A simulator-based autoencoder for focal plane wavefront sensing

Maxime Quesnel, Gilles Orban de Xivry, Olivier Absil, Gilles Louppe

SPIE.

A Simulator-based Autoencoder
for Focal Plane Wavefront Sensing

Maxime Quesnel[a,b], Gilles Orban de Xivry[b], Olivier Absil[b], and Gilles Louppe[a]

[a]Montefiore Institute of Electrical Engineering and Computer Science, University of Liège
[b]Space sciences, Technologies and Astrophysics Research (STAR) Institute, University of Liège

ABSTRACT

Instrumental aberrations strongly limit high-contrast imaging of exoplanets, especially when they produce quasi-static speckles in the science images. With the help of recent advances in deep learning, we have developed in previous works an approach that applies convolutional neural networks (CNN) to estimate pupil-plane phase aberrations from point spread functions (PSF). In this work we take a step further by incorporating into the deep learning architecture the physical simulation of the optical propagation occurring inside the instrument. This is achieved with an autoencoder architecture, which uses a differentiable optical simulator as the decoder. Because this unsupervised learning approach reconstructs the PSFs, knowing the true phase is not needed to train the models, making it particularly promising for on-sky applications. We show that the performance of our method is almost identical to a standard CNN approach, and that the models are sufficiently stable in terms of training and robustness. We notably illustrate how we can benefit from the simulator-based autoencoder architecture by quickly fine-tuning the models on a single test image, achieving much better performance when the PSFs contain more noise and aberrations. These early results are very promising and future steps have been identified to apply the method on real data.

Keywords: Deep learning, convolutional neural networks, autoencoders, focal plane wavefront sensing, phase retrieval

1. INTRODUCTION

Imaging exoplanets requires instruments to have the capability of probing high contrasts between planets and their host star. Telescopes incorporating a coronagraph and adaptive optics (AO) tremendously help in this matter, but residual wavefront aberrations, in particular non-common path aberrations (NCPA) between the scientific and wavefront sensing arms, still greatly limit the detection of exoplanets.[1] Focal-plane wavefront sensing (FPWFS) methods[2] are particularly beneficial to measure NCPAs since the signature of these aberrations can be found in the focal-plane images.

Deep learning with convolutional neural networks (CNN) has recently been implemented for focal-plane wavefront sensing[3–8] with good results. CNNs offer great performance and robustness as well as fast inference, making them particularly adequate for NCPA correction during observing nights. One limitation of supervised learning using CNNs however is that the models require a ground truth to be trained. As building models with on-sky data is particularly challenging due to the lack of precise knowledge about the NCPAs themselves, it may be difficult to build reliable labelled datasets. Unsupervised learning approaches are thus particularly compelling, since they work without knowing the true phase aberrations. By reconstructing the input data, deep learning autoencoders are a well proven unsupervised approach. For focal-plane wavefront sensing the targets are the NCPAs instead of the point spread functions (PSF), and one way to extract this phase information with an autoencoder is to add physical knowledge to the model.

Several works in the field of optics have incorporated a physical simulator into their deep learning architecture. Bostan et al. 2020[9] and Wang et al. 2020[10] for instance proposed each an unsupervised approach with an untrained neural network that reconstructs images on-the-fly based on an optical model. Peng et al. 2020[11] also

---

Further author information: (Send correspondence to Maxime Quesnel)
E-mail: maxime.quesnel@uliege.be

incorporates an optical wave propagation model in the architecture to improve the quality of holographic images. Automatic differentiation of optical models has also been performed in the context of wavefront aberrations, notably for PSF reconstruction (Liaudat et al. 2022[12]) as well as for phase retrieval and design (Wong et al. 2021[13]).

Motivated by the benefits a differentiable simulator can offer for an unsupervised approach, we define an autoencoder architecture that contains a CNN as the encoder and a optical simulator as the decoder. This allows to constrain the encoder to predict Zernike coefficients corresponding to the probed NCPAs. The method is described in Sec. 2. This unsupervised approach is tested on simulated data, and the results are presented in Sec. 3. We notably evaluate how the method handles photon noise and atmospheric turbulence residuals compared to using only a CNN. The possibility of fitting the models on an observed PSF that lies outside the training data distribution is also investigated. Developments for the approach to work with real instruments are finally discussed in Sec. 4.

