Decoding optical aberrations of low-resolution Instruments from PSFs: machine learning and Zernike polynomials perspectives

Lucas Sauniere, William Gillard, Julien Zoubian

Decoding Optical Aberrations of Low-Resolution Instruments from PSFs: Machine Learning and Zernike Polynomials Perspectives

L. Sauniere[1], W. Gillard[1], and J. Zoubian[1]

[1]Aix Marseille Université, CNRS/IN2P3, CPPM, Marseille, France

ABSTRACT

In this paper, we explore a way to extract the Zernike coefficients from low-resolution PSF images using a neural network model. The goal of this study is to obtain accurate reconstructions of instrumental responses even with under-sampled PSFs.

We used the Python module POPPY to simulate a Newtonian telescope system with a primary mirror diameter of 1 meter and a secondary mirror diameter of 0.2 meter. PSFs were simulated over a wavelength range from 200 nm to 1000 nm. Detector sampling parameters included a pixel scale of 0.05 arcseconds/pixel and a 32x32 pixel grid. The ZerNet model was developed based on the Inception architecture. The input is a 32x32 pixel PSF image and the output is a set of Zernike coefficients. The model has three blocks of convolutional kernels of different sizes. These are combined and flattened, then pass through several layers of dense neurons before being activated by a hyperbolic tangent function to predict Zernike coefficients.

The model was trained using the Adam optimizer with a learning rate of 0.001 over 40 epochs and a batch size of 64. The data was divided into three sets: training (70%), validation (20%), and test (10%). The ZerNet model showed good accuracy in predicting Zernike coefficients from PSF images. The accuracy improved with increasing wavelength, reaching 95.34% at 1000 nm. The PSF reconstruction error was measured using the Frobenius norm and showed a reduction in error with higher order Zernike coefficients. The model showed that the median error decreased with increasing order, proving that including higher order Zernike coefficients helps with PSF reconstruction.

Keywords: Zernike Polynomials, Point Spread Function, Deep Learning, Neural Network, Optics

1. INTRODUCTION

The accurate modeling of the Point Spread Function (PSF) is essential to understand the instrumental impact on images taken by optical systems. As a key indicator of an instrument's optical quality, it reveals the nature of its aberrations which are fundamental parameters in evaluating image quality and systematics. Knowledge of the instrumental response plays a crucial role in applications such as image deconvolution, where it enables the extraction of improved images from observed data. Using the PSF to deconvolve images can improve their sharpness and extract finer details, but it's essential to control the process carefully because deconvolution can also magnify noise. Fortunately, these general image quality enhancements can greatly contribute to our comprehension of physical phenomena [1].

To generate the most realistic PSF of an optical system, numerous methods have been employed, with either data-driven strategies [2] or model-based approaches [3]. For example, model-based adaptive optics [4] combine optical system models with iterative algorithms that evaluate the optical system's performance and identify phase aberrations to adjust mirrors shapes as a correction.

In this study, our objective is to develop a machine learning model that can establish a connection between Zernike coefficients, which define the Wavefront Error (WFE) of optically deformed systems, and the related PSF images. This algorithm final purpose is the accurate modeling of optical aberrations, especially in cases where PSFs are under-sampled, to allow an accurate reconstruction of instrumental response. Resulting from the combination of neural networks and Zernike Polynomials, a brand-new model was created: ZerNet.

Further author information: (Send correspondence to L.S.) E-mail: sauniere@cppm.in2p3.fr

2. METHODS

2.1 Zernike Polynomials and Wavefront error

2.1.1 Theoretical background

A powerful mathematical framework for explaining optical aberrations is offered by Zernike polynomials. These orthogonal functions depict various aberration modes found in optical systems. They are denoted as $Z_n^m(\rho,\theta)$, where (n,m) is a couple of non-negative integers defining radial and azimuthal frequency, and (ρ,θ) is a couple of normalized radial distance and azimuthal angle. The equation for Zernike polynomials in polar coordinates is given by:

$$Z_n^m(\rho,\theta) = \sum_{k=0}^{\frac{n-|m|}{2}} \frac{(-1)^k (n-k)!}{k!\left(\frac{n+|m|}{2}-k\right)!\left(\frac{n-|m|}{2}-k\right)!} \rho^{n-2k} \cdot \begin{cases} \cos m\theta \\ \sin m\theta \end{cases} \tag{1}$$

As it can be seen in this equation, the polynomials depend on either a sine or a cosine part, which is complementary to the description of the pupil distribution. Thanks to Z_n^m, the WFE of an optical system can be represented as a linear combination of Zernike polynomials. Let $\Phi(\rho,\theta)$ denote the WFE, and let α_n^m represent the Zernike coefficients. The WFE can be written as:

$$\Phi(\rho,\theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_n^m \cdot Z_n^m(\rho,\theta) \tag{2}$$

where α_n^m are the expansion coefficients.

Among indexation systems for Zernike polynomials, Noll's indexation [5] (typically denoted by j), which is used by the *poppy* [6, 7] library, provides a common and standardized method to easily reference particular polynomials within the Zernike basis. Using this indexation, the equation (2) then becomes :

$$\Phi(\rho,\theta) = \sum_{j=1}^{\infty} \alpha_j \cdot Z_j(\rho,\theta) \tag{3}$$

This new expression now allows us to have a simple linear description of the Zernike polynomials coefficients, which allows to produce the Figure 1 :

