Research Article Vol. 16, No. 5 / 1 May 2025 / Biomedical Optics Express 2135
Biomedical Optics EXPRESS

Wavefront estimation through structured detection in laser scanning microscopy

FRANCESCO FERSINI,[1,2,7] ALESSANDRO ZUNINO,[1] PIETRO MORERIO,[3] FRANCESCA BALDINI,[4] ALBERTO DIASPRO,[4,5] MARTIN J. BOOTH,[6] ALESSIO DEL BUE,[3] AND GIUSEPPE VICIDOMINI[3,*]

[1]*Molecular Microscopy and Spectroscopy, Istituto Italiano di Tecnologia, Genoa, Italy*
[2]*Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, University of Genoa, Genoa, Italy*
[3]*Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Genoa, Italy*
[4]*Nanoscopy, and NIC@IIT, Istituto Italiano di Tecnologia, Genoa, Italy*
[5]*Department of Physics, University of Genoa, Genoa, Italy*
[6]*Department of Engineering Science, University of Oxford, Oxford, UK*
[7]*Current affiliation: OncoBreast, Inserm - CRCT, Toulouse, France*
[*]*giuseppe.vicidomini@iit.it*

Abstract: Laser scanning microscopy (LSM) is the base of numerous advanced imaging techniques, including confocal laser scanning microscopy (CLSM), a widely used tool in life sciences research. However, its effective resolution is often compromised by optical aberrations, a common challenge in all optical systems. While adaptive optics (AO) can correct these aberrations, current methods face significant limitations: aberration estimation, which is central to any AO approach, typically requires specialized hardware or prolonged sample exposure, rendering these methods sample-invasive, and less user-friendly. In this study, we propose a simple and efficient AO strategy for CLSM systems equipped with a detector array – image-scanning microscopy – and an AO element for beam shaping. We demonstrate, for the first time, that datasets acquired with a detector array inherently encode aberration information. As a proof-of-concept of this important property, we designed a custom convolutional neural network capable of decoding aberrations up to the 11^{th} Zernike coefficient, directly from a single acquisition. While this data-driven approach represents an initial exploration of the aberration content, it opens the door to more advanced decoding strategies – including model-based methods. This work establishes a new paradigm for aberration sensing in LSM and is designed to work synergistically with conventional AO approaches such as phase diversity, enabling faster, less invasive, and more accessible high-resolution imaging.

1. Introduction

Confocal laser-scanning microscopy (CLSM) is a pivotal imaging tool in life science research [1–5]. While its beam-scanning architecture results in lower imaging temporal resolution (frames *per* second) compared to its wide-field counterparts, this design also introduces significant advantages. By employing a physical pinhole to eliminate out-of-focus fluorescence signals, CLSM provides exceptional optical sectioning [6], making it ideal for three-dimensional imaging of thick samples. Additionally, CLSM can be combined with spectroscopy (hyperspectral imaging) to analyze the fluorescence emission spectrum and/or fluorescence lifetime of the sample [7]. This integration allows researchers to correlate the sample's structural features with its functional properties and phenotyping [8].

However, like all optical microscopy techniques, CLSM is susceptible to system- or specimen-induced aberrations, which can distort the optical wavefronts of both the illumination and

#559899 https://doi.org/10.1364/BOE.559899
Journal © 2025 Received 25 Feb 2025; revised 15 Apr 2025; accepted 16 Apr 2025; published 29 Apr 2025

Biomedical Optics EXPRESS

detection light, significantly degrading image quality. Adaptive optics (AO) systems address this issue by employing reconfigurable elements, such as deformable mirrors (DMs), spatial light modulators (SLMs), or adaptive lenses, to dynamically adjust the optical wavefronts and effectively compensate for aberrations [9–17]. A key element for the success of an adaptive optics approach is its ability to detect optical aberrations using simple and non-invasive methods, both for the optical architecture and the sample. There are two primary classes of aberration detection methods: direct wavefront sensing and indirect image-based estimation – often referred to as "sensorless" approaches. In wavefront sensor methods, phase aberrations are measured directly using a Shack-Hartmann sensor or an interferometer, offering speed in the estimation of the aberration but suffering from complexity in optical design and non-common path errors [18–20]. Conversely, sensorless methods utilize simpler optical designs but depend on iterative estimation of aberrations by evaluating specific image quality metrics, such as total signal intensity or spatial frequency-based sharpness, in response to phase modulations of the optical wavefront [21–24]. Phase diversity, a sensorless technique widely used in astronomy, has also been successfully adapted to fluorescence microscopy [25]. By acquiring images under known phase modulations, phase diversity constructs an inverse problem to estimate the underlying aberrations. While effective, these indirect methods are often time-consuming and require repeated exposures, which can induce photodamage or motion artifacts, particularly in live or sensitive samples.

