Research Article | Vol. 33, No. 6 / 24 Mar 2025 / *Optics Express* 14431

Optics EXPRESS

Practical aberration correction using deep transfer learning with limited experimental data

YONG EN KOK,[1,4] **ALEXANDER BENTLEY,**[2] **ANDREW J. PARKES,**[1] **MICHAEL G. SOMEKH,**[2,3] **AMANDA J. WRIGHT,**[2,5] **AND MICHAEL P. POUND**[1,6]

[1] *School of Computer Science, University of Nottingham, Nottingham NG8 1BB, UK*
[2] *Optics and Photonics Group, Department of Electrical and Electronic Engineering, University of Nottingham, Nottingham NG7 2RD, UK*
[3] *Research Center for Humanoid Sensing, Zhejiang Laboratory, Hangzhou 3111100, China*
[4] *yong.kok@nottingham.ac.uk*
[5] *amanda.wright@nottingham.ac.uk*
[6] *michael.pound@nottingham.ac.uk*

Abstract: Adaptive optics is a technique for correcting aberrations and improving image quality. When adaptive optics was first used in microscopy, it was common to rely on iterative approaches to determine the aberrations present. It is advantageous to avoid iteration, and therefore there has been a shift to deep learning for aberration prediction. However, issues remain regarding the practicalities of machine learning for adaptive optics, an important one being the requirement for a large training dataset. Here, we explore transfer learning to overcome this need for data by pre-training a network on a large simulated dataset and fine-tuning it with reduced experimental data for application in an experimental setting. We demonstrate that the pre-trained network can make noticeable improvements with fine-tuning on just 24 experimental samples. To further enhance practicality, we significantly extend the range of aberrations present, predicting up to 25 Zernike modes with each coefficient ranging from -1 to 1, and perform a thorough analysis of the type and magnitude of phase-diversity required in the input data for a successful network. Our approach demonstrates substantial aberration reduction on experimental data for 10 Zernike modes, with an average 73% decrease in RMS wavefront error from 1.81 to 0.48 rad when correction is applied. This method achieves complete experimental image capture and aberration inference at rates comparable to the image acquisition time of a typical laser scanning microscope. Additionally, we consider the benefits of further improvements via an iterative step. As such, this work addresses some of the key practical hurdles that remain in the use of deep learning for aberration prediction and correction.

1. Introduction

Adaptive Optics (AO) is a technique used to enhance image quality in optical systems by correcting aberrations that arise from both imperfections in the optical system, and variations within the sample being imaged [1,2]. AO works by dynamically shaping the wavefront of the light using correction devices such as spatial light modulator [3] or deformable mirror [4] to cancel out the aberrations present [5,6]. To obtain the appropriate wavefront correction, and so optimise the system reducing aberration, it is necessary to either measure the aberration directly or use an indirect measure, such as image intensity. Traditional wavefront sensors, including the Shack Hartmann [7] or interferometer [8], require access to a guide star. Modal wavefront sensing uses intensity as an indirect measure and therefore no additional hardware is required; instead, Zernike modes are applied in turn and intensity is measured, here iteration is required [9].

#557993 https://doi.org/10.1364/OE.557993
Journal © 2025 Received 28 Jan 2025; revised 14 Mar 2025; accepted 14 Mar 2025; published 21 Mar 2025

By comparison, optimisation methods search for the appropriate wavefront correction by using algorithms such as stochastic parallel gradient descent [10], simulated annealing [11], random search [12], genetic algorithm [13], Gerchberg-Saxton [14] or the Extended Nijboer-Zernike theorem [15]. These conventional search algorithms are iterative, slow and computationally demanding. Recently, Johnson et al. [16] demonstrated a promising phase diversity approach for fluorescence microscopy, using multiple images with known diversity aberrations and Gauss-Newton optimisation to estimate wavefronts. This advancement significantly enhances correction efficiency, while operating within the broader framework of optimisation approaches where computational considerations naturally scale with aberration complexity. Speed is particularly crucial when working with biological samples, where minimizing photon dose is essential to reduce photo-damage risk. Moreover, samples are often not static and aberrations can vary spatially across the field of view, so for a practical application it is vital to make the correction quickly.

Alongside the development of computer vision, deep learning has emerged as a promising technique for addressing a range of image-based inverse problems. A comprehensive review summarizing the advances of classical and modern computational AO techniques applied to optical microscopy can be found in [17]. Researchers have applied deep learning approaches for aberration retrieval through either direct reconstruction of the phase of the aberrated wavefront [18,19] or recovery of Zernike coefficients [20–31] from the input aberrated images.

However, there remain significant challenges in this field. Collecting a large amount of experimental data for training deep learning models, where often hundreds of thousands of data are needed, is impractical and time-consuming. The domain gap between simulated and experimental data, primarily due to imaging with a high numerical aperture lens [32] and minor imperfections in alignment, although small, hinders the direct application of models trained purely on simulated data to an experimental setting. Jin et al. addressed this challenge by proposing a weight-sharing dual-stream Convolutional Neural Network (CNN) framework for Zernike coefficient prediction, training on a large amount of labeled simulated data with domain adaptation to unlabeled target-domain experimental data [33]. Their approach showed promising results for 9 Zernike modes in simulations and was validated on phantom samples. However, they did not report overall average accuracy for the experimental data, limiting insights into real-world performance. In this paper, we explore the use of transfer learning to address the requirement for a large training dataset and bridge the domain gap without comprehensive system characterisation or computationally expensive simulations. A network is pre-trained using hundreds of thousands of simulated samples and then fine-tuned with decreasing amounts of experimental data before being tested in an experimental setting. As such, we systematically explore the impact of using varying amounts of experimental Point Spread Function (PSF) data for fine-tuning, aiming to optimise performance on real-world aberrations with minimal data requirements.

Beyond the challenge of data requirements, deep learning models may encounter additional complexities in aberration retrieval due to the inherent ambiguity associated with determining the Zernike coefficients solely from an aberrated image without phase information. In particular, angularly-even Zernike modes can produce identical PSF images at focus for oppositely-signed Zernike coefficients, thus posing difficulties in model training and prediction due to this ambiguity [27,34]. To address this issue, Siddik *et al.* [34] proposed to predict the absolute values for the specific Zernike modes that are susceptible to this ambiguity. However, this method does not represent a complete wavefront sensing approach, because the sign of the coefficient is not recovered. Alternatively, Nishizaki [20] explored the use of a single image acquired through preconditioners such as overexposed, defocused, or scattered images to enhance estimation accuracy. To further improve accuracy, researchers [19,22–29,35] have leveraged using several phase-diverse images typically obtained from two or more image planes. Notably, Zhang et al. [35] have shown that using two symmetrical defocused planes demonstrated higher accuracy than traditional phase diversity methods using focused and single defocused planes, especially for large-scale and high-frequency wavefront sensing. When working with extended 2D images, Xin et al. [30] created a metric that was independent of the object being imaged and enabled aberration retrieval from any extended images. Building on Xin et al.'s concept [30], Hu et al. [31] extracted the PSF information from astigmatic image pairs and introduced a CNN to demonstrate the method's generalisability on various microscopy modalities. Xin et al.'s and Hu et al.'s work [30,31] both benefit from the input to the network being independent of the object being imaged.