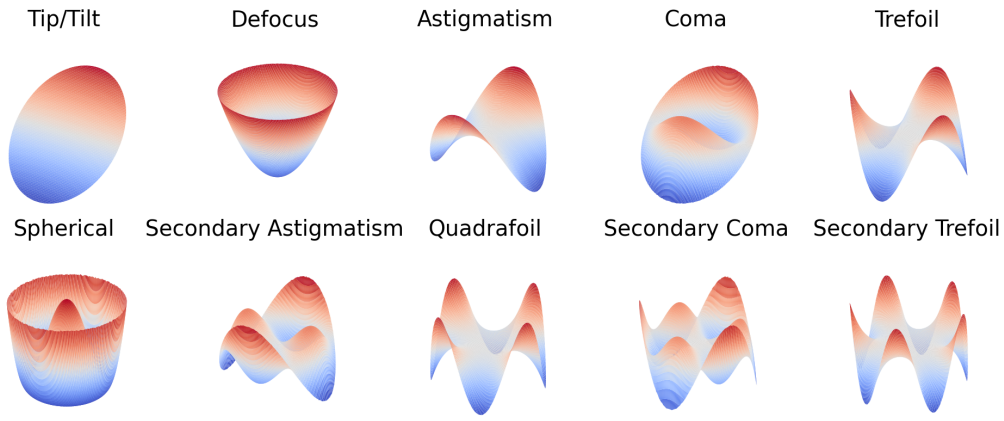


Figure 1: Visualization of Zernike polynomials and optical aberrations.

Understanding the wavefront error (WFE) in optical systems, which is represented by the above equations, is particularly important in this context.

2.1.2 Budgeting and Constraints

However, in real-world situations, optical systems are tuned to reduce higher-order aberrations, thereby limiting Noll's index j to N_{orders}. This allows us to express the maximum values for each Zernike coefficient α_j (in nanometers of optical path difference) as a finite WFE budget, noted B, in our simulations :

$$B = [0,\ 100, 100,\ 50, 50, 50,\ 36, 36, 36, 36,\ 18, 18, 18, 18, 18,\ 9, 9, 9, 9, 9, 9]$$

Regarding the usage of Zernike polynomials, we made a number of important choices in this work. First, by restricting the number of orders to 21 (including Piston), we were able to focus our future analysis of the optical system on the optical aberrations that mattered most. Furthermore, in order to lessen the effect of higher-order aberrations, the amplitude of each order was selected to decrease as the order index increased. This method guarantees that we reduce errors in estimating low-order aberration contributions, which are crucial to our analysis. Additionally, small coefficient high order aberrations would hardly affect the PSF at all, making them undetectable, which is why we have not considered optical aberrations having only lower amplitude than 9 nm.

Another important aspect of this budget is that it assigns equal values for each radial order. This implies that all Zernike modes, regardless of their azimuthal order, with the same radial order will have values constrained within the same range. For example, regardless of the azimuthal index, the Zernike coefficients for radial order 2 are defined to lie between -50 and 50 nm. This method mimics the physical tuning frequently performed in optical systems to minimize aberrations uniformly across modes of the same radial order, but it also makes managing aberration budgets easier.

Equalizing the values within each radial order is a strategy that accomplishes several goals:

1. **Uniformity in Aberration Management**: By treating all modes within the same radial order equally, it becomes easier to predict and control the system's overall aberration profile.
2. **Simplification of Error Budget Allocation**: Equal values within radial orders allow for a straightforward allocation of the WFE budget, ensuring that no single mode within a radial order disproportionately affects the optical performance.
3. **Practical Reflectance of Real-world Tuning**: Optical systems in practice are often adjusted to minimize aberrations consistently across similar radial orders, which aligns with our simulation approach.
4. **Focus on Low-order Aberrations**: This method prioritizes the reduction of low-order aberrations, as the decreasing amplitude strategy reduces the influence of higher-order aberrations, which typically have a lesser impact on image quality.

To simplify the expression of B, we created a shortened version named B_s, of the form : $B_s = [0, 100, 50, 36, 18, 9]$
This new definition makes it possible to write:

$$\forall (n, m) \in \mathbb{N} \times \mathbb{Z}, \quad |\alpha_n^m| \leq B_{s_n} \tag{4}$$

A way to describe the relationship between B and B_s would be :

$$\forall n \in \mathbb{N}, \forall n' \in n, n+1, \ldots, 2n, \quad B_{n'} = B_{s_n} \tag{5}$$

In establishing these constraints on Zernike coefficients through our budgeting approach, we ensure a focused analysis of optical aberrations crucial to our system's performance. Since these limitations are in place, our next task involves simulating Point Spread Functions (PSFs) to investigate how these limited optical aberrations appear during the imaging process.

2.2 PSFs simulation

As our research centers on the computation and analysis of PSFs from optical systems with low resolution, all image plots are oversampled by a factor of 4 in order to facilitate visualization.

2.2.1 Designated optical system

The Python module *poppy* allowed to model a telescope with a Newtonian shape, consisting of two mirrors, a primary and a secondary supported by three struts. In this configuration the mirrors' sizes have been set to $D_{primary} = 1\text{m}$ and $D_{secondary} = 0.2 \cdot D_{primary} = 0.2\text{m}$, and the number of secondary supports to 3.

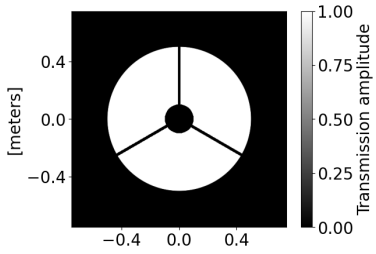


Figure 2: Optical system pupil function represented as transmission amplitude.

We selected this specific optical design because it is commonly used, particularly in astronomy, and its configuration remains basic allowing to be customized and adapted for specific research purposes. In our ML approach which requires extensive samples of PSF, simulating accurately PSFs takes a significant amount of time, which is why approximations used within *poppy* make it an ideal tool.

Additionally, this module includes a ZernikeWFE class that, by taking as input the coefficients for each optical aberration, defines the wavefront error contribution of any optical element in terms of its Zernike components.

