Paper Designing a Representation Learning Method for Wavefront Estimation from Focal Plane Speckle Images

Nick Murphy
Georgia State University

Vignesh Sathia
Georgia State University

Dr. Berkay Aydin
Georgia State University

Dr. Dustin Kempton
Georgia State University

Dr. Stuart Jefferies
Georgia State University

Dr. Fabien Baron
Georgia State University

ABSTRACT

Ground-based telescopes are used to record extraterrestrial images with great scientific, commercial, and military importance. Yet the images they collect are blurred due to ever-changing conditions and turbulence in the Earth's atmosphere. Current methods to mitigate this issue include adaptive optics, which measures the current conditions with a wavefront sensor and then deforms a mirror in the telescope's optical path. This process produces clearer images, but requires precise adjustments which may not scale as the number of mirror segments increases. In this work, we propose a multi-task learning approach coupled with representation learning to determine monochromatic wavefront aberrations at a single telescope aperture in real time. This approach utilizes shared weights for learning relevant, and inter-related parameters of wind layers, which allows for better scaling and control of multi-segment mirror adaptive optics. In other words, we seek to uncover aberration-related information about the characteristics of each layer of atmospheric turbulence from the distorted image. To create this model, we first develop a training dataset of 3D rasters – each instance corresponds to a set of speckle images of size NxNxP where the NxN is the size of individual images. These rasters are distorted by a Point Spread Function that simulates the effects of relevant atmospheric conditions. The atmospheric conditions in the form of wind layer velocity vectors will be used to label these instances. This considerably sparse dataset will then be processed using self-supervised learning models to learn representations in sufficiently low dimensions and extract important features. The resulting encoded representations will then be used to estimate the multi-layer wind velocities, where we will utilize multi-task learning approaches. The proposed model is envisioned to provide the necessary adjustment values for adaptive optics systems, at comparable levels of precision to existing physical wavefront sensors. It can then be used as a tool to manage a large multi-segment mirror adaptive optics system with less reliance on expensive hardware and environmental conditions.

1. INTRODUCTION

There is an increasing need to collect and process high quality images of satellites and other objects in outer space. Orbital telescopes like the James Webb Space Telescope can record clear, high resolution observations of objects light-years away from Earth. However, it is more expensive to build, launch, and maintain such platforms compared

to ground-based telescopes. Terrestrial telescopes have a different problem – they are separated from the extraterrestrial objects of interest by layers of atmosphere. These layers move at different speeds and directions and contain ever-changing pockets of turbulence. These results in a distortion effect on the observed images. This problem is currently mitigated by using additional highly calibrated sensors and mirrors to further alter the signal as it travels to the telescope receiver, reversing the effects of the atmosphere's distortion. Such solutions introduce more maintenance costs and complexity into the existing systems however. Furthermore, they introduce new challenges when designing smaller or more modular telescopes.

Ideally, the necessary corrections could be derived from data contained within the observed images, rather than relying on corrections through physical hardware. While there do exist efficient methods to recreate and invert the point spread function that blurs these images, the algorithms require sufficient knowledge of the environment to work effectively. As the atmosphere is constantly changing during the period of observation, it is difficult to collect and update the formulas with the current conditions. There are other "blind" deconvolution methods which do not require this amount of prior information, but these tend to be time and processing intensive operations which are not suitable for real-time corrections.

In this paper, we discuss a novel approach to use representation learning to predict relevant information about multiple wind layers based on simulated telescope observations. This information can be used to augment and supplement existing sensor systems, to reduce the amount of physical hardware needed to provide adjustments to ground-based telescope observations.

2. RELATED WORKS

The current approach to real-time image deconvolution is Adaptive Optics [2]. Beams of light are used to create a reference point near the object of interest. This reference is then used to carefully adjust additional deformable mirrors in the telescope, which cause another distortion in the received light. This new distortion is intended to negate the effects of the atmosphere, leaving a clear view of the original object. As said earlier, this is a very effective approach that produces good quality observations. However, it also adds additional expense, complexity, and maintenance requirements to the ground-based telescope it is fitted to. Furthermore, hardware designers need to accommodate it when developing new types of telescopes, which adds additional size constraints to the work.

