# Image-based wavefront sensing for astronomy using neural networks

**Torben Andersen,**[a,*] **Mette Owner-Petersen,**[a] **and Anita Enmark**[b]
[a]Lund Observatory, Lund, Sweden
[b]Luleå University of Technology, Department of Computer Science,
Electrical and Space Engineering, Kiruna, Sweden

**Abstract.** Motivated by the potential of nondiffraction limited, real-time computational image sharpening with neural networks in astronomical telescopes, we studied wavefront sensing with convolutional neural networks based on a pair of in-focus and out-of-focus point spread functions. By simulation, we generated a large dataset for training and validation of neural networks and trained several networks to estimate Zernike polynomial approximations for the incoming wavefront. We included the effect of noise, guide star magnitude, blurring by wide-band imaging, and bit depth. We conclude that the "ResNet" works well for our purpose, with a wavefront RMS error of 130 nm for $r_0 = 0.3$ m, guide star magnitudes 4 to 8, and inference time of 8 ms. It can also be applied for closed-loop operation in an adaptive optics system. We also studied the possible use of a Kalman filter or a recurrent neural network and found that they were not beneficial to the performance of our wavefront sensor. [DOI: 10.1117/1.JATIS.6.3.034002]

**Keywords:** optics; astronomy; telescope; wavefront sensor; neural network; image sharpening.

Paper 20014 received Feb. 6, 2020; accepted for publication Jun. 25, 2020; published online Jul. 10, 2020.

## 1 Introduction

Unaided, astronomical telescopes generally have a resolution that is limited by the atmosphere. The theoretical resolution of a 4-m optical telescope at 545 nm is 0.034 arcsec, but due to atmospheric blurring, the real resolution is often in the range 0.5 to 1 arcsec. To correct for atmospheric seeing, adaptive optics with a wavefront sensor and a deformable mirror are often used. Such systems are expensive and complex to operate.

With the advent of powerful computers and fast readout, science-grade CMOS sensors, it has been speculated that efficient real-time image sharpening can be achieved in software, without a traditional wavefront sensor and a deformable mirror. Recent developments in artificial intelligence, and in particular in the field of convolutional neural networks, could pave the way for implementation of such systems.

The field of computational image sharpening by postprocessing was developed in detail 20 to 30 years ago and is well covered in the literature.[1,2] Some more recent publications deal with the same aspects of phase retrieval.[3–6] They mostly focus on determination of static or quasistatic aberrations that are in noncommon mode with an adaptive or active optics system. Due to computational limitations, it has been difficult until recently to perform image sharpening in real time. However, with the advent of affordable faster computers and the development in neural network technology, it now seems feasible to apply neural networks for affordable, fast computational image-sharpening.

Until recently, neural networks have found only limited use for astronomical image sharpening. In 1990, Angel et al.[7] published important work in the field. Most of the studies related to the use of neural networks for astronomy have focused on neural networks in tandem with a traditional wavefront sensor, for instance, for tomography, mapping the influence of the atmosphere on the light reaching the telescope over a field larger than the isoplanatic angle.[8]

*Address all correspondence to Torben Andersen, E-mail: torben.andersen@astro.lu.se

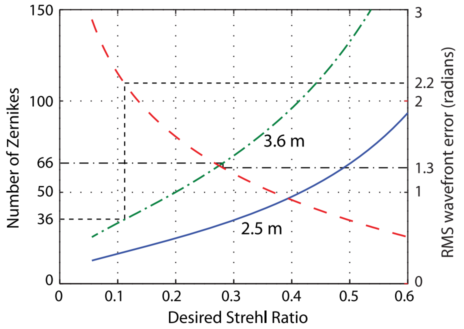

**Fig. 1** Number of Zernike terms (dash-dot green and solid blue) and RMS wavefront error (dashed red) at the entrance pupil needed for a given Strehl ratio for two different primary mirror diameters and Fried's parameter $r_0 = 0.17$ m.

There has been some limited work related to image-based wavefront sensing.[9–12] A recent paper[13] deals with some of the same topics as the present publication but focuses on laser communication with monochromatic light and does not touch upon the issues of phase screen generation, noise, blurring by wide-band imaging, closed-loop operation, and bit depth, which all are important for our application. Another recent article[14] briefly describes an interesting image sharpening approach using a neural network based on unsupervised learning.

As a beginning, we studied the possibility of image-based wavefront sensing. Telescopes above 4 m almost all have adaptive optics, so our initial range of interest was for telescopes in the 3- to 4-m class in the V-band. Narrow-band image-based wavefront sensing is easier than wide-band sensing, but for astronomy, wide-band operation is attractive for reasons of light economy even though an atmospheric dispersion compensator may be needed. With the advent of even faster computers and, in particular, faster graphical processor units (GPUs), methods could be applicable also for larger telescopes in the near future, not the least in the near-IR, which is less demanding than the V-band. Although our motivation is primarily related to real-time image sharpening, we also looked into the possibility of using image-based neural network wavefront sensors in a conventional adaptive or active optics system.

We studied whether neural networks can estimate the incoming wavefront from a natural guide star on the basis of in-focus and out-of-focus point spread functions (PSFs). To characterize the wavefront, we use Zernike polynomials. This is beneficial for subsequent wavefront-assisted deconvolution because a finite Zernike representation has few high-frequency components. Hence, our aim is to design a modal wavefront sensor, based upon a neural network, that takes a PSF-pair as input and gives a vector of Zernike coefficients as output. Neural networks need to be trained on a large amount of realistic data. Atmospheric optics is a well-developed discipline, and it is relatively easy (albeit computationally intensive) to generate a large number of simulated phase screens for the incoming light to the telescope after passing through the atmosphere. We generated the corresponding in-focus and out-of-focus PSFs for training of neural networks and decomposed the phase screens into Zernike terms.

Using Maréchals equation for the relation between the Strehl ratio and wavefront error[15] and Noll's approach for the relation between the number of Zernikes included and the wavefront error,[16] the number of Zernikes needed to achieve a certain image quality can be estimated, as shown in Fig. 1, for two different telescope diameters and for Fried's parameter $r_0 = 0.17$ m, representing typical seeing conditions at a good site. For a 3.6-m telescope, using between 36 and 66 Zernikes (up to radial degree 8 to 11) is a good choice.