In this example, every optical aberration coefficient is set to $100nm$ to display its impact on the PSF shape, while all the other coefficients are set to 0. This change in the PSF is best seen in its central region for *defocus* and *astigmatisms* where we distinguish similar shapes to the WFE in fig.1 (gaussian shape for *defocus* and cross shape for *astigmatism*).

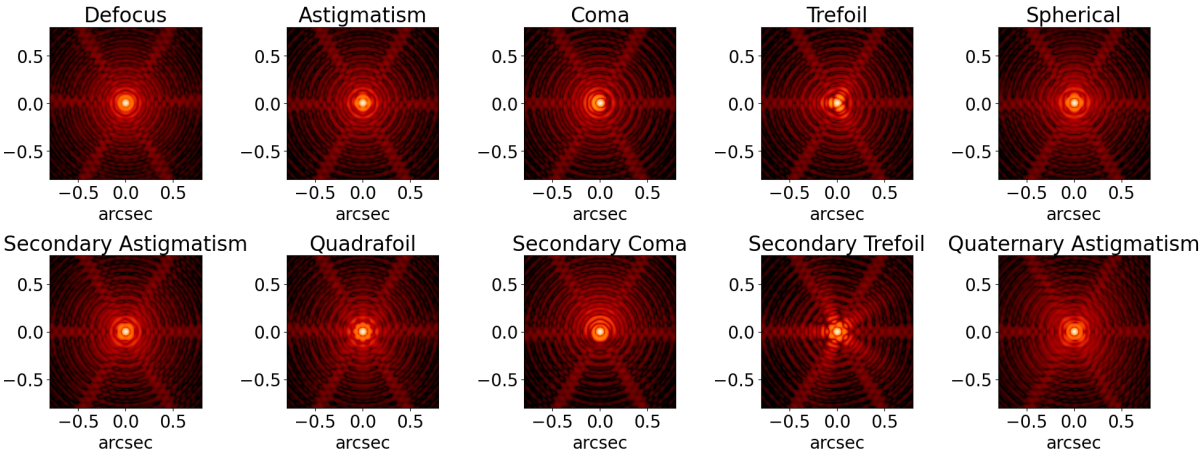


Figure 3: Impact of optical aberrations on the shape of the PSF.

The wavelength range of 0.2 μm to 1 μm was strategically chosen because it includes the visible light spectrum, which is important for many optical imaging applications. This range enables in-depth examination of PSF behavior at various wavelengths, guaranteeing a complete comprehension of optical system performance.

2.2.2 Fraunhofer approximation

Involved in light propagation, diffraction occurs in many applications and places a fundamental restriction on the amount of resolution and detail that imaging systems can achieve.

The Fraunhofer approximation applies when the light waves have traveled a sufficiently large distance from the diffracting aperture or when the light is focused through a lens, making the source appear at infinity. In these circumstances, the wavefronts are regarded as planar, and the mathematical treatment becomes simpler, making the computation and analysis of diffraction patterns easier.

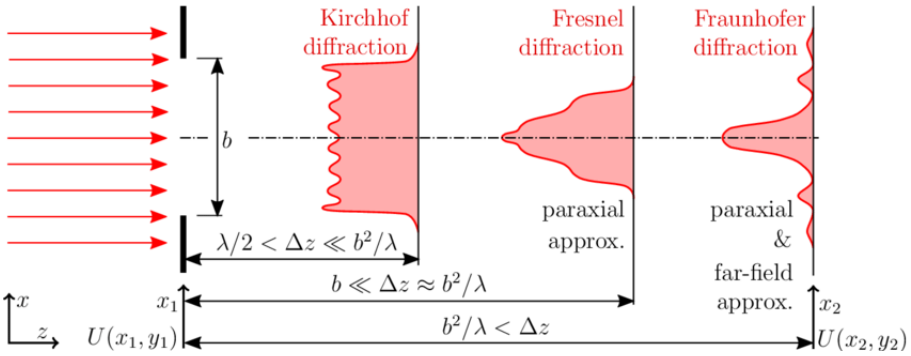


Figure 4: Diffraction regions after aperture / Credit : Figure from Störkle [8]

As objects in astronomical imaging are usually located at great distances, the Fraunhofer approximation's far-field assumption is valid. Furthermore, the optical design and calculations are made simpler by the paraxial approximation, which makes the assumption that light rays make small angles with the optical axis. High precision and clarity are crucial in the design and analysis of telescopes and other astronomical instruments, so this is especially helpful in those areas.

As a result, the Fraunhofer diffraction region in *poppy* offers a compromise between the accuracy required for high-resolution astronomical imaging and computational efficiency.

2.2.3 Detector sampling

In addition to chosen approximation mentioned in previous section, other physical parameters limiting our accuracy to a finite value, were set in our simulation, such as the pixel-scale at 0.05 arcseconds/pixel and a wavelength range. We also set the number of pixels to 32x32, thus in a square shape. This size choice simply relies on the powers of 2 being commonly used in ML algorithms.

It is critical to comprehend the implications for image resolution and analysis given this configuration. The specific choice of 32x32 pixels is adequate for allowing efficient computation made possible by tiny datasets, while preserving detail within images. Considering our image size of 32x32 pixels, we end up with an apparent angular size observed of 32 pixels $= 32 *$ pixel-scale $= 1.6$".

When it comes to resolving fine features within an image, high-aperture optical systems are a necessity. Two point sources are deemed just resolved by the Rayleigh criterion if their separation is at least $r/2$, where r is the Airy disk's radius. This criterion was developed in the 19th century by Lord Rayleigh and offers a basic framework for determining an optical system's resolving power.

This criterion's mathematical expression is as follows :

$$\theta_{\min} = 1.22\frac{\lambda}{D} \tag{6}$$

where $\theta_{\min}$ is the angular resolution (in radians), λ is the wavelength of incident light, and D the diameter of the aperture (in our case, the primary mirror diameter $D_{primary}$).