2. SIMULATOR-BASED AUTOENCODER

2.1 Data generation

For the present study we consider a simple optical propagation between a pupil plane and a focal plane, so that the PSF is expressed as:

$$PSF(x,y) = |\mathscr{F}[A(x,y)\,e^{\,i\,\Theta(x,y)}]|^2, \tag{1}$$

where $A(x,y)$ is the pupil function and $\Theta(x,y)$ the pupil-plane phase aberrations.

We construct the phase maps using Zernike polynomials following the Noll convention,[14] starting from the tip mode, and based on an annular aperture with a central obstruction of 30% of the total diameter. The sets of Zernike coefficients are randomly generated to approximate a $1/f^2$ power spectral density profile, and scaled to a given median root-mean-square (rms) wavefront error (WFE). In our simulations we consider two aberration regimes: NCPAs around 70 nm rms covering 20 Zernike modes, and 350 nm rms over 100 modes. The PSFs are computed using the PROPER[15] optical propagation package. Our data is generated in the K band at $\lambda = 2200$ nm, while we consider an aperture diameter of 10 m, a pixel scale of $\simeq 11.43$ mas/pix, and a field-of-view of $\simeq 1.47$". The phase maps as well as the PSFs contain 64×64 pixels. In addition to the in-focus PSFs, we also generate out-of-focus PSFs with a defocus of $\lambda/4 = 550$ nm rms, which are used to lift the sign ambiguity that exists for the modes of even radial order, as typically done for focal-plane wavefront sensing.[16,17] We add photon noise to the PSFs so that SNR $= \sqrt{N_{ph}}$, with N_{ph} the number of photons. We also apply a square-root stretching operation to the PSFs, before normalizing the flux in the range [0,1]. This operation is done inside the encoder for the autoencoder architecture.

For the experiments done in Sec. 3.2.4, we also generate data containing atmospheric turbulence residuals. This is achieved by simulating an extreme AO system using the COMPASS library,[18] giving a wavefront error of about 50 nm rms. To simulate a 1-s exposure in presence of a given amount of static NCPA, we use a sequence of 10 consecutive AO phase screens, and we sum up the corresponding PSFs. These AO simulations are described in more details in Quesnel et al. 2022[8] (submitted).

2.2 Autoencoder architecture

In this work we propose an autoencoder architecture, which is composed of an encoder and a decoder. The encoder is a CNN that maps Zernike coefficients from PSFs x, i.e., it approximates a non-linear function f such that $z \approx f(x)$. We choose to use a state-of-the-art deep CNN called EfficientNet.[19] This type of architecture stands out from other ones by using a new scaling technique, with all dimensions of the CNN (depth, width, and resolution) being scaled similarly. We choose to use the EfficientNet-B4 flavour, which offers the best trade-off between performance and runtime.

For our unsupervised approach, we also add a decoder which corresponds to a differential optical simulator, i.e. the decoder reconstructs the input PSFs based on the predicted Zernike coefficients, allowing to work without labels. This decoder is the same as the data generator defined in Eq. 1, with the exception that it produces noiseless PSFs. All the simulator's parameters are here fixed, while it may be beneficial in the future to consider
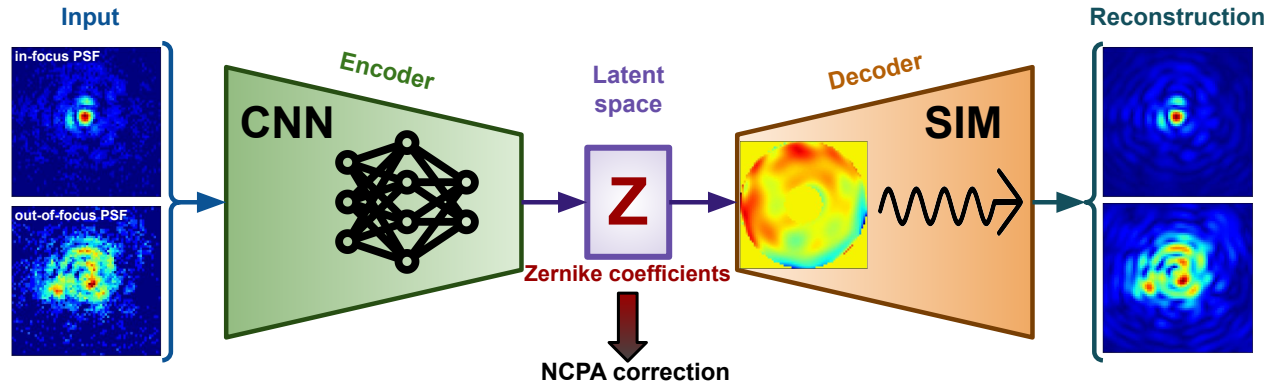
Figure 1: The proposed simulator-based autoencoder architecture (SimAE), using a CNN as the encoder and a differentiable optical simulator as the decoder. The latent space contains the Zernike coefficients which are used during training to reconstruct noiseless PSFs, while during inference these coefficients can be applied for NCPA correction.