Later, we aim to create a system using more than one guide star to get a larger field than the isoplanatic patch. For such a system, a CMOS image sensor with 2k × 2k pixels, a frame rate of 50 fps, and 12- or 16-bit resolution is commercially available with science-grade specifications, so we selected that frame rate for our study. With typical atmospheric layer characteristics, this leads to temporal undersampling, so higher frame rates would be beneficial when working only with bright guide stars.

A first, brief report was given in a previous article.[17] Here, we go into more detail to demonstrate the possibilities and limitations of the use of neural networks for image-based wavefront sensing. We first describe the neural networks selected for our studies and how we generated the training and validation data by simulation. Next, we present our simulated tests of the neural networks for both timewise uncorrelated input images and for time series of correlated input images. Finally, conclusions are given.

## 2 Method

### 2.1 *Neural Networks*

Neural networks have been dealt with extensively in recent literature.[18–20] Because we use PSFs as inputs, we apply "convolutional" neural networks that can handle image information with a reasonable amount of network parameters. Our networks have two images as inputs and a vector of Zernike coefficients as outputs. We have a "regression" problem because the networks are estimating numbers, in contrast to "classification" networks, for instance, looking for cars or persons.

We do not show network diagrams here but refer to the references for the networks selected. We first tested a modest neural network similar to the VGG16 network from 2014,[21] but it turned out to be insufficient for our purpose, so we turned to deeper networks, the Inception v. 3[22] from 2014 and the ResNet[23] from 2015. Both networks were originally designed for three-channel input (RGB images), but we modified them to two-channel networks for our application. We also removed the top classification layers and added instead one or two linear, dense layers on top for regression. The ResNet network can be set up with various depths (i.e., number of layers), generally in the range 20 to 164, although even a network with as many as 1001 layers has been described. As a compromise between computation time and accuracy, we opted for 50 layers (ResNet50).

It is often difficult to train deep networks because of the "vanishing gradient problem" in which parameters of the first layers become practically unreachable when training the network using backpropagation. However, both the Inception and ResNet have approaches for overcoming the vanishing gradient problem.

Recently, more advanced networks, such as the SENET[24] and NASNET,[25] have been made available by their developers. These networks are not dramatically more accurate than the Inception and ResNet, but they are lighter and therefore particularly suited for mobile applications. So far, estimation (inference) time has not been critical for our application, so we decided initially not to go through the added complication of implementing the new, advanced networks in our study. They may, however, be considered at a later stage.

The Inception and ResNet neural networks have been trained with a huge amount of data by their developers, and the network parameters are publicly available. For some applications, it has been useful to reuse the pretrained networks with only minor additions. For our application, this turned out not to be beneficial. We believe that our input data differ too much from those originally used for training of the networks (such as images of cats and dogs), and we have only two input channels, whereas the pretrained networks have three channels.

As a metric for the success of the Zernike coefficient estimates, we use the RMS of the difference between the original wavefront and the estimated wavefront. This is the same as the residual wavefront RMS after correction by an "ideal" deformable mirror. We only include a limited number of Zernikes, so the residual wavefront error stems both from estimation errors, for instance, due to noise, and from omission of high order Zernikes (fitting error).

We applied the "Adam" solver[26] to determine network parameters by backpropagation. We tested other solvers, but Adam turned out to be more precise and faster, as reported by many authors. For training using backpropagation, a loss function must be defined; it is minimized by the solver while iterating for network parameters. We used mean square errors of the estimated Zernike coefficients as loss function, which corresponds to minimizing the RMS of the residual wavefront for the Zernikes included. All Zernikes were weighted equally because they play an equal role for the quality of wavefront correction. They were all normalized to provide a wavefront RMS of 1, so each Zernike coefficient was equal to the RMS of the wavefront of that Zernike mode (see Sec. 2.2). Also, during the training process, we scaled the input images

**Table 1** Baseline parameters for neural network training.

| Primary mirror diameter (m) | 3.6 |
|---|---|
| Wavelength range | V-band |
| Star magnitudes | 4 to 14 |
| PSF bit depth | 16 |
| Zernikes estimated | 2 to 36 |
| PSF size (pixels$^2$) | $257 \times 257$ |
| PSF field (arsec$^2$) | $3.46 \times 3.46$ |
| Oversampling[a] | 1.22 |
| Data sets for training | 200 k |
| Data sets for validation | 80 k |
| PSF defocus[b] | $4.5 \times \lambda_c$ at edge of pupil |

[a] With respect to Nyquist sampling of the diffraction limited PSF.
[b] $\lambda_c = 545$ nm is the center wavelength of the V-band.

to have pixel values in the range 0 to 1, as is general practice, and we used dropout to reduce overfitting.

In most cases, we trained and validated the neural networks using the baseline case shown in Table 1, with one collection of data sets for training and another for validation. The input to the neural network is a pair of in-focus and out-of-focus PSFs, and the output is a vector of Zernike coefficients. We chose a relatively large collection of data sets for validation, thereby reducing unintentional "information leakage" from the validation data to the neural network parameters. The Inception model had 19.8 million trainable parameters, and training typically required eight to nine epochs (i.e., the network was presented to the same training data eight to nine times). It took a total of about 20 h to train the network on an Intel i7-8750H CPU with an NVIDIA GeForce GTX1070 GPU. The ResNet50 network for 35 Zernike coefficients had 29.3 million trainable parameters, and we generally ran 12 to 14 epochs with a total training time for each case of about 32 h. In total, we generated ~11 million files for training and validation, and we spent over 3000 CPU-hours on data set generation and network training, both only related to this publication. We used Python with Tensorflow and Keras libraries for the neural network part and for many calculations and plots.

## 2.2 *Training and Validation Data*

For training and validation, a large number of representative data sets of simulated PSF-pairs and Zernike coefficient vectors are needed. To find these, we followed established procedures.[15] For each data set, we first determined a phase screen at the entrance pupil of the telescope. The atmosphere was modeled with the "seven-layer model" with Fried's parameter and coherence time values of

$$r_0 = \{0.28, 0.30, 0.79, 1.17, 1.30, 1.19, 3.41\}\ \text{m},$$

$$\tau_0 = \{0.005, 0.0076, 0.0093, 0.0206, 0.0505, 0.0137, 0.0456\}\ \text{s},$$

at 545 nm for the seven layers and assuming a Kolmogorov atmosphere model.