Due to prior decisions, the criterion from the equation (6) simplifies as : $\theta_{min} = 1.22 \cdot \lambda$ thanks to primary's diameter of 1 meter. We may compute the resolving power in this case : $\lambda = 200nm \Rightarrow \theta_{min} = 1.22 \cdot 10^{-6}\ rad \approx 0.25$".

This leads us to the link between the resolving power and the field of view (FoV) of the selected optical system :

$$FoV = 1.6" \wedge \theta_{min} \approx 0.25" \implies FoV \approx 6.4 \cdot \theta_{min}(\lambda = 1000nm) \tag{7}$$

Using similar development, we are able to compute $\theta_{min}(\lambda)$ for all chosen wavelengths within this study, see table 1.

Wavelength (nm)	200	400	600	800	1000
Resolving Power (")	0.050	0.101	0.151	0.201	0.252
Fraction of FoV (FoV/θ_{min})	31.79	15.90	10.60	7.95	6.36

Table 1: Characteristics of considered optical systems.

As the wavelength considered increases, the PSF size within our region of interest becomes greater, making our box cut more and more PSF information.

We use the metric of encircled energy (EE) to estimate and quantify PSF loss during the cropping process, so as to guarantee that no meaningful information is lost from the Point Spread Function (PSF). By calculating the percentage of the total PSF energy contained within a specified radius, encircled energy provides a clear image of how much of the PSF's information is contained in the area of interest.

The ideal cropping boundaries that preserve the majority of the PSF's energy while preserving its crucial characteristics and features can really be precisely determined thanks to EE computing. We can balance the need to focus on a specific location with the necessity to maintain the integrity of the PSF data using this strategy. Thus, this method ensures the precision and effectiveness of our analysis and the subsequent imaging processes, protecting vital information for better results.

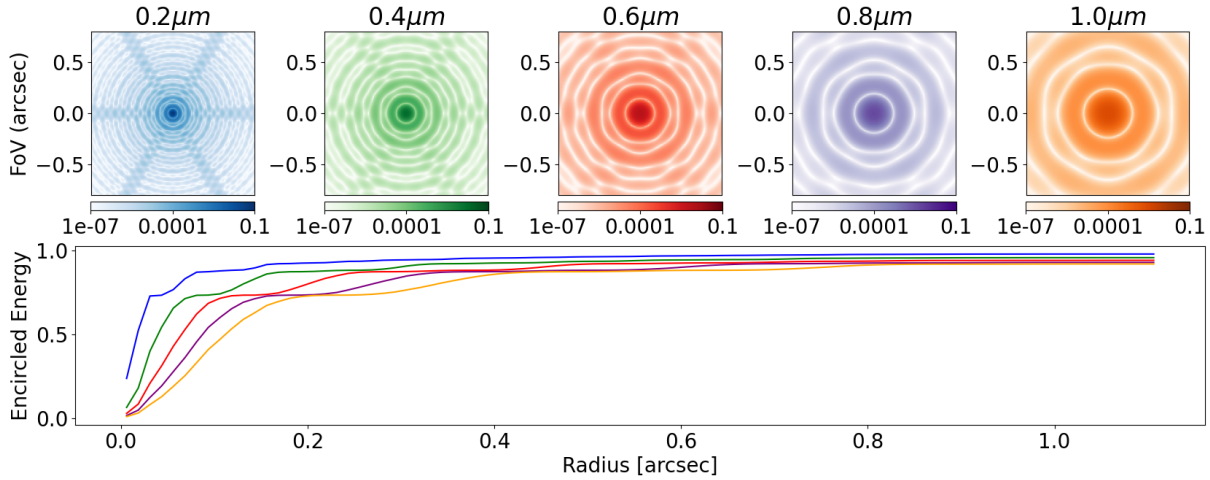


Figure 5: Point Spread Functions and Encircled Energies for λ between 0.2 μm and 1 μm.

The cumulative fraction of total PSF energy enclosed within increasing radii is shown by the previous EE plots in fig.5. It is evident from the quick initial growth and convergence of each EE curve towards an asymptote close to unity that nearly all of the PSF energy (up to 98%) is contained within our field of view.

Now that we have covered the fundamentals of PSF analysis, we will apply machine learning approaches to improve our comprehension of WFE impact on PSF data.

2.3 Machine Learning

2.3.1 Data simulation and pre-processing

As was already mentioned, a range of wavelengths from 200 nm to 1000 nm was chosen to be taken into consideration. We did thorough simulations of the Point Spread Functions (PSFs) over this spectrum in order to accomplish this. With 100,000 monochromatic PSFs in each dataset, the behavior of the optical system at various wavelengths is fully represented.

The large amount of data produced by these simulations required management, thus we saved each dataset in HDF5 format. This format was selected because it meets our demands perfectly and is effective for storing massive datasets with intricate structures. We made sure the datasets stayed accessible and manageable by structuring them into HDF5 files, which made it easier to train machine learning models and do additional research.

We used two different normalization methods for the PSF images and the associated Zernike polynomial parameters. For the PSF images, we applied a min-max normalization to adjust the scale of the pixel values to the range $[0, 1]$. This was chosen to standardize the scale and preserve the relative distribution of pixel intensities. For the Zernike polynomial parameters, we used a max normalization to scale the parameters so that they have a maximum of one. This normalization approach was chosen to preserve the zero point of the original parameter.

The data set was divided into training set (70%), validation set (20%), and test set (10%). Here we refer to the standard naming of the dataset in machine learning, the training set is used to train the model, the validation set is used while tuning the model to prevent overfitting, and the test set is used to evaluate the model performance.

2.3.2 ZerNet model

To link the images and their corresponding coefficients, we made use of Zernet. This study's model is built on the Inception model's premise (Szegedy et al. 2014 [9]), which concatenates convolution block outputs to extract different image features, repeated three times consecutively. In this instance, we have a $2D$ tensor as the input and a $1D$ tensor of N_{orders} values as the output.