learning parameters of the simulator for in-lab and on-sky applications. Regarding the latent space that exists between the encoder and the decoder, the number of Zernike modes has to be predefined. During inference on a trained model, the encoder containing the CNN allows to obtain the Zernike coefficients that correspond to the NCPAs of interest. Fig. 1 summarizes our simulator-based autoencoder architecture, which we call "SimAE" in the following sections.

The loss function, which is used to optimize the neural network's weights in order to fit the model to the data, needs to be adapted to our SimAE architecture. Since the decoder reconstructs noiseless PSFs, the typical mean-square-error loss that makes the assumption of Gaussian noise is not appropriate on the PSF residuals. Since we are working on a simplified case where only photon noise is present in the images, i.e. each pixel follows a Poisson distribution, we can define our loss so that the log-likelihood of the Poisson probability distributions is maximized, giving the expression:

$$\mathcal{L}_{SimAE}(x; \phi) = -\mathbb{E}_{x \sim p(x)} \left[\log \left(\frac{\lambda(x; \phi)^x}{x!} \exp(-\lambda(x; \phi)) \right) \right], \tag{2}$$

where $\lambda(x; \phi)$ are the rates of the Poisson distributions, which depend on both the input PSFs x sampled from the dataset distribution $p(x)$, and the parameters ϕ of the encoder (CNN). These rates correspond to the noiseless PSFs that are constructed by the decoder.

The baseline on which we compare our SimAE models is a standard CNN architecture. The loss function is in this case defined as the root-mean-square error of the phase residuals. It can be expressed using the Zernike coefficients as:

$$\mathcal{L}_{CNN}(z, \widehat{z}(x; \phi)) = \sqrt{\frac{1}{N} \sum_{i}^{N} (z_i - \widehat{z}_i(x; \phi))^2}, \tag{3}$$

where z and $\widehat{z}$ correspond to the true and predicted Zernike coefficients respectively, while N is the number of Zernike modes (20 or 100).

3. EXPERIMENTS

3.1 Protocol

For a given training, a dataset composed of 10^5 PSF pairs is randomly split into training (90%) and validation (10%) sets, with a batch size of 64. For the supervised learning approach with a CNN, each sample also contains
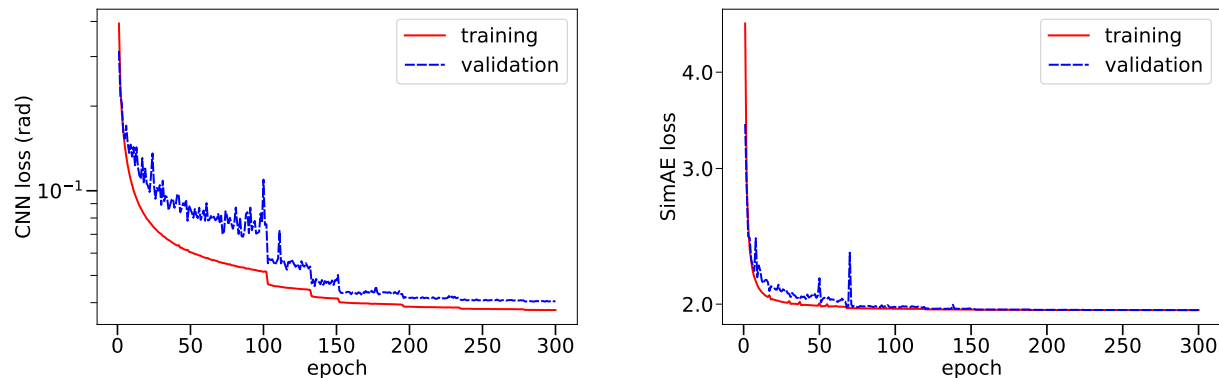
Figure 2: Examples of training and validation losses over the epochs for each approach: CNN (**left**) and SimAE (**right**). In these examples the models are trained on datasets containing 350 nm rms of aberrations distributed over 100 modes, and a SNR equal to 300.

