Machine learning for image-based wavefront sensing

Pierre-Olivier Vanberg University of Liège
Gilles Orban de Xivry University of Liège
Olivier Absil University of Liège
Gilles Louppe University of Liège

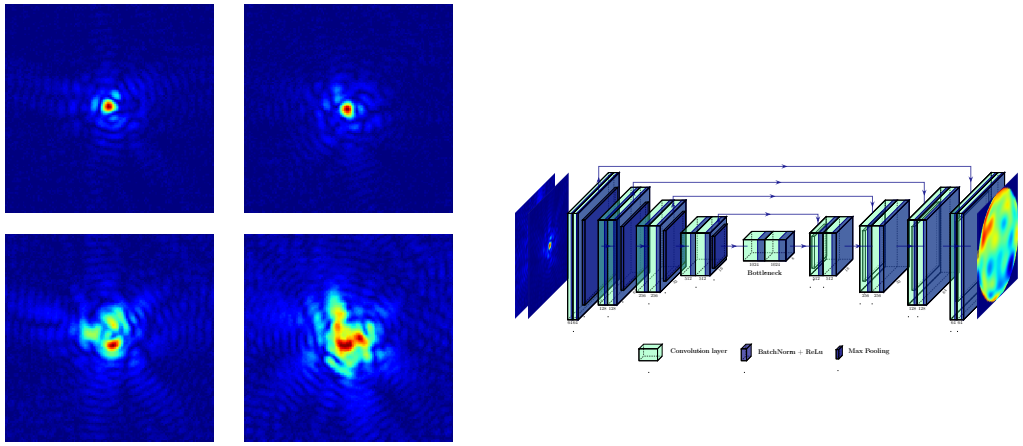
Abstract

High-contrast imaging systems in ground-based astronomy rely on a precise control of the wavefront. On one hand, atmospheric turbulence distorts the wavefront which is corrected by a dedicated adaptive optics system. On the other hand, non-common path aberrations between wavefront sensor and scientific paths can also overwhelm a putative scientific signal and need to be corrected. Hence, measuring precisely the wavefront at the scientific focal plane is of prime interest for applications such as direct imaging of exoplanets. While early attempts using neural networks showed some successful results, the latest advances in machine learning have yet to be fully exploited for the problem of focal plane wavefront sensing. In this paper, we explore the use of convolution neural networks to perform image-based wavefront sensing. Based on simulated data, we evaluate neural architectures on two different data sets, one with only low order aberrations (20 Zernike modes) and one including higher orders modes (100 Zernike modes). We discuss the accuracy reached in both cases, and we show that direct phase map reconstruction outperforms classical modal approaches. The precision achieved ranges typically between 1% and 10% of the injected wavefront. Finally, we explore the impact of phase diversity, and we compare our optimized CNN model to a standard iterative phase retrieval algorithm.

1 Wavefront sensing

Exoplanet direct imaging with ground-based telescopes requires extreme adaptive optics (AO) system to correct the corrugated wavefront caused by atmospheric turbulence. Those extreme AO systems provide near-perfect diffraction limited point spread functions, which can be effectively suppressed by a coronagraph. Ultimately, the achievable contrast, or likewise the exoplanet detectability, may be limited by non-common path aberrations (NCPA) between the AO wavefront sensor and the scientific imager. In fact, NCPA still represent a real challenge for exoplanet direct imaging using extreme AO observations [16, 13, 7]. Typical wavefront sensors encode phase information into intensity and allow the direct measurement of phase gradient. However, NPCA are best measured at the scientific focal plane, which typically means determining the phase in the pupil plane from its Fourier-domain magnitude, i.e., from the intensity measurement provided by the detector. This problem of phase retrieval is a very active field of research in several areas of physics, e.g., in microscopy, X-ray crystallography, and astronomical imaging [20, 15, 18]. Phase retrieval generally consists of an iterative algorithm calibrated using a parameterized physical model [3]. The iterative nature of these algorithms often requires time and intensive computational resources. Furthermore, the convergence is not guaranteed, as the optimization procedure may encounter different modes of stagnation such as twin images or stripes [4, 17]. Gradient-based optimization algorithms may also stagnate in local minimum when starting from an initial guess significantly different from the true solution [12]. In parallel, machine learning algorithms have been developed and applied to phase retrieval. Often based on neural networks, these methods have the advantage of providing an almost immediate estimate of the optical aberrations. Historically, fully connected neural networks were first investigated [9, 2, 1]. Despite promising results, the lack of generalization power and the poor scaling of the networks limited the achievable performance.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.