The model architecture is illustrated in Figure 9 in appendix A and described below:

- **Input Layer:** Input a PSF image of dimensions 32×32 pixels.
- $3\times$ **Inception Blocks:** Parallel blocks with 1x1, 3x3, and 5x5 convolutional kernels, extracting features and then concatenated and flattened into a single feature vector.
- **Fully connected layers:** After flattening the concatenated features, multiple dense layers result in an output layer that uses a hyperbolic tangent function tuned to predict the N_{orders} coefficients of the Zernike polynomials associated with each PSF image.

We used the mean squared error (MSE) as the loss function :

$$MSE = \frac{1}{N_{orders}} \sum_{j=1}^{N_{orders}} \left(Y_j - \hat{Y}_j\right)^2, \tag{8}$$

where Y are the true values of each coefficient α_j and $\hat{Y}$ are the true values.

The model was trained using the Adam optimizer (Kingma & Ba, 2014 [10]), with the default values of the Keras implementation: learning rate $\alpha = 0.001$, exponential decay rates $\beta_1 = 0.9$ and $\beta_1 = 0.999$, and the precision threshold $\hat{\epsilon} = 1e^{-7}$. The models were trained on the NVIDIA GPU GeForce RTX 2080 Ti. Training was performed over 40 epochs with a batch size of 64. The batch size was optimized based on memory constraints and to maintain a stable loss reduction. After 40 epochs, the score of validation dataset stabilized around $3.5e^{-3}$, $2e^{-3}$, $5e^{-4}$, $3e^{-4}$, and $3e^{-4}$ at wavelengths of $200nm$, $400nm$, $600nm$, $800nm$ and $1000nm$, respectively, without showing signs of overfitting. The score on the test dataset are resumed in the tabular 2.

Wavelength (nm)	200	400	600	800	1000
Accuracy (%)	83.60	87.51	94.03	95.28	95.34
MSE	3.5e-03	1.9e-03	5.11e-04	3.22e-04	2.94e-04

Table 2: Performance of ZerNet with aberrated monochromatic PSFs across varying wavelengths.

3. RESULTS

3.1 Coefficients estimation

In this section, we delve into the performance evaluation of ZerNet to estimate Zernike coefficients from PSF images. The figure 6 provides an overview of a unitary test case, illustrating the aberrated PSF alongside the accuracy of ZerNet's estimations.

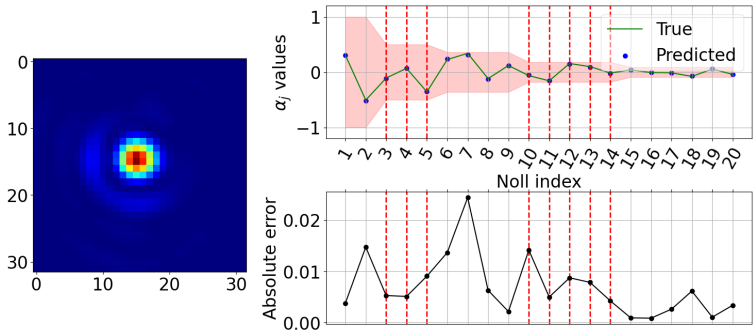


Figure 6: The left plot presents an aberrated PSF. The right plots illustrate the accuracy of ZerNet estimations, with the red zone indicating the WFE budget. Vertical dashed red lines denote even radial orders.

Even though ZerNet accuracy seems satisfactory within figure 6, this result does not rely on robust statistics. For that reason, the following figure 7 has been implemented to represent the final outcome of an analysis of 1000 nm PSFs (wavelength leading to the best convergence of ZerNet in table 2). For simplification purposes, the identical plots for other wavelengths will be displayed in the figure 10 in Appendix B.

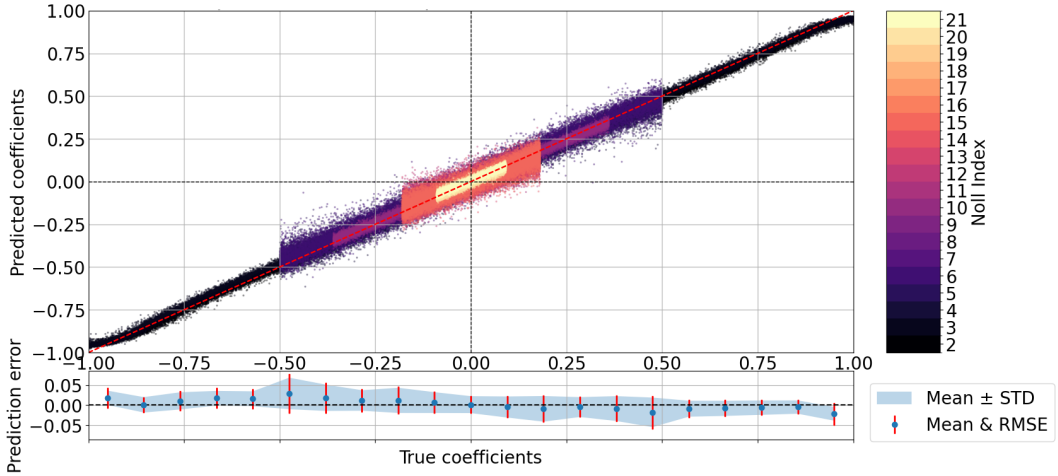


Figure 7: Comparison of ZerNet predictions versus true Zernike coefficients where the color indicates the Zernike Polynomial order. The ideal y = x relationship is represented by the dashed red line. The prediction error is displayed in the lower subplot, where the RMSE and standard deviation are represented by red bars and the mean error is indicated by blue dots.