the true NCPA phase maps as labels. The Adam optimizer[20] with a starting learning rate of 10^{-3} is used to update the weights of the CNN. The learning rate is decreased by a factor two as soon as the validation loss reaches a plateau over 15 epochs to improve the performance. Efficientnet-b4 weights are initialized with models pretrained on ImageNet. An example of the resulting training and validation losses for each method can be found in Fig. 2. The models are then systematically evaluated on 1000 new samples.

3.2 Results

3.2.1 Performance compared to a standard deep CNN

We first compare the performance regarding each Zernike mode. The metric used is the root-mean-square error (RMSE) per mode, computed over the entire test set:

$$\sigma_z = \sqrt{\frac{1}{N_{\text{test}}} \sum_i^{N_{\text{test}}} (\hat{c}_i - c_i)^2} , \qquad (4)$$

where N_{test} is the number of test samples, while $\hat{c}$ and c are the estimated and true Zernike coefficients, respectively.

As shown in Fig. 3, the Zernike coefficients are very well reconstructed, even at high aberrations contents (350 nm rms over 100 modes). There is no particular distinction between even and odd modes, and the SimAE and CNN manifest very similar performance.

Since the SimAE is defined to take into account photon noise through its loss function, it is worth assessing the performance for different SNR levels. We thus compare the residual errors in terms of rms WFE on the total phase residuals. The metric is therefore defined for each test sample as:

$$\sigma_\phi = \sqrt{\frac{1}{N_{\text{pix}}} \sum_i^{N_{\text{pix}}} (\hat{\phi}_i - \phi_i)^2}, \qquad (5)$$

where N_{pix} is the number of pixels within the pupil area, while $\hat{\phi}$ and ϕ are the estimated and true pupil phases, respectively.

The errors on the phase retrieval for a range of SNR levels are shown in Fig. 4, where each point corresponds to a model trained and evaluated at the same given SNR. The two approaches show almost identical performance, and a plateau is reached around a SNR of 3000 in the high aberration regime. This has been observed and discussed in previous works[7,8] as well. Only at very low SNRs the SimAE appear to differ from the CNN approach. This difference could be due to the loss function of the SimAE that is defined on the PSFs and not
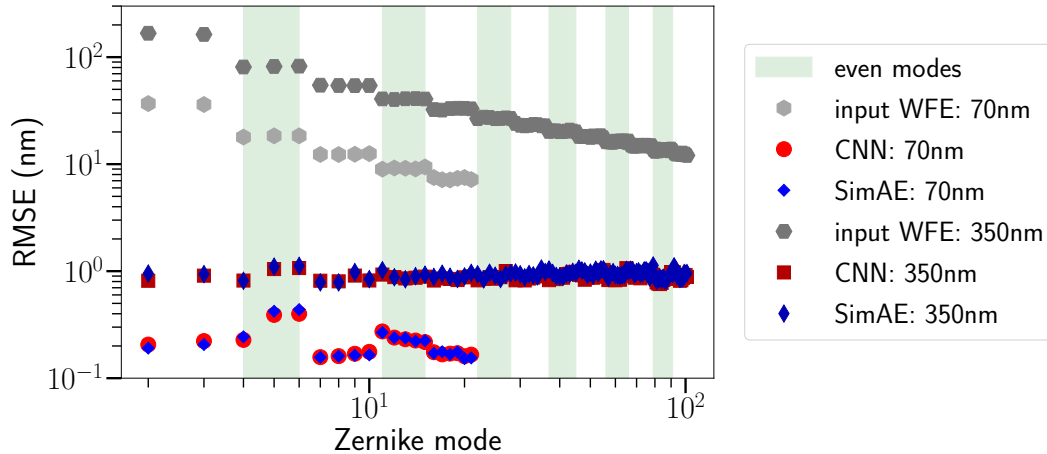
Figure 3: Performance per Zernike mode, for a low aberration regime (70 nm over 20 modes, light shades) and a high aberration regime (350 nm over 100 modes, darker shades), at a SNR of 1000. The CNN and SimAE architectures are compared (blue and red, respectively).
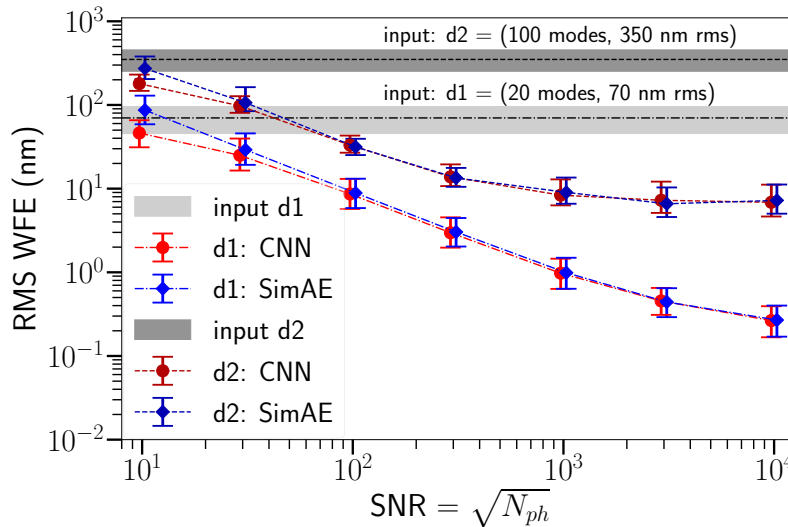