(a) (b)

Figure 1: (a) Examples of point spread functions used for training. Top: In-focus, bottom: Out-of-focus. The two images are concatenated along the Z-axis and form an input feature map of size $128 \times 128 \times 2$ pixels. (b) Unet architecture. The output is the phase map estimate.

In this paper, we apply modern convolution neural networks (ResNet, Inception V3, Unet and Unet++ [8, 19, 14, 22]) to the estimation and correction of non-common path aberrations. Two approaches are considered. First, neural networks are trained to estimate the Zernike coefficients from one or multiple distorted point spread functions. Secondly, a direct phase map reconstruction is proposed based on deconvolution layers. Both approaches are compared on low and high number of modes, in the presence or the absence of phase diversity [6].

2 Methods

Data The data consist of a set of numerically simulated and aberrated point spread function (PSF) pairs: in-focus and out-of-focus. Phase diversity helps resolve the possible phase ambiguity while being easy to set up in practice. The image size is fixed to 128×128 pixels. Two distinct data sets of $100,000$ images are constituted, with sample illustrations provided in Figure 1. The first one is based on the first 20 Zernike modes, while the second extends up to 100 Zernike modes (Noll convention [11]). The observation wavelength λ has been chosen in the near-infrared at $2.2\mu m$ for a telescope diameter of 10m. The pixel scale of the science camera is fixed to 0.01 arcseconds. Finally, the defocus has been fixed to $\lambda/4$. Those physical parameters are representative of existing instruments such as, e.g., NIRC2 at the Keck Observatory.

In order to achieve distorted PSFs representative of non-common path aberrations, the phase map must reflect typical errors of conventional optics. This is achieved by uniformly generating the Zernike coefficients between $-\lambda$ and λ and dividing each coefficient by its radial order. This procedure approximates a power spectral density profile, $S \approx 1/f^2$ (where f is the spatial frequency). Such profile is for instance frequently produced by polishing errors [10]. The resulting phase maps have on average 1 rad RMS WFE (or equivalently $350nm$ considering the $2.2\mu m$ wavelength). The phase map $\theta(x, y)$ is then obtained as a linear combination of the Zernike polynomials modulated by the previously generated coefficients. Next, the phase map characterizing the optical aberrations is numerically propagated through the system. In this paper, we consider bright point source objects for which the noise is dominated by shot noise. Hence the detection is modelled by a Poisson process, which in practice consists of a Gaussian distribution with mean N (photon counts) and standard deviation $\sigma = \sqrt{N}$. Assuming a long enough exposure time, the signal-to-noise ratio was adjusted to an average value of 100 over the main PSF peak.

Network architectures Two estimation approaches are considered: i) the estimation of a set of modal coefficients (20 or 100 values) and ii) the direct reconstruction of the phase map (128×128 images). For the first approach, we consider Inception V3 [19] and ResNet-50 [8] neural networks.

Architecture	20 Zernike	100 Zernike	Inference time
Inception V3	0.0240 ± 0.0051	0.1094 ± 0.0154	$0.1182s$
ResNet 50	0.0187 ± 0.0039	0.1145 ± 0.0138	$0.1090s$
Unet	0.0132 ± 0.0019	0.0976 ± 0.0109	$0.1102s$
Unet++	0.0130 ± 0.0023	0.0943 ± 0.0133	$0.1358s$

Table 1: Summary of the final performance expressed as the RMSE between the exact and the estimated phase map in radians. The original phase maps have on average 1 rad RMS WFE. Inference times are determined using a single input (batch size=1) on CPU, Intel Xeon e3-1230v5.

The second approach requires the networks to directly output a phase map, for which we consider the Unet and Unet++ architectures initially developed for image segmentation. As illustrated in Figure 1b, the encoding part is made of successive convolution layers followed by max pooling layers. The input PSF images are thus progressively downsampled while the most relevant features are extracted. The output phase map is then reconstructed by successive deconvolution operations in a symmetric way to the downsampling part; the input and the output having the same spatial dimensions.