We used two different approaches for generation of phase screens. For studies in which time correlation was not a concern, we combined the seven layers into one with $r_0 = 0.17$ m and used the fast "power spectrum method." The spatial phase power spectrum is

$$\Phi_{\varphi}(f) = 0.0229 r_0^{-5/3}(f^2)^{-11/6}, \tag{1}$$

where $f$ is a spatial frequency. Fried's parameter, $r_0$, scales with the wavelength raised to $6/5$. A phase screen sample can then be generated by

$$\varphi(\vec{r}) = \mathcal{F}^{-1}\left[\sqrt{\Phi_{\varphi}(\vec{f})}H(\vec{f})\right], \tag{2}$$

where $\vec{f}$ is a spatial frequency vector, $\vec{r}$ is a vector in the phase screen plane, $\mathcal{F}^{-1}$ is the inverse Fourier transform, and $H(\vec{f})$ is a complex filter function with unit amplitude and a random phase with uniform distribution between 0 and $2\pi$. The lowest frequency included in our phase screen was the reciprocal of the primary mirror diameter. We accepted this limitation for calculation speed and because neural networks are quite good at estimating low-frequency components.

For the case in which PSF time correlation was needed for the neural network study, we generated the phase screens using the slower "Skylight" method, also described in Ref. 15. For each of the layers, the phase screen was simply computed as a sum of two-dimensional harmonic functions with random directions and phases:

$$\varphi(x, y) = \sum_{j=1}^{n_\mathrm{f}} A(f_j)\cos[2\pi f_j(x \cos \alpha_j + y \sin \alpha_j) + \psi_j], \tag{3}$$

where $n_\mathrm{f}$ is the number of absolute frequency samples, $\alpha_j$ is a random angle, $x$ and $y$ are coordinates in the phase screen, and $\psi_j$ is a random phase angle. We used $n_\mathrm{f} = 500$ and we sampled the absolute frequencies logarithmically. The values for coefficients $A(f_j)$ depend on $r_0$ and are given in Ref. 15. We used a Taylor approximation and computed "frozen" phase screens for all seven atmospheric layers, and before adding all seven phase screens, we moved them with $x$- and $y$-velocities given by the following two vectors for $x$ and $y$ directions: $v_x = \{12.05, 8.6, -18.6, -12.4, 8{,}25, 23.2\}$ m/s and $v_y = \{12.05, -8.6, 18.6, 12.4, 0{,}10, 0\}$ m/s. The Skylight method can be set up to include time components of arbitrarily low frequency corresponding to large spatial wavelengths.

We compared phase screens found using the power spectrum and the Skylight methods by numerically computing the phase structure function. The structure function is a metric for correlation between different points in the phase screen as a function of the distance, $r_s$, between them.[15] We computed the structure functions numerically by randomly choosing one million start and end points and averaging the squares of the phase differences for 100 phase screens as a function of the separation of the two points. The result is shown in Fig. 2. The figure also shows the theoretical structure function ($\Phi = 6.88(r_s/r_0)^{5/3}$) for a Kolmogorov phase screen. As expected, the "power spectrum method" phase screens drop off at low spatial frequencies compared with the two other curves.

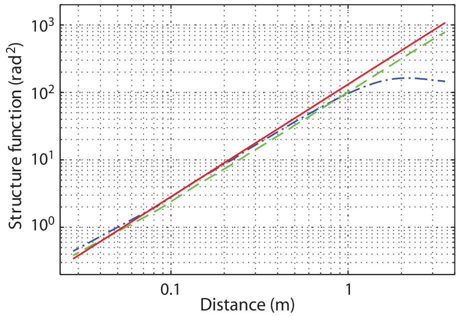

**Fig. 2** Mean phase structure functions for 100 phase screens derived with each of the power spectrum (dash-dot blue) and Skylight (dashed green) methods, together with the theoretical structure function (solid red) for a Kolmogorov atmosphere. $r_0 = 0.17$ m at 545 nm.

Once a phase screen, $\Phi(x, y)$, was determined, the PSFs were found by first Fourier transforming $\exp[i\Phi(x, y)]$ masked by the entrance pupil of the telescope and zero padded. Then, the resulting complex amplitudes were numerically squared to get the intensity distribution in the focal planes.[15] In addition to the in-focus PSF, we determined the out-of-focus PSF by adding a paraboloidal phase over the entrance pupil before determination of the out-of-focus PSF. The pupil mask was circular, and for simplicity we disregarded any central obstruction in the telescope and spider shadows.

The above considerations were applicable for a single wavelength. However, to make the study more realistic, we covered the entire V-band (470 to 700 nm) by the resampling and weighted addition of individual PSFs with wavelength increments of 10 nm in the V-band with the weights given in the reference. As an approximation, we assumed the guide star spectrum to be flat over the V-band. The addition of the many PSFs led to a smearing of the PSFs. The center wavelength is $\lambda_c = 545$ nm. We then computed the theoretical number of photons in each pixel by a radiometric calculation, taking both typical atmospheric and telescope losses into account. The Zenith angle was assumed to be 30 deg. Subsequently, we took photon noise of each pixel into account by drawing an integer number from a Poisson distribution based upon the theoretical number of incident photons in each pixel. Finally, for the PSFs, we included the effect of discretization of the A/D readout, assuming 16-bit conversion for the cases presented here. Figure 3 shows three different pairs of PSFs, with a logarithmic scale to better show faint details.