Figure 4: Performance at different SNR levels for the SimAE (blue) and the standard CNN (red) models, at low and high aberration regimes (darker and brighter shades respectively). The points correspond to the median value and the error bars to the 2-98th percentiles.

on the NCPAs, so that there is not the implicit regularization occurring with supersized CNNs, which has been described by Orban de Xivry et al. 2021.[7] The residual errors thus tend to follow the theoretical limit for lower SNRs with the SimAE.

3.2.2 Robustness of the trained models

Deep learning models may show limitations when evaluated on data lying outside the training data distribution. It is therefore important to assess the stability of the models in such conditions, which would probably occur when applying the algorithm on sky.

Here we compare again the SimAE method with the CNN, this time varying the testing SNR (Fig. 5, left), on a model trained at a median WFE of 70 nm rms distributed over 20 modes. Giving data containing less noise to the models always provides stable predictions, and we even obtain a marginal improvement with the SimAE when trained at a SNR of 30. If we introduce more noise to the test data, the error quickly increases, at about the same rate with both approaches. We also experiment by changing the testing input WFE in the data
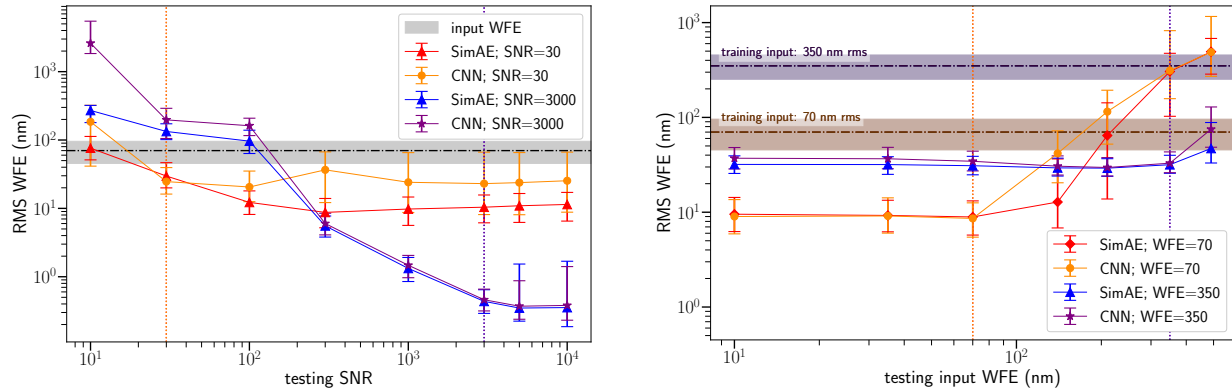
Figure 5: **Left:** Performance with changing the SNR of the test data for two models trained on SNRs of 30 and 300 (indicated by the vertical dotted lines). The median input WFE is 70 nm with 20 modes. **Right:** Performance for different testing input WFE, on models trained at 70 nm and 350 nms (20 and 100 modes respectively).
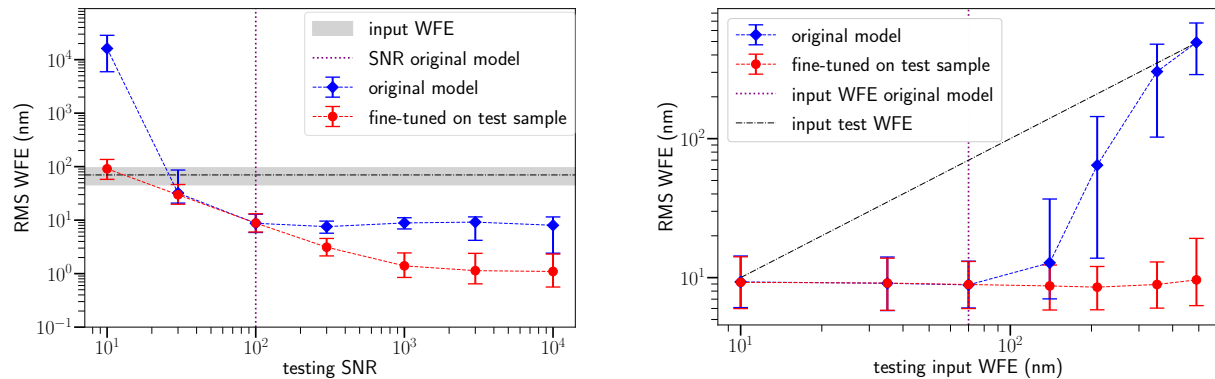