In recent years, sensorless AO methods driven by machine learning (ML) have emerged as alternatives to traditional techniques. These approaches significantly reduce the number of sample exposures (i.e., phase modulations) required for aberration correction, offering gentler solutions for specimens while maintaining robustness. Convolutional neural networks (CNNs), in particular, are powerful tools for automatically extracting relevant features directly from raw data without requiring manual intervention. Additionally, CNNs are capable of handling large and complex datasets and can be trained on simulated data, overcoming the challenge of limited experimental data availability. Early implementations of these methods relied on access to the three-dimensional point spread function (PSF) of the system, either directly or through imaging point-like structures, such as fluorescent beads [26–29]. However, this dependence on PSFs limited their applicability to a broader range of imaging scenarios. More recently, advances in machine learning have enhanced the efficiency and adaptability of neural network-based AO methods by removing the need for point-like source imaging. For example, transfer learning frameworks and novel neural network architectures that integrate physical principles of microscope image formation directly into their models have shown promising results [30,31]. Despite these improvements, some degree of phase modulation is still required, and there remains significant potential to further enhance efficiency by providing more informative datasets for neural network training.

Here, we demonstrate how leveraging an emerging class of fast detector arrays can enhance adaptive optics (AO) capabilities for confocal laser-scanning microscopy (CLSM), enabling more versatile and sample-friendly aberration correction. Unlike conventional single-element detectors, such as photomultiplier tubes, these arrays capture spatial information that is otherwise lost. We first show that this spatial information inherently encodes details about optical aberrations affecting image quality. Building on this, we present a proof-of-principle neural network–based sensorless approach that estimates aberrations directly from a single acquisition using data collected by these fast detector arrays. Notably, these array detectors are the same ones emerging as the gold standard for image-scanning microscopy (ISM), which is gradually replacing CLSM. ISM enables effective super-resolved imaging while retaining all the functionalities of traditional confocal microscopy. In this context, we demonstrate that our approach introduces a novel CLSM implementation that synergistically and straightforwardly combines the advantages of AO and ISM. We named this approach image scanning microscopy with adaptive optics (ISMAO).

Biomedical Optics EXPRESS

Image scanning microscopy [32–35] encompasses a class of techniques that achieve effective super-resolution by leveraging the theoretical ability of confocal microscopy to double the resolution of conventional microscopy. In confocal microscopy, reducing the size of the pinhole decreases the point spread function (PSF), potentially improving resolution. However, this approach also significantly reduces the signal from the focal point, deteriorating the image signal-to-noise ratio (SNR) and negating the resolution gain. In laser-scanning architectures, such as the CLSM, ISM overcomes this limitation by substituting the pinhole and single-element detector with a fast array detector. Examples of such detectors include the AiryScan [36] and asynchronous read-out single-photon avalanche diode (SPAD) arrays [37,38]. These detectors capture microimages for each focal region probed by the laser during raster scanning, thus preserving spatial information that is discarded by the single-element detector. The captured microimages are processed to produce a final image with effectively doubled spatial resolution and improved signal-to-noise ratio (SNR). Specifically, by treating each element of the pixellated detector as a small pinhole, the microimages recorded during the scanning are transformed into a series of CLSM images (referred to as scanned images), but overall the array detector collects all fluorescence photons. Using either a simple reconstruction algorithm known as pixel reassignment [39,40] or a more computationally demanding approach such as multi-image deconvolution [41,42], the scanned images are fused in a final high-resolution and high-SNR image. Essentially, the scanned images are spatially registered, using a phase-correlation approach, before integration. More recently, novel ISM reconstruction algorithms – in the context of pixel assignment [43] or multi-image deconvolution [44] – have been introduced, enabling resolution enhancement without compromising the optical sectioning capability of CLSM.

In this work, we demonstrate that the ISM dataset inherently contains information not only to enhance resolution but also to extract the optical wavefront of the illumination and detection beams. The array detector functions similarly to a scanning Shack-Hartmann sensor, enabling the decoding of the aberrations without the need for extensive phase-modulation approaches or dedicated sensors. In other words, the same detector used for imaging also performs wavefront sensing. To this end, we developed a pipeline to pre-process the data to create a specimen-free dataset for a custom CNN. This latter decodes the phase information from the data, returning the coefficient of the Zernike polynomials required to describe the wavefront error that corrupts the images. We trained the CNN on synthetic datasets generated through a comprehensive image formation simulation framework, parametrizing a variety of imaging conditions. We improve this approach by fine-tuning the network with just a limited amount of experimental data, thanks to the transfer learning approach, which enhances the network's performance in practical applications. Finally, we demonstrated the capability of ISMAO to correct aberrations in a custom-built ISM system featuring a SPAD array detector and a DM integrated into both the excitation and detection paths. Our results represent a step forward toward a deeper understanding of the physics of structured detection in laser-scanning microscopy and its potential for applications in adaptive optics.

2. Results

2.1. Principle of ISMAO: aberrations encoding

In this section, we demonstrate that ISM imaging datasets inherently capture information about the optical aberrations affecting the images, offering new insights for aberration sensing in laser-scanning microscopy.