The spreadiness of the scatter plot around the dashed red line demonstrates how the model might miscalculate the coefficients values. Quite as anticipated, the lowest orders coefficients are precisely predicted due to their simple but major impact on PSFs. This representation contains features such as an upper plot comparing the Zernike coefficient of the model with the reconstructed ZerNet predictions, colors corresponding to different orders. The lower plot shows the error of predictions over a range of real coefficients values.

Yet, we were expecting the estimation of coefficients to be lowered for high-order aberrations (due to small contributions with respect to the WFE budget), and surprisingly the highest order considered is better estimated than orders colored in orange/red. It turns out the orders causing this scattering around the y = x relationship are even radial orders, which means their properties are somehow different from the other polynomials'.

Two significant facts while dealing with them are their radial symmetry and their reflection symmetry, meaning that the polynomial is symmetric about the origin and either symmetric or anti-symmetric with respect to reflections across both the x-axis and the y-axis, respectively. The reason why we chose a radially asymmetric optical system, is to prevent these symmetries in the WFE to prevent ZerNet from diverging, as they would be ambiguous. In the meantime, the figure 7 forecasts a need on using more sophisticated optical models (with more struts for example), to counterbalance the symmetries of the wavefront.

3.2 Reconstruction error

The PSF reconstruction error is crucial to the propagation of uncertainty in scientific applications, and it must be minimized to limit systematics. In this instance, the error ϵ resulting from coefficients' estimation until order j_{max} was measured using the Frobenius norm [11]:

$$\epsilon_{j_{max}} = \left\| \left(PSF_{True} - PSF(\alpha_{:j_{max}}) \right) \right\|_{\mathrm{F}} \tag{9}$$

This norm was selected because it can quantify the overall difference between matrices, which makes it powerful for evaluating the difference between the PSF reconstructed using predicted coefficients and the true PSF. The Frobenius norm is frequently used in image comparison tasks to measure how similar or different two matrices representing images are from one another. We can easily compare and analyze the results because the Frobenius norm computation yields a single scalar value that captures the total reconstruction error.

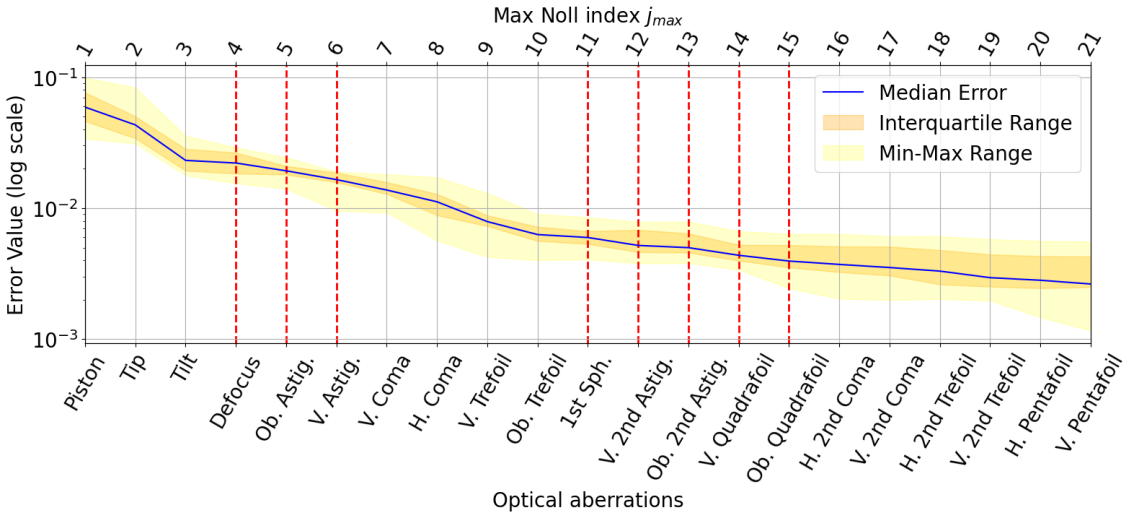


Figure 8: PSF reconstruction log-scale error values across a range of optical aberrations, arranged according to Zernike polynomial order (shown at the top). The interquartile range is represented by the orange area, the min-max by the yellow area, and the median error by the blue line on the plot. Red dashed lines denote even radial orders.

The figure 8 showcases, through computation of $\epsilon_{j_{\max}}$, the ability of ZerNet to precisely assess optical system aberrations. Even though the min-max range seems broad and fluctuating, the median error's constant decline indicates that using predictions from multiple orders is beneficial. This finding emphasizes how useful it is to use higher-order Zernike coefficients in PSF reconstruction since, even at the highest orders, there is a notable decrease in reconstruction error with no diminishing returns.

4. DISCUSSION & CONCLUSION

This novel approach demonstrated that our machine learning algorithms could establish a relationship between PSF images and the corresponding coefficients without requiring any knowledge of the WFE. Yet, it is crucial to remember the many assumptions and constraints that were selected to arrive at these findings, including the Fraunhofer region approximation, the specific WFE budget, the radial asymmetry of the optical system and its resolving power.

Among the peculiarities of radially symmetrical optical systems under consideration is the failure of the models to converge on even radial orders. The benefit of employing radially asymmetric optical systems is emphasized by this problem. The conditions leading to such results shed the light on future possible improvements :

- Adding Noise to Match Real-case Scenarios → Diffusion Denoising Probabilistic (or Implicit) Model:

  Real-world situations frequently include noise, which can have an impact on imaging results. Our machine learning algorithms can become more resilient and useful by adding noise to our simulations. Effective methods for handling noisy data include diffusion denoising probabilistic models, which model the data distribution while denoising, and autoencoders, which can learn effective representations of input data. By using these techniques, our model will be more resilient to noise in real-world situations.

- Using Polychromatic PSFs:

  In reality, optical systems frequently use polychromatic light sources, which causes the Point Spread Function (PSF) to vary with respect to the source observed. We can account for these spectral variations and gain a more thorough understanding of the imaging process by taking polychromatic PSFs into consideration. In order to capture the entire range of optical aberrations present in the system, this approach entails simulating PSFs across various wavelengths or spectral bands.