One of the earliest methods to restore a blurred image via image processing instead of physical hardware is blind deconvolution [9]. This process involves performing a Fourier Transform on the affected raster, calculating the logarithmic magnitude of its subsections and averaging them to correct the distortions in the original image. While this was effective for its time, it was a slow and resource-intensive operation that would not be suitable for real time image correction. Increases in computational processing power led to an improvement to this method called Iterative Blind Deconvolution [5] [4]. Unlike the original single-pass approach, Iterative Blind Deconvolution performed multiple gradual adjustments to the convolved raster until a pre-defined constraint was reached. The final image quality was often much improved; however, the time and computational resources necessary to run the algorithm still did not allow for real time processing for sequences of blurred image observations. Additional improvements have been made to the baseline Iterative Blind Deconvolution algorithm, such as pre-processing the data further [8]. However, the previously stated challenges of applying it to near-real-time correction of frequent telescope observations remains to various degrees.

In recent years, machine learning (ML) has been incorporated to mitigate some of the noted issues with Blind Deconvolution. Bolbasova, Andrakhanov, and Shikhovtsev utilized a polynomial neural network to predict the ground layer refractive index at the Baikal Astrophysical Observatory [1]. Trained on data collected from a specialized thermometer, they achieved correlation between 0.79 and 0.91 of the actual and predicted C_N^2 value during their experiments. However, their study was limited to a single wind layer under the specific conditions of Lake Baikal, and specifically under stable atmospheric conditions. Hart et. al. studied the effects of simulated atmospheric turbulence on sound waves [3]. They collected both empirical results and generated output from a Crank-Nicholson parabolic equation, and trained three ML models to predict sound propagation over long ranges. They found that with a single layer of atmospheric turbulence, their simulated sound waves were predicted within $5 – 7$ dB of the experimental values. However, they noted that they had worse performance with multiple layers of turbulence. Additionally, their maximum test range of 8KM is much smaller than the length of the Earth's atmosphere and their data was one-dimensional sound waves rather than two-dimensional image observations. In that domain, Wang et. al. focused on predicting the next

observation frame of a telescope based on prior frames using a residual learning fusion network [10]. After training the model, they were able to estimate using the six prior frames with structural similarity index measure (SSIM) scores above 0.94 for several different telescope diameters and r0 values. As with Hart et. al., the test range was much smaller than the full range of the atmosphere, at only 1 KM. Additionally, the model is trained on the phase screen rather than the raster observations directly.

Our contribution in this paper is studying how to predict relevant characteristics of multiple layers of atmospheric turbulence. Prior work often focused on a single ground layer of atmospheric turbulence or a composite of all layers at once, using empirical or simulated sensor data. We believe the necessary information can be learned from the raster data itself.

3. METHODOLOGY

In the following section, we describe the main components of the proposed process: the generation of speckled image rasters with simulated layers of atmospheric turbulence, the encoding of this data to reduce dimensionality and extract relevant features, and finally training a model to predict the wind speeds of each layer based on this input.
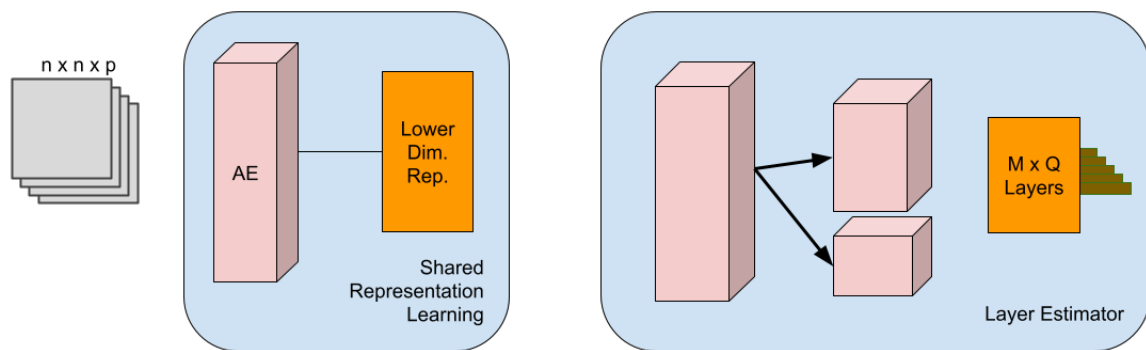