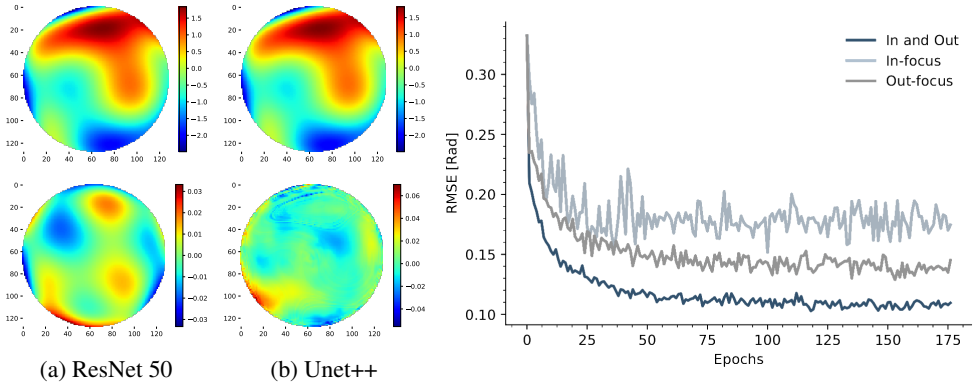
Training is carried out using stochastic gradient descent with momentum. The learning rate is determined by grid search and tuned to each architecture. Inception V3 and ResNet 50 are initialized using pretrained weights while Unet and Unet++ are instantiated using Xavier initialization [5]. For the first approach, the loss function is the root mean squared error between the exact c_i and the estimated c_i' Zernike coefficients. For the second approach, the networks are trained with respect to pixel-wise root mean squared error between the exact θ and the estimated θ' phase map.

3 Results

In order to compare both approaches, i.e., the modal coefficient estimation and the direct estimation, we either reconstruct the phase map using the Zernike basis and compare the phase maps or alternatively we project the phase map onto the basis and compare the coefficients. Table 1 summarizes our results and shows that direct map reconstruction outperforms modal coefficient estimation. A thorough evaluation of our best architecture for each approach (ResNet vs Unet++) shows an overall improvement of 36% (or 2 nm rms) on the first data set generated with 20 Zernike modes and an improvement of 19% (or 7 nm rms) on the second data set generated using 100 Zernike modes. The second advantage of the direct phase estimation is the independence with respect to a particular representation, in our case the Zernike polynomial basis.

20 Zernike modes Our first set of experiments focuses on low order mode aberrations, with wavefront phase maps generated from the first 20 Zernike modes. Each neural network is trained independently using the same training examples. The dataset is respectively split as $90,000$ training images, $5,000$ test images and $5,000$ validation images. Figure 2 highlights the various network phase estimation given an in-focus and an out-of-focus PSF. It is first demonstrated that each of the architectures is able to correctly retrieve the overall wavefront shape. The high quality estimation of the phase map, with an accuracy ~ 0.01 rad (residual RMSE), makes the differences between the modal and direct output approaches hardly noticeable. Nevertheless, the residual phase maps clearly show the fundamental distinction. Based on the Zernike polynomials reconstruction, the phase maps are perfectly smooth and exhibit very small pixel-to-pixel variation. On the opposite, the phase maps directly inferred by the CNN tend to demonstrate a larger level of discontinuity as shown in Figure 2 (b). It should be noted however that this level of discontinuity is very small compared to the initial pixel values and does not affect the estimation quality (see Table 1).

100 Zernike modes This second experiment demonstrates, for the first time to our knowledge, that CNNs can be trained to predict a large number of aberration modes. Following the same procedure as the previous experiment, the various architectures are trained based on $90,000$ training samples. Numerical values over the test set are provided in Table 1. The phase estimation reaches an accuracy of ~ 0.1 rad RMS WFE. On average, the higher order modes are more likely to be misestimated. While the absolute error over the first 20 modes and the last 20 modes is respectively given by 0.021 and 0.017 rad, the relative error is much more indicative with a mean value of 13% for the first modes



(a) ResNet 50 (b) Unet++

Figure 2: (Left) On top, examples of phase map estimates (20 Zernike modes), true RMS WFE of 1.093 rads. On the second line, residual wavefront after correction. The color scales are expressed in radians. (Right) Validation learning curve of the Unet architecture trained with only in-focus PSFs, only out-of-focus PSFs and both simultaneously.

and 44% for the last modes. This effect is mainly related to the NCPA profile, which favors the low order modes, i.e., larger magnitudes. This typically results in phase maps whose edges are poorly estimated.

Phase diversity Figure 2 shows that the simultaneous use of in-focus and out-of-focus images significantly improves the final accuracy. We also notice that when working with a single intensity measurement, it is preferable to use out-of-focus images rather than in-focus images. Defocused images indeed increase the number of informative pixels as shown in Figure 1. This conclusion is consistent with previous studies [21]. Finally, it is worth noting that CNNs do not seem to suffer from the twin-images problem, even from single intensity measurements and unlike iterative algorithms.