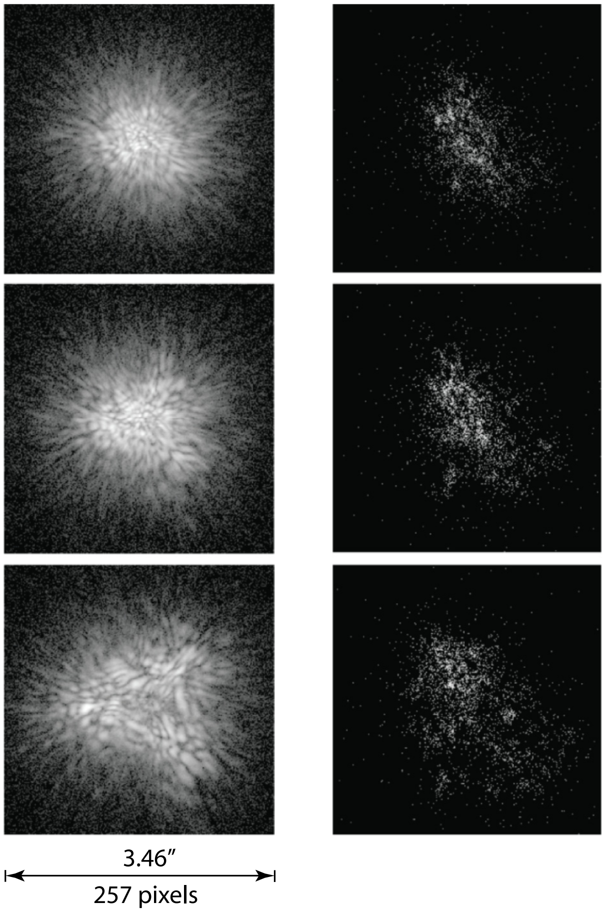

**Fig. 3** Representative PSFs for the V-band. The left column is for a magnitude 8 guide star and the right column for a magnitude 14 star with a different seeing realization. The top row shows the in-focus PSFs, the middle row PSFs with a defocus of $1.5 \times \lambda_c$, and the bottom row PSFs with a defocus of $4.5 \times \lambda_c$, where $\lambda_c$ is the center wavelength of the V-band (545 nm). The defocus is measured as a wavefront aberration at the edge of the entrance pupil. All PSFs were logarithmically scaled for clarity.

To find the Zernike coefficients for a data set, we first sampled the Zernike polynomials for a circular aperture[15] following the numbering scheme of Noll.[16] We disregarded the first Zernike term, which gives a phase piston of no consequence for image quality. Zernike polynomials are mutually orthogonal, but the spatially sampled Zernikes are not, so we modified the Zernikes slightly by Gram–Schmidt orthogonalization[15] to ensure orthogonality. We then found the Zernike coefficients by least-squares fitting to the incident wavefront:

$$\mathbf{w} = \sum_{i=1}^{n_z} a_i \mathbf{v}_i + \mathbf{r}, \tag{4}$$

where $\mathbf{w}$ is the total wavefront arranged as a vector, $\mathbf{v}_i$ is a normalized Zernike vector, $a_i$ is a Zernike coefficient, $\mathbf{r}$ is a vector holding residuals, and $n_z$ is the number of Zernikes included in the approximation.

With the power spectrum method running Matlab on an Intel i7-8750H CPU, it took about 26 h to generate a total of 280,000 data sets for training and validation, and with the Skylight method, it took about 8 days. The Skylight method is inherently slow and, in addition, our Skylight coding is not yet fully optimized for speed.

Scintillation effects were not included in our models. It would be difficult to do from the point of view of computation time, so the influence of scintillation is more efficiently studied by directly recording PSF time series with an actual telescope. We hope to do so in the future.

## 3 Single-Shot Wavefront Sensing

For basic studies of neural networks for wavefront sensing, we first focus on wavefront sensors operating in "single-shot" mode, with no memory of previous wavefronts and estimates, using training and validation data sets that are uncorrelated over time. In Sec. 4, we return to the more complex issue of multishot operation, dealing with time series of wavefronts that are mutually correlated over time.

### 3.1 *Inception*

We first trained two different versions of the Inception v.3 network using the baseline parameters of Table 1. The first version included two "auxiliary" Zernike outputs from intermediate layers as suggested by the original network designers.[22] The purpose of the auxiliary outputs is to force the training algorithm to pay attention to the lowest layers during the training process. We gave the auxiliary outputs weight factors of 0.3 during network training, whereas the weight factors of the top layer Zernike outputs were 1. The residual wavefront RMS is shown for network NN1 in Table 2.

The second version studied did not have auxiliary outputs from the intermediate layers. The result is shown as NN2 in Table 2. There is not a big difference between NN1 and NN2, and NN2 is, in fact, slightly better than NN1, so for our application there is no gain in training with auxiliary outputs.

**Table 2** RMS of wavefront residuals for different neural networks.

| | Residual RMS (nm) | Defocus factor | Network | Note |
|---|---|---|---|---|
| NN1 | 196 | 4.5 | Inception v. 3 | Aux. outputs |
| NN2 | 185 | 4.5 | Inception v. 3 | — |
| NN3 | 182 | 4.5 | ResNet50 | Baseline |
| NN4 | 211 | 1.5 | ResNet50 | Two out-of-focus PSFs |
| NN5 | 211 | 1.5 | ResNet50 | Small defocus |
| NN6 | 178 | 4.5 | ResNet50 | Zernikes 2 to 66 |

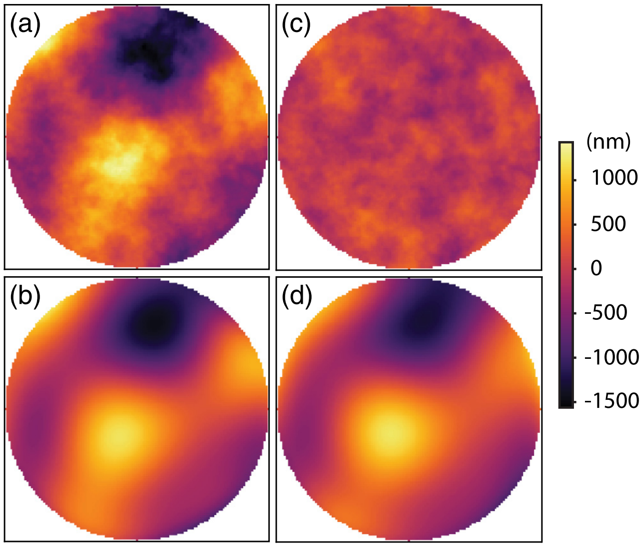

**Fig. 4** Example showing (a) entrance pupil wavefront, wavefront approximation using (b) “true” Zernikes 2 to 36 only, wavefront approximation based on (d) estimated Zernikes 2 to 36 using NN3, and (c) difference between true wavefront and estimated wavefront.

