Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **505,** 5702–5713 (2021)
Advance Access publication 2021 June 9

https://doi.org/10.1093/mnras/stab1634

Focal plane wavefront sensing using machine learning: performance of convolutional neural networks compared to fundamental limits

G. Orban de Xivry ,[1★] M. Quesnel,[1,2] P.-O. Vanberg,[1,2] O. Absil[1] and G. Louppe[2]

[1] *Space sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, Allée du Six Août 19c, B-4000 Sart Tilman, Belgium*
[2] *Montefiore Institute, Université de Liège, Allée de la découverte 10, 4000 Liège, Belgium*

Accepted 2021 June 2. Received 2021 June 1; in original form 2021 January 24

ABSTRACT

Focal plane wavefront sensing (FPWFS) is appealing for several reasons. Notably, it offers high sensitivity and does not suffer from non-common path aberrations (NCPAs). The price to pay is a high computational burden and the need for diversity to lift any phase ambiguity. If those limitations can be overcome, FPWFS is a great solution for NCPA measurement, a key limitation for high-contrast imaging, and could be used as adaptive optics wavefront sensor. Here, we propose to use deep convolutional neural networks (CNNs) to measure NCPAs based on focal plane images. Two CNN architectures are considered: ResNet-50 and U-Net that are used, respectively, to estimate Zernike coefficients or directly the phase. The models are trained on labelled data sets and evaluated at various flux levels and for two spatial frequency contents (20 and 100 Zernike modes). In these idealized simulations, we demonstrate that the CNN-based models reach the photon noise limit in a large range of conditions. We show, for example, that the root mean squared wavefront error can be reduced to $<\lambda/1500$ for 2×10^6 photons in one iteration when estimating 20 Zernike modes. We also show that CNN-based models are sufficiently robust to varying signal-to-noise ratio, under the presence of higher order aberrations, and under different amplitudes of aberrations. Additionally, they display similar to superior performance compared to iterative phase retrieval algorithms. CNNs therefore represent a compelling way to implement FPWFS, which can leverage the high sensitivity of FPWFS over a broad range of conditions.

Key words: instrumentation: high angular resolution, adaptive optics – methods: numerical.

1 INTRODUCTION

High-contrast imaging instruments are now routinely used in ground-based astronomy to explore circumstellar environments and to detect exoplanets. To achieve such a feat, they must reach high contrast at small angular separation and thus rely on a precise control of the wavefront. Extreme adaptive optics (AO) systems correct the corrugated wavefront caused by atmospheric turbulence and provide near-perfect diffraction limited point spread functions (PSFs), which can then be effectively suppressed by a coronagraph. However, the contrast, or likewise the exoplanet detectability, may still be limited by non-common path aberrations (NCPAs) between the wavefront sensor arm and the scientific path. These NCPAs are quasi-static with minute to hour time-scales due to slowly evolving instrumental aberrations and beam wander related to temperature, humidity, and mechanical changes. Because of their nature, they appear essentially at low spatial frequencies. These properties make them challenging to remove in post-processing and detrimental to the final contrast. In this respect, focal plane wavefront sensing (FPWFS) with the scientific detector is an appealing solution. In addition to getting rid of NCPAs and chromatic errors, FPWFS offers high sensitivity that is only surpassed by the Zernike wavefront sensor (ZWFS; Guyon 2005). It is also simple opto-mechanically and necessitates few to no modifications of the optics. This low complexity means less risk of failure and less maintenance. An overview of existing FPWFS techniques for high-contrast imaging instruments can be found in Jovanovic et al. (2018) but the interest for FPWFS goes well beyond NPCA measurement. Some aberrations are not well measured by pupil WFS such as phase discontinuities caused by the presence of spiders (so-called petalling and low wind effect; e.g. Vievard et al. 2019). Other applications range from co-phasing segmented mirrors (e.g. Delavaquerie, Cassaing & Amans 2010) to real-time AO systems (e.g. Keller et al. 2012; Korkiakoski et al. 2012). However, the price to pay of FPWFS is typically a high computational burden, as the problem is non-linear, and it generally requires a source of diversity to effectively lift any phase ambiguity.

In parallel, machine learning algorithms have been developed and applied to phase retrieval and wavefront sensing, in many different fields including astronomy. Neural networks were first used for real-time atmospheric compensation and co-phasing (Angel et al. 1990; Sandler et al. 1991), and retrieval of static aberration in the *Hubble Space Telescope* (Barrett & Sandler 1993). These techniques have then been used more broadly in the field of AO, to reduce Shack–Hartmann WFS slope errors (Montera et al. 1996), to perform open-loop AO tomographic reconstruction (Osborn et al. 2014), or to predict wavefront and reduce temporal errors (e.g. Jorgenson & Aitken 1992; McGuire et al. 1999; Liu et al. 2020). The non-linear nature of neural networks makes them good candidates to solve the non-linear phase retrieval problem. Despite these early results, the lack of generalization power and the poor scaling of the networks ultimately limited the achievable performance. Later on,

[★] E-mail: gorban@uliege.be

Published by Oxford University Press on behalf of Royal Astronomical Society

convolutional neural networks (CNNs) were introduced (LeCun et al. 1990; Krizhevsky, Sutskever & Hinton 2017). Specifically designed for images, CNNs use successive convolution operations, learning from group of pixels and assembling progressively more complex patterns. More recent works have applied such CNNs to non-linear wavefront reconstruction (Swanson et al. 2018; Landman & Haffert 2020), wavefront prediction (Swanson et al. 2018, 2021), to extend the usable range of Lyot-based low-order wavefront sensors (Allan et al. 2020a) and of Zernike phase-contrast wavefront sensors (Allan et al. 2020b), and to FPWFS (e.g. Paine & Fienup 2018; Andersen, Owner-Petersen & Enmark 2020).

Deep learning techniques are in fact burgeoning in all optical applications using phase retrieval, ranging from biomedical microscopy (e.g. Cumming & Gu 2020; Krishnan et al. 2020) to holography (e.g. Peng et al. 2020) and astronomy. The specific application to image-based wavefront sensing has been investigated in several recent works that we attempt to summarize in the following. Naik et al. (2020) used a compact CNN for object-agnostic wavefront sensing, inferring up to six Zernike coefficients, but reported a poorly sensed coma. Wu et al. (2020) trained a CNN for fast inference of 13 Zernike coefficients and obtained mild improvements for input aberrations of around 2 rad root mean square (rms). Nishizaki et al. (2019) proposed to extend the design space of wavefront sensor using deep learning where the inputs are preconditioned images such as overexposed, defocused, or scattered images. Guo et al. (2019) used a direct phase-map-output CNN model for the inference of up to 64th Zernike mode with input wavefront error (WFE) from about 1.5 up to 4.5 rad rms. They obtained residual errors in the range of approximately 0.45–0.82 rad and validated their approach experimentally. Paine & Fienup (2018) used the Inception v3 architecture to expand the capture range of gradient-based optimization methods. They applied their approach to the *JWST* aperture and consider up to 18 Zernike coefficients with input WFE in the range of 1.57–25.1 rad rms. The residual WFE after the CNN is on average 2.3 rad rms. The trained CNN provides here good initial estimates to a second stage gradient-based method. Andersen, Owner-Petersen & Enmark (2019) and Andersen et al. (2020) studied the potential of real-time image sharpening using ResNet and Inception v3 models to estimate Zernike coefficients from pairs of in-focus and out-of-focus images that are blurred by the atmospheric turbulence. They included the effect of noise, guide star magnitude, polychromaticity, and bit depth. They explored an aberration regime of about 8–13 rad rms ($D/r_0 = 12$ and 21, respectively, where D is the telescope diameter and r_0 is the Fried parameter) and obtained a residual error of about 1.4 and 2 rad rms, respectively, by correcting for 36 Zernike modes. When increasing the number to 66 modes, they obtained only a marginal improvement.

While those studies demonstrated the validity of the approach, i.e. using a CNN-based framework for FPWFS, it is unclear what really limits the performance reported, or if CNNs can leverage fully the sensitivity of FPWFS and if they can be applied effectively in a lower aberration regime relevant to NCPA measurements. We first investigated this lower aberration regime in a short report (Vanberg et al. 2019) demonstrating its working principle. We then explored in Quesnel et al. (2020) different numerical aspects and applied our framework to vortex coronagraphic imaging.

In this paper, our focus is to better understand the limitations of such a CNN-based framework for FPWFS in the context of NCPA measurements. More specifically, we study a regime of up to 1 rad rms input WFE and up to 100 Zernike modes. We also deliberately limit the number of simulated effects, such as e.g. noise sources or higher order disturbances, to systematically explore the achievable performance of CNNs for FPWFS and compare it to the fundamental limit for wavefront sensing. First, in Section 2, we describe our simulation set-up, i.e. the data simulator and the CNN architectures used for this work. In Section 3, we analyse our CNN models under idealized and degraded conditions and compare them with the expected photon noise limit. We also consider the implication of the sign ambiguity and the pixel sampling. Finally, in Section 4 we compare the CNN model to iterative phase retrieval and discuss numerical considerations. Overall, we demonstrate that CNN-based algorithms can efficiently solve the inverse problem posed by FPWFS. In particular, our framework is shown to be readily applicable for the measurement of NCPA, and we discuss throughout the paper different considerations for laboratory and on-sky applications, and for a broader usage such as, e.g. AO.

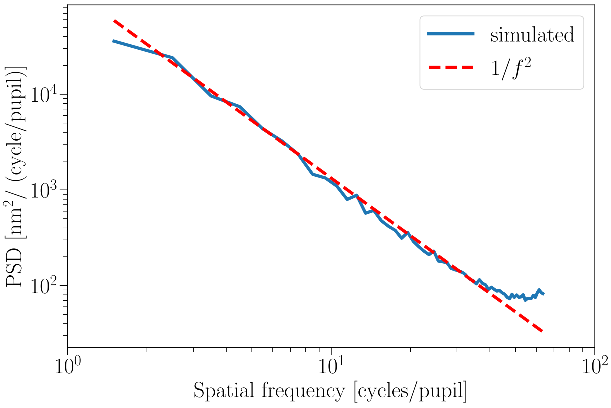

Figure 1. Spatial PSD of the generated phase maps reproducing high-quality optical surfaces.

