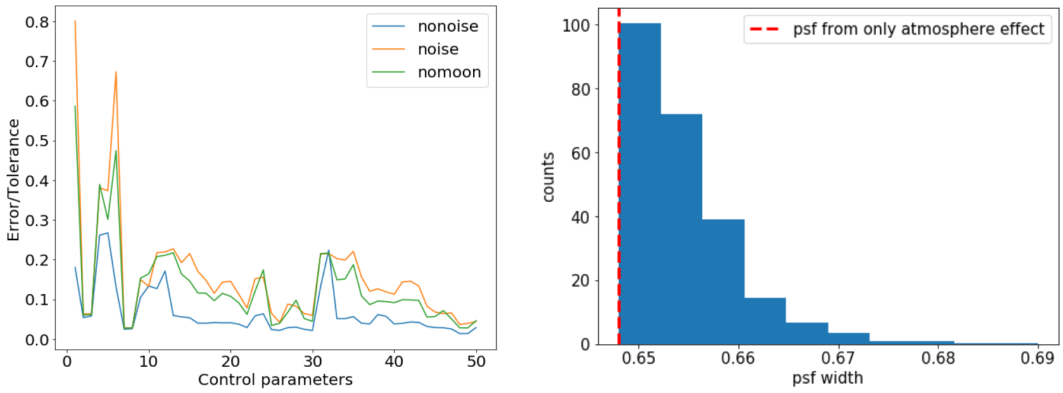

Figure 4: **Left:** Prediction RMSE over error tolerance for the 50 control parameters. The RMSE of each CP output by the neural network is within the tolerance. **Right:** Selecting the 10% worst cases based on PSF, we construct the PSF for typical seeing of ~ 0.65 arcsec, measure its FWHM, and find that the prediction error of the CPs makes only a small contribution to the FWHM.

In summary, a straightforward CNN with anti-aliasing pooling and a problem-specific loss function performs well on simulated data even in the presence of realistic noise and image shifts. As with

most machine learning problems, significant up-front computational expense is rewarded with fast evaluation. The training data requires 4000 CPU hours to generate, and the CNN trains in 6 GPU hours, however, it evaluates in a few milliseconds.

An operational system for Rubin would need to address the upstream problem of donut selection, tolerate overlaps of faint donuts, and deal with real-life issues like saturation, cosmic rays, and missing data. This proof of concept on simulated data makes it plausible that such a system could be built, and that neural networks can play a key role in keeping the Rubin images sharp.

4 Broader Impact

This paper presents our work on a machine-learning (ML) based method that can be used by the Active Optics System (AOS) of the Vera C. Rubin Observatory (Rubin). Rubin has the world's largest digital camera and will take approximately 5 million exposures during its 10 year mission, probing the sky with a sensitivity and cadence superior to any previous wide-area survey, and visiting each section of the sky every 3-4 days. The ML-based method reduces the time required to process the data obtained by the wavefront sensors and are used to determine the optical adjustments that must be made to correct for instrumental changes that affect the shape and size of the PSF. The broader impact of this work is related to the potential improvement in image resolution. These enhancements allow Rubin to understand the mysterious dark matter and dark energy, to study the formation of the Milky Way galaxy, to map out objects in the solar system, and to observe a wide variety of transient phenomena. The study of dark matter uses weak gravitational lensing (WL). In WL, the shape of a distant galaxy is weakly distorted by the foreground masses and this shearing is a key observable. Because the shear effects are small (a few percent), the distortions in the images due to imperfections of the optical system must be minimized. Our method enhances the ability to control the PSF of the Rubin optical system and thus will help unveil the mystery of dark matter.

5 Acknowledgements

It is a pleasure to acknowledge Chuck Claver, Sandrine Thomas, Aaron Roodman, and Bo Xin who provided insight into the VCRO optical system and help with simulation software. Josh Meyers provided advice on the use of GalSim. Michelle Ntampaka provided feedback on the paper draft. Parts of this work were developed in a class at MIT taught by Phillip Isola.

J.E.Y. is partially supported by U.S. Department of Energy grant DE-SC0007881 and the Harvard Data Science Initiative. D.J.E. is supported by U.S. Department of Energy grant DE-SC0013718 and as a Simons Foundation Investigator. D.P.F. is partially supported by NSF grant AST-1614941. C.W.S. is supported by U.S. Department of Energy grant DE-SC0007881.

References

[1] Paul A Abell, Julius Allison, Scott F Anderson, John R Andrew, J Roger P Angel, Lee Armus, David Arnett, SJ Asztalos, Tim S Axelrod, Stephen Bailey, et al. Lsst science book, version 2.0. *arXiv preprint arXiv:0912.0201*, 2009.

[2] Željko Ivezić, Steven M Kahn, J Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F Anderson, John Andrew, et al. Lsst: from science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, 2019.

[3] A. Roodman, K. Reil, and C. Davis. Wavefront sensing and the active optics system of the dark energy camera. In *Ground-based and Airborne Telescopes V*, volume 9145, page 914516, July 2014.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *arXiv preprint arXiv:1506.01497*, 2015.

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[6] Richard Zhang. Making convolutional networks shift-invariant again. In *ICML*, 2019.

[7] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. *CVPR*, 2018.

[8] Cihang Xie, Yuxin Wu, Laurens van der Maaten, Alan L. Yuille, and Kaiming He. Feature denoising for improving adversarial robustness. *CoRR*, abs/1812.03411, 2018.

[9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.