Figure 6: Performance of models fine-tuned on each test sample individually, for varying SNR (**left**) and input WFE levels (**right**). The model used for both plots was initially trained on data with 70 nm rms, 20 modes, and a SNR of 100, as indicated by the vertical dotted lines.

(Fig. 5, right). Here the CNN and SimAE showcase identical behaviour, with constant performance for lower testing WFE than the training one, and strong degradation for higher testing WFE, although with the SimAE the performance does not degrade as quickly as with the CNN. Similar trends regarding the CNN has been found in previous works.[7,8] Overall the SimAE method offers only slight improvement in terms of robustness compared to the CNN. To benefit the most from such a physics-based approach, it would be useful to include more knowledge about the data in the models, which could be provided by a variational autoencoder[21] approach for instance.

3.2.3 Transfer learning

Instead of directly applying a trained model on observed data, we propose to fine-tune the model using a single "on-sky" image, initializing with the trained weights of the original model. This is done to illustrate how we could quickly adapt the model from simulated to on-sky data, which may slightly differ from the training data. To test this on our SimAE method, we use a model trained on data containing 70 nm rms and a SNR of 100. The levels of NCPA and SNR are changed on the test data, and the results are shown in Fig. 6. The models perform much better when transfer learning is done on each test sample, in particular when their SNR and WFE are higher than for the original training data. This second training on a single test sample is performed over 200 epochs, which takes about 20 seconds to complete on one GPU (Nvidia GeForce RTX 2080 Ti). For data relatively close to the ones used for pre-training, convergence is naturally achieved after much less epochs and the final prediction can be made in only a few seconds.
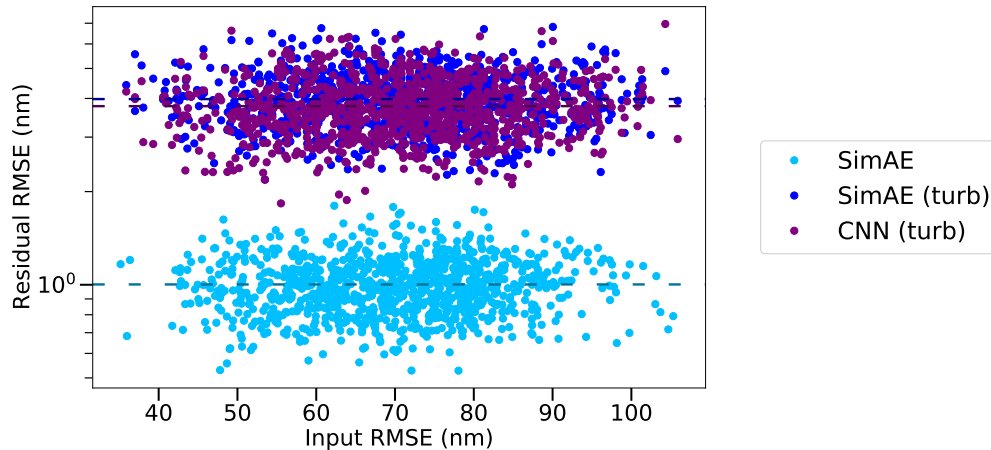
Figure 7: Residual errors as a function of the input WFE, for a model trained on a dataset with WFEs distributed around 70 nm (20 modes). A SimAE model with data containing only photon noise (cyan), is compared to CNN and SimAE models (purple and blue, respectively) trained on data containing also AO residuals, which are not provided to the decoder during training and evaluation.