The majority of the studies discussed above have concentrated on relatively small aberrations with either a limited number of Zernike modes present, or Zernike coefficients of small amplitude. Here we look to push this boundary by considering cases of severe aberration. We work with 25 Zernike modes where *each* Zernike coefficient is in the range between -1 to 1 rad. We consider this problem for PSF images, predicting directly from the images without any pre-processing steps. As noted by others, we find that using a set of phase-diverse input images is important, and we present a detailed study of the different bias modes, the amplitude of bias mode, and the minimum number of phase-diverse images required for a satisfactory prediction of the Zernike coefficients.

This paper aims to address some of the practical issues that remain when using machine learning to predict and ultimately correct aberrations in an imaging system. In particular, we address: i) the requirement for a large dataset when training a network, ii) the ability to correct for large amounts of aberrations and improve severely distorted images, and iii) the type and magnitude of phase-diversity required, and the minimum number of input images needed. In addition, we explore the potential of iterative correction to further enhance image quality, particularly in cases where significant residual aberrations persist after the initial correction.

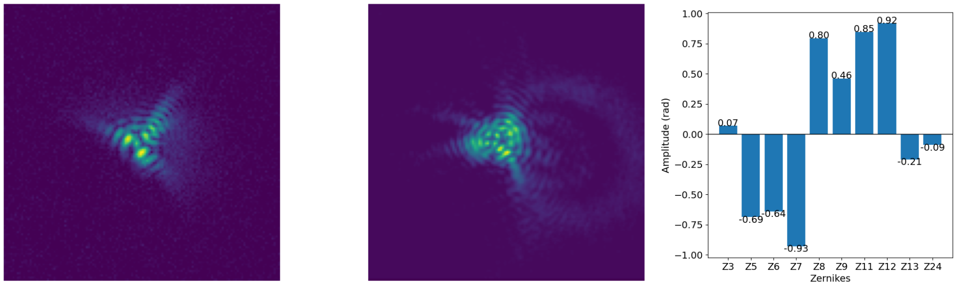
By addressing these challenges collectively, we move towards more practical approaches to deep learning for aberration prediction. These combined improvements enable faster aberration correction with adaptive optics, reducing both sample photon exposure and multiple corrections per image.

2. Materials

2.1. Simulated dataset

We represented aberrations as a series of Zernike modes randomly sampled from a uniform distribution within the amplitude range of $[-1,1]$ rad per Zernike. The Zernike modes were used to simulate an aberrated point source image. To simulate severe aberrations, 25 Zernike modes were used. These were $Z_3 - Z_{27}$ (ANSI indices), excluding Z_0 (piston), Z_1 (tip), and Z_2 (tilt), as these could be easily corrected in hardware or software using centroiding algorithms or other registration methods. A bias mode (e.g. defocus Z_4) with coefficient 1 rad was added or subtracted from the Zernike coefficients to generate the phase-diverse images I_{-1}, I_0, I_{+1}. This first data set, consisting of 128x128 pixel resolution images synthesised using Python code, was referred to as SD25 throughout the paper and served to explore the optimal parameters of different bias modes, the amplitude of the bias mode, and the minimum number of phase-diverse images required for accurate Zernike coefficient prediction.

A second set of simulated PSF images, referred to as SD10, was generated for pre-training purposes to be used for fine-tuning on experimental data. This dataset included 10 Zernike modes ($Z_3, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{11}, Z_{12}, Z_{13}, Z_{24}$). We did not include tip, tilt, piston and defocus, and added second-order spherical. This subset was selected due to its common use in literature [31,36]. Following the same procedure as SD25, the Zernike modes were randomly sampled from $[-1,1]$ rad, and here the phase-diverse images were generated using only the defocus mode, Z_4. The generated images were processed to better simulate real experimental conditions through multiple transformations: multiplication with a Gaussian kernel of size 128, sigma 30, the addition of Poisson noise (noise level = 1000), and Gaussian noise (zero mean and standard deviation of 0.02). An example of the simulated PSF image and the experimental data with the same nominal aberration coefficients applied are shown in Fig. 1.



(a) Simulated data (b) Experimental data (c) Zernike coefficients

Fig. 1. An example of (a) simulated and (b) experimentally aberrated PSFs captured at focus when a set of Zernike modes, presented in (c) are applied.

2.2. Experimental dataset

Two datasets of experimental images were acquired from the adaptive optics system described in Section 3.2. The first dataset, which was referred as ED10, comprised 3000 aberrated PSF sets. Each aberrated PSF set was represented by three phase-diverse images captured above, below, and at the focal plane. This dataset was randomly generated from 10 Zernike modes $Z_3, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{11}, Z_{12}, Z_{13}, Z_{24}$.

The second set of experimental images, denoted as ED24, was created to predict a larger set of 24 Zernike modes $Z_3, Z_5 - Z_{27}$. The defocus mode Z_4 was excluded from the datasets due to the 3D nature of real samples and to avoid the network returning values relating to an unwanted imaging plane. This dataset contained 15,000 PSF sets, each of 3 images.

Similar to the simulated dataset, for both datasets, each value of Zernike modes was randomly sampled from a uniform distribution within the range of $[-1, 1]$ rad. To create the phase-diverse images, a defocus mode, Z_4 with coefficient 1 rad was added or subtracted from the Zernike coefficients.

3. Methodology

3.1. Framework overview

We adopted a residual network (ResNet50) [37], a widely used backbone network for various computer vision tasks, to predict the Zernike coefficients from the phase-diverse images. To accommodate the specific requirements of each experiment and the differing number of input images, we adapted the network to accept n_z-channel input, with the value of n_z not exceeding three in this work. The last fully connected layer was modified to output N Zernike coefficients, corresponding to the specific task. The overall framework is illustrated in Fig. 2. More training details can be found in Section 1 of Supplement 1.