2 METHODS AND SIMULATION SET-UP

2.1 Focal plane imaging and data set generation

One of the keys to the success of deep learning is the availability of a large and representative labelled data set. In this paper, the data consists of a set of numerically simulated, aberrated PSF pairs: in-focus and out-of-focus. The introduction of this phase diversity ensures the uniqueness (Foley & Butts 1981; Gonsalves 1982; Paxman, Schulz & Fienup 1992) of the solution while being easy to implement in practice, either by introducing defocus on a deformable mirror or by displacing the detector itself. The well-known sign ambiguity in the absence of diversity is discussed in Section 3.6.

Since our work is primarily motivated by the measurement and correction of NCPAs, we generate phase maps to reflect typical errors of high-quality optical surfaces (e.g. Dohlen et al. 2011), with a spatial power spectral density (PSD) profile $S \approx 1/f^2$, where f is the spatial frequency. The PSD is illustrated in Fig. 1. This is achieved by drawing random Zernike coefficients from a uniform distribution between $[-1, 1]$ and dividing each coefficient by its Zernike radial order. The coefficients are then scaled to obtain the desired median rms WFE, where the median is calculated over the data set. The rms WFE distribution over one of our data sets is illustrated in Fig. 2. This procedure leads to a uniform density distribution for each individual Zernike mode, where the minimum and maximum depend on the Zernike index and are a function of the desired median rms WFE and the number of Zernike coefficients.

Finally, the phase maps $\varphi(x, y)$ are obtained as a linear combination of the Zernike polynomials weighted by the previously generated coefficients. Each one of them is then propagated through the system

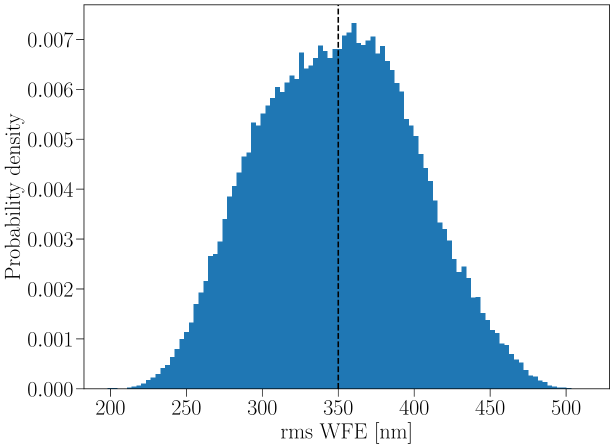

Figure 2. Distribution of the rms WFE of one data set with a median of 350 nm.

to produce the corresponding PSFs, *PSF*(x, y),

$$\mathrm{PSF}(x, y) \propto |\mathcal{F}[A(x, y)\exp(i\varphi(x, y))]|^2, \tag{1}$$

where $A(x, y)$ is the pupil function. The pupil function considered here is a simple uniformly illuminated circular pupil. The measurement, i.e. the PSF, is finally affected by noise. Here, we limit ourselves to photon noise, and disregard, for example, detector noises that are technological in nature. Hence, the signal-to-noise ratio of our image is $\mathrm{SNR} = \sqrt{N_{\mathrm{ph}}}$, where N_{ph} is the total number of photons in the image.

The image sizes are fixed to 128 × 128 pixels. The PSFs are sampled by 4.5 pixels over $1\lambda/D$ and the corresponding field of view is $\sim 28.5\lambda/D$. Such PSF sampling can be obtained, for instance, for a wavelength of 2.2 μm, a pixel scale of 0.01 arcsec per pixel, and a telescope diameter of 10 m. These parameters are representative of existing instruments such as, e.g. NIRC-2 at the Keck Observatory. Before being saved, the focal plane images are formatted in half-precision (float 16 bits). This step ensures that the theoretical sign ambiguity is perfectly reproduced numerically, i.e. that the PSFs generated from phase maps that only differ by the sign of their odd Zernike modes are numerically identical.

In this work, we consider median rms WFE of 70 and 350 nm at a wavelength of 2.2 μm, thus 0.2 and 1 rad rms, respectively. For convenience, we will often refer to these two levels as ‘low’ and ‘high’ aberration regime. We also consider two different numbers of Zernike modes, 20 and 100. We have thus four different scenarios for our following analyses. The resulting PSFs are illustrated in Fig. 3. The introduced phase diversity for the second PSF is a defocus term set to $\lambda/4$, i.e. 550 nm rms. The motivation to limit our training data to those four regimes is two-fold. First, the number of modes and the aberration level represent what is typically considered for NCPA correction on 8–40 m class telescopes. Second, increasing the number of modes and the aberration level increases the dimensionality of the problem. Thus, defining different data sets (with different dimensionalities), rather than a single one containing all the studied cases, allows to better understand the performance obtained, i.e. fundamental limit for wavefront sensing versus limitations of the CNN models (e.g. generalization power or suboptimal training). Nevertheless, in Section 3.4, we consider other appropriate data sets: one drawn from a uniform rms WFE distribution, and several with higher level of aberrations.

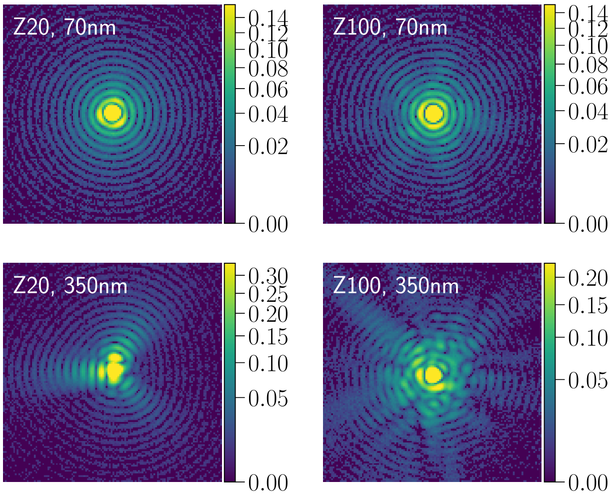

Figure 3. Illustration of the simulated PSFs with a square root stretch and 99 per cent interval. The signal-to-noise ratio equals 1000. Aberrations distributed over 20 (left) and 100 (right) Zernike modes. ‘Low’ (top) and ‘high’ (bottom) aberration levels.

2.2 Network architectures and training

We consider two approaches to our problem: one where the CNN is trained to estimate Zernike coefficients and one where the CNN is trained to do a direct phase map estimation. During the onset of this work, we considered a number of architectures with good ranking at ImageNet classification challenges: VGG-16, Inception v3, ResNet-50, U-Net, and U-Net++. Eventually, and in this paper, we only use ResNet-50 and U-Net, the other architectures either did not work well for our application or do not add further insights to the topics discussed here. It is worth noting, however, that Inception v3 has shown promising results in different simulation studies (Paine & Fienup 2018; Andersen et al. 2019).

Residual neural networks (He et al. 2016), or ResNet, are very deep networks where skip connections are introduced to improve gradient flow during the training steps. We use ResNet-50, which is 50 layers deep, and we initialize it with the parameters pre-trained on ImageNet. In order to adapt the architectures to the prediction of Zernike coefficients, the softmax activation and the last fully connected layers were replaced to match the output requirements.

For the direct phase estimation approach, we focused on an architecture initially developed for biomedical image segmentation: U-Net (Ronneberger, Fischer & Brox 2015). The overall network structure follows a U-shaped geometry. The encoding part is made of successive 3 × 3 convolution layers followed by 2 × 2 max pooling layers. The input PSF images are thus progressively downsampled while the most relevant features are extracted. The contracting part is followed by an expansion part replacing pooling operators by upsampling operators. Importantly, there are skip connections combining features from the contracting path with the upsampling part. Since we perform regression rather than segmentation, the last softmax layer was removed. In our implementation, the input PSF images and the output phase maps have the same grid sizes.

For the optimization, we used Adam (Kingma & Ba 2015) with an initial learning rate of 10^{-3} and a scheduler dividing the learning rate by two every 75 epochs. Our typical training procedure consists of a data set of 100 000 entries, each consisting of two focal plane images and one phase map. Before being fed into the CNN, photon noise is added to the images, a square root stretch is applied, and each image is normalized by its maximum. Our data set is typically split in a 90:10 ratio, i.e. 90 000 entries are used for training and 10 000 for validation. We use a batch size of 64 entries, all batches constitute one epoch, and we train for 200 epochs. The results and analyses, as presented in Section 3, use different data sets based on different random seeds for the Zernike coefficient generation. The loss function corresponds to the rms error, i.e.

$$\mathrm{loss}(\varphi, \hat{\varphi}) = \sqrt{\frac{1}{N}\sum_{i,j}^{N}\left(\varphi(x_i, y_j) - \hat{\varphi}(x_i, y_j)\right)^2}, \quad (2)$$

where N is the total number of pixels per phase map, $\hat{\varphi}$ is the estimated phase, and φ is the true phase map.

3 RESULTS AND ANALYSIS

In this section, we explore the performance of our CNN models under idealized and degraded conditions. Both architectures, ResNet-50 and U-Net, give similar results. Therefore, we will only compare them when appropriate and we will use them interchangeably otherwise.

3.1 Fundamental limit for wavefront sensing

The wavefront estimation is fundamentally limited by the information contained in the measurement process. The Fisher information matrix and the Cramér–Rao (CR) bound are typically used to quantify the information ultimately extractable from the measurement whatever the estimation method. More specifically, the CR bound gives the lower bound on the error variance, and its reciprocal is the Fisher information for an unbiased estimator. Several studies rely on this lower bound to get the fundamental limit of different wavefront sensor's performances (Lee, Roggemann & Welsh 1999; Schulz, Sun & Roggemann 1999; Noethe & Adorf 2007; Paterson 2008, 2013; Plantet et al. 2015).