## 3.2 *ResNet*

The ResNet network is more recent than the Inception and solves the vanishing gradient problem more elegantly. To compare the two networks, we trained a ResNet with 50 layers (ResNet50) on the same data set that was used for the NN1 and NN2 of the Inception, using the baseline case of Table 1. The results are shown for NN3 in Table 2. The ResNet50 performed slightly better than the Inception. The ResNet was found to be more robust than the Inception network, with more consistent convergence during training. Hence, we selected the ResNet as the candidate for our studies. As an example, Fig. 4 shows the true wavefront over the entrance pupil together with the wavefront based on Zernike estimates by NN3. As expected, the estimates have less high-frequency content than the true wavefront.

The phase diversity method has generally been used with narrow-band light and a defocus of the order of a few wavelengths at the edge of the entrance pupil.[27,28] Curvature wavefront sensors operate in a different regime with a defocus of $\gg$10 wavelengths, where the transport equation assumed for the curvature sensor holds.[27,29] In our baseline, the out-of-focus PSFs were defocused by $4.5 \times \lambda_c$. We also tested a defocus of $1.5 \times \lambda_c$ (NN5). Representative PSFs for the two cases are shown in Fig. 3. Comparing NN3 and NN5 in Table 2 shows that for our wide-band application, the network working with a defocus of $4.5 \times \lambda_c$ clearly was better, so we kept that as our baseline approach. A defocus much larger than $4.5 \times \lambda_c$ would call for a larger field of view and then more pixels than can be handled conveniently by today’s neural networks.

A traditional curvature wavefront sensor works with a pair of two defocused PSFs,[29] and the phase diversity method can work with either two defocused PSFs or one in-focus and one defocused PSF. To study whether it is better for our neural network to read two out-of-focus PSFs rather than one in-focus and one out-of-focus PSF, we made a training run (NN4) using pairs of PSFs that were defocused in opposite directions by the same amount, $1.5 \times \lambda_c$, on the edge of the entrance pupil. The results of NN4 and NN5 are very similar, and because it is easier to implement a system with one in-focus PSF, we find it preferable to stay with in-focus and out-of-focus PSFs as selected in our baseline.

It was speculated that it would be preferable to train the neural network on clear PSFs from bright guide stars, so we studied the influence of guide star magnitude on ResNet performance. We trained a ResNet50 network (NN6) with baseline parameters, except that the training guide stars were in the range 4 to 8. We compared that network with the baseline NN3 network estimates of Zernike coefficients for different guide star magnitudes. We validated both networks with 80,000 guide stars with magnitudes in the range 4 to 14. We then binned the RMSs of

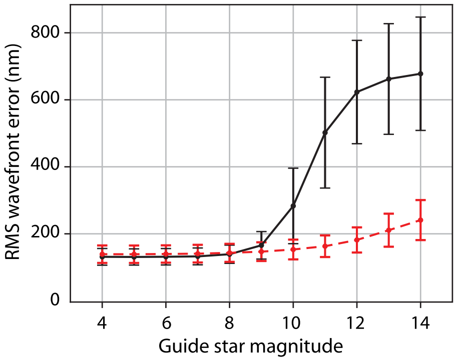

**Fig. 5** Wavefront RMS residuals and one standard deviation of uncertainty for different guide star magnitudes with NN6. Black, solid curve: network trained with guide star magnitudes in the range 4 to 8. Red, dashed curve: network trained with guide star magnitudes in the range 4 to 14.

the wavefront residuals in intervals around integer guide star magnitudes in the range 4 to 14. As is apparent from Fig. 5, which shows both the mean values and standard deviations for the binned intervals, the network trained with magnitudes 4 to 8 performed only slightly better in that range than the one trained with magnitudes 4 to 14, but much worse in the range 8 to 14. The overall mean RMS error for the network trained with the magnitude 4 to 8 and tested with magnitude 4 to 14 stars was 357 nm, and for the network trained with magnitude 4 to 14 guide stars, it was 182 nm. Hence, the network should be trained for its final application range and not in the range with the clearest PSFs.

To study how the estimation errors of the ResNet50 are distributed over the various Zernike terms and how they relate to the true Zernike coefficient values, we again submitted a validation set of 80,000 pairs of in-focus and out-of-focus PSFs to NN3. We determined the mean and standard deviation of each Zernike coefficient and plotted them together with mean absolute values of the true Zernike coefficients in Fig. 6. The usefulness of the neural network clearly decreases for Zernikes of higher order. We also computed the correlation coefficient matrix for the estimation errors of the Zernike coefficients as shown in Fig. 7. There is a significant (negative) correlation between several of the estimation errors. For instance, estimation error of Zernike no. 11 (third-order spherical aberration) is negatively correlated with that of Zernike no. 4 (defocus), meaning that when one is overestimated, the other will be underestimated. This is to be expected and is even desirable because of the similarity of the two aberrations.

Further, we checked whether it would be useful to increase the number of Zernikes included. Based on 80,000 validation datasets, we found that the theoretical, maximum achievable correction for $r_0 = 0.17$ m could go from an RMS error of 134 to 101 nm by increasing the number

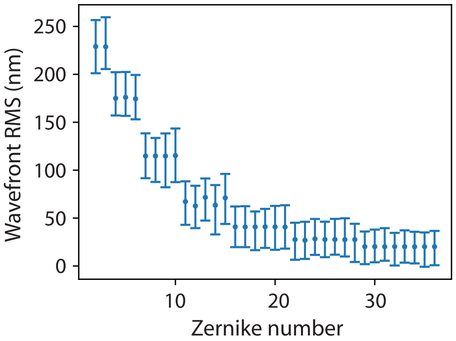

**Fig. 6** The dots show the mean absolute values of the true Zernike coefficients for a validation data set with 80,000 pairs of in-focus and out-of-focus PSFs for NN3. The error bars indicate the one standard deviation limits of the corresponding estimates of the Zernikes by neural network NN3. As can be seen, there is only a little bias in the estimates.