Fig. 1: Overview of Layer Estimator Process

3.1 Data Generation

We first developed a series of wind profiles to represent each of the four layers in the simulated atmosphere. These profiles were created using formulas designed by Roberts and Bradford [7] as well as Montillaa and Montoyaa [6]. While Roberts and Bradford provide a more detailed explanation in their paper, the applicable formula used in our data generation process was:

$$v(z) = v_G + v_T \exp\left(-\left[\frac{z\cos\zeta - h_T}{L_T}\right]^2\right)$$

where $v(z)$ represents the wind velocity v at a specified elevation z, v_G represents the ground layer wind profile and v_T represents the wind velocity at the atmosphere's tropopause boundary, ζ represents the angle of observation, h_T represents the height of the tropopause and L_T represents its thickness. The elevations of the wind layers were 1KM, 5KM, 12KM, and 25KM respectively, based on the average values used in studies such as Montillaa and Montoyaa[6]. Montillaa and Montoyaa's work was also used to adjust the formula's layer effects to handle the higher altitudes, and an additional clamp was performed in code to prevent the possibility of negative wind speeds that the formula might produce. We also added a small amount of additional noise to the wind vector values, to increase our generator's coverage of possible wind layer profiles.

Figure 2 shows a histogram of the velocities and layer heights for four simulated wind layers over Hawaii. The equation generates this information for seven different locations: Hawaii, Oakland, San Diego, Tucson, Flagstaff, Tenerife, and Antofagasta. As seen in Figure 4, we also produced different wind profiles for all twelve months of the
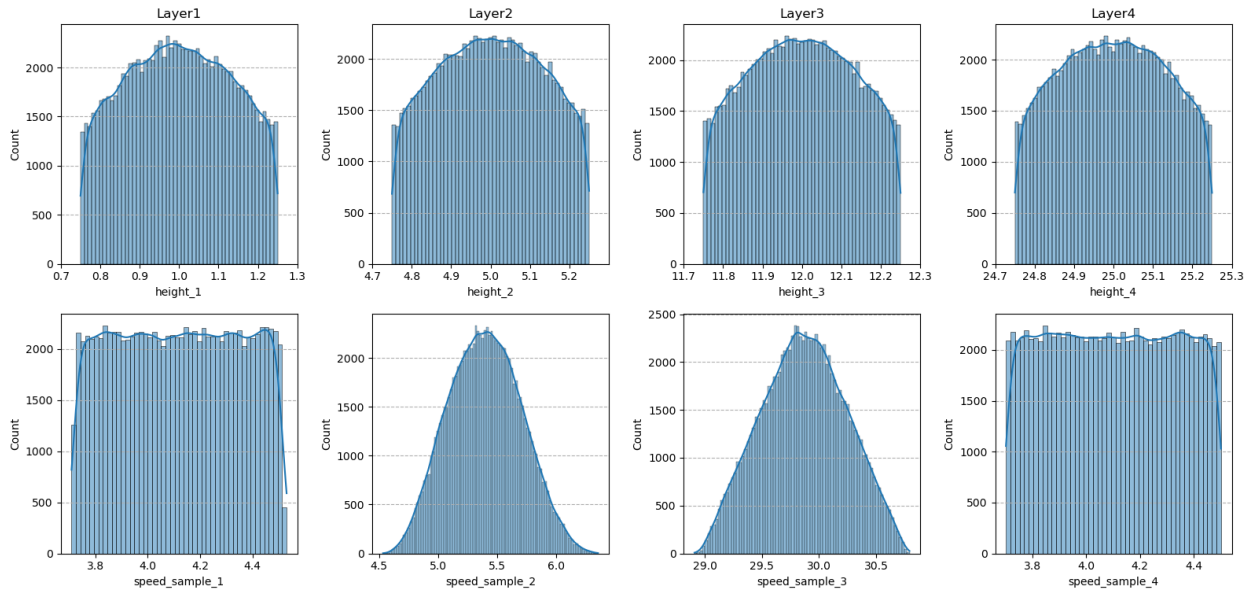
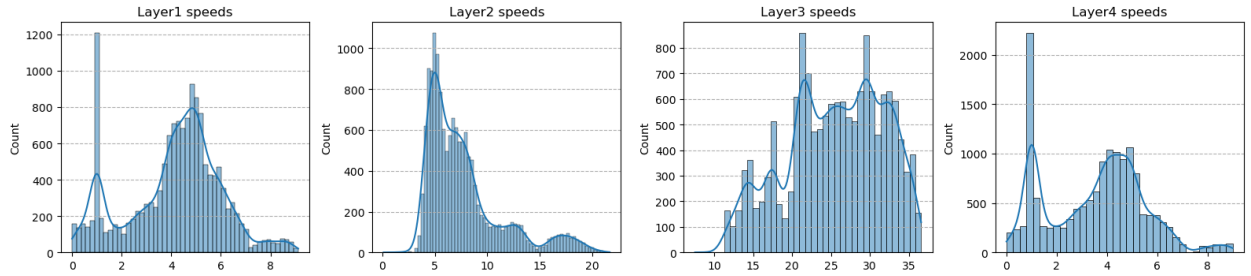Fig. 2: Histograms of sampled layer heights and wind speeds for Hawaii