Considering photon noise only, i.e. the ultimate noise limit since it pertains to the nature of light, Paterson (2008) derives a fundamental limit for wavefront sensing without any assumption on the optics, if only that the wavefront sensor transforms the pupil phase in a measureable intensity. He derives the CR bound and finds that the measurement error of an aberration mode j must satisfy $\sigma_j^2 \geq 1/(4N_{\mathrm{ph}})$. Interestingly, this limit can also be derived from the uncertainty principle in the form $\Delta\varphi\ \Delta N_{\mathrm{ph}} \geq 1/2$, or $\Delta\varphi \geq 1/\sqrt{4N_{\mathrm{ph}}}$ for N_{ph} independent photon probes. However, the existence of this bound does not guarantee that it is actually possible to reach it. The most sensitive wavefront sensors, among existing concepts, are the ZWFS (Guyon 2005; N'Diaye et al. 2013) and the iQuad (Fauvarque et al. 2019). Both concepts rely on a $\pi/2$ phase shift between different parts of the focal plane and differ in its tessellation. Under photon noise only, the measurement error of the ZWFS[1] is $\sigma_j^2 = 1/(2N_{\mathrm{ph}})$ for the aberration mode j (N'Diaye et al. 2013). The loss of a factor two with respect to the fundamental limit is the result of the diffraction by the phase mask distributing half of the light outside the exit pupil. Since only the exit geometrical pupil is used for the measurement, the signal is effectively half of the input. Recently, however, a variation on the ZWFS with an enlarged central dot has been proposed improving its sensitivity at the expense of the lower spatial frequencies (Chambouleyron et al. 2021). In those conditions, the measurement error can almost reach the fundamental limit of $1/4N_{ph}$.

Second after the ZWFS, FPWFS offers a high sensitivity with a measurement error known to be $\propto 1/N_{\mathrm{ph}}$ (e.g. Meynadier et al. 1999; Guyon 2005; Paul et al. 2013; Bos et al. 2019). Classical focal plane imaging does not maximize the phase contrast, like the ZWFS, which explains a loss of sensitivity compared to the fundamental limit (Paterson 2008). Nevertheless, FPWFS still provides a very high sensitivity across a wide range of spatial frequencies, in contrast to, e.g. the Shack–Hartmann wavefront sensor, which has a poor sensitivity at low spatial frequencies (e.g. Guyon 2005). In practice, and for the analysis in this paper, the expected total theoretical residual error for N_{modes} statistically independent aberration modes is expressed by

$$\sigma_{\mathrm{th}}^2 = N_{\mathrm{modes}} \frac{1}{n_{\mathrm{img}} N_{\mathrm{ph}}} \quad (\mathrm{rad}^2), \quad (3)$$

where n_{img} and N_{ph} are the number of images and the number of photons per images, respectively.

3.2 Performance limit of CNNs

To analyse the capability of our CNN in terms of sensitivity, we train and evaluate different models over a broad range of flux levels: from 10^2 to 10^7 photons per image. This corresponds to a range of star magnitudes $m_H = 7$–19.5 (for 10^7 and 10^2 photons, respectively) assuming the following parameters: an integration time of $T_i = 1$ s, a transmission and quantum efficiency equal to 50 per cent, a telescope diameter of 10 m, and a filter bandwidth of 50 nm. We examine the two levels of aberrations (70 and 350 nm rms) and the two different spatial frequency contents (20 and 100 Zernike modes) described in Section 2.1. This allows us to study the limit of our trained models as a function of an increased dimensionality of the wavefront sensing problem.

The results are illustrated in Fig. 4. Each point is the median residual error of 100 evaluations and the error bars are the 5–95 per cent percentiles.[2] Note that the 100 evaluations refer to different phase screens, and not just to different photon noise realizations. These figures show where the performance is limited by photon noise and where it is limited by the model accuracy.

In the low aberration regime distributed over 20 Zernike coefficients, we can observe that the CNN reaches the sensitivity limit defined in equation (3) over a broad range of photon levels. The only exception is the low flux regime ($\lesssim 1e4$) where the error does not become arbitrarily large but reaches a saturation level at around 70 nm rms. A similar saturation level is observed at low flux with COFFEE, a coronagraphic phase diversity method based on a maximum a posteriori approach, when the appropriate regularization is used (Paul et al. 2013). In our case, this saturation can be interpreted as an implicit regularization originating from the training data distribution. The limit is reached when no aberration can be distinguished from the noise, in which case the predicted phase tends to zero. When the level of aberration is increased to 350 nm WFE, we can observe a plateau at the high signal end. This saturation level is a numerical limitation and can be reduced by increasing the data set size. See also Section 4.2.2 for a dedicated discussion.

The same analysis is performed for 100 Zernike modes (see Fig. 4, bottom). The low aberration regime exhibits a very similar behaviour, while the high aberration regime case is more strongly influenced by

[1] For a phase error close to zero and phase shift of $\pi/2$ over the central $1.06\lambda/D$ (N'Diaye et al. 2013).

[2] Note that this is applicable to all the following figures with error bars.

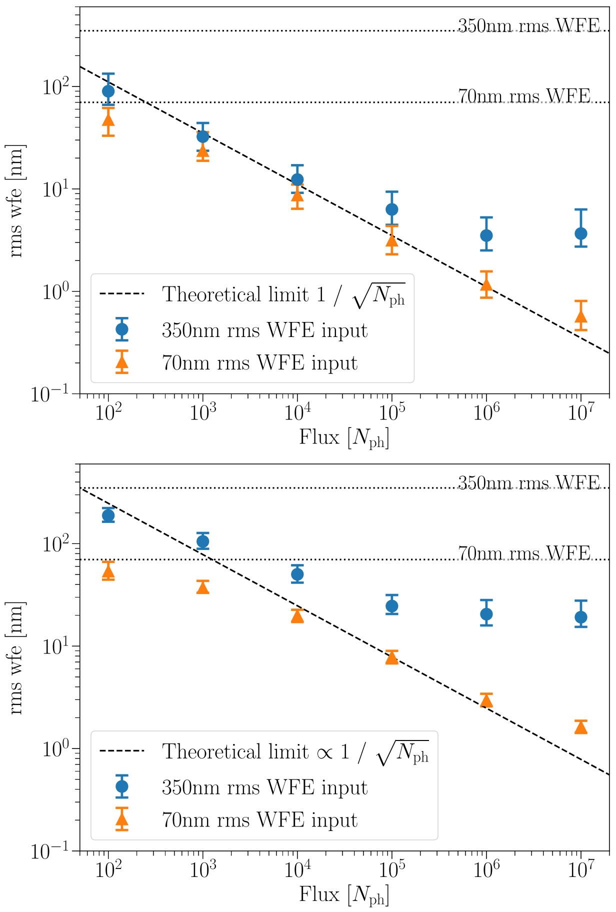

Figure 4. Residual rms WFE as a function of the flux per image for an input median rms WFE of 70 nm (orange) and 350 nm (blue). Each point uses a model specifically trained for that flux and aberration regime. (Top) Low spatial frequency content with 20 Zernike modes. (Bottom) Higher spatial frequency content with 100 Zernike modes.

the saturation level due to a suboptimal training. Hence, the accuracy rarely reaches the theoretical one. Again, this can be mitigated by increasing the data set size.

3.3 Robustness to changing signal-to-noise ratio

In this section, we explore the robustness of the CNN models under varying signal levels. The network architectures are trained at a specific SNR and then, during evaluation, exposed to a range of SNR. Fig. 5 is the result of six different trainings and 36 different evaluations, for a median WFE of 70 nm rms and 20 Zernike modes. At low signal level, the performance moves progressively away from the photon noise limit as the SNR used in the evaluation decreases, and converges to 70 nm rms due to the intrinsic noise regularization. At high signal level, the performance first degrades slightly as the SNR used in the evaluation increases, and then stagnates to a given WFE level, which depends on the training SNR. In all cases, the minimum WFE is reached at the training SNR. It is noteworthy that the performance degradation is mild in the vicinity of the training signal level so that a single model might suffice for a range of observing conditions.

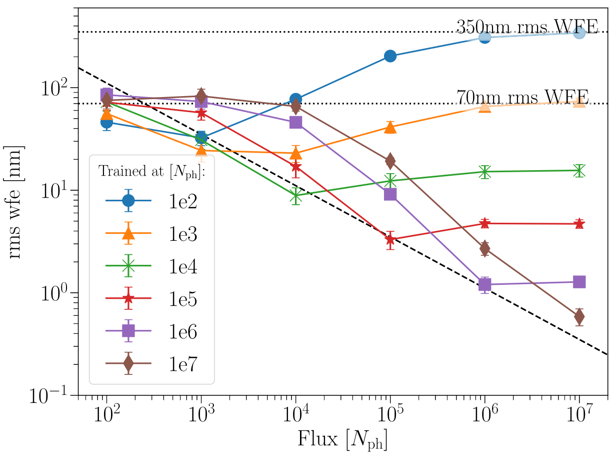

Figure 5. Robustness to changing photon flux levels. Each curve uses a different model and is evaluated at six different flux levels. We use here ResNet-50 trained on data sets with input median rms WFE of 70 nm rms distributed over 20 Zernike modes.