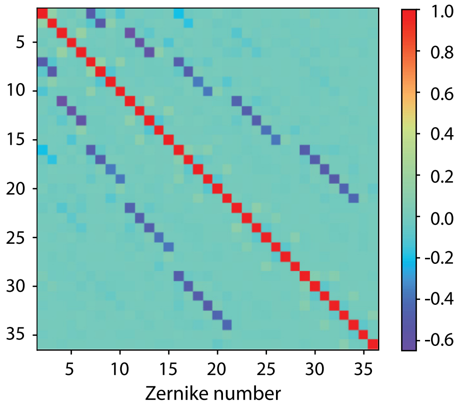

**Fig. 7** Correlation coefficient matrix for the Zernike coefficient estimation errors for neural network NN3.

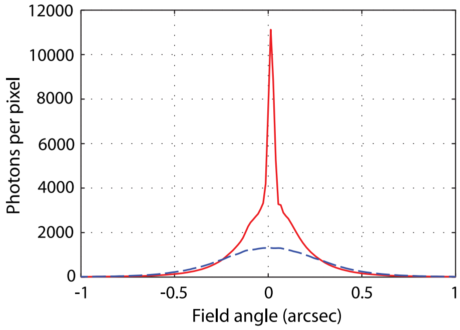

**Fig. 8** Long-exposure PSFs. The solid, red curve is the PSF after correction by an "ideal" deformable mirror, and the dashed, black curve is the PSF before correction.

of Zernikes included from 2 to 36 to 2 to 66. Hence, we trained the ResNet50 with the baseline parameters, but now with Zernikes 2 to 66. The network is listed as NN6 in Table 2. Because the high-order Zernike estimates are not precise, the performance of NN6 is similar NN3, so for our application there is no gain in including more Zernikes, at least when working in the noisy range of guide stars up to magnitude 14 and with single-shot wavefront sensors.

Our networks were trained for Fried's parameter, $r_0 = 0.17$ m at 545 nm. To see whether the network is sensitive to seeing quality, we tested the NN3 network with 80,000 PSFs computed for a seeing of $r_0 = 0.30$ m. The mean residual RMS value for guide stars in the range 4 to 14 was 130 nm, and for guide stars below magnitude 8, it was 123 nm. Hence, the network is also satisfactory with moderately better seeing.

To further study the neural network performance, we ran 20,000 PSF pairs on neural network NN3 and determined the long-exposure in-focus PSFs after correction by taking the mean of the PSFs after correction with the estimated Zernikes. We included only guide stars in the range 4 to 12. The result can be seen in Fig. 8 together with the mean in-focus PSF before correction. The long-exposure PSF is not even nearly diffraction limited, but there is a significant improvement in image quality compared with the raw PSF.

The inference time of the Inception network, NN1 and NN2, was around 6 ms and of the ResNet50 was 8 ms for NN3, NN4, and NN5 and 10 ms for NN6.

## 4 Multishot Wavefront Sensing

The previous section dealt with wavefront sensing from PSF pairs and with no knowledge of the previous history of the atmosphere. In the following, we present results from studies of the use of sequential neural network wavefront sensing based on a time series of PSFs that are correlated over time.

**Table 3** Result from closed-loop wavefront sensing with three runs.

| | |
|---|---|
| RMS of original wavefront | 674 nm |
| RMS after one run and correction | 204 nm |
| RMS after two runs and corrections | 159 nm |
| RMS after three runs and corrections | 152 nm |

## 4.1 *Closed-Loop Wavefront Sensing*

As an introduction, we first studied what happens when the same neural network is run more than once in a closed loop with the same phase screen. As before, our simulated wavefronts were generated by the "power-spectrum" method and with the baseline parameters of Table 1. However, this time, the neural network included 65 Zernikes and had been trained on guide stars with $r_0$ randomly selected in the range 0.17 to 0.5 m to ensure that it could handle narrow PSFs well. We tested the neural network with 10,000 simulated PSF-pairs computed for $r_0 = 0.17$ m. The procedure was as follows: (1) submit a PSF-pair to the neural network and find Zernike coefficients, (2) subtract Zernikes from the original wavefront, and (3) compute new, simulated PSF-pairs and return to (1). We repeated this three times for each dataset. Finally, we computed the mean residual wavefront RMS values for all datasets. The results are shown in Table 3.

Although the correction is not perfect, repeated use of the neural network indeed improves the result significantly.

We then turned to closed-loop operation of the neural network wavefront sensor in an adaptive optics system with a "perfect" deformable mirror and a sampling frequency of 50 Hz as shown in Fig. 9. For this, we calculated the simulated PSF-pairs for $r_0 = 0.17$ m at 545 nm and a magnitude 8 guide star using phase screens generated by the Skylight method as described in Sec. 2.2. The PSFs are then correlated over time. The neural network was the NN3 of Table 2. The procedure was as follows: (1) submit a PSF-pair to the neural network and find the Zernike coefficients, (2) subtract Zernikes from the subsequent wavefront 20 ms later, and (3) compute a new, simulated PSF-pair from the residual wavefront 20 ms later, and return to (1). We averaged the residual wavefront RMS after correction over 2500 time steps and got the value 151 nm, which should be compared with the value found for a magnitude 8 guide star in Fig. 5, which is 173 nm. Hence, working in a closed loop indeed is an advantage because, in the null-seeking mode, the wavefront to be estimated is on average smaller.

A note of caution is needed. We also tried the same simulation but with a neural network estimating 65 Zernikes. However, we found that when including more than 45 Zernike terms, the adaptive optics became unstable. The network commands a Zernike correction that it does not "see" well itself, and it is integrated by the inherent integrator of the closed loop (see Fig. 9), leading to Zernike runaway.