- Element-wise Analysis → WFE Description per Optical Element:

  Several optical components make up optical systems, and each one adds to the total wavefront error (WFE). The WFE can be analyzed element by element to provide a more thorough description of aberrations and their causes. We are able to identify individual sources of aberrations and adjust corrective measures by providing a WFE description for each optical element. This method makes it possible to analyze the system's performance more thoroughly and makes focused optimization efforts easier.

- Physics-related Loss → Physics Integrated Neural Networks:

  There could be differences between model predictions and actual physical reality if physics-based constraints were not explicitly included in current machine learning models. Neural networks that incorporate physics-related loss functions can impose physical constraints during training, guaranteeing that the learned representations follow basic optics principles. Models produced by physics-integrated neural networks are more in line with underlying physical phenomena because they use domain knowledge to direct the learning process. We find that adding physics-based constraints to our machine learning algorithms improves their interpretability and accuracy when applied to optical aberration modeling.

These potential solutions address various challenges encountered in our study and pave the way for future advancements in optical aberration modeling and correction.

In the future, ZerNet holds promise for applications across diverse fields such as astronomy, microscopy, or medical imaging, where understanding wavefront errors may present significant challenges. By leveraging machine learning algorithms like ZerNet, researchers and practitioners can gain deeper insights into optical aberrations and improve imaging quality in a wide range of scientific and medical applications.

ACKNOWLEDGMENTS

This proceeding is part of a project funded by the Agence Nationale de la Recherche (ANR) under the ANR-22-CE46-0009 grant, hosted by the Centre de Physique des Particules de Marseille. We take this opportunity to express our gratitude to the French space agency (CNES) for co-funding Lucas Sauniere's PhD.

This research made use of *poppy*, an open-source optical propagation Python package originally developed for the James Webb Space Telescope project [12].

References

[1] Makarkin, M. and Bratashov, D., "State-of-the-Art Approaches for Image Deconvolution Problems, including Modern Deep Learning Architectures," *Micromachines* **12**, 1558 (Dec. 2021).

[2] Liaudat, T., Starck, J.-L., Kilbinger, M., and Frugier, P.-A., "Rethinking data-driven point spread function modeling with a differentiable optical model," *Inverse Problems* **39**, 035008 (Mar. 2023).

[3] Pakdel, A., Mainprize, J. G., Robert, N., Fialkov, J., and Whyne, C. M., "Model-based PSF and MTF estimation and validation from skeletal clinical CT images: PSF and MTF estimation in skeletal CT," *Medical Physics* **41**, 011906 (Dec. 2013).

[4] Vishniakou, I. and Seelig, J. D., "Differentiable model-based adaptive optics for two-photon microscopy," *Optics Express* **29**, 21418 (July 2021).

[5] Niu, K. and Tian, C., "Zernike polynomials and their applications," *Journal of Optics* **24**, 123001 (Dec. 2022).

[6] Perrin, M., Long, J., Douglas, E., Sivaramakrishnan, A., Slocum, C., and others, "POPPY: Physical Optics Propagation in PYthon," *Astrophysics Source Code Library* , ascl:1602.018 (Feb. 2016). ADS Bibcode: 2016ascl.soft02018P.

[7] Perrin, M., Long, J., Douglas, E., Sivaramakrishnan, A., Slocum, C., and others, "POPPY: Physical Optics Propagation in PYthon." Astrophysics Source Code Library, record ascl:1602.018 (Feb. 2016).

[8] Störkle, J., [*Dynamic simulation and control of optical systems*], no. Band 58 (2018) in Schriften aus dem Institut für Technische und Numerische Mechanik der Universität Stuttgart, Shaker Verlag, Aachen (2018).

[9] Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions," in [*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 1–9, IEEE, Boston, MA, USA (June 2015).

[10] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *CoRR* **abs/1412.6980** (2014).

[11] Herzog, R., Köhne, F., Kreis, L., and Schiela, A., "Frobenius-Type Norms and Inner Products of Matrices and Linear Maps with Applications to Neural Network Training," (Nov. 2023). arXiv:2311.15419 [cs, math].

[12] Perrin, M., "poppy documentation ; release 0.9.0," (2019).