3.2.4 Atmospheric turbulence residuals

In the previous experiments, the decoder, i.e the differentiable optical simulator, was able to exactly reconstruct what our data generator produced to build our training and test sets, except for photon noise, but for which the loss function was adapted. In order to test the method on data the decoder cannot totally reconstruct, we introduce atmospheric turbulence residuals to the input data, without giving this information to the decoder. The way these AO residuals are added is described in Sec. 2.1. Even with the presence of this additional source of noise, the training is very stable, and the evaluation results shown in Fig. 7 indicate that the SimAE suffers a reasonable degradation in performance compared to a case without turbulence. Using a CNN with labels that does not account for these AO residuals gives the same results.

4. CONCLUSIONS AND FUTURE WORK

We have developed an unsupervised deep learning method for focal-plane wavefront sensing by incorporating an optical simulator into a CNN-based autoencoder architecture (SimAE). We observe very similar performance compared to using a simple CNN, both approaches reaching the expected theoretical limit over a large range of conditions. The SimAE however has the major advantage of not requiring labels during the training of the models. One promising opportunity with an approach like this SimAE is that we can quickly fit the model to the currently observed PSF using a pre-trained model. We have shown in this paper that such fine-tuning of the models can greatly improve the results for data showing different noise or NCPA levels. Dealing with data the decoder cannot reproduce is naturally a potential limiting factor for the method. We have trained models with atmospheric turbulence residuals present on top of NCPAs, and the loss of performance that we obtain stays well constrained. In practice these turbulence residuals may even be given to the decoder, by using the wavefront sensor telemetry to learn what the AO residuals effectively are, with the help of a small neural network for example.

Several components of our architecture can be improved in the future. We may first benefit from variational inference, i.e. predicting posterior distribution for Zernike coefficients as well as to incorporate prior knowledge about them. This could yield better robustness for the models thanks to the additional information given during training and inference. Regarding the decoder, a more complete simulator, e.g. including a coronagraph for high-contrast imaging and other relevant instrumental components, will be essential to envision applying the technique on-sky with a real instrument.

ACKNOWLEDGMENTS

REFERENCES

[1] Guyon, O., "Extreme Adaptive Optics," **56**, 315–355 (Sept. 2018).

[2] Jovanovic, N., Absil, O., Baudoz, P., Beaulieu, M., Bottom, M., Cady, E., Carlomagno, B., Carlotti, A., Doelman, D., Fogarty, K., Galicher, R., Guyon, O., Haffert, S., Huby, E., Jewell, J., Keller, C., Kenworthy, M. A., Knight, J., Kühn, J., Miller, K., Mazoyer, J., N'Diaye, M., Por, E., Pueyo, L., Riggs, A. J. E., Ruane, G., Sirbu, D., Snik, F., Wallace, J. K., Wilby, M., and Ygouf, M., "Review of high-contrast imaging systems for current and future ground-based and space-based telescopes: Part II. Common path wavefront sensing/control and coherent differential imaging," in [*Adaptive Optics Systems VI*], Close, L. M., Schreiber, L., and Schmidt, D., eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **10703**, 107031U (July 2018).

[3] Paine, S. W. and Fienup, J. R., "Machine learning for improved image-based wavefront sensing," *Optics Letters* **43**, 1235 (Mar. 2018).

[4] Andersen, T., Owner-Petersen, M., and Enmark, A., "Neural networks for image-based wavefront sensing for astronomy," *Optics Letters* **44**, 4618 (Sept. 2019).