Bit depth of the image sampling is another issue of importance for a closed-loop system. For the photon numbers related to the single-shot studies reported in Sec. 3, bit depths of 8, 12, and

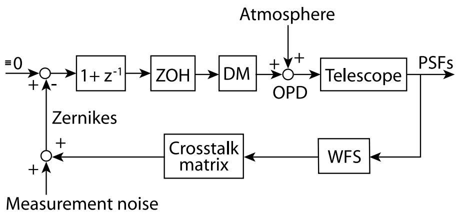

**Fig. 9** Principle of wavefront sensing in a closed-loop adaptive optics system, where *z* is the time-shift operator, ZOH is a zero-order hold, DM is a deformable mirror, and WFS is a wavefront sensor. The deformable mirror form accumulates commands from earlier steps, corresponding to an integrator in a sampled-data system.

16 all give a similar performance of the neural network. However, in a closed-loop system, the PSFs will often be narrower, so the maximum number of photons in a pixel at the peak of the PSF may easily be above $2^8$, leading to loss of information in the faint part of a PSF. A radiometric evaluation is needed for the actual case to determine the bit depth needed for the A/D-conversion in a closed-loop system.

### 4.2 *Kalman Filter*

Variations in atmospheric seeing can be seen as "plant errors" and neural network errors as "measurement errors."[30] A Kalman filter was set up using numerically determined covariance matrices for plant noise and measurement noise, and taking the Zernike coefficients and their time derivatives as state variables. The Zernike coefficients were determined running the NN3 neural network described in Sec. 3.2 together with time-correlated validation data generated with the Skylight method and the atmosphere model with seven frozen layers. However, the Kalman filter did not improve the estimates much because the neural network measurement errors are not truly uncorrelated Gaussian but are correlated over time. We believe that better results might be obtained using a plant model that includes the effect of frozen phase screens, but such a Kalman filter would be highly complex and potentially slow. A major effort would be needed to clarify whether the filter could work and whether computation time could be made acceptable.

### 4.3 *Recurrent Neural Network*

We also looked into the possibility of using a recurrent neural network (RNN), which has "memory" of previous estimates and inputs. The phase screens of the incoming light are correlated over time, and it was thought that the neural network might be able to find the nature of the correlation.

We used a ResNet50 network as the front end to the RNN as shown in Fig. 10. The task of the RNN is then to improve the Zernike estimates from the ResNet50. We studied different architectures and got the best results with the one shown. It has a direct Zernike feed from the ResNet50 to the output, to which a correction is added by the RNN. We used the long short-term memory (LSTM) RNN building block,[31] which overcomes the vanishing gradient problem for long input sequences. As an example, we looked back 16 s, taking a total of $n = 200$ samples from the past, including only every fourth sample over time. The guide star had a magnitude of 8, and the time-correlated PSFs were the same as those used for the Kalman filter described above. The input Zernikes were scaled to have a standard deviation of 1 for the network, and the output Zernikes were rescaled accordingly. As can be seen in the example for Zernike no. 7 in the lower part of in Fig. 11, there is indeed an improvement with respect to the ResNet50 estimates.

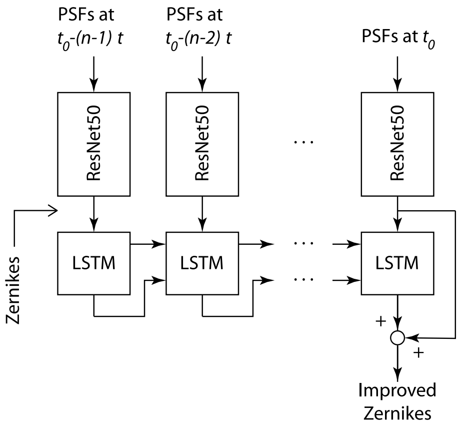

**Fig. 10** Architecture of the recurrent network using NN3 as a building block. Here, $\Delta t$ is the sampling period of the network, $n$ is the number of samples taken into account, and $t_0$ is the time.

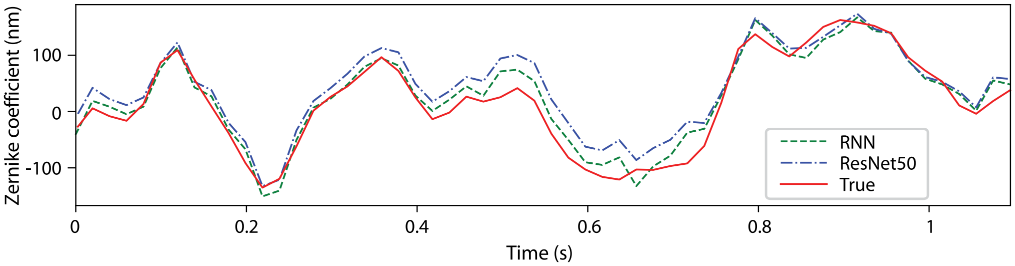

**Fig. 11** Part of a simulation run with the RNN. The solid, red curve is the track of the true Zernike no. 7. The dot-dashed blue curve is the estimate given by the ResNet50, and the dashed, green curve is the estimate from the recurrent network improving the ResNet50 estimate.

However, the improvement is smaller for high-order Zernikes, and in total, the recurrent network only improved the wavefront RMS averaged over 80,000 samples from 171 to 164 nm.

## 5 Conclusion

Motivated by the potential of real-time computational image sharpening with neural networks in astronomical telescopes, we studied whether convolutional neural networks can be used for image-based wavefront sensing. We generated by simulation a large amount of in-focus and out-of-focus PSFs and trained several neural networks to estimate Zernike polynomial approximations for the incoming wavefront. We included the effects of photon noise, guide star magnitude, bit depth, and PSF blurring from wide-band imaging. We compared the performance in the V-band of neural networks for a 3.6-m telescope and found that the "ResNet" gave the best results. The best wavefront residual RMS obtained for guide star magnitude 4 to 8 and $r_0 = 0.3$ m was 123 nm for wide-band PSFs over the wavelengths 470 to 700 nm. Although not even nearly diffraction limited, real-time image sharpening based on the ResNet PSF-based wavefront sensor seems promising for medium-sized telescopes.

We also studied how the neural network works with sequential PSF-pairs that are correlated over time and found that closed-loop operation together with a deformable mirror seems feasible, provided that a reasonable number of Zernikes are included. We compared the performance of a ResNet neural network combined with a Kalman filter and a recurrent network, both working with time-correlated PSF-pairs. We found that a Kalman filter of manageable size did not improve the ResNet estimates, and the RNN led to only a modest improvement.