3.2. Adaptive optics imaging system and model deployment

To demonstrate the capabilities in an experimental setting and to collect the experimental training datasets, we used the optical layout shown in Fig. S2 of Supplement 1. This system used a Ti:Sapphire laser (Spectra-Physics) operating at 690 nm in pulsed mode to minimize laser speckle. A spatial light modulator (SLM) was placed at a conjugate plane to the back aperture of

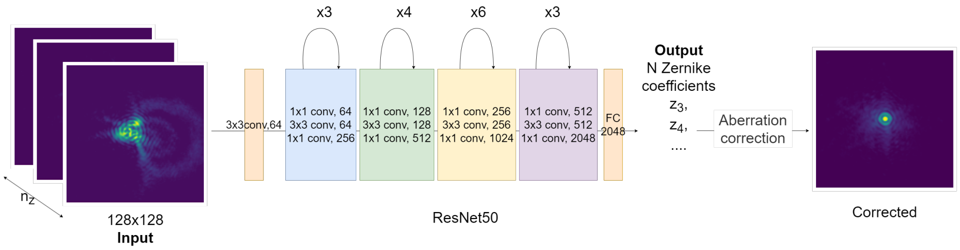


Fig. 2. Overview of framework: The ResNet50 is trained to predict the Zernike coefficients from n_z phase-diversity images captured at different planes (n_z not more than three in this study). The predicted Zernike coefficients are used to correct the aberration on the image.

the microscope objective lens, serving as an adaptive correction device. To obtain quantitative performance of the proposed network, the SLM was used to both introduce and correct for aberrations, ensuring ground truth values were known. A mirror was placed at the focal plane of the objective lens, and images of the reflective focal spot were collected using a CMOS camera (Thorlabs) placed at the exit port of the microscope body. The shape of the focal spot was used to represent the PSF of the microscope.

For aberration correction using the network deployed on the imaging system, we first captured a dataset of phase-diverse images using the CMOS camera. To do this known, but randomly generated, Zernike modes were applied to the SLM referred to as the ground truth Zernike coefficients. These images and the corresponding ground truth Zernike modes were used to either fine-tune a network pre-trained on a simulated dataset or to train a network from scratch. Once trained, the network reads in three phase-diverse images of the focal spot with unknown aberrations and predicted the Zernike coefficients present. To determine how successful an aberration correction would be, the residual Zernike coefficients were determined by subtracting the predicted coefficients from the ground truth coefficients. These residual Zernike coefficients could then be applied as a phase map to the SLM to obtain an image of the corrected focal spot for comparison. In the case of the iteration experiments, these corrected focal spot images served as the starting point for further correction. This implementation was automated through a MATLAB script that controlled all optical hardware and obtained Zernike coefficient predictions through integration of the models in Python.

3.3. Evaluation metrics

To facilitate future comparisons across different implementations, we have chosen to report both Strehl ratio and Root Mean Square Wavefront Error ($\mathrm{RMS_{WFE}}$). These metrics are universal, relevant regardless of the type of microscope modality being used, and their values can be directly linked to diffraction-limited resolution, enabling standardised evaluation of aberration correction effectiveness across diverse optical systems.

To assess the accuracy of the Zernike predictions, we used $\mathrm{RMS_{WFE}}$ to quantify the residual wavefront error by subtracting the predicted Zernike modes, y_{pred} from the ground truth Zernike modes, y_{true}:

$$RMS_{WFE} = \sqrt{\sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} \left(y_{\mathrm{pred}_{ij}} - y_{\mathrm{true}_{ij}} \right)^2} \tag{1}$$

where N is the number of samples and M is the number of Zernike modes.

We calculated the Strehl ratio to quantify the quality of the corrected PSF. The Strehl ratio was obtained by comparing the peak intensity of the corrected PSF to that of the ideal, aberration-free, PSF. A Strehl ratio exceeding 0.8 and a $\mathrm{RMS}_{\mathrm{WFE}} < 0.5$ are generally considered indicative of diffraction-limited imaging performance [38].

4. Experimental design

4.1. Exploring optimal phase-diversity parameters

Building upon prior works that primarily employed defocus as the standard bias mode for phase-diverse imaging [19,22–29], as well as the work by Hu et al. [31] that proposed the astigmatism mode as the most effective bias mode, we explored the use of different bias modes. Given the inherent ambiguity associated with single-channel inputs, we examined the minimum number of phase-diverse images required for good performance and investigated the different values of amplitude of the imposed phase applied. We further verified the performance of the model on experimental data. The experiments were as follows:

- **Bias mode**: Using the simulated PSF dataset, SD25, we compared the bias modes employed to generate our standard 3-channel phase-diverse images, I_{-1}, I_0, I_{+1}. The considered bias modes included oblique astigmatism Z_3, defocus Z_4, primary spherical Z_{12}, and vertical secondary trefoil Z_{16}. These modes were selected to cover a range of circularly symmetric and non-symmetric modes plus higher and lower-order distributions. For comparison purposes, we also present a baseline model that utilised a 1-channel input without any bias modes.

- **Amplitude of bias mode**: Following our identification of defocus as the optimal bias mode, we compared the different amplitudes of defocus within a 3-channel input on the simulated PSF dataset, SD25.

- **Number of input channels**: We investigated the minimum number of phase-diverse images required to predict the Zernike coefficients and explored the different combinations of phase-diverse inputs (negative, positive or at focus) for the 2-channel input on the simulated PSF dataset, SD25. We then trained separate networks from scratch on the experimental dataset, ED10, to validate the findings and confirm the generalisability of our approach.

4.2. Correcting for severe aberrations

To validate the model's performance with severe aberrations in an experimental setting, we evaluated the network using the optimal phase-diversity parameters determined from experiments in subsection 4.1 on the ED24 experimental dataset, which encompassed 24 Zernike modes.

4.3. Minimizing experimental data for aberration correction and the role of transfer learning approach on limited experimental data

Here we investigated the minimum amount of experimental data required to achieve satisfactory aberration correction and explored the potential benefits of pre-training a model on a large simulated dataset followed by fine-tuning on experimental data. We compared models with and without pre-training, fine-tuned on varying training subsets of the ED10 experimental dataset for 300 epochs: 24 (1%), 72 (3%), 120 (5%), 240 (10%), 480 (20%), 720 (30%), 1200 (50%), and the full training set of 2400 (100%) samples.