In an ISM setup (Fig. 1 a and Supplement 1, Fig. S1, Fig. S2), a focused laser beam excites the sample point-by-point at the scan coordinates $\mathbf{x_s} = (x_s, y_s)$. Then, the microscope collects and descans the fluorescence light, imaging the probed region onto the sensitive area of the detector array and generating a microimage with detector coordinates $\mathbf{x_d} = (x_d, y_d)$. Each pixel of the detector array acts as a small pinhole, located in a unique position of the detector plane.

Biomedical Optics EXPRESS

With a full scan, the signal collected by each pixel arranged along the scan coordinates forms a confocal-like image of the same specimen observed from a slightly different point of view [41]. Thus, the ISM dataset $i(\boldsymbol{x}_s \mid \boldsymbol{x}_d)$ is four-dimensional and can be described either as a large collection of microimages (Fig. 1(b)) or as a small collection of scanned images (Fig. 1(c)). The latter perspective allows for the following description of the ISM image formation:

$$i(\boldsymbol{x}_s \mid \boldsymbol{x}_d) = o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s \mid \boldsymbol{x}_d) \tag{1}$$

where $*$ is the convolution operator, $o(\boldsymbol{x}_s)$ is the object function – namely the distribution of fluorophores in the specimen – and $h(\boldsymbol{x}_s \mid \boldsymbol{x}_d)$ is the point spread function (PSF) of the imaging system. The latter is given by

$$\begin{aligned} h(\boldsymbol{x}_s \mid \boldsymbol{x}_d) &= h_{\text{exc}}(-\boldsymbol{x}_s) \cdot h_{\text{det}}(\boldsymbol{x}_s \mid \boldsymbol{x}_d) \\ &= h_{\text{exc}}(-\boldsymbol{x}_s) \cdot [h_{\text{em}}(\boldsymbol{x}_s) * p(\boldsymbol{x}_s - \boldsymbol{x}_d)] \end{aligned} \tag{2}$$

where h_{exc}, h_{em}, and h_{det} are the excitation, emission, and detection PSF of the system, respectively. The first describes the focused excitation beam, while the second describes the collected fluorescence light. Finally, p is the characteristic function of the sensitive element of the array detector located at coordinate $\boldsymbol{x}_d$.

In an ideal scenario, both excitation and detection PSFs are diffraction-limited. Thus, their product is narrower than the diffraction limit, enabling super-resolution. However, in a realistic scenario, the optical performances might be compromised by wavefront distortions, which degrade the imaging quality and limit the effective image resolution and signal-to-noise ratio (SNR). Using scalar diffraction theory [45], the relation between the PSF and the wavefront at the pupil plane is given by

$$h_\alpha(\boldsymbol{x}_s) = \left| \mathcal{F}\left\{ P(\boldsymbol{x}_p) \cdot \exp[iW_\alpha(\boldsymbol{x}_p)] \right\} \left(\frac{\boldsymbol{x}_s}{\lambda_\alpha f} \right) \right|^2 \tag{3}$$

where $\mathcal{F}$ is the Fourier transform operator, f the focal length of the objective lens, $\boldsymbol{x}_p$ are the coordinates of the pupil plane, P is the amplitude pupil function and W is the wavefront at the pupil plane of the objective lens. The label α is a place-holder for either excitation or emission. While vectorial diffraction theory is more appropriate for high numerical aperture lenses, aberrations are still described as wavefront distortion at the pupil plane [46].

The specimen itself is the most common source of optical aberrations. As light travels through the layers of a thick sample (Fig. 1(d)), wavefront distortions accumulate, affecting both the excitation and fluorescence light. These distortions lead to aberrations in the optical imaging system, which are directly reflected in the shape of the PSF. Consequently, conventional wavefront sensing methods require access to the system PSF, typically obtained by imaging point-like structures [47]. Notably, also different sensorless ML-based methods use three-dimensional imaging of point-like structures to estimate the wavefront and, by extension, the aberrations. In contrast, we demonstrate that wavefront information is inherently encoded within a single-plane ISM dataset of a generic structure, thanks to the unique capabilities of array detection. This innovation eliminates the need for imaging three-dimensional point-like structures, offering a more efficient and accessible approach to aberration estimation.

The rationale behind the ISMAO approach is that all scanned images originate from the same source and represent the same specimen. Therefore, any variations between these images must be attributed to differences in the PSF. Since the structure of the PSF uniquely encodes aberrations, we can leverage the diversity of these images to recover the wavefront, eliminating the need for a dedicated sensor or direct measurement of the PSF, relying solely on the detector array itself. Nonetheless, we need a strategy to decode the wavefront information independently from the

Biomedical Optics EXPRESS

observed specimen. To this end, we need to measure a quantity that depends uniquely on the PSFs of the system.