Fig. 3: Histogram of wind speed distributions for the seven cities combined

year. We intended to utilize the recorded metadata about the location and month in future studies to further categorize and extract features from the training data.

The outputted profiles are stored as 2D vectors consisting of the wind speed (m/s) and direction (degrees). Additional metadata on the simulated location, month, and year are also stored for future labeling, along with an array of coefficient values which are used later in model creation.

The wind profiles were then used to generate sequences of convolved rasters that simulate observations of a blurred image over time. This involves creating a digital model of a telescope, and then simulating the viewing of a solar object through multiple layers of atmosphere over a period of time. For this study, we used a 512px by 512px grayscale raster of a single source of light as the ground truth image.

The simulated telescope has an aperture diameter of 3.6 meters, with a minimum pupil sampling rate of 0.04 meters per pixel, an exposure time of 0.01 seconds, and a minimum and maximum observing wavelength of 550.0e-9 meters. The object is treated as if it were 35,786,000 meters away from the telescope lens. The distance from the zenith is simulated as $\frac{34}{360}\pi$ radians.

The ground truth raster of the object is then loaded from a FITS image file. The raw pixels are multiplied by a series of co-efficients, to simulate the effects that materials such as kapton and aluminized mylar would have on the observation as they collide with the solar panels and parts of the telescope. It is then converted into the frequency domain via recursive Fourier transforms to approximate the photon as it travels from the object to the telescope.
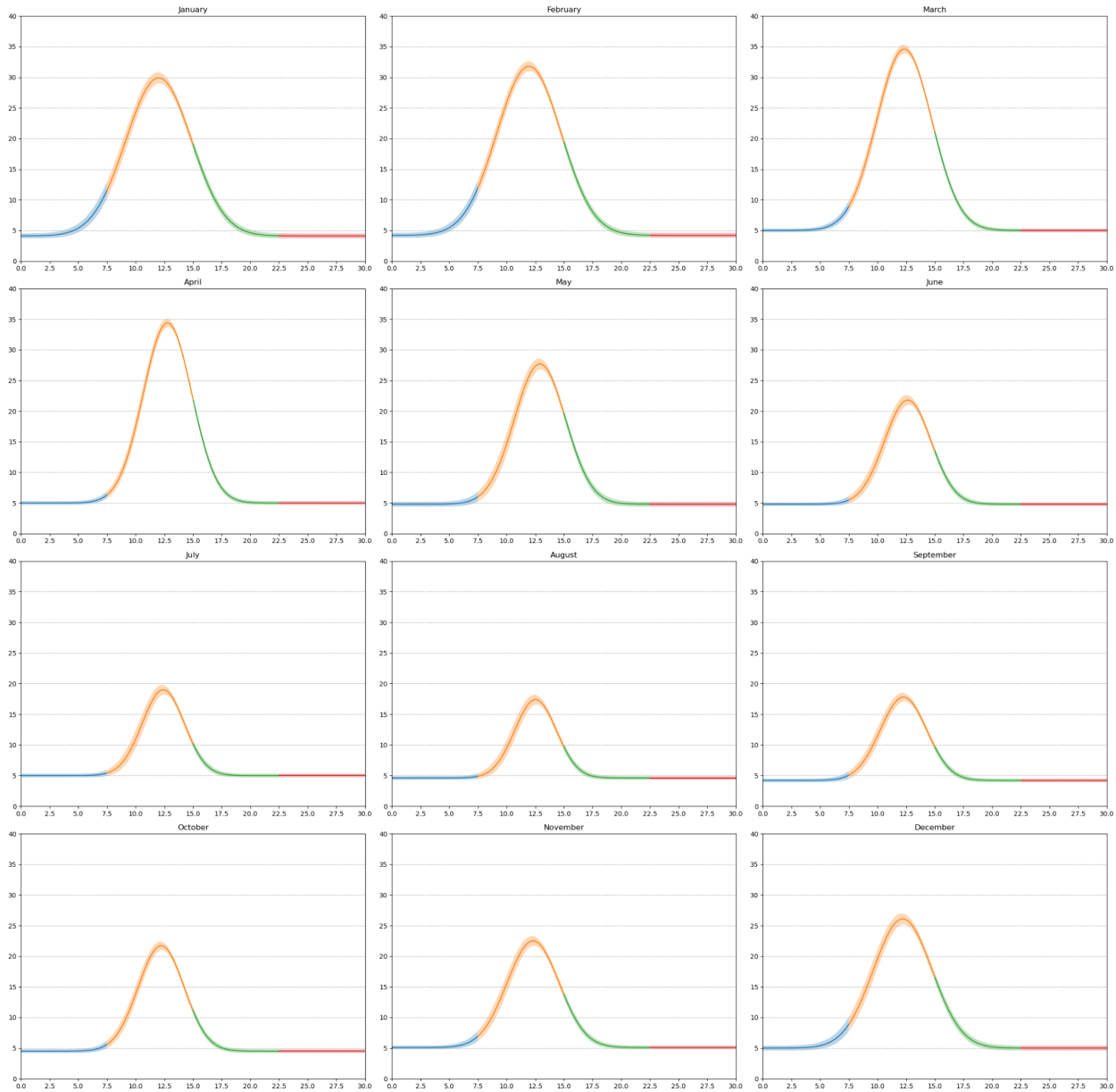
Fig. 4: Seasonal variation of wind speeds along with range for 12 months in Hawaii

Once the simulated telescope is initialized and the object is converted in a spectral representation, they are used to generate a C_N^2 value using the specified layer elevations. The phase screen and atmospheric model is then built using this information and the inputted wind layer profiles. This atmospheric model is finally used to generate and apply a convolution over time to the ground truth observation image. We assumed the atmospheric flow was frozen for the length of the simulated observation. The final output of the data generation process is a sample of 2,000 raster frames of the blurred image observed over time. Each frame is 256 pixels wide and tall with eight bit integer precision. Metadata including the wind profiles, configurations, and Zernike Polynomials are also recorded in the H5 output file for current and future studies.