4.4. Iterative aberration correction

To further enhance aberration correction, we evaluated the potential benefits of an additional iterative correction step, wherein phase-diverse images of the residual aberrations were subsequently captured and used as the input for a second correction iteration. This step aimed to remove any remaining aberrations that were not fully corrected for the initial iteration. As a proof of concept, we applied this approach to the best-performing model trained on the ED10 experimental dataset to demonstrate the potential effectiveness of iterative correction.

5. Results and discussion

5.1. Exploring optimal phase-diversity parameters

- **Bias Mode:** Based on Table 1, the network using defocus to introduce phase-diversity demonstrates the best performance, exhibiting an improvement of approximately 82% in terms of $\mathrm{RMS}_{\mathrm{WFE}}$ compared to the pre-correction level (2.89 rad). In contrast, 1-channel input without phase diversity shows the poorest performance, consistent with previous studies [27,34] that identify the challenges associated with single-image ambiguity. Our investigation of higher-order bias modes, including primary spherical (Z_{12}) and vertical secondary trefoil (Z_{16}), revealed no additional performance benefits and in fact, the vertical secondary trefoil mode is only slightly better than the 1-channel baseline. These findings differ from others, where astigmatism was found to perform slightly better when imaging a 3D object rather than a 2D plane [31].

- **Amplitude of bias mode:** As shown in Table 2, the 3-channel defocus phase-diverse input with an amplitude of ±1 rad achieves the best performance. This suggests that an amplitude of ±1 rad provides a good balance between introducing enough distortion to predict the Zernike coefficients and maintaining a manageable defocus effect. Interestingly, the performance for the amplitude of 0.5 is very close to the optimal value, while the amplitude of 1.5 leads to a significantly worse performance. This might indicate that a range of amplitudes below the optimal value remains informative for prediction, while larger amplitudes introduce excessive distortions, making it difficult to accurately determine the Zernike coefficients.

- **Number of input channels:** Results in Table 3 demonstrate that for pre-correction levels of 2.89 and 1.81 rad in SD25 simulated data and ED10 experimental data, respectively, the model employing 3-channel phase-diverse input with an amplitude of 1 rad achieves superior aberration correction compared to both 1-channel input and all 2-channel input combinations with varying plane configurations. However, in the ED10 experimental dataset, the improvement when increasing the number of channel inputs is less pronounced compared to SD25, possibly due to factors such as noise and real-world variations.

  As previously discussed, the 1-channel input has the worst performance due to the ambiguity inherent in using a single image for aberration correction. In the SD25 simulated dataset, further analysis reveals that with 1-channel input, the network predominantly predicts values around zero mean, indicating its inability to effectively correct image aberrations. However, in the ED10 experimental dataset, the model with 1-channel input achieves a Strehl ratio approaching diffraction-limited quality (0.76). While the reduced number of Zernike modes likely contributes to this improved performance by reducing ambiguity, we suggest that the rich information present in the experimental data is the primary factor. This is supported by additional evaluation on the ED24 experimental data with a larger number of Zernike modes, which shows that the model with 1-channel input successfully corrects much of the image aberration (see Section 2 of Supplement 1 for details).

  Across both datasets, the 2-channel input showed slightly lower performance compared to the 3-channel input. For the more aberrated starting point (SD25), it is important that one of the two input channels is recorded at focus. Indeed, when using 2-channel input for SD25, the symmetric defocus [−1,1] configuration performed the worst, contrasting with findings where symmetric defocused planes have sometimes yielded higher accuracy than using a single defocused plane plus focus [35]. This discrepanc ymay be attributed to the distinct nature of our deep learning approach compared to conventional phase diversity algorithms. In contrast, for the less challenging starting point (ED10), all 2-channel configurations performed very similarly, with the configuration [−1,0] performing best, while the symmetric defocus is no longer the worst. This similarity in performance across configurations for lower aberrations suggests that the PSF distortion is sufficiently moderate to allow informative signal capture across multiple defocus configurations, thereby diminishing any significant performance differences. Conversely, in the case of very severe aberrations (i.e. SD25), less information can be extracted from the out of focus planes as the PSF becomes highly distorted, causing the intensity to spread across the entire field of view—potentially missing the detector—and leading to low signal-to-noise ratios.

In summary, when image acquisition time and sample photon exposure are critical constraints, the 2-channel input ([−1, 0]) presents a viable alternative, achieving Strehl ratios of 0.78 and 0.82 for SD25 and ED10, respectively. For applications prioritizing precision, particularly with increased ambiguity from a higher number of Zernike modes, the 3-channel input provides substantial advantages in terms of robustness and correction accuracy.

Table 1. Comparison of using the different bias modes for the 3-channel phase-diverse input on the simulated PSF dataset, SD25, where the average uncorrected $\mathrm{RMS}_{\mathrm{WFE}}$ and Strehl ratio are 2.89 rad and 0.06 respectively. The average $\mathrm{RMS}_{\mathrm{WFE}}$ and average Strehl ratio of the models on the test set are reported.

Bias mode	$\mathrm{RMS}_{\mathrm{WFE}}$ (rad)	Strehl ratio
Oblique Astigmatism, Z_3	0.73	0.71
Defocus, Z_4	**0.51**	**0.82**
Primary spherical, Z_{12}	0.85	0.66
Vertical secondary trefoil, Z_{16}	2.41	0.09
None/1-channel (baseline)	2.51	0.08

Table 2. Comparison of using the different amplitudes of the defocus mode within the 3-channel input on the simulated PSF dataset, SD25, where the average uncorrected $\mathrm{RMS}_{\mathrm{WFE}}$ and Strehl ratio are 2.89 rad and 0.06 respectively. The average $\mathrm{RMS}_{\mathrm{WFE}}$ and average Strehl ratio of the models on the test set are reported.

Amplitudes of the defocus mode, Z_4 (channels)	$\mathrm{RMS}_{\mathrm{WFE}}$ (rad)	Strehl ratio
[−0.5, 0, 0.5]	0.52	0.82
[−1, 0, 1]	**0.51**	**0.82**
[−1.5, 0, 1.5]	0.72	0.73

Overall, the 3-channel phase-diverse input with a ±1 rad defocus amplitude represents the optimal configuration for robust and accurate aberration correction. When evaluated using experimental data characterised by 10 Zernike modes, these phase-diversity settings achieved an average reduction in $\mathrm{RMS}_{\mathrm{WFE}}$ of 73% from 1.81 to 0.48 rad, which would result in a diffraction-limited image (see Table 3). The complete image acquisition process required an average of 2.2 seconds per image, while neural network inference on a CPU averaged 0.9 seconds. A key advantage of our approach is its consistent inference time, unaffected by aberration magnitude or complexity. When accelerated with GPU hardware, inference time is reduced to 0.02 seconds, enabling rapid and efficient processing even in the presence of severe aberrations. These timings

Table 3. Comparison of using different number of channels on the simulated PSF dataset, SD25 and experimental dataset, ED10. The average $\mathrm{RMS_{WFE}}$ and average Strehl ratio of the models on the test set are reported. Note the average uncorrected $\mathrm{RMS_{WFE}}$ and Strehl ratio are 2.89 rad/0.06 for SD25 and 1.81 rad/0.15 for ED10, respectively.