A natural choice would be the fingerprint $f(\boldsymbol{x}_d)$, which is given by the sum of all the micro-images collected during the scan (Fig. 2(a)). Assuming that the signal is largely in focus, the fingerprint is independent of the structure of the specimen and is given by

$$f(\boldsymbol{x}_d) = \int_{\mathbb{R}^2} i(\boldsymbol{x}_s \mid \boldsymbol{x}_d)\, \mathrm{d}\boldsymbol{x}_s \propto h_{exc}(\boldsymbol{x}_d) \star h_{det}(\boldsymbol{x}_d) \tag{4}$$

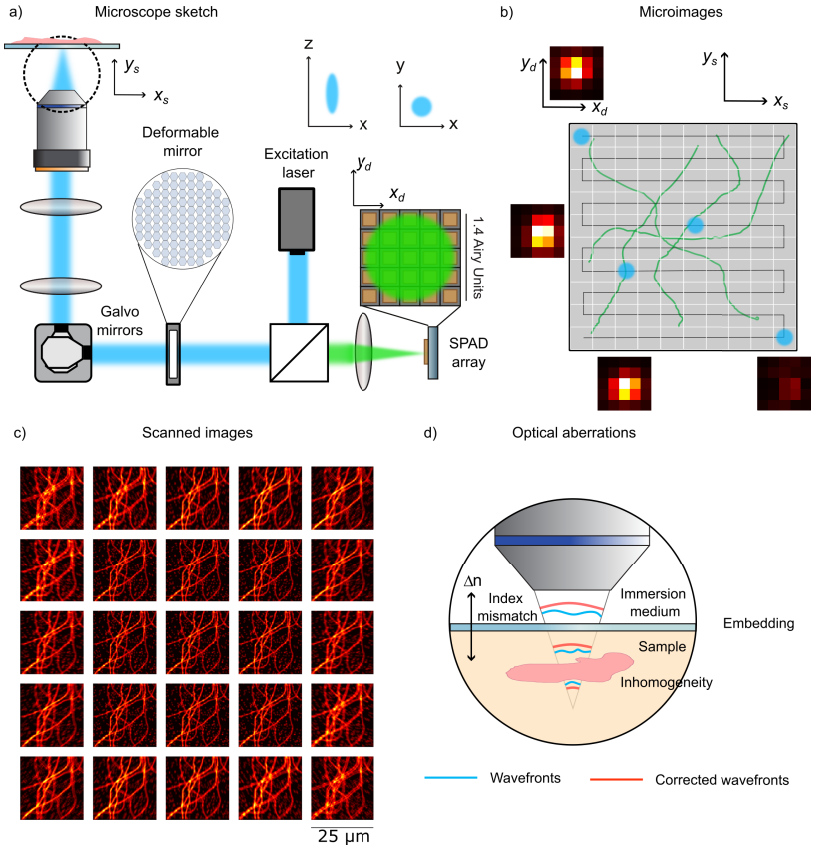

Fig. 1. Image scanning microscopy with adaptive optics. a) Schematic of the ISMAO setup. The system features a fast detector array in the detection path and an Alpao deformable mirror with 97 actuators, positioned in a plane conjugate to the objective lens. ISMAO produces a multidimensional output consisting of both micro-images and scanned images. b) Representation of the micro-images: during the raster scanning of the sample, each scanned point generates a micro-image composed of a 5×5 intensity image. These images represent the distribution of light among the SPAD sensors, providing direct information about the status of the alignment, distortion, and PSF. c) The 25 confocal-like scanned images collected by each element of the detector array. These images capture the same portion of the sample but are observed from different points of view. Following the intensity distribution described by the micro-images, the peripheral images appear dimmer compared to the central one. For better visualization, the images are plotted with rescaled intensity values. d) An example of wavefront deformation caused by a thick sample with a refractive index mismatch, which generates an ill-defined diffraction-limited excitation spot.

Biomedical Optics EXPRESS

where ⋆ is the cross-correlation operator. In other words, the fingerprint describes how light is distributed on the detector plane with a full scan [41].

Through simulations, we illustrate how aberrations alter the intensity distribution of the fingerprint, becoming no longer sharply centered in the detector array but spread across the entire detector plane (Fig. 2(a), and Supplement 1, Fig. S3, Fig. S4, Fig. S5, Fig. S6). Therefore, aberrations directly affect the structure of the fingerprint distribution, enabling almost direct access to the PSF of the microscope of the system without the need for point-like sources, acting as guide-stars. Nonetheless, to maintain a temporal resolution compatible with fast scanning,