3.2 Model Design

Once the data was generated, we next needed to encode it to reduce dimensionality and associate it with relevant labels. For the scope of this project, our main labels were the per-layer wind speed and direction. We also evaluated
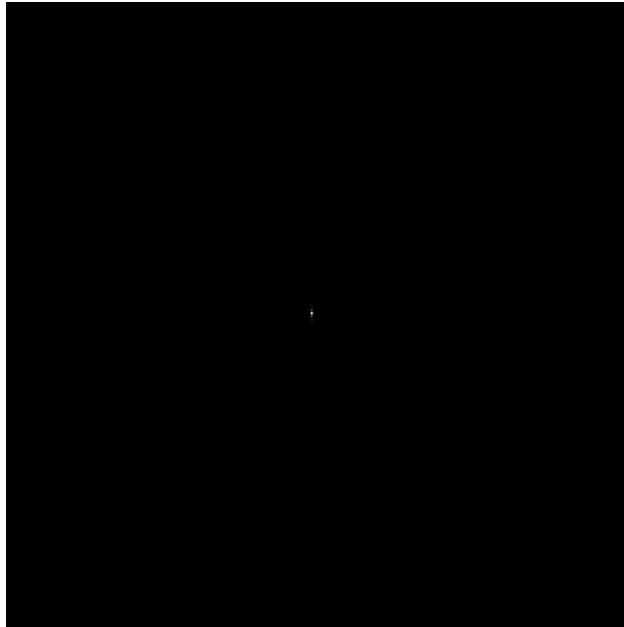
Fig. 5: Raster representing a single point of light with no convolution applied
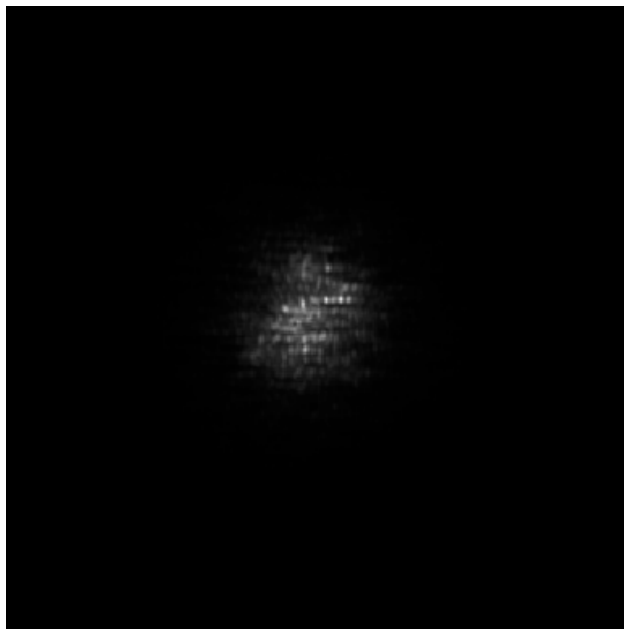

Fig. 6: One frame from a sequence of convolved rasters generated from Fig. 5

the model's ability to reconstruct the original image as a second task.

During experimental design, we considered three candidate autoencoding architectures: a Convolutional Neural Network (CNN), Variational autoencoders (VAE), and Fourier Neural Operators (FNO) based encoder. Table 1 shows an initial comparison of validation Mean Absolute Error (MAE) and Mean Square Error (MSE) trained on 1,740 samples (containing four wind layers) with a training/testing/validation split of 1290/250/200 for 5,000 epochs.

As seen in Table 1, the highest initial MAE and MSE values were observed in instances of CNN and FNO. Given the limitations of our current data generation method, we chose to focus on one model type for the initial study. Through

Table 1: Comparison of Candidate Architectural Models

Model Type	Data Mode	Latent Dimensions	Validation MAE	Validation MSE
CNN	Slice	1000	0.27	0.12
CNN	Slice	100	0.22	0.09
CNN	Default	1000	0.25	0.11
CNN	Default	100	0.26	0.11
VAE	Slice	1000	0.24	0.09
VAE	Slice	100	0.26	0.11
VAE	Default	1000	0.23	0.09
VAE	Default	100	0.23	0.09
FNO	Slice	1000	0.24	0.10
FNO	Slice	100	0.26	0.10
FNO	Default	1000	0.25	0.10
FNO	Default	100	0.27	0.12

multiple tests, we found the Fourier Neural Operator model to most consistently have the highest R^2 skill score and Peak-Signal-to-Noise ratio and the lowest Mean Square Error (MSE) and intensity-filtered MSE, therefore we focused on this model.

The loss function was defined as:

$$\ell(I_{(D,c,h,w,d)}, L) = h * w * MSE_I + MSE_L$$

where I represents the image, L represents the labels, c representing the color channels of the image, h and w represent the height and width of the image in pixels, d represents the depth of the image. As each frame is 256 pixels tall by 256 pixels wide, in the initial study the value of $h * w$ is always 2^{16}.