Defocus mode,		SD25		ED10	
Z_4 (channels)		$\mathrm{RMS_{WFE}}$ (rad)	Strehl ratio	$\mathrm{RMS_{WFE}}$ (rad)	Strehl ratio
1-channel	[0]	2.51	0.08	0.59	0.76
2-channel	[−1,0]	0.61	0.78	0.49	0.82
	[0,1]	0.62	0.77	0.53	0.79
	[−1,1]	0.89	0.65	0.50	0.81
3-channel	[−1,0,1]	**0.51**	**0.82**	**0.48**	**0.83**

are faster or comparable to image acquisition times in the majority of microscopes, and help to limit photodamage whilst enabling several corrections per imaging volume.

5.2. Correcting for severe aberrations

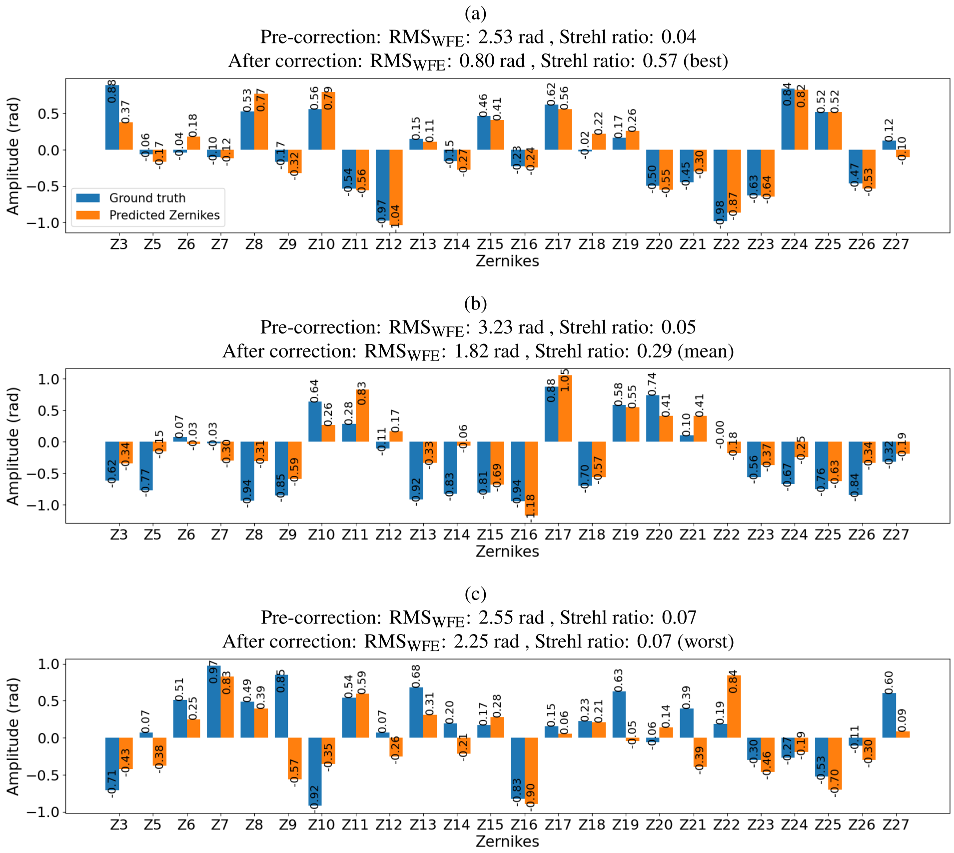
Evaluation on ED24 experimental dataset shows that using the model with optimal parameters (3-channel defocus phase-diverse input) effectively corrects substantial aberrations present in the distorted images. The model reduces the average uncorrected $\mathrm{RMS_{WFE}}$ from 2.82 to 1.68 rad when corrected, and the Strehl ratio increases from 0.06 to 0.29. While aberrations remain on this more challenging dataset with higher-order Zernike modes (24 Zernike modes), the model consistently reduces the wavefront error across the best, mean and worst performance scenarios, as demonstrated in Fig. 3.

5.3. Minimizing experimental data for aberration correction and the role of transfer learning approach on limited experimental data

Analysis of the model performance across varying experimental training dataset sizes depicted in Fig. 4 reveals that the best-performing model requires a minimum of 2,400 training sets to achieve near diffraction-limited imaging quality ($\mathrm{RMS_{WFE}} < 0.5$ rad). Furthermore, the analysis highlights the advantages of transfer learning in data-limited scenarios: models initialised with pre-trained weights consistently achieve lower $\mathrm{RMS_{WFE}}$ values compared to models trained from scratch when the training set size is less than 720 sets of images. Despite noticeable differences between the simulated and experimental aberrated images (as shown in Fig. 1), the transfer learning approach remains effective, demonstrating its robustness even with large amplitude aberrations. This transfer learning approach effectively overcomes inevitable day-to-day experimental variations—in focal plane position, SLM alignment, sample flatness, and illumination conditions-by allowing an existing network to be quickly adapted to new conditions using a small number of experimental images.

The advantages of transfer learning are most evident when training with extremely limited experimental data. With only 24 experimental training samples, the pre-trained model demonstrates more meaningful learning by reducing wavefront error through capturing directional trends in Zernike coefficients despite occasional overshooting (Fig. 5(a)). In contrast, model without pre-training exhibits zero-mean predictions rather than actual aberration correction capability. The pre-trained models effectively leverage representations learned from simulated data to provide robust initial estimates, maintaining their performance even with reduced experimental training sets of up to 720 (Fig. 4 and 5(a),(b)).