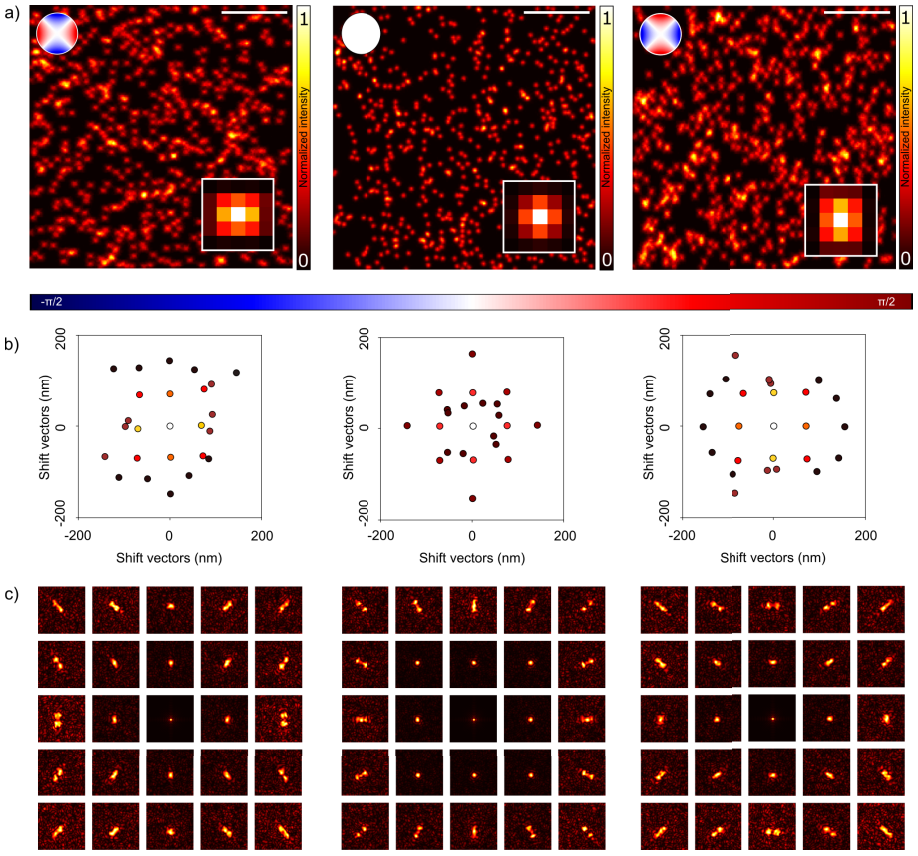

Fig. 2. This figure demonstrates how an array detector inherently encodes optical aberration information. a) The sum of scanned images is displayed, with phase modulation introduced by a deformable mirror. The phase mask used in each simulation is indicated in the top-left corner. Aberrations in the PSF alter the intensity distribution across the active area of the sensor, modifying the fingerprint shown in the bottom-right corner of each simulation. Scale bar 10 μm. b) Shift vectors provide an alternative method for estimating optical aberrations. These vectors are calculated from the intensity values corresponding to the maximum correlation between the central image and others. While less comprehensive than fingerprints, they offer a direct measure of aberrations. Each sensor element is assigned a shift vector, represented using the same color coding as the fingerprint. c) Correlograms encode optical aberration information directly and are independent of the sample structure. Unlike scanned images, correlograms preserve the physical and optical properties of the PSF, making them a more robust and generalizable approach for aberration analysis compared to shift vectors.

Biomedical Optics EXPRESS

the ISM detector array possesses only a few sensitive elements (twenty-five for the SPAD array detector used in this work), which results in an inadequate sampling to robustly retrieve a complex wavefront.

A second approach for quantifying the aberrations lies in the measurement of the shift vectors, namely the translations required to maximize the similarities between the raw images of the ISM dataset to the one generated by the central element of the detector array. Indeed, the scanned images are approximately identical, but shifted and rescaled, as long as the magnification of the system is sufficiently high [48]. A well-established algorithm, known as adaptive pixel reassignment (APR), exploits this concept to shift back and sum the raw images, reconstructing a single super-resolution image from the raw data [49,50]. Given the linearity of fluorescence image formation, a shift in the images uniquely relates to a shift in the PSFs. Therefore, if one or multiple aberrations distort the shape of the PSFs, shift vectors change accordingly, encoding the wavefront information in their pattern (Fig. 2(b)). However, shift-vectors are well-defined only for single-peaked PSFs, a condition fulfilled only without strong aberrations and using large magnification values. Given the limitations of the previous two approaches, we developed a more comprehensive strategy that inherently contains all the information provided by the two approaches. In detail, we leverage the phase cross correlation between the scanned images to calculate the correlograms, which depend uniquely on the PSFs of the microscope (Fig. 2(c)).