3.3 Model Training

The model was implemented using the Python programming language and the PyTorch library.

As noted earlier, a limiting factor in training our model was the speed and quantity of samples produced by our data generation process. This was especially challenging because our models are data intensive and require a large amount of training samples. Thus, rather than wait to train on a large data set, we instead trained on small sets of 1,500 total samples with a persistent model. We also utilized data slicing, with 10 layer skip connections.

As the frames are time-series data, we used overlapping data windows of 100 and 1,000 per sample for training, shifting the window by ten frames each time. The batch size was eight and the optimizer used was Adam, and a Cosine Annealing Learning Rate of 0.0001 served as the scheduler. A pre-trained ResNet was used as a regression model for the prediction task. The models were trained for 5,000 epochs. We compared reducing the latent space dimensions to 100 and 1,000 during encoding in our evaluations.

4. RESULTS AND DISCUSSION

The results described in this section were derived from training the model on two sets of 1,500 data samples.

All three candidate architectures (CNN, VAE, and FNO) were able to learn the image reconstruction tasks within fifty epochs. However, it took the full 5,000 epochs for the FNO model to gain experience with label prediction. Furthermore, neither the VAE nor CNN (with frame skipping) models showed much ability to correctly predict labels even after 5,000 epochs. CNN and VAE models are commonly used with image input data, so they may be better adapted to tasks related to recreating an image.

In Table 2, we see slight improvement in the Peak Signal-to-Noise Ratio (PSNR) in the model after training on the second dataset compared to only on the first. While there is not a significant difference, it does indicate a possibility that further improvement is possible with an increase in training data.

Table 2: Peak Signal-to-Noise Ratio (PSNR)

Dataset	PSNR
Dataset 1	43.88
Dataset 2	44.15

Table 3: Mean Square Error (MSE)

Dataset	MSE
Dataset 1	0.06894
Dataset 2	0.06516

Table 4: Intensity Filtered Mean Square Error (MSE)

Dataset	Intensity Filtered MSE
Dataset 1	0.06717
Dataset 2	0.07089

Table 3 for MSE and Table 4 for intensity-filtered MSE show similar gradual improvements from the first to the second set of training data samples.

Table 5: R^2 Error Metric for Wind Speed Prediction

Wind Layer	R^2
Layer 1	0.014114022
Layer 2	-0.11379993
Layer 3	0.76637888
Layer 4	0.206390917

As seen in Table 5, the third and fourth layers have a significantly larger R^2 value than the first and second layers. This result may be due to the differences in velocity, as the higher layers have faster average wind speeds than the layers closer to the ground. The elapsed time between each simulated frame in a sample is five milliseconds, for a total observation time of two seconds per sample. It may be that the simulated pocket of atmospheric turbulence are not fully moved through the observed region at lower wind speeds. This would reduce the amount of information that the lower wind layers would contain, which could affect the quality of the prediction.

Table 6: R^2 Error Metric for Wind Direction Prediction

Wind Layer	R^2
Layer 1	-0.292384267
Layer 2	-0.175454259
Layer 3	-0.196609616
Layer 4	-0.053709745

Table 6 shows that in contrast to wind velocity, the R^2 values for wind direction were negative for all layers. There are several possible explanations for this performance. A larger training data set may be required to capture sufficient detail about this feature. It may also be that a blurred raster of a single point of light does not contain sufficient information to estimate wind layer direction, and that we would need to repeat the experiment with a more complex object or multiple reference points in order to better learn this information.

When examining the loss function, it is noted that the $h * w$ coefficient of MSE_I is much larger than that of MSE_L. This imbalance may give a larger importance to the image prediction rather than the labels, which would impact our model's ability to converge . Improvements may be gained by decreasing the size of the MSE_L co-efficient after each new batch during the training process. Another avenue to explore would be measuring the MSE of individual features rather than the error of the labels as a whole. So MSE_L could become $\alpha_1 * MSE_{WVL1} + \alpha_2 * MSE_{WVL2} + \alpha_3 * MSE_{WVL3} + \alpha_4 *$

MSE_{WVL4} where $WVL(n)$ represents the velocity of wind layer n.