In the overall performance, as the experimental training set size expands beyond 720 images, as shown in Fig. 4, the performance gap between models with and without pre-training narrows, with the latter eventually surpassing the former. This crossover in performance can be attributed



(a)
Pre-correction: $\mathrm{RMS}_{\mathrm{WFE}}$: 2.53 rad , Strehl ratio: 0.04
After correction: $\mathrm{RMS}_{\mathrm{WFE}}$: 0.80 rad , Strehl ratio: 0.57 (best)

(b)
Pre-correction: $\mathrm{RMS}_{\mathrm{WFE}}$: 3.23 rad , Strehl ratio: 0.05
After correction: $\mathrm{RMS}_{\mathrm{WFE}}$: 1.82 rad , Strehl ratio: 0.29 (mean)

(c)
Pre-correction: $\mathrm{RMS}_{\mathrm{WFE}}$: 2.55 rad , Strehl ratio: 0.07
After correction: $\mathrm{RMS}_{\mathrm{WFE}}$: 2.25 rad , Strehl ratio: 0.07 (worst)

Fig. 3. The representative examples of Zernike coefficient predictions on the ED24 test set, correspond to the (a) best, (b) mean, and (c) worst Strehl ratio performances. The ground truth and predicted Zernike coefficients are compared for each example.

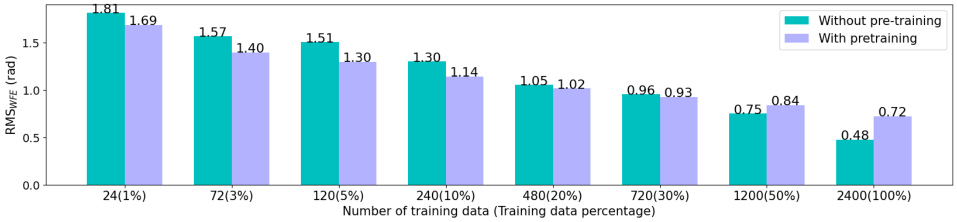
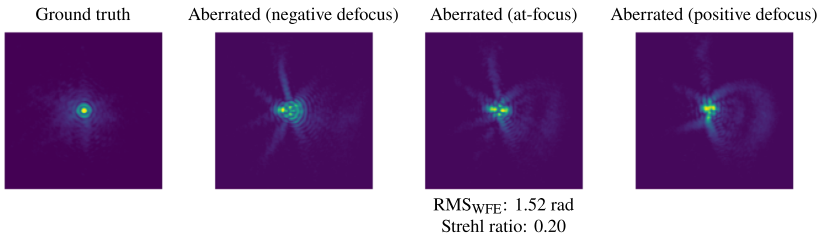


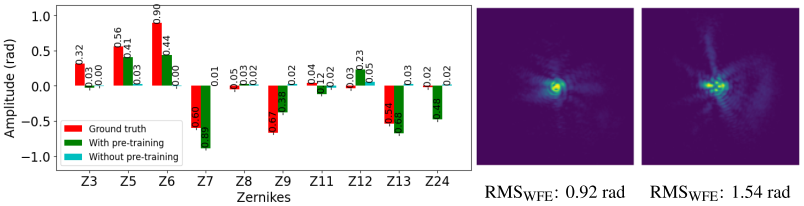
Fig. 4. Comparison of $\mathrm{RMS}_{\mathrm{WFE}}$ on test set using the models with and without pre-training fine-tuned on varying subsets of ED10 experimental dataset. Note the average uncorrected $\mathrm{RMS}_{\mathrm{WFE}}$ and Strehl ratio are 1.81 rad and 0.15 respectively.



RMS$_{\text{WFE}}$: 1.52 rad
Strehl ratio: 0.20

Zernikes prediction **With pre-training** **Without pre-training**

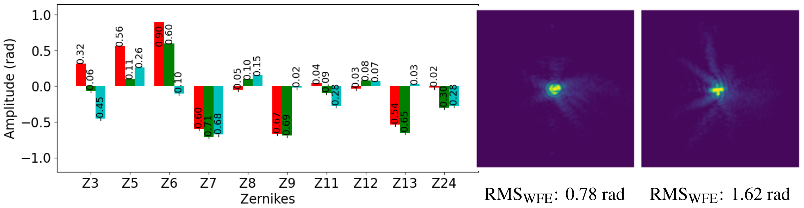
a) No. of training data: 24 (1%)



RMS$_{\text{WFE}}$: 0.92 rad
Strehl ratio: 0.43

RMS$_{\text{WFE}}$: 1.54 rad
Strehl ratio: 0.09

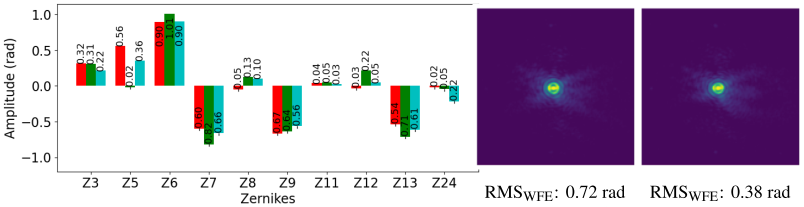
b) No. of training data: 240 (10%)



RMS$_{\text{WFE}}$: 0.78 rad
Strehl ratio: 0.55

RMS$_{\text{WFE}}$: 1.62 rad
Strehl ratio: 0.07

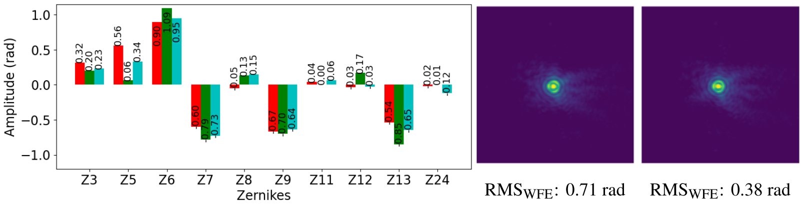
c) No. of training data: 1200 (50%)



RMS$_{\text{WFE}}$: 0.72 rad
Strehl ratio: 0.59

RMS$_{\text{WFE}}$: 0.38 rad
Strehl ratio: 0.87

d) No. of training data: 2400 (100%)



RMS$_{\text{WFE}}$: 0.71 rad
Strehl ratio: 0.60

RMS$_{\text{WFE}}$: 0.38 rad
Strehl ratio: 0.87

Fig. 5. Real-time aberration correction: the models, with and without pre-training on simulated data, were fine-tuned on varying training subsets of ED10 experimental dataset prior to evaluation on the unseen experimental data. The models predict the Zernike coefficients from the phase-diverse aberrated images (negative defocus, at-focus and positive defocus) to correct the aberrations. The top row shows examples of ground truth and aberrated images, with the corresponding ground truth Zernike coefficients shown in the bar graphs in subsequent rows. Each subsequent row presents (from left to right columns): the predicted Zernike coefficients, the corrected image using a pre-trained model, and the corrected image without pre-training.

to initialisation bias, where simulation-based pre-training weights may constrain the model to a suboptimal local minimum reflecting the simulated data rather than the characteristics of the real-world data.