The cross-correlation between an image of the ISM dataset and the reference image at the center of the array detector ($\boldsymbol{x}_d = \mathbf{0}$) is the following:

$$\begin{aligned} \boldsymbol{x}_s \mid \boldsymbol{x}_d) = i(x_\mathrm{s} \mid x_\mathrm{d}) \star i(x_\mathrm{s}|\mathbf{0}) = \\ = [o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d)] \star [o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\mathbf{0}] \end{aligned} \tag{5}$$

where $*$ and $\star$ are the convolution and the cross-correlation operator, respectively. Then, we apply the Fourier transformation with respect to the scan coordinates $\boldsymbol{x}_s$. Using the convolution theorem, we obtain

$$\begin{aligned} C(\boldsymbol{k}_s \mid \boldsymbol{x}_d) = I(\boldsymbol{k}_s \mid \boldsymbol{x}_d) \cdot \bar{I}(\boldsymbol{k}_s \mid \mathbf{0}) = \\ = |O(\boldsymbol{k}_s)|^2 \cdot H(\boldsymbol{k}_s|\boldsymbol{x}_d) \cdot \bar{H}(\boldsymbol{k}_s|\mathbf{0}) \end{aligned} \tag{6}$$

where the overline represents the complex conjugate operation, $\boldsymbol{k}_s$ is the scan spatial frequency, and the uppercase letters indicate the Fourier transform of the quantities previously indicated with the lowercase letter. To discard the object contribution, we need to normalize the correlation in Fourier space.

$$R(\boldsymbol{k}_s \mid \boldsymbol{x}_d) = \frac{C(\boldsymbol{k}_s \mid \boldsymbol{x}_d)}{|C(\boldsymbol{k}_s \mid \boldsymbol{x}_d)|} = \frac{H(\boldsymbol{k}_s|\boldsymbol{x}_d) \cdot \bar{H}(\boldsymbol{k}_s|\mathbf{0})}{|H(\boldsymbol{k}_s|\boldsymbol{x}_d)| \cdot |H(\boldsymbol{k}_s|\mathbf{0})|} \tag{7}$$

Finally, we obtain the correlograms by inverting the Fourier transform and we normalize with the fingerprint

$$r(x_s \mid x_d) = \frac{\mathscr{F}^{-1}\{R(k_s \mid x_d)\}}{f(x_d)} \tag{8}$$

The result depends uniquely on the PSFs of the microscope, regardless of the structure of the specimen. Notably, the coordinates of the maximum of the correlograms are, by definition, the shift vectors. Therefore, we designed a quantity that encodes the highest amount of information about the wavefront without the bias of a specific sample.

2.2. Principle of ISMAO: aberrations decoding

In this section, we demonstrate how the optical aberrations encoded within the ISM dataset can be effectively decoded and used to control an AO element for precise aberration correction.

Two approaches can be envisioned for this task: a model-based approach and a data-driven approach. The model-based approach involves inverting the correlogram formation model by

Biomedical Optics EXPRESS

solving a minimization problem to minimize the error between the model and the data. While this method leverages an explicit physical model, it requires a relatively slow and computationally intensive minimization procedure. In contrast, the data-driven approach uses a convolutional neural network (CNN) to directly retrieve the Zernike polynomial coefficients from the data. This method does not rely on a predefined model; instead, the network learns the necessary relationships through the training procedure. We trained the CNN using physically accurate numerical simulations of correlograms. Additionally, the deformable mirror (DM) in our custom setup allows for the straightforward acquisition of experimental data, which can further refine the network.

Given the ease of data generation for the training and the speed of the CNN predictions, we opted for the CNN approach. Indeed, the speed of a trained neural network better fits the needs of real-time adaptive optics systems, where aberrations must be estimated rapidly to drive the corrective optical elements.

We implemented two closely related CNNs that differ mainly in their input datasets. The first CNN (CNNx1) receives as input the normalized correlograms obtained from a single focal plane ISM acquisition. In contrast, the second CNN (CNNx3) processes normalized correlograms derived from a three-planes ISM acquisition. To acquire more information, we introduce a defocus as large as half of the depth-of-field (DoF), where DoF = $(2\lambda n)/(\mathrm{NA}^2)$, λ is the wavelength of light, n is the refractive index of the immersion medium, and NA is the numerical aperture of the objective lens. For CNNx3, two additional focal planes, positioned at $-\mathrm{DoF}/2$ and $+\mathrm{DoF}/2$ relative to the original plane, are included as inputs. This multi-plane approach enhances aberration decoding, particularly for aberrations that affect axial symmetry, such as astigmatism or spherical aberrations. The design of CNNx3 draws inspiration from the phase-diversity strategy [25], in which multiple images are acquired with controlled, known wavefront distortions to estimate the original wavefront. Unlike phase diversity, where multiple images with additional aberrations are used to refine the wavefront estimation, our method introduces defocus to provide complementary information, improving robustness and capturing a wider range of aberration types.

To generate the training set, we used the open-source BrightEyes-ISM Python package [42]. To generate ISM images under varying optical aberrations, we parameterize the wavefront distortions at the objective pupil plane – common for the excitation and emission PSFs – using a Zernike polynomial basis:

$$W(\rho, \varphi) = \sum_{i=5}^{11} c_i \cdot Z_i(\rho, \varphi) \tag{9}$$

Here, ρ is the normalized radius of the pupil, φ is the azimuthal angle of the pupil, c_i is an amplitude expressed in radians, and Z_i is the Zernike polynomial i (using the Noll index notation). Note that we neglect the first four orders of aberrations (piston, tip, tilt, and defocus) because they describe a simple misalignment that we are not interested in correcting through adaptive optics. Furthermore, we interrupt the series expansion at the 11^{th} polynomial to take into account only the primary aberrations. Neglecting higher-order aberrations balances computational efficiency while still addressing the most common aberrations encountered in microscopy. Once the ISM dataset is generated, we calculate the normalized correlograms to feed the CNN (Eq. (8)).