APPENDIX A. ZERNET MODEL ARCHITECTURE

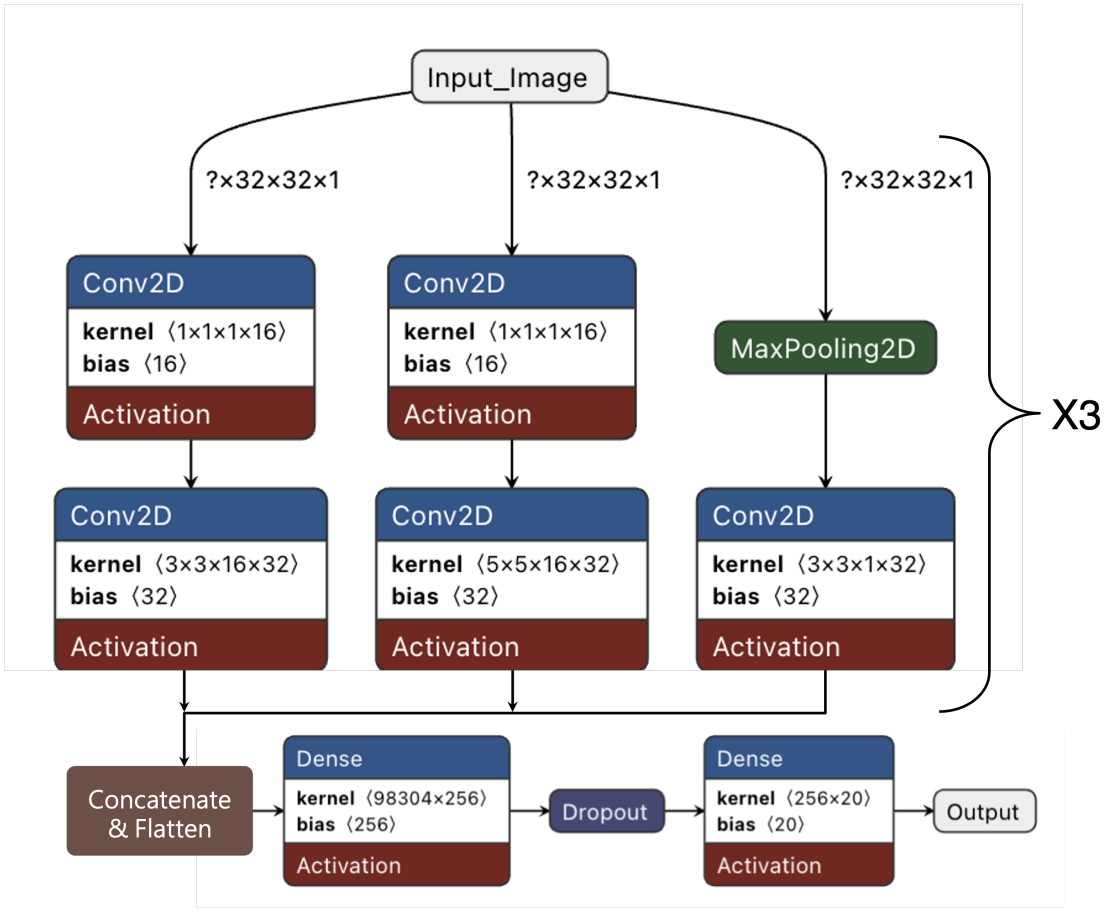
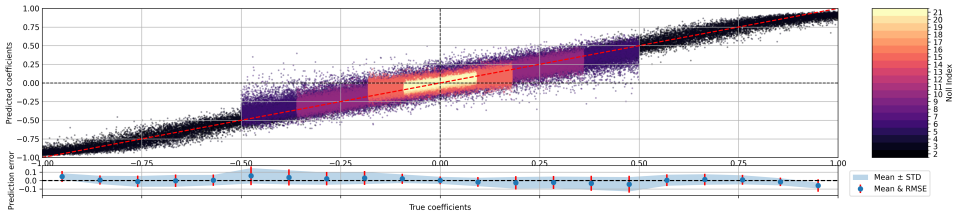
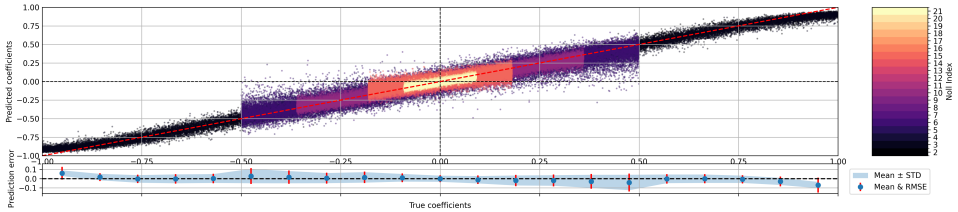


Figure 9: Architecture of the ZerNet model, based on the enhancement of an inception model.

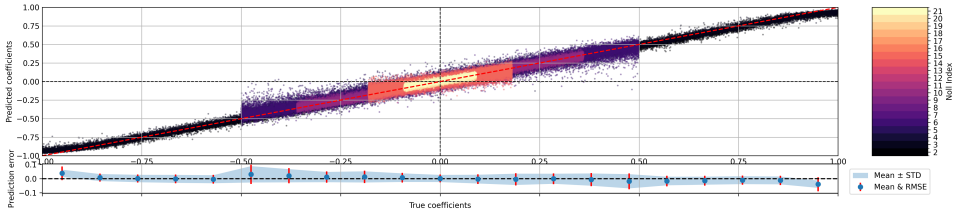
APPENDIX B. ZERNET COEFFICIENTS PREDICTIONS



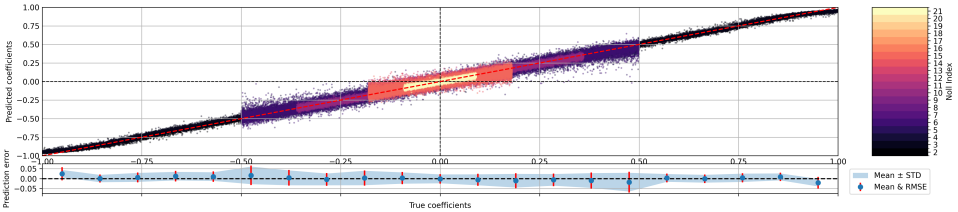
(a) $\lambda = 200$nm



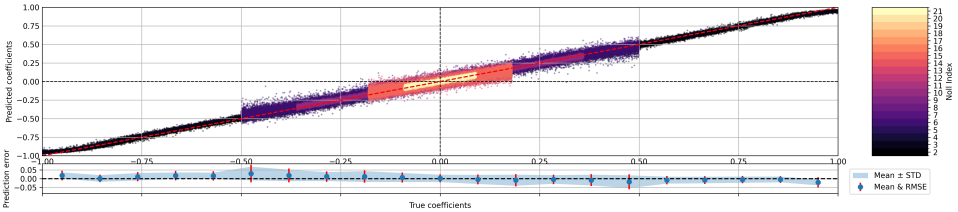
(b) $\lambda = 400$nm



(c) $\lambda = 600$nm



(d) $\lambda = 800$nm



(e) $\lambda = 1000$nm

Figure 10: ZerNet coefficients predictions for three wavelengths evenly spaced.