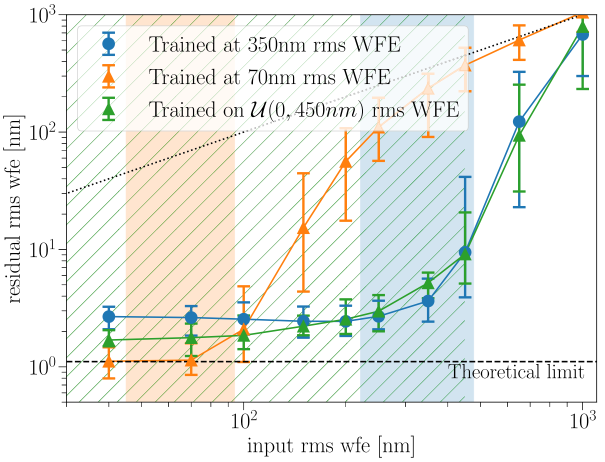

Figure 6. Residual rms WFE as a function of input rms WFE (40–450 nm rms WFE) distributed over 20 Zernike modes. In blue, for a ResNet-50 model trained around 350 nm WFE, in orange around 70 nm WFE, and in green trained on a uniform distribution of wavefront aberrations ranging from 0 to 450 nm WFE. The dashed line gives the fundamental limit as discussed in Section 3.1. The dotted line gives the one-to-one relation. The shaded areas represent the 2–98 per cent rms WFE percentiles in the respective training data sets.

If robustness under a wide range of SNR levels is desired, the training could be adapted. The distribution of flux in the training data set should be established based on the range of expected stellar magnitudes versus desired accuracy, and observing variability. Alternatively a number of models could be used and selected as a function of the current stellar magnitude.

3.4 Dynamic range

Here, we analyse the accuracy of the estimated wavefront for different levels of aberrations. In Fig. 6, we use two different models trained at low (70 nm rms) and high (350 nm) aberration levels. We test these models on eight data sets with aberration level ranging from 40 up to 450 nm rms, corresponding to 0.11 up to 1.28 rad rms. The results show a similar or better accuracy for aberration level below the trained one, and a rapidly decreasing accuracy at higher aberration levels.

The behaviour at low aberration level results from the way our data sets are generated, where each Zernike coefficient is drawn from a uniform distribution around zero. Hence, the training implicitly includes low aberration samples. It is interesting to note that the evaluation at the trained aberration level (70 and 350 nm, respectively) is actually not where the best performance can be found. Again this may be explained by the way our data sets are generated, with a distribution of WFE around a given median. In the case of a median WFE of 350 nm rms, the aberration distribution extends down to ∼225 nm (see blue shaded area in Fig. 6), which approximately coincides with the lowest residual error in Fig. 6. The performance is here limited by the size of the data set (100 000 entries), and since the effective number of entries with ≥350 nm WFE is simply lower than for ≥225 nm WFE in the training data set, the results are better at lower aberrations than at the median WFE.

We then train a model on a uniform distribution of wavefront aberrations ranging from 0 to 450 nm rms. To obtain a good training, we set the weight decay of the Adam optimizer to 1e-7, which is otherwise set to 0 by default. The result is also shown in Fig. 6. We observe an intermediate behaviour where the model performs better at low aberration but slightly worse at higher aberration compared to the model trained at 350 nm rms.

At higher aberration level, the models are less efficient in picking up features and extracting useful information. Nevertheless, it is worth noting that a valid correction extends well beyond the trained aberration level, and despite a rapidly decreasing accuracy.

Following the results in Fig. 6, we apply iteratively the CNN model for aberrations well beyond the trained aberration level. At each iteration, the CNN infers a wavefront, which we subtract from the input to produce a new pair of PSFs based on the residual wavefront. We test this iterative approach for 40 different aberration levels ranging from 500 to 1750 nm. To limit the effect of field-of-view cropping, the tip-tilt modes are removed from all phase maps (and the quoted WFE is also calculated without tip-tilt). The training range of the CNN is here around 350 nm rms. The results are presented in Fig. 7. We can observe that the CNN properly converges in a few iterations for initial aberration levels well beyond its training range. Also, once it has converged, the correction stays stable. It is only for initial aberration levels $\gtrsim 1.1$ µm that the rms WFE either stagnates or starts to diverge. In Fig. 7 (right), two PSFs are illustrated, with rms WFE of ∼320 and ∼1060 nm, respectively. Both are well corrected after a few iterations. The morphology of these images is very different, yet the CNN is able to converge, which is remarkable. Overall these are very encouraging results for real applications, typically running in closed-loop, starting with high levels of aberrations and with the objective of stabilizing the WFE to the lowest level.

An alternative approach would be of course to train the CNN model on a wider aberration range for which the application would not need to be iterative. As a comparison, we trained three additional CNNs[3] with larger aberrations: 535, 800, and 1070 nm rms (or 700, 1050, and 1400 if the tip-tilt was not removed), and we test them on a range of aberrations from 250 to 1070 nm, similarly to the results presented in Fig. 6. The bottom line of this comparison is that the residual WFE increases with the level of aberrations the CNN was trained with, in particular for training at 350, 535, 800, and 1070, the residual error is 3.5, 5.2, 13.6, and 27.4 nm rms. While the training can certainly be improved with, e.g. larger data sets or a better distribution of the input WFE in the data sets, the results presented in Fig. 7 show that we can also benefit from the generalization power of CNN, for instance, for the bootstrapping phase in a closed-loop system, without the need for more demanding training strategy or different network architectures, such as recurrent neural networks.

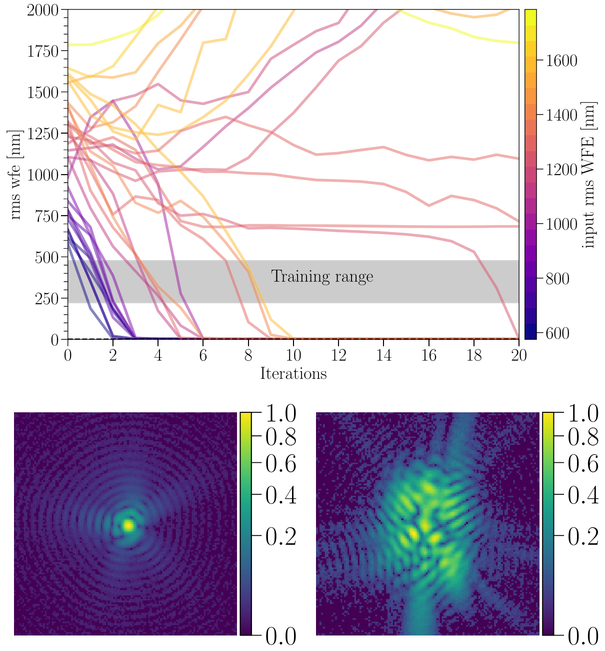

Figure 7. (Top) Iterative application of the CNN to different levels of aberrations. (Bottom) Illustration of two PSFs, with rms WFE of 320 (top) and 1060 nm (bottom); in both cases the CNN converges after a few iterations.

3.5 Robustness to higher order disturbances

In realistic conditions, a wavefront sensor measures the projection of the phase aberration on a finite set of values, such as Zernike coefficients or zonal values corresponding to deformable mirror actuators. Higher order aberrations are thus not sensed and can be considered as an additional source of disturbance. This is the case, for example, with the Shack–Hartmann wavefront sensor, where a limited sampling of the wavefront leads to an aliasing effect increasing the measurement error. It is therefore of practical interest to study how this unwanted signal affects the inference of the trained CNN.

To explore the effect of higher order disturbances, we use different models trained on phase maps constructed with 20 Zernike modes and evaluate their accuracy on a data set with 100 Zernike modes. We consider both ResNet-50 and U-Net models, with input WFE of 70 and 350 nm rms. The residual WFEs ($\Delta\varphi = \varphi - \hat{\varphi}$) are projected on 100 Zernike modes and the rms (over 100 different evaluations) of the obtained Zernike coefficients are calculated. The results are illustrated in Fig. 8. We can observe a higher loss of accuracy for higher aberration levels (see Fig. 8, left versus right), and for higher flux levels (not shown). The simple interpretation is that the CNN models are more affected by higher order aberrations as they become more prominent, and therefore distinctive, with respect to the photon noise. Although a minor effect, it is interesting to note that the direct phase estimation done by U-Net provides a valid correction beyond the 20 Zernike modes for which it was trained, while the ResNet-50 is bounded to the first 20 Zernike modes by construction.

[3] With 100 000 entries in our data set, and for a photon flux of 1e6.