To account for experimental uncertainties during CNN training, we introduced several variations. The number of simultaneously occurring aberrations was randomly chosen, with their amplitudes sampled from a distribution defined as the sum of two Gaussian, each with $\mu = \pm 0.75, \mathrm{rad}$ and $\sigma = 0.3$, and restricted to an RMSE range of $\pm\pi/2, \mathrm{rad}$ (Fig. 3(a)). Additionally, the emission photon flux and pixel dwell time were adjusted to reflect diverse signal-to-noise imaging conditions (Fig. 3(b)-c). Variations in the numerical aperture (NA) of the objective were also introduced to simulate additional experimental conditions.

Biomedical Optics EXPRESS

2.3. ISMAO validation on synthetic images

To evaluate the performance of the ISMAO approach, we implemented and tested two CNNs models. Both models were trained on the same synthetic dataset to ensure consistency and reproducibility. The validation protocol was conducted in two phases.

In the first phase, the CNNs were employed to identify individual aberrations, with only one aberration introduced at a time. This phase aimed to assess the baseline capability of the models to predict aberrations individually. For this test, the validation dataset consisted of 490 simulations, corresponding to 7 aberration types, each evaluated at 7 different amplitude points, with 10 repetitions for each aberration amplitude. Specifically, the amplitude of the test dataset ranged from -1.5 to 1.5 rad, with a step size of 0.5 rad. This experiment demonstrated that CNNx3

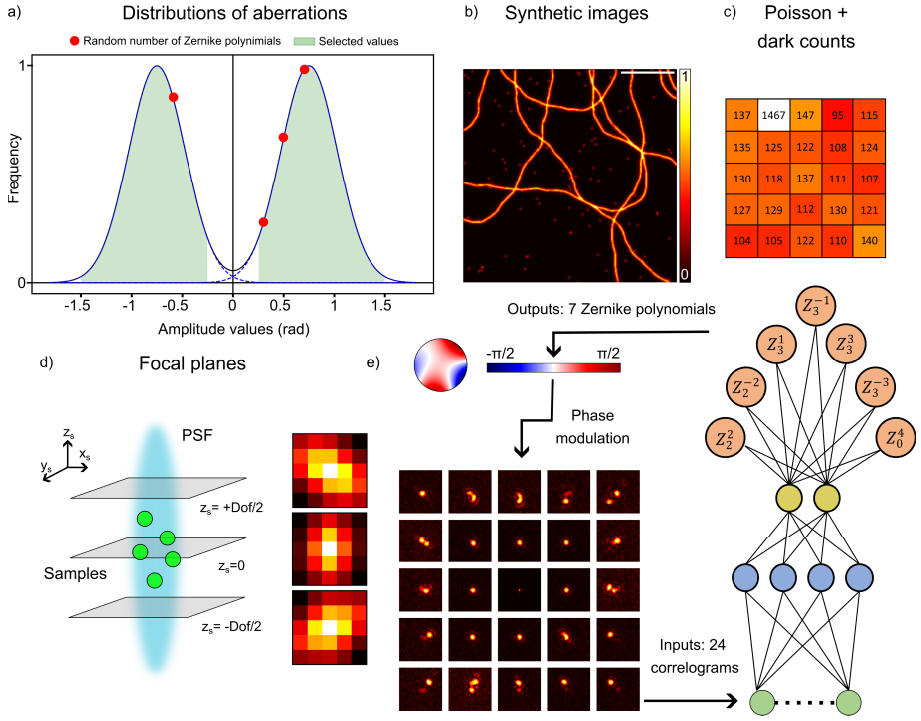

Fig. 3. a) The first panel illustrates the density distribution of optical aberrations randomly selected to generate the simulated phase mask of the PSF. The green area under the curves represents the density distribution range considered for the simulations. The number of simultaneous aberrations is defined randomly; an example in the graph shows four selected aberrations. b-c) Show an example of a simulated phantom with scale bar 25 μm, including 3D tubulin filaments and beads, which are used for testing and the photon dark count collected during a 1-second acquisition with the array detector, replicating realistic detector noise during pixel dwell time acquisition. d) The fourth panel represents the different focal planes simulated. The axial asymmetry of certain aberrations highlights the advantage of collecting data from multiple focal planes to improve aberration identification. e) ISMAO approach. The correlograms (normalized with the fingerprint) are used as input data for the CNN, which is based on the AlexNet structure. It has seven outputs corresponding to seven Zernike polynomial coefficients from 5^{th} to 11^{th}. Alternatively, the CNN can process 3×24 correlograms if data is collected from three focal planes. The outputs are then used to drive an adaptive element in the microscope to apply phase modulation to the experimental PSF and compensate for the predicted aberrations.