5.4. *Iterative aberration correction*

Figure 6 illustrates the iterative aberration correction process using the best model trained solely on ED10 experimental dataset. In cases where the initial correction by the deep learning model is insufficient, a second correction can further improve image quality, bringing it closer to the diffraction limit, as demonstrated in Fig. 6(a). However, when the initial correction is highly effective, leaving minimal residual aberration, a second iteration may lead to overcorrection, as depicted in Fig. 6(b). This phenomenon likely occurs because small residual aberrations produce subtle PSF deformations that become difficult to distinguish from noise and imaging artifacts, resulting in less accurate prediction. In future, training a separate model specifically for small aberrations and using this network for a second iteration, could potentially address this issue.

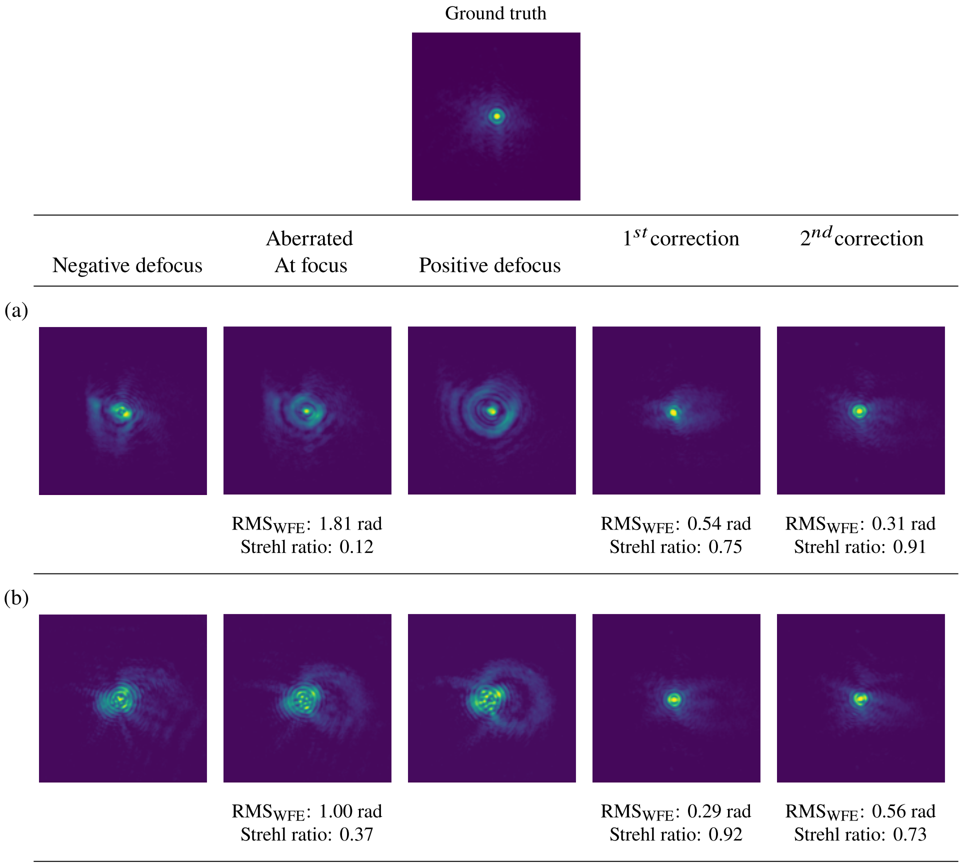


Fig. 6. Iterative aberration correction using the best model trained solely on ED10 experimental dataset. The fourth and fifth columns show the corrected images after Zernike prediction at 1^{st} and 2^{nd} iterations. (a) shows an example of improvement through iterative corrections, while (b) demonstrates potential overshoot in the 2nd iteration, as diffraction-limited quality is already achieved after the 1^{st} correction.

6. Conclusion

This study addressed some of the underlying challenges when using deep learning to predict and correct aberrations in practical experimental settings. We demonstrated the ability of deep learning to tackle severe aberrations from highly distorted wavefronts and investigated a number of different phase-diversity options to establish best practices. The successful application of deep-learning for aberration prediction in experimental data involving up to 24 Zernike modes validates its practical utility in real-world scenarios. Our investigation into phase-diversity identified that a 3-channel input using defocus-based phase-diversity with an amplitude of ±1 rad provides the most robust configuration for accurate Zernike coefficient prediction. Additionally, while our exploration of iterative correction demonstrated potential for further image quality improvement, our results indicated the need for careful consideration to avoid overcorrection effects.

A key contribution of this work addressed the challenge of large datasets required for training deep networks through transfer learning. Using a network pre-trained on simulated data and fine-tuned with experimental datasets, we demonstrated successful prediction of aberrations with minimal experimental data. Interestingly, while experimental datasets containing more than 720 image sets showed minimal benefit from pre-training, datasets with fewer than 720 image sets demonstrated clear advantages from the pre-trained network's initial representations.

Moving forward, the use of machine learning for wavefront correction shows promising potential to speed up the correction process and enable real-time adaptive optics of live biological samples. With our developed approaches, the entire experimental image capture and inference process completes in under 8 seconds, paving the way for practical computational-based aberration prediction that reduces both correction time and optical system complexity. Although the current implementation achieves substantial improvements in PSF image quality, future work could extend this approach to extended two-dimensional objects, consider a double pass through the aberrating medium and develop more sophisticated strategies for aberration correction with minimal data requirements.

Funding. Engineering and Physical Sciences Research Council (EP/T020997/1); University of Nottingham (PhD studentship).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [39]. The source code for data generation and neural network training and evaluation is available in Ref. [40].

Supplemental document. See Supplement 1 for supporting content.