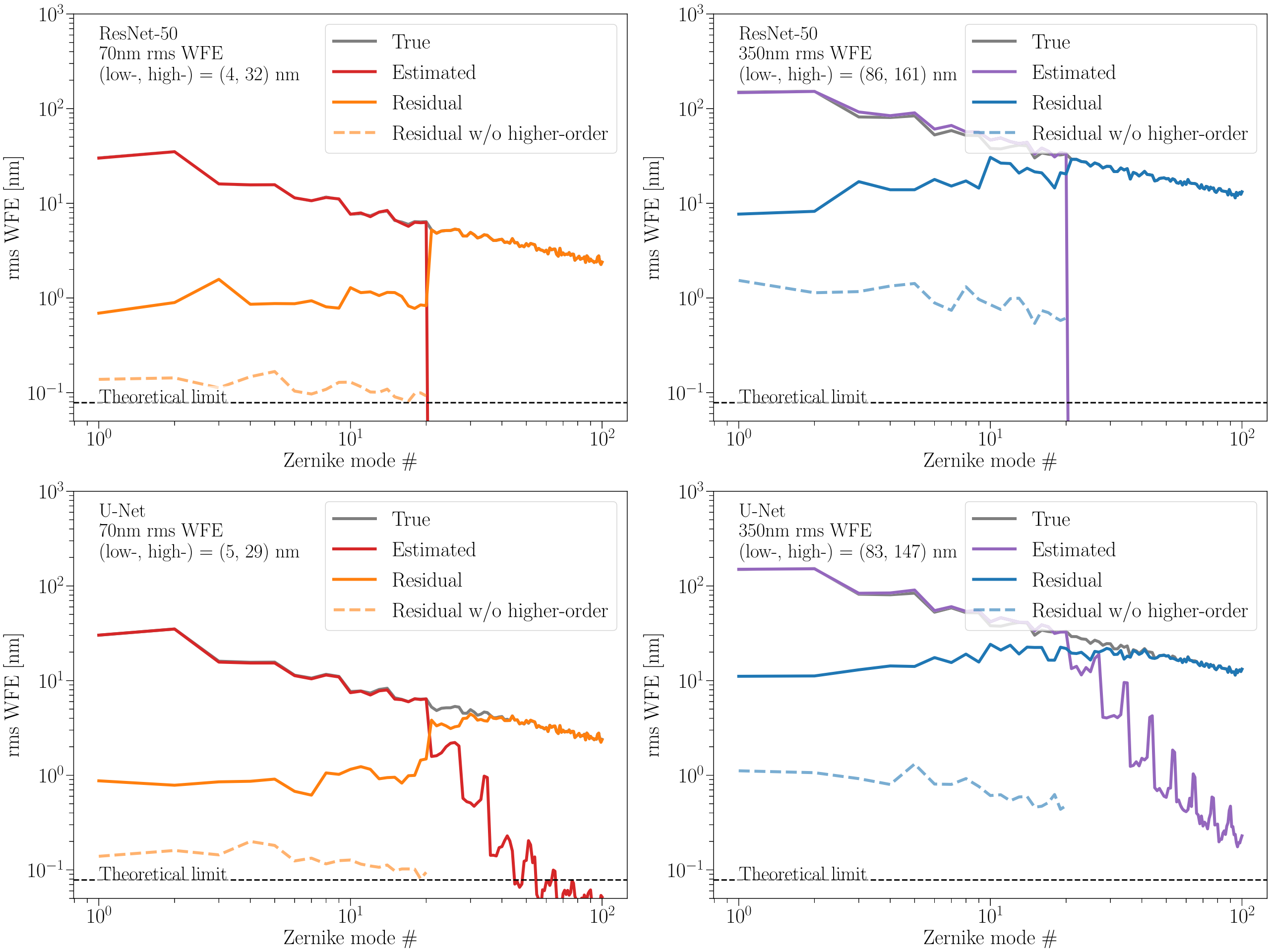

Figure 8. Modal rms WFE for two different levels of aberrations: 70 nm rms (left; red and orange) and 350 nm rms (right; purple and blue); and two architectures: ResNet-50 (top) and U-Net (bottom). The models are exposed to aberrated PSFs with higher order aberrations compared to the training data set (solid lines). The estimated phase rms WFE (red and purple) and the residual errors (orange and blue) are compared to the true phase rms WFE (grey). The residual errors, when no higher order aberrations are present, are also plotted in dashed lines. In all cases, the signal per image is 10^7 photons. For a median input of 70 nm rms, both ResNet and U-Net reduce the WFE of the low-order modes to about 4 nm rms, while the higher order modes are essentially unaffected and have an rms of about 30 nm. When the aberrations are larger (350 nm rms), the models start to be significantly affected and the rms error increases to about 80 nm (140 nm) for the low (high) orders, respectively.

While for NCPA correction, behind an extreme AO system where residual atmospheric aberrations are kept to a minimum, our training strategy might be appropriate; for AO application this degradation might be a showstopper (although we do not fully explore this here). In fact, beyond the mild robustness offered by the CNN-based models, these results illustrate the following rule-of-thumb: the training data should always be as representative as possible of the real observing conditions.

3.6 Phase diversity: implication of sign ambiguity

To properly recover and avoid any ambiguity on the phase in the pupil plane using focal plane images requires a unique intensity measurement for a given phase aberration. This uniqueness is not generally guaranteed, in particular, for circularly symmetric pupils. Ambiguities occurring in phase retrieval is extensively discussed in the literature for applications from image reconstruction to wavefront sensing (see e.g. Bos et al. 2019, for a review relevant to astronomy). The non-uniqueness in the case of circularly symmetric pupils is the well-known sign ambiguity, which results from the Hermitian properties of Fraunhofer propagation. Indeed, the pupil-plane electric field $E_{\mathrm{pup}}(x)$ and the same field flipped and conjugated $E^{*}_{\mathrm{pup}}(-x)$ have the same Fourier transform, and therefore lead to the same intensity distribution in the focal plane. For an even phase aberration (defined by $\varphi_{\mathrm{even}}(x) = \varphi_{\mathrm{even}}(-x)$), and assuming an even amplitude distribution across the pupil – which we omit in the following – we can write

$$E^{*}_{\mathrm{pup}}(-x) = \exp(-j\varphi_{\mathrm{even}}(-x)) \tag{4}$$

$$= \exp(-j\varphi_{\mathrm{even}}(x)), \tag{5}$$

and thus

$$\mathcal{F}\{\exp(-j\varphi_{\mathrm{even}}(x))\} = \mathcal{F}\{\exp(j\varphi_{\mathrm{even}}(x))\}, \tag{6}$$

i.e. $\varphi_{\mathrm{even}}(x)$ and $-\varphi_{\mathrm{even}}(x)$ produce the same PSFs. Expressing φ_{even} as a sum of even Zernike modes, $\varphi_{\mathrm{even}} = \sum_{n,\,m;\,n\,=\,\mathrm{even}} a_{n,\,m} Z_{n,\,m}$, one easily understands that the relative sign between even modes is not ambiguous and the degeneracy reduces to a single sign ambiguity, regardless of the number of even modes to be evaluated.

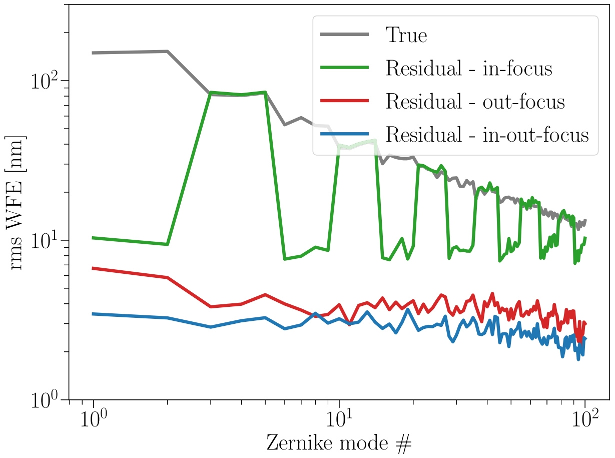

Figure 9. Modal WFE evaluated over 100 different phase maps. In blue, the input phase map following a spatial power spectrum with a slope of −2. The other curves give the residual error for three different models: using a single in-focus PSF, using a single out-of-focus PSF, and using both PSFs.

Many approaches exist to lift this ambiguity (see e.g. Jovanovic et al. 2018, for high-contrast imaging). Two natural (but not necessarily desired in the way they may affect the observations) ways are either to introduce a phase diversity as done in this paper or to use a non-centro symmetric pupil support (e.g. Bos et al. 2019). Here, we simply illustrate how the CNN behaves with respect to this sign ambiguity, by comparing the modal rms error for CNN models trained using two images (as in the previous sections), one defocused image only and one in-focus image only. The results are shown in Fig. 9. In the case of a single in-focus PSF, we see that the residual errors on the even modes correspond to the input error (the estimation is close to zero) as a result of the ambiguity, while odd modes are properly sensed. However, the accuracy in the odd modes is degraded compared to the single defocused case. This is the result of a suboptimal training and translates in overfitting, which can be identified by the training and validation curves. This could be circumvented by adapting the loss function and replacing equation (2) by the rms error on the odd coefficients only. The single defocused PSF case does not suffer from sign ambiguity. This is the result of an implicit prior: the introduced defocus is equal to λ/4 or 550 nm rms, while the focus term in the aberration to be sensed is drawn from a uniform distribution in the range ∼[−150, 150] nm rms (for 350 nm rms WFE and 20 Zernike modes), so that the total focus term is always positive. The factor $\sqrt{2}$ between the single defocused image and the two images case is solely due to the increased SNR.

3.7 Effect of field-of-view and PSF sampling

In the results presented in the previous sections, the PSF was sampled with 4.5 pixels over λ/D, or a pixel scale of $0.22\lambda/D$/px, with a grid size of 128 × 128 pixels. Here, we study how the PSF sampling influences the performance. In order to preserve the exact same network architectures, we keep a fixed grid size of 128 × 128 pixels. Therefore, increasing the PSF sampling means reducing the field of view and potentially leaving out information. Conversely, a coarser sampling, in particular, below the Nyquist sampling (i.e. <2 pixel/λ/D), may lead to a loss of information.

To examine this effect, we generate different data sets with pixel scales between $0.1\lambda/D$ and $1\lambda/D$ and we train a new model for each case. An example of the generated PSF is given in Fig. 10. The results are illustrated in Fig. 11 for a median rms WFE of 350 nm, 10^7 photons per image, and 20 Zernike modes. We notice a mild degradation with increasing pixel scale (for pixel scale $>0.5\lambda/D$), and very similar results for small pixel scales ($<0.5\lambda/D$). Above the Nyquist limit (pixel scale $<0.5\lambda/D$), the information loss due to the field-of-view cropping is negligible at this level of aberrations and flux. One can expect that in a more aberrated regime, this may start to have a noticeable impact. Below the Nyquist limit (pixel scale $>0.5\lambda/D$), we notice a loss of accuracy of a factor two. That this degradation is not more severe may be related to the limited number of Zernike modes used here, which produce extended signatures in the focal plane. With an increased spatial frequency content and higher level of aberrations, the PSFs would break into many speckles, which are expected to lead to a more serious degradation of the performance.