Comparison against Hybrid Input-Output To demonstrate the efficiency and the limitations of CNNs, a comparison with the standard Hybrid Input-Output [3] iterative algorithm is carried out. The first and main advantage of CNNs is the almost instantaneous prediction unlike iterative algorithms that require several iterations before reaching the same level of correction. In counter part, CNNs require a lot of upstream preparation and training time. Secondly, CNNs do not suffer from convergence issues or stagnation modes. However, our experimental results (not shown here) also suggest that CNN predictions are generally outperformed by the iterative algorithm after typically 40 to 60 iterations.

4 Conclusions

In this work, we demonstrated that deep CNNs can perform accurate wavefront sensing based on focal plane images. We compared two approaches, either the estimation of modal Zernike coefficients or the direct phase recovery. The later approach with its associated architectures (Unet and Unet++) provided the best accuracy. It is also not limited to circular pupils. We then studied the influence of phase diversity and showed that, while inevitably less precise, CNNs are able to perform a valid wavefront estimation on single intensity measurements. Finally, we demonstrated the ability of machine learning to compete with a classical iterative algorithm. A combination of both machine learning and iterative algorithm could provide higher inference speed, better precision and avoid any stagnation in local minima. While this work shows that CNNs are a promising avenue for the problem of non-common path aberrations measurement, the robustness and limitations of CNNs need to be further characterized both through simulations and experimental application.

References

[1] T. K. Barrett and D. G. Sandler. Artificial neural network for the determination of hubble space telescope aberration from stellar images. *Appl. Opt.*, 32(10):1720–1727, Apr 1993.

[2] D. G. S. et al. Use of a neural network to control an adaptive optics system for an astronomical telescope. *Nature*, 351:300–302, 1991.

[3] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758–2769, Aug 1982.

[4] J. R. Fienup and C. C. Wackerman. Phase-retrieval stagnation problems and solutions. *J. Opt. Soc. Am. A*, 3(11):1897–1907, Nov 1986.

[5] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh and M. Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.

[6] R. A. Gonsalves. Phase retrieval and diversity in adaptive optics. *Optical Engineering*, 21(5): 829 – 832 – 4, 1982.

[7] O. Guyon. Extreme Adaptive Optics. *ARAA*, 56:315–355, Sep 2018. doi: 10.1146/annurev-astro-081817-052000.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[9] M. L.-H. J. R. P. Angel, P. Wizinowich and D. Sandler. Adaptive optics for array telescopes using neural-network techniques. *Nature*, 348:221–224, 1990.

[10] M. Lamb, D. R. Andersen, J.-P. Véran, C. Correia, G. Herriot, M. Rosensteiner, and J. Fiege. Non-common path aberration corrections for current and future ao systems. volume 9148, page 914857, 07 2014.

[11] R. J. Noll. Zernike polynomials and atmospheric turbulence. *J. Opt. Soc. Am.*, 66(3):207–211, Mar 1976.

[12] S. W. Paine and J. R. Fienup. Machine learning for improved image-based wavefront sensing. *Opt. Lett.*, 43(6):1235–1238, Mar 2018.

[13] D. Ren, B. Dong, Y. Zhu, and D. J. Christian. Correction of non–common-path error for extreme adaptive optics. *Publications of the Astronomical Society of the Pacific*, 124(913):247–253, 2012. ISSN 00046280, 15383873.

[14] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[15] G. Rousset et al. Naos, the first ao system of the vlt: on-sky performance, 2003.

[16] J.-F. Sauvage, T. Fusco, G. Rousset, and C. Petit. Calibration and precompensation of noncommon path aberrations for extreme adaptive optics. *J. Opt. Soc. Am. A*, 24(8):2334–2346, Aug 2007.

[17] J. H. Seldin and J. R. Fienup. Numerical investigation of the uniqueness of phase retrieval. *J. Opt. Soc. Am. A*, 7(3):412–427, Mar 1990.

[18] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, May 2015.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.

[20] X. Tao, O. Azucena, M. Fu, Y. Zuo, D. C. Chen, and J. Kubby. Adaptive optics microscopy with direct wavefront sensing using fluorescent protein guide stars. *Opt. Lett.*, 36(17):3389–3391, Sep 2011. doi: 10.1364/OL.36.003389.

[21] R. H. K. K.-M. S. J. T. Yohei Nishizaki, Matias Valdivia and E. Vera. Deep learning wavefront sensing. *Opt. Express*, 27(1):240–251, Jan 2019.

[22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018.