References

1. M. Schwertner, M. J. Booth, M. A. Neil, *et al.*, "Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry," J. Microsc. **213**(1), 11–19 (2004).
2. M. J. Booth, "Adaptive optics in microscopy," Philos. Trans. R. Soc., A **365**(1861), 2829–2843 (2007).
3. J. D. Schmidt, M. E. Goda, and B. D. Duncan, "Aberration production using a high-resolution liquid-crystal spatial light modulator," Appl. Opt. **46**(13), 2423–2433 (2007).
4. L. Zhu, P.-C. Sun, D.-U. Bartsch, *et al.*, "Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation," Appl. Opt. **38**(1), 168–176 (1999).
5. J. M. Girkin, S. Poland, and A. J. Wright, "Adaptive optics for deeper imaging of biological samples," Curr. Opin. Biotechnol. **20**(1), 106–110 (2009).
6. M. J. Booth, "Adaptive optical microscopy: the ongoing quest for a perfect image," Light: Sci. Appl. **3**(4), e165 (2014).
7. B. C. Platt and R. Shack, "History and principles of Shack-Hartmann wavefront sensing," (2001).
8. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, vol. 16 (Oxford University Press, 1998).
9. M. A. Neil, M. J. Booth, and T. Wilson, "New modal wave-front sensor: a theoretical analysis," J. Opt. Soc. Am. A **17**(6), 1098–1107 (2000).
10. M. A. Vorontsov, G. W. Carhart, M. Cohen, *et al.*, "Adaptive optics based on analog parallel stochastic optimization: analysis and experimental demonstration," J. Opt. Soc. Am. A **17**(8), 1440–1453 (2000).
11. S. Zommer, E. Ribak, S. Lipson, *et al.*, "Simulated annealing in ocular adaptive optics," Opt. Lett. **31**(7), 939–941 (2006).

12. A. Wright, S. Poland, J. Girkin, *et al.*, "Adaptive optics for enhanced signal in CARS microscopy," Opt. Express **15**(26), 18209–18219 (2007).
13. P. Yang, M. Ao, Y. Liu, *et al.*, "Intracavity transverse modes controlled by a genetic algorithm based on Zernike mode coefficients," Opt. Express **15**(25), 17051–17062 (2007).
14. R. W. Gerchberg, "A practical algorithm for the determination of plane from image and diffraction pictures," Optik **35**(2), 237–246 (1972).
15. A. J. Janssen, "Extended Nijboer–Zernike approach for the computation of optical point-spread functions," J. Opt. Soc. Am. A **19**(5), 849–857 (2002).
16. C. Johnson, M. Guo, M. C. Schneider, *et al.*, "Phase-diversity-based wavefront sensing for fluorescence microscopy," Optica **11**(6), 806–820 (2024).
17. Q. Zhang, Q. Hu, C. Berlage, *et al.*, "Adaptive optics for optical microscopy," Biomed. Opt. Express **14**(4), 1732–1756 (2023).
18. J. Liu, P. Wang, X. Zhang, *et al.*, "Deep learning based atmospheric turbulence compensation for orbital angular momentum beam distortion and communication," Opt. Express **27**(12), 16671–16688 (2019).
19. H. Guo, Y. Xu, Q. Li, *et al.*, "Improved machine learning approach for wavefront sensing," Sensors **19**(16), 3533 (2019).
20. Y. Nishizaki, M. Valdivia, R. Horisaki, *et al.*, "Deep learning wavefront sensing," Opt. Express **27**(1), 240–251 (2019).
21. Y. Xu, D. He, Q. Wang, *et al.*, "An improved method of measuring wavefront aberration based on image with machine learning in free space optical communication," Sensors **19**(17), 3665 (2019).
22. H. Ma, H. Liu, Y. Qiao, *et al.*, "Numerical study of adaptive optics compensation based on convolutional neural networks," Opt. Commun. **433**, 283–289 (2019).
23. T. Andersen, M. Owner-Petersen, and A. Enmark, "Neural networks for image-based wavefront sensing for astronomy," Opt. Lett. **44**(18), 4618–4621 (2019).
24. T. Andersen, M. Owner-Petersen, and A. Enmark, "Image-based wavefront sensing for astronomy using neural networks," J. Astron. Telesc. Instrum. Syst. **6**(03), 1 (2020).
25. Y. Wu, Y. Guo, H. Bao, *et al.*, "Sub-millisecond phase retrieval for phase-diversity wavefront sensor," Sensors **20**(17), 4877 (2020).
26. A. P. Krishnan, C. Belthangady, C. Nyby, *et al.*, "Optical aberration correction via phase diversity and deep learning," BioRxiv (2020).
27. D. Saha, U. Schmidt, Q. Zhang, *et al.*, "Practical sensorless aberration estimation for 3d microscopy with deep learning," Opt. Express **28**(20), 29044–29053 (2020).
28. H. Ma, W. Zhang, X. Ning, *et al.*, "Turbulence aberration restoration based on light intensity image using googlenet," in *Photonics*, vol. 10 (MDPI, 2023), p. 265.
29. J. B. Shohani, M. Hajimahmoodzadeh, and H. Fallah, "Using a deep learning algorithm in image-based wavefront sensing: determining the optimum number of zernike terms," Opt. Continuum **2**(3), 632–645 (2023).
30. Q. Xin, G. Ju, C. Zhang, *et al.*, "Object-independent image-based wavefront sensing approach using phase diversity images and deep learning," Opt. Express **27**(18), 26102–26119 (2019).
31. Q. Hu, M. Hailstone, J. Wang, *et al.*, "Universal adaptive optics for microscopy through embedded neural network control," Light: Sci. Appl. **12**(1), 270 (2023).
32. J. Pawley, "Fundamental limits in confocal microscopy," in *Handbook of Biological Confocal Microscopy* (Springer, 2006), pp. 15–26.
33. Y. Jin, J. Chen, C. Wu, *et al.*, "Wavefront reconstruction based on deep transfer learning for microscopy," Opt. Express **28**(14), 20738–20747 (2020).
34. A. B. Siddik, S. Sandoval, D. Voelz, *et al.*, "Deep learning estimation of modified Zernike coefficients and recovery of point spread functions in turbulence," Opt. Express **31**(14), 22903–22913 (2023).
35. P. Zhang, C. Yang, Z. Xu, *et al.*, "High-accuracy wavefront sensing by phase diversity technique with bisymmetric defocuses diversity phase," Sci. Rep. **7**(1), 15361 (2017).
36. V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils," Appl. Opt. **33**(34), 8121–8124 (1994).
37. K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Conference on computer vision and pattern recognition* (IEEE, 2016), pp. 770–778.
38. C. Dainty, "Adaptive optics," in *Optical Imaging and Microscopy: Techniques and Advanced Systems* (Springer, 2003), pp. 235–255.
39. Y. E. Kok, A. Bentley, A. Parkes, *et al.*, "Dataset for Phase-Diverse Imaging and Models for Real-World Aberration Correction," Zenodo (2024), https://zenodo.org/records/14023331.
40. Y. E. Kok, M. P. Pound, A. J. Wright, *et al.*, "zernike-prediction," GitHub (2024), https://github.com/janetkok/zernike-prediction.git.