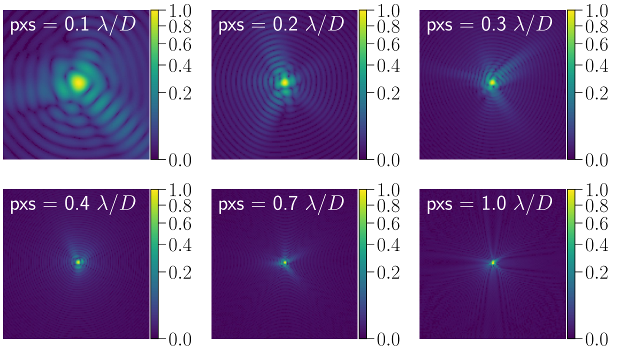

Figure 10. Illustration of different PSF sampling. Starting from (top left), the pixel scale is [0.1, 0.2, 0.3, 0.4, 0.7, 1] λ/D. Since the PSF grid size is kept fixed to 128 × 128 pixels, the field of view changes accordingly with, respectively, [12.8, 25.6, 38.4, 51.2, 89.6, 128] λ/D.

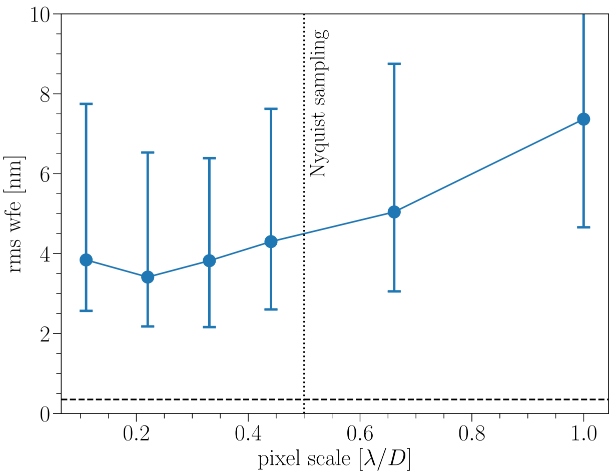

Figure 11. Residual rms WFE for different pixel scales. The same pixel scale is used for training and evaluation. Each point is a different model. The Nyquist sampling ($0.5\lambda/D$) is indicated by the vertical dotted line. The photon noise limit is indicated by the horizontal dashed line.

4 DISCUSSION

4.1 Comparison to Gerchberg–Saxton phase retrieval

While CNNs can solve the capture range problem and provide an initial estimate for gradient-based optimizers in the case of large

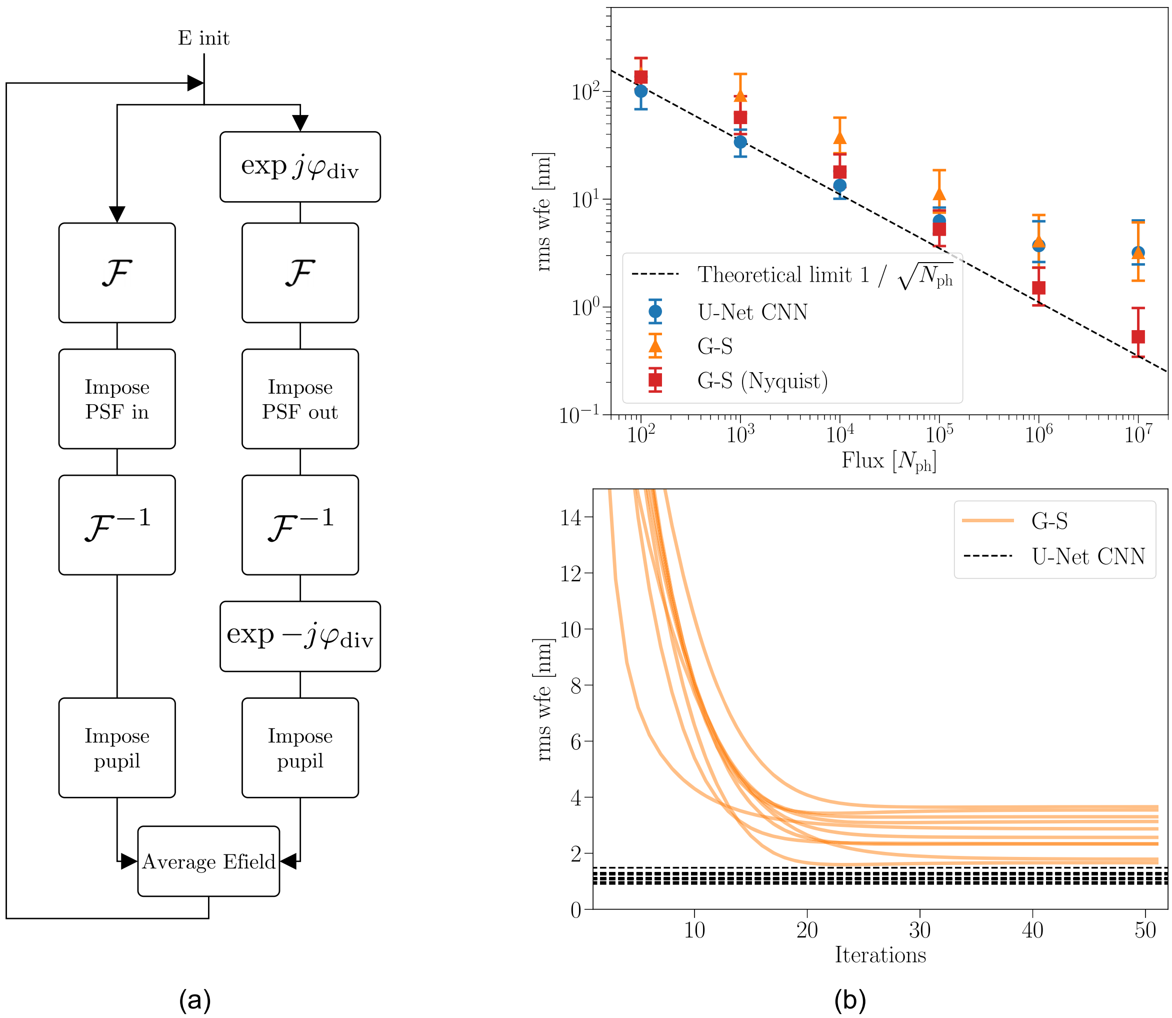

Figure 12. Comparison of an iterative algorithm to the CNN model. (Left) Implemented GS algorithm for the phase retrieval using two images, one with a phase diversity. (Top right) Residual rms WFE as a function of photon number per image for the CNN model and for the GS algorithm for two different pixel sampling of the PSFs: in orange, the same sampling used with the CNN, and in red, Nyquist sampling. (Bottom right) Illustration of the GS convergence for 10 different evaluations with the rms WFE as a function of iteration, and for a sampling of $0.22\lambda/D$ per pixels. It takes approximately 20 iterations for the GS algorithm to converge.

aberrations (Paine & Fienup 2018), it is also interesting to see how it competes with 'classical' approaches in a lower aberration regime such as explored in this paper. A detailed comparison with the most efficient algorithms that may exist is outside the scope of this paper. Rather we compare the CNN models with a standard iterative phase retrieval algorithm to illustrate where the CNN may be superior. The Gerchberg–Saxton (GS) iterative algorithm (Gerchberg & Saxton 1972; Fienup 1982) is relatively simple to implement, widely used, and can be easily adapted to a specific application. To exploit our two images, one in-focus and one defocused, we implement an algorithm that uses multiple images in parallel (Milster 2020). The algorithm is depicted in Fig. 12 (left). Since the phase diversity $\varphi_{\rm div}$ is known, it can be appropriately added or removed at different steps of the algorithm. At the end of each iteration, once the phase diversity is removed, the pupil plane electric fields are averaged and the output is used for the next iteration. We compared this parallel approach to a serial one (e.g. Guyon 2010; Milster 2020) and found it to be superior.

Since the phase is inferred from the complex exponent, the phase output from the iterative algorithm needs to be unwrapped. With phase maps of about 1 rad rms, phase wrapping can occur (the phase can locally be larger than π) and we need to take it into consideration. Phase unwrapping can be challenging and it is interesting to note that it can also be solved by CNNs (Wang et al. 2019). To avoid unnecessary complications, we assess the performance of the iterative algorithm by analysing the phase residual directly, calculated by $\angle \exp{(i(\varphi - \hat{\varphi}))}$, which we expect to be in the range of $[-\pi, \pi]$.

We reproduce the analysis in Section 3.1 and evaluate both approaches at different signal levels. For the iterative algorithm, we consider two different pixel scales: $0.5\lambda/D$ per pixel (Nyquist sampling) and $0.22\lambda/D$ per pixel (same sampling as used for the CNN, see Section 2.1). With Nyquist sampled PSFs, the GS algorithm provides an accuracy close to the theoretical limit over the full range of photon levels (see Fig. 12, top right). It surpasses the CNN model at high flux, where it does not reach a plateau. With the finer sampling, the GS algorithm becomes less accurate at all signal levels and reaches a plateau at high flux similar to the CNN. We can only suspect that the cropping of the PSFs and the way we impose the amplitude in the image plane have a detrimental effect on the GS

Table 1. Computational cost of the two CNN architectures for 100 Zernike modes.

Architectures	Number of parameters (10^6)	MACs (10^9)	Model size (MB)
ResNet-50	23.71	4.11	91
U-Net	13.40	7.77	52

algorithm. While the plateau seen with the CNN is apparently not due to pixel sampling (see Section 3.7), it may be alleviated by a larger training data set (see Section 4.2.2).

We also monitor the convergence of the GS algorithm (see Fig. 12, bottom right). We can see that the iterative algorithm needs approximately 20 iterations to converge where the CNN performs a similar or superior inference in just one iteration. While we do not quantify the computational gain, this likely translates into a speed advantage for the CNN. For example, Paine & Fienup (2018) indicated a gain of a factor ∼80 when comparing to non-linear optimization methods.