5. CONCLUSION

This paper outlines a process for estimating wind profile information for multiple atmospheric layers from simulated blurred images of solar objects. While our work was limited by the initially slower output of our data generation process, we have seen some predictive results for fast-moving wind layers at the higher elevations. Layer 3 had the best performance with a 0.76 R^2 skill score, followed by Layer 4 with a 0.2 R^2 skill score. We have also observed gradual improvements in several of the model's wind speed predictions as we have trained it on subsequent collections of training data. This lends credence to the idea that the model's performance will improve further with a sufficiently large amount of training data. We believe there is value in further exploration of this approach to learn multi-layer wind profile information for use in ground-based telescope observation correction.

6. FUTURE WORK

Our immediate focus is increasing the output of our training and data generation software. We are porting the code to the HCIPy library and Python programming language so that we can produce more samples in a shorter amount of time. We believe that community-maintained dependencies will be faster and more memory-efficient than our in-house software, and the efficiency boost will allow us to more quickly test hypotheses and train our models with much larger and diverse datasets.

As we develop a larger training dataset, we will explore predicting more information about the wind layers in the blurred images. In addition to determine the R0 values, we are also interested in determining relevant Zernike Polynomials which could be used to further correct the distortions in the image. These Zernike Polynomials are already included as metadata within the datasets, but not yet utilized by the model. We also intend to test different configurations, such as larger numbers of wind layers, more complex objects, and different telescope aperture diameters.

The next phase of the research will be using the Layer Estimator models as part of a larger system to deblur observed images directly.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Lidiia Bolbasova, A Andrakhanov, and Artem Shikhovtsev. The application of machine learning to predictions of optical turbulence in the surface layer at baikal astrophysical observatory. *Monthly Notices of the Royal Astronomical Society*, 504:6008–6017, 05 2021.

[2] Richard Davies and Markus Kasper. Adaptive optics for astronomy. *Annual Review of Astronomy and Astrophysics*, 50(Volume 50, 2012):305–351, 2012.

[3] Carl R. Hart, D. Keith Wilson, Chris L. Pettit, and Edward T. Nykaza. Machine-learning of long-range sound propagation through simulated atmospheric turbulencea). *The Journal of the Acoustical Society of America*, 149(6):4384–4395, 06 2021.

[4] Stuart M. Jefferies and Julian C. Christou. Restoration of Astronomical Images by Iterative Blind Deconvolution. , 415:862, October 1993.

[5] R. G. Lane. Blind deconvolution of speckle images. *J. Opt. Soc. Am. A*, 9(9):1508–1514, Sep 1992.

[6] Luz Montoya and I. Montilla. The multi-conjugate ao system of the est: Dm height determination for best performance using real daytime statistical turbulence data. In *Adaptive Optics for Extremely Large Telescopes 5*, 01 2017.

[7] Lewis C. Roberts and L. William Bradford. Improved models of upper-level wind for several astronomical observatories. *Opt. Express*, 19(2):820–837, Jan 2011.

[8] Zhu Shi-qing, Yang Ling, Cong Wen-sheng, Yang Rong, and Hua Jun. Iterative blind deconvolution algorithm for support domain based on information entropy. In *Proceedings of the 2019 International Conference on Blockchain Technology*, ICBCT '19, page 39–44, New York, NY, USA, 2019. Association for Computing Machinery.

[9] T.G. Stockham, T.M. Cannon, and R.B. Ingebretsen. Blind deconvolution through digital signal processing. *Proceedings of the IEEE*, 63(4):678–692, 1975.

[10] Ning Wang, Licheng Zhu, Shuai Ma, Wang Zhao, Xinlan Ge, Zeyu Gao, Kangjian Yang, Shuai Wang, and Ping Yang. Deep learning-based prediction algorithm on atmospheric turbulence-induced wavefront for adaptive optics. *IEEE Photonics Journal*, 14(5):1–10, 2022.