Our studies focused on a 3.6-m telescope in the V-band. With today's technology, it may be difficult to apply our approach for telescopes in the range 4 to 8 m, but that may change in the near future due to the availability of more powerful graphical processors. Also, in near IR, which is much easier than V, the technology may apply to telescopes in the 4 to 8 m class.

For the case studied, the inference times are manageable and the computational burden sufficiently low enough that a conventional work station with one good graphics card suffices. However, for a future system with multiple guide stars and on-the-fly deconvolution, a number of high-performance graphic cards will be needed.

We now plan to study how neural networks can be used for wavefront-assisted deconvolution for astronomy, making the way for real-time computational image sharpening. Other interesting possibilities include using a neural network computational image sharpening system on top of traditional adaptive optics for additional image improvement or for image-based tomography to determine wavefront error contributions from different atmospheric layers.

## Acknowledgments

This work was supported by the Walter Gyllenberg Foundation in Sweden. The authors have no conflicts of interest related to this paper.

## References

1. R. G. Paxman, "Overview of post-detection techniques for fine-resolution astronomical imaging," in *Wavefront Supported Post-Facto Image Correction*, T. Andersen, Ed., pp. 1–8, Nordic Optical Telescope Scientific Association, Roskilde, Denmark (1993).
2. R. A. Gonsalves, "Perspectives on phase retrieval and phase diversity in astronomy," *Proc. SPIE* **9148**, 91482P (2014).
3. L. M. Mugnier, T. Fusco, and J.-M. Conan, "Mistral: a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images," *J. Opt. Soc. Am. A* **21**(10), 1841–1854 (2004).
4. C. Z. Bond et al., "Iterative wave-front reconstruction in the Fourier domain," *Opt. Express* **25**, 11452–11465 (2017).
5. B. N. Sitarski et al., "Modeling instrumental field-dependent aberrations in the NIRC2 instrument on the Keck II telescope," *Proc. SPIE* **9148**, 91486T (2014).
6. M. Lamb et al., "Exploring the operational effects of phase diversity for the calibration of non-common path errors on NFIRAOS," *Proc. SPIE* **9909**, 99096E (2016).
7. J. R. P. Angel et al., "Adaptive optics for array telescopes using neural-network techniques," *Nature* **348**, 221–224 (1990).
8. C. Gonzalez et al., "Multi-GPU development of a neural networks based reconstructor for adaptive optics," *Complexity* **2018**, 5348265 (2018).
9. R. Swanson et al., "Wavefront reconstruction and prediction with convolutional neural networks," *Proc. SPIE* **10703**, 107031F (2018).
10. S. W. Paine and J. R. Fienup, "Machine learning for improved image-based wavefront sensing," *Opt. Lett.* **43**, 1235–1238 (2018).
11. S. Gómez et al., "Hybrid artificial intelligent systems," *Lect. Notes Comput. Sci.* **9121**, 411–421 (2018).
12. Y. Nishizaki et al., "Deep learning wavefront sensing," *Opt. Express* **27**(1), 240–251 (2019).
13. H. Guo et al., "Improved machine learning approach for wavefront sensing," *Sensors* **19**(16), 3533 (2019).
14. P. Jia et al., "Astronomical image restoration and point spread function estimation with deep neural networks," *Astron. J.* **159**, 183 (2020).
15. T. Andersen and A. Enmark, *Integrated Modeling of Telescopes*, Springer, New York, Dordrecht, Heidelberg, London (2011).
16. R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**(3), 207–211 (1976).
17. T. Andersen, M. Owner-Petersen, and A. Enmark, "Neural networks for image-based wavefront sensing for astronomy," *Opt. Lett.* **44**, 4618–4621 (2019).
18. F. Chollet, *Deep Learning with Python*, Manning Publications Co., Shelter Island, New York (2017).
19. A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed., O'Reilly UK Ltd., Farnham, Surrey (2019).
20. B. Planche and E. Andres, *Hands-On Computer Vision with TensorFlow 2*, Packt Publishing, Birmingham (2019).
21. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd Int. Conf. Learn. Represent., ICLR 2015*, San Diego, California (2014).
22. C. Szegedy et al., "Going deeper with convolutions," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–9 (2015).
23. K. He et al., "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
24. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 7132–7141 (2018).
25. B. Zoph et al., "Learning transferable architectures for scalable image recognition," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 8697–8710 (2018).
26. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd Int. Conf. Learn. Represent., ICLR 2015*, San Diego, California (2014).

27. J. Fienup et al., "Comparison of phase diversity and curvature wavefront sensing," *Proc. SPIE* **3353**, 930–940 (1998).
28. R. L. Kendrick, D. S. Acton, and A. L. Duncan, "Phase-diversity wave-front sensor for imaging systems," *Appl. Opt.* **33**(27), 6533–6546 (1994).
29. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," *Appl. Opt.* **27**, 1223–1225 (1988).
30. R. R. Labbe, Jr., *Kalman and Bayesian Filters in Python*, https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python (2018).
31. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* **9**(8), 1735–1780 (1997).

**Torben Andersen** designed the ESO 1.5-m CAT Telescope, and led design of the 2.5-m Nordic Optical Telescope and the 32-m EISCAT Svalbard Antenna. He was a player in the design of the planned LEST solar telescope and ALMA antennas, and for a period he was a system engineer and head of telescope engineering for the ESO VLT. Later, as a professor at Lund University, he has focused on design of extremely large telescopes.

**Mette Owner-Petersen** received her PhD in physics from the Technical University of Denmark (DTU). She was an associate professor in the Physics Department at DTU and later at Lund University, Sweden. She has teaching experience in physics and optics. Her main research areas have been within low-temperature physics, optical design, and adaptive optics for large telescopes and for high-resolution imaging of the human retina.

**Anita Enmark** has a master's degree in computer science and engineering and a PhD in modeling of adaptive optics for astronomy. She is a senior lecturer at Luleå University of Technology (image processing, signal processing for radar applications and solar observations, microcomputers and on-board computers for space applications). She is the co-author of the book *Integrated Modeling of Telescopes*, by T. Andersen and A. Enmark, Springer, 2011.