Finally, we also explore the performance obtained with a gradient-based optimizer. More specifically, similarly to Peng et al. (2020), we implement the image formation in PYTORCH and use its autodifferentiation capabilities to optimize the objective function[4] using the same variant of stochastic gradient descent, i.e. the Adam optimizer. Our results, not illustrated here for the sake of conciseness, suggest a similar performance to the GS algorithm but requiring a higher number of iteration. Hence, in the context of our paper, gradient-based optimization neither appears to be outperforming the GS algorithm (whose performance is already close to the theoretical limit) nor does it require fewer iterations. A thorough exploration and comparison of iterative methods is, however, outside the scope of this paper.

4.2 Numerical considerations

In this last section, we discuss two numerical aspects relevant for practical applications: the computational cost associated with the CNN models and the influence of the training data set size on performance.

4.2.1 Computational cost

Computational cost is often measured by the number of floating-point operations (FLOPs) or the number of multiply accumulated operations (MACs). The number of MAC, and the number of parameters and the memory size of the two CNN models are estimated with the package THOP[5] and are given in Table 1. The number of FLOPs is about twice the number of MACs.

In the case of NCPAs, where a correction is expected at best on a time-scale of a second, the computational cost given in Table 1 is perfectly acceptable with at most 8–16 GFLOPs per second. In contrast, for an AO system typically running at 1 kHz, about 8–16 TFLOPs per second would be required. Considering that a good GPU RTX 2080Ti provides >13 TFLOPs per second in single precision, it can be considered as a feasible approach from the computational power standpoint. Its practical implementation, ranging from an appropriate software implementation to keeping latencies to a minimal level and synchronizing the estimation time from multiple GPUs, if needed, might, however, not be trivial.

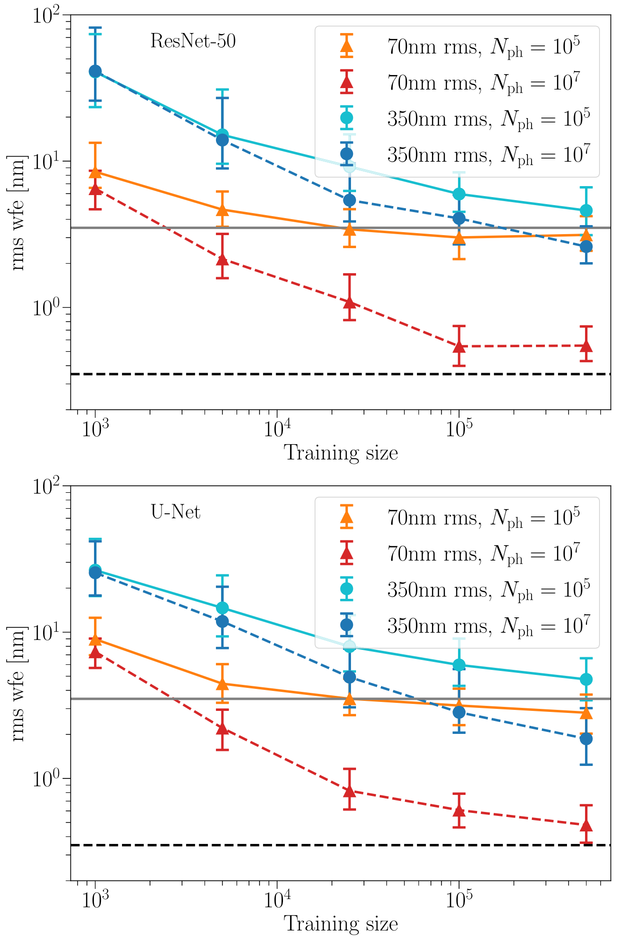

Figure 13. Residual rms WFE as a function of training data set size for different flux levels and amplitude of the input aberrations. The phase errors are distributed over 20 Zernike modes. The horizontal lines (solid and dashed) represent the theoretical limit for a flux of 10^5 and 10^7 photons, respectively. (Top) ResNet-50. (Bottom) U-Net.

However, the numbers given in Table 1 should only be considered as upper bounds. Indeed, the deployment of CNN-based models on a real set-up could use compression and acceleration techniques to reduce the memory and computational cost (e.g. Cheng et al. 2020). Compression and speed-up ratios of about 5–10 have actually been reported for ResNet and similar architectures (e.g. Wang et al. 2018).

4.2.2 Influence of training data set size

To try and understand what may limit the CNN models accuracy, we study the delivered accuracy as a function of the training data set size. We compare two levels of aberrations (distributed over 20 Zernike modes) and two levels of photon noise for the two architectures, ResNet-50 and U-Net. The training data set sizes range from 1000 to 500 000 entries, while in the previous sections we used 100 000 entries. The results are illustrated in Fig. 13.

In the low aberration (70 nm rms) and low flux (10^5 photons per image) regime, the accuracy quickly converges with increasing data set size and requires only >5000 entries. Increasing the flux by two orders of magnitude (10^7 photons per image), the SNR is boosted

[4] We use here the mean square error of the amplitudes in the focal plane.

[5] See also `https://github.com/Lyken17/pytorch-OpCounter`, which only considers the number of multiplication operations to evaluate the computational costs of our two CNN architectures.

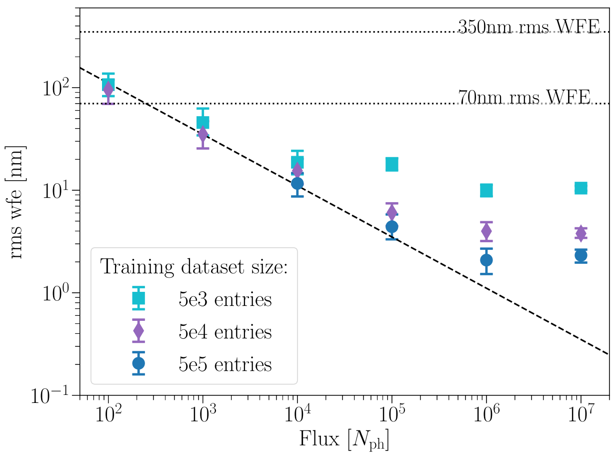

Figure 14. Residual rms WFE as a function of flux level for three training data set sizes: 5000, 50 000, and 500 000 entries, respectively, using the U-Net model. Increasing the data set size by a factor 10 reduces the model error by a factor of two approximately, at high flux levels.

by a factor 10 and finer details can be picked up during training. The data set then needs to be larger with >100 000 entries for the training to fully converge and to reach the sub-nm theoretical floor. For aberrations five times larger (350 nm rms), the larger parameter space requires a substantially larger data set with $\gtrsim$500 000 entries to attain the theoretical limit in the low flux case. Finally, at higher flux, we do not reach the theoretical limit, which would require a much larger data set, i.e. >>500 000 entries. Comparing ResNet-50 and U-Net, we can observe that both architectures reach very similar performance.

To emphasize the effect of flux, which directly relates to the amount of extractable information, we plot in Fig. 14 the rms WFE as a function of photon level for three different training set sizes. The interpretation is relatively straightforward: in each case the accuracy is close to the theoretical limit until it reaches a plateau. The level of this plateau depends on the training data set size, where a 10-fold increase leads to approximately a factor two improvement of the rms WFE.

The analysis performed here illustrates the need for large training sets when the problem parameter space increases if one wants to reach the finest accuracy, i.e. the photon noise limit. For NCPA correction, the range of data set sizes explored here is presumably large enough. For AO that has likely a larger parameter space (larger range of aberration amplitudes and larger number of modes to be controlled) but possibly a lower accuracy requirement, one may trade-off accuracy with data set size for practical reasons, in particular, if the training set is composed of experimental data.

5 CONCLUSIONS

In this paper, we explored the use of deep CNNs to perform image-based wavefront sensing. We focused on low level of aberrations (0.1–1 rad rms WFE) and a limited number of spatial frequencies (20–100 Zernike modes). These parameters are characteristic of NCPA measurements on large ground-based telescopes (8–40m; e.g. VLT or ELT). Our simulations suggest that the CNN models are able to leverage the high sensitivity of FPWFS over a broad range of signal levels. In terms of dynamic range, we have demonstrated successful correction for aberration levels up to λ/6 rms WFE in one iteration, and $\gtrsim$λ/2 in 5–10 iterations. The models are robust under reasonable flux changes, with a mild departure from the photon noise limit with changing SNR level. The prediction accuracy of the trained models is, however, affected by unknown disturbances such as higher order aberrations, and the training strategy should be adapted accordingly.

The type of architecture, and in particular the choice of approach between the Zernike coefficients (ResNet) versus the direct phase map (U-Net) estimations, has a negligible impact in our experiments (contrary to, e.g. Guo et al. 2019). In fact, we have used ResNet-50 and U-Net interchangeably. U-Net does have a slight edge in terms of generalization power, as illustrated in Fig. 8, but this advantage is too marginal to justify alone a preference for this architecture. When compared to an iterative phase retrieval algorithm, the CNN models display similar and often superior accuracies in just one iteration, while the iterative algorithm requires 10–20 iterations to converge. Hence, in addition to a close-to-optimum estimation accuracy, CNN models are expected to be faster.

While using CNN-based FPWFS for NCPA measurement seems readily applicable, its utilization for AO appears more challenging for an equivalent telescope diameter. Indeed, AO calls for the sensing of a larger number of modes (at least a factor 10) and larger aberration levels (a factor ∼10 in the bootstrapping phase of the AO closed-loop) compared to NCPA measurement, hence the dimensionality is largely increased compared to NCPA measurement. Finally, AO also requires sub-ms inference speed potentially constraining the CNN architecture and its implementation. In the prospect of real application, an encouraging result of our simulation is that CNNs can be applied iteratively, in closed-loop, to reduce the WFE to a low level. In particular, the WFE stays at a stable low level, close to the expected theoretical limit, and is able to converge in just a few iterations for initial aberration levels well beyond its training range.