[5] Andersen, T., Owner-Petersen, M., and Enmark, A., "Image-based wavefront sensing for astronomy using neural networks," *Journal of Astronomical Telescopes, Instruments, and Systems* **6**, 034002 (July 2020).

[6] Quesnel, M., Orban de Xivry, G., Louppe, G., and Absil, O., "Deep learning-based focal plane wavefront sensing for classical and coronagraphic imaging," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **11448**, 114481G (Dec. 2020).

[7] Orban de Xivry, G., Quesnel, M., Vanberg, P. O., Absil, O., and Louppe, G., "Focal plane wavefront sensing using machine learning: performance of convolutional neural networks compared to fundamental limits," **505**, 5702–5713 (Aug. 2021).

[8] Quesnel, M., Orban de Xivry, G., Absil, O., and Louppe, G., "A deep learning approach for focal-plane wavefront sensing using vortex phase diversity," (submitted, 2022).

[9] Bostan, E., Heckel, R., Chen, M., Kellman, M., and Waller, L., "Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network," *Optica* **7**, 559–562 (Jun 2020).

[10] Wang, F., Bian, Y., Haichao, W., Lyu, M., Pedrini, G., Osten, W., Barbastathis, G., and Situ, G., "Phase imaging with an untrained neural network," *Light: Science Applications* **9**, 77 (05 2020).

[11] Peng, Y., Choi, S., Padmanaban, N., and Wetzstein, G., "Neural holography with camera-in-the-loop training," *ACM Trans. Graph.* **39** (nov 2020).

[12] Liaudat, T., Starck, J.-L., Kilbinger, M., and Frugier, P.-A., "Rethinking data-driven point spread function modeling with a differentiable optical model," (2022).

[13] Wong, A., Pope, B., Desdoigts, L., Tuthill, P., Norris, B., and Betters, C., "Phase retrieval and design with automatic differentiation: tutorial," *Journal of the Optical Society of America B* **38**, 2465 (aug 2021).

[14] Noll, R. J., "Zernike polynomials and atmospheric turbulence.," *Journal of the Optical Society of America (1917-1983)* **66**, 207–211 (Mar. 1976).

[15] Krist, J. E., "PROPER: an optical propagation library for IDL," in [*Optical Modeling and Performance Predictions III*], Kahan, M. A., ed., **6675**, 250 – 258, International Society for Optics and Photonics, SPIE (2007).

[16] Gonsalves, R. A., "Phase Retrieval And Diversity In Adaptive Optics," *Optical Engineering* **21**(5), 829 – 832 (1982).

[17] Vievard, S., Bos, S. P., Cassaing, F., Ceau, A., Guyon, O., Jovanovic, N., Keller, C., Lozi, J., Martinache, F., Mary, D., Montmerle-Bonnefois, A., Mugnier, L., N'diaye, M., Norris, B., Sahoo, A., Sauvage, J.-F., Snik, F., Wilby, M. J., and Wong, A., "Overview of focal plane wavefront sensors to correct for the Low Wind Effect on SUBARU/SCExAO," in [*6th International Conference on Adaptive Optics for Extremely Large Telescopes, AO4ELT 2019*], (June 2019).

[18] Ferreira, F., Gratadour, D., Sevin, A., and Doucet, N., "Compass: An efficient gpu-based simulation software for adaptive optics systems," *2018 International Conference on High Performance Computing & Simulation (HPCS)* , 180–187 (2018).

[19] Tan, M. and Le, Q., "EfficientNet: Rethinking model scaling for convolutional neural networks," in [*Proceedings of the 36th International Conference on Machine Learning*], Chaudhuri, K. and Salakhutdinov, R., eds., *Proceedings of Machine Learning Research* **97**, 6105–6114, PMLR (09–15 Jun 2019).

[20] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," (2017).

[21] Kingma, D. P. and Welling, M., "Auto-encoding variational bayes," (2013).

[22] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S., "Pytorch: An imperative style, high-performance deep learning library," in [*Advances in Neural Information Processing Systems*], Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., eds., **32**, 8026–8037, Curran Associates, Inc. (2019).

[23] Klaus Greff, Aaron Klein, Martin Chovanec, Frank Hutter, and Jürgen Schmidhuber, "The Sacred Infrastructure for Computational Research," in [*Proceedings of the 16th Python in Science Conference*], Katy Huff, David Lippa, Dillon Niederhut, and Pacer, M., eds., 49 – 56 (2017).