Real-life data are, however, more complex than our simulated images, including their finite wavelength range, the different detector noises, the residual atmospheric perturbations, imperfect optical alignment, etc. Some of these effects can be anticipated and simulated, such as the polychromaticity, which will wash out part of the high spatial frequency information, but some others may not be. While the instrument model can be improved to generate more realistic labelled data sets, ultimately experimental data are needed for the training of the CNN models. How to best exploit a limited experimental data set is thus one of the key challenge for future applications. Another key aspect in the context of NCPA measurements is to maintain a 100 per cent science duty cycle despite the phase diversity required to lift any ambiguity in the FPWFS measurement. An interesting avenue in that context is to leverage the diversity introduced by the changing atmosphere using, for example, long short term memory networks to exploit its temporal structure.

ACKNOWLEDGEMENTS

This research used different PYTHON packages and libraries: PYTORCH (Paszke et al. 2019), ASTROPY (Astropy Collaboration 2013), AOTOOLS (Townson et al. 2019), PROPER (Krist 2007), NUMPY (Harris et al. 2020), and MATPLOTLIB (Hunter 2007). The research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 819155), and by the Federation Wallonia-Brussels (grant for Concerted Research Actions).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Allan G., Kang I., Douglas E. S., Barbastathis G., Cahoy K., 2020a, Opt. Express, 28, 26267

Allan G., Kang I., Douglas E. S., N'Diaye M., Barbastathis G., Cahoy K., 2020b, in Lystrup M., Perrin M. D., Batalha N., Siegler N., Tong E. C., eds, Proc. SPIE Conf. Ser. Vol. 11443, Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave. SPIE, Bellingham, p. 1144349

Andersen T., Owner-Petersen M., Enmark A., 2019, Opt. Lett., 44, 4618

Andersen T., Owner-Petersen M., Enmark A., 2020, J. Astron. Telesc. Instrum. Syst., 6, 034002

Angel J. R. P., Wizinowich P., Lloyd-Hart M., Sandler D., 1990, Nature, 348, 221

Astropy Collaboration, 2013, A&A, 558, A33

Barrett T. K., Sandler D. G., 1993, Appl. Opt., 32, 1720

Bos S. P. et al., 2019, A&A, 632, A48

Chambouleyron V. et al., 2021, A&A, 650, L8

Cheng Y., Wang D., Zhou P., Zhang T., 2020, preprint (arXiv:1710.09282)

Cumming B. P., Gu M., 2020, Opt. Express, 28, 14511

Delavaquerie E., Cassaing F., Amans J. P., 2010, in Clénet Y., Conan J.-M., Fusco Th., Rousset G., eds, 1st AO4ELT conference - Adaptative Optics for Extremely Large Telescopes. EDP Sciences, Les Ulis, France, p. 05018

Dohlen K., Wildi F. P., Puget P., Mouillet D., Beuzit J.-L., 2011, 2nd AO4ELT conference - Adaptive Optics for Extremely Large Telescopes. Victoria, Canada, p. 75

Fauvarque O. et al., 2019, 6th AO4ELT conference - Adaptive Optics for Extremely Large Telescopes, Québec, Canada

Fienup J. R., 1982, Appl. Opt., 21, 2758

Foley J. T., Butts R. R., 1981, J. Opt. Soc. Am., 71, 1008

Gerchberg R. W., Saxton W. O., 1972, Optik, 35, 237

Gonsalves R. A., 1982, Opt. Eng., 21, 829

Guo H., Xu Y., Li Q., Du S., He D., Wang Q., Huang Y., 2019, Sensors, 19, 3533

Guyon O., 2005, ApJ, 629, 592

Guyon O., 2010, PASP, 122, 49

Harris C. R. et al., 2020, Nature, 585, 357

He K., Zhang X., Ren S., Sun J., 2016, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, p. 770

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Jorgenson M. B., Aitken G. J. M., 1992, Opt. Lett., 17, 466

Jovanovic N. et al., 2018, in Close L. M., Schreiber L., Schmidt D., eds, Proc. SPIE Conf. Ser. Vol. 10703, Adaptive Optics Systems VI. SPIE, Bellingham, p. 107031U

Keller C. U., Korkiakoski V., Doelman N., Fraanje R., Andrei R., Verhaegen M., 2012, in Ellerbroek B. L., Marchetti E., Véran J.-P., eds, Proc. SPIE Conf. Ser. Vol. 8447, Adaptive Optics Systems III. SPIE, Bellingham, p. 844721

Kingma P. D., Ba L. J., 2015, The 3rd International Conference on Learning Representations

Korkiakoski V., Keller C. U., Doelman N., Fraanje R., Andrei R., Verhaegen M., 2012, in Ellerbroek B. L., Marchetti E., Véran J.-P., eds, Proc. SPIE Conf. Ser. Vol. 8447, Adaptive Optics Systems III. SPIE, Bellingham, p. 84475Z

Krishnan A. P., Belthangady C., Nyby C., Lange M., Yang B., Royer L. A., 2020, bioRxiv

Krist J. E., 2007, in Kahan M. A., ed., Proc. SPIE Conf. Ser. Vol. 6675, Optical Modeling and Performance Predictions III. SPIE, Bellingham, p. 66750P

Krizhevsky A., Sutskever I., Hinton G. E., 2017, Commun. ACM, 60, 84

Landman R., Haffert S. Y., 2020, Opt. Express, 28, 16644

LeCun Y., Boser B., Denker J., Henderson D., Howard R., Hubbard W., Jackel L., 1990, in Touretzky D., ed., Advances in Neural Information Processing Systems. vol. 2. Morgan-Kaufmann, Burlington, Massachusetts, p. 396

Lee D. J., Roggemann M. C., Welsh B. M., 1999, J. Opt. Soc. Am. A, 16, 1005

Liu X., Morris T., Saunter C., de Cos Juez F. J., González-Gutiérrez C., Bardou L., 2020, MNRAS, 496, 456

McGuire P. C., Sandler D. G., Lloyd-Hart M., Rhoadarmer T. A., 1999, Adaptive Optics: Neural Network Wavefront Sensing, Reconstruction, and Prediction, Springer Berlin Heidelberg, p. 97

Meynadier L., Michau V., Velluet M.-T., Conan J.-M., Mugnier L. M., Rousset G., 1999, Appl. Opt., 38, 4967

Milster T. D., 2020, in Blanche P.-A., ed., Optical Holography. Elsevier, Amsterdam, The Netherlands, p. 61

Montera D. A., Welsh B. M., Ruck D. W., Roggemann M. C., 1996, Appl. Opt., 35, 4238

N'Diaye M., Dohlen K., Fusco T., Paul B., 2013, A&A, 555, A94

Naik K. R., Wright R. H., Claveau D. D., Acton D. S., Knight J. S., 2020, in Schreiber L. Schmidt D. Vernet E., eds, Proc. SPIE Conf. Ser. Vol. 11448, Adaptive Optics Systems VII. SPIE, Bellingham, p. 114481H

Nishizaki Y., Valdivia M., Horisaki R., Kitaguchi K., Saito M., Tanida J., Vera E., 2019, Opt. Express, 27, 240

Noethe L., Adorf H. M., 2007, J. Mod. Opt., 54, 3

Osborn J. et al., 2014, MNRAS, 441, 2508

Paine S. W., Fienup J. R., 2018, Opt. Lett., 43, 1235

Paszke A. et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E., Garnett R., eds, Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Curran Associates, Inc.

Paterson C., 2008, J. Phys.: Conf. Ser., 139, 012021

Paterson C., 2013, Imaging and Applied Optics. Optical Society of America, p. OM2A.1

Paul B., Mugnier L. M., Sauvage J. F., Ferrari M., Dohlen K., 2013, Opt. Express, 21, 31751

Paxman R. G., Schulz T. J., Fienup J. R., 1992, J. Opt. Soc. Am. A, 9, 1072

Peng Y., Choi S., Padmanaban N., Wetzstein G., 2020, ACM Trans. Graph., 39, 1

Plantet C., Meimon S., Conan J. M., Fusco T., 2015, Opt. Express, 23, 28619

Quesnel M., Orban de Xivry G., Louppe G., Absil O., 2020, in Schreiber L., Schmidt D., Vernet E., eds, Proc. SPIE Conf. Ser. Vol. 11448, Adaptive Optics Systems VII. SPIE, Bellingham, p. 114481G

Ronneberger O., Fischer P., Brox T., 2015, in Navab N., Hornegger J., Wells W. M., Frangi A. F., eds, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, p. 234

Sandler D. G., Barrett T. K., Palmer D. A., Fugate R. Q., Wild W. J., 1991, Nature, 351, 300

Schulz T. J., Sun W., Roggemann M. C., 1999, in Roggemann M. C., Bissonnette L. R., eds, Proc. SPIE Conf. Ser. Vol. 3763, Propagation and Imaging through the Atmosphere III. SPIE, Bellingham, p. 23

Swanson R., Lamb M., Correia C., Sivanand am S., Kutulakos K., 2018, in Close L. M., Schreiber L., Schmidt D., eds, Proc. SPIE Conf. Ser. Vol. 10703, Adaptive Optics Systems VI. SPIE, Bellingham, p. 107031F

Swanson R., Lamb M., Correia C. M., Sivanandam S., Kutulakos K., 2021, MNRAS, 503, 2944

Townson M. J., Farley O. J. D., Orban de Xivry G., Osborn J., Reeves A. P., 2019, Opt. Express, 27, 31316

Vanberg P.-O., Orban de Xivry G., Absil O., Louppe G., 2019, Machine Learning and the Physical Sciences. Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada, p. 107

Vievard S. et al., 2019, 6th AO4ELT conference - Adaptive Optics for Extremely Large Telescopes. Québec, Canada

Wang Y., Xu C., Qiu J., Xu C., Tao D., 2018, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18. Association for Computing Machinery, New York, NY, USA, p. 2476

Wang K., Li Y., Kemao Q., Di J., Zhao J., 2019, Opt. Express, 27, 15100

Wu Y., Guo Y., Bao H., Rao C., 2020, Sensors, 20, 4877

This paper has been typeset from a TEX/LATEX file prepared by the author.