Biomedical Optics EXPRESS

more accurately tracked the trends of imposed aberrations compared to CNNx1 (Fig. 4(a), and Supplement 1, Fig. S7). While CNNx1 provided predictions comparable to CNNx3 for smaller aberrations, however, its performance suffered for larger aberrations. Additionally, CNNx3 exhibited enhanced robustness in handling aberrations that disrupted axial symmetry, such as astigmatism and spherical aberration. This trend was further confirmed by the correlation matrix (Fig. 4(b)), which evaluates the ability of the model to accurately identify specific aberrations. The correlation matrix reflects the strength of the relationship between real and predicted data across the set of aberration variables.

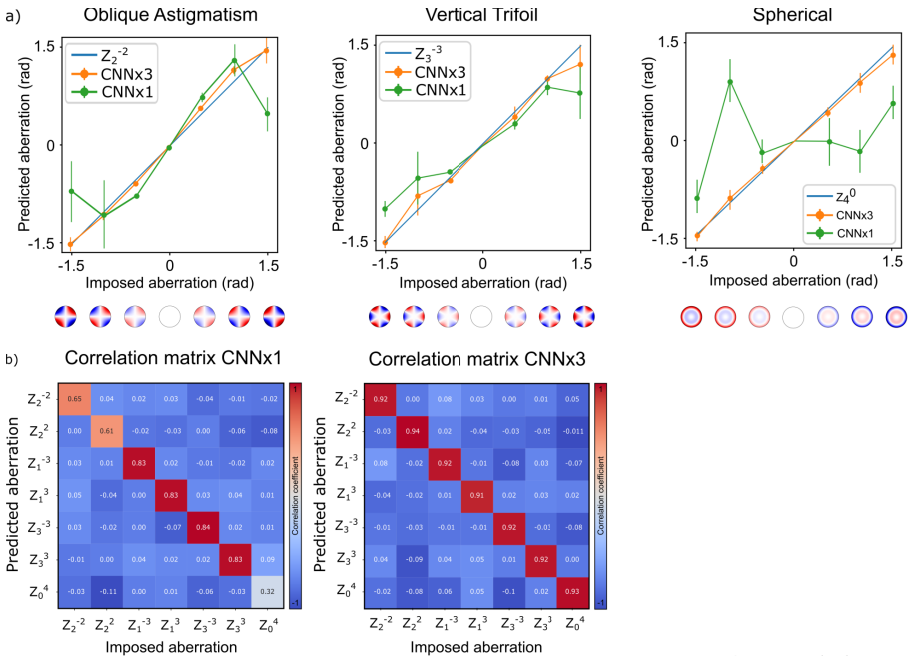

Fig. 4. This figure presents a portion of the results from 490 simulations performed after phase retrieval using CNNx1 and CNNx3. Each simulation includes only one aberration at a time. a) The plots display a comparative analysis of phase retrieval performance between the two CNN models. The analysis focuses on three specific aberrations: oblique astigmatism (Z_2^{-2}), vertical trefoil (Z_3^{-3}), and spherical aberration (Z_4^0), each varying within the range of $[-1.5, 1.5]$ radians in 0.5 radian increments. The results demonstrate that CNNx3 (orange) more accurately approximates the imposed aberration (blue) compared to CNNx1 (green). b) The correlation maps summarize the full dataset of 490 simulations for both CNNs. This representation highlights that CNNx3 achieves a correlation score of 92%, significantly outperforming CNNx1, especially in estimating aberrations with axial asymmetry.

In the second phase, a more complex scenario was introduced, wherein multiple aberrations were simultaneously imposed. This phase allowed for a comprehensive evaluation of each CNN ability to manage the compounded effects of overlapping aberrations, providing an assessment of their robustness and generalizability under more realistic conditions.

For this phase, we employed simulated datasets designed to replicate realistic experimental scenarios, where multiple aberrations commonly occur due to factors such as sample thickness, refractive index mismatches, and system-induced aberrations. To simulate this complexity, we generated a dataset of 700 simulations for each focal plane, with the number and amplitude of simultaneously occurring aberrations chosen randomly based on a double Gaussian distribution, as during CNN training.

Biomedical Optics EXPRESS

We summarize the results for both CNN models using a scatter plot of the imposed RMS aberration versus the residual RMS error (Fig. 5(a)). The RMSE is computed as the difference between the imposed aberration and the phase retrieved by CNNx1 and CNNx3. The imposed RMS values ranged from 0.25 to 1.5 rad (values below 0.25 rad were discarded due to their negligible effect on image quality). The scatter plot clearly illustrates that CNNx3 consistently yielded lower residual errors compared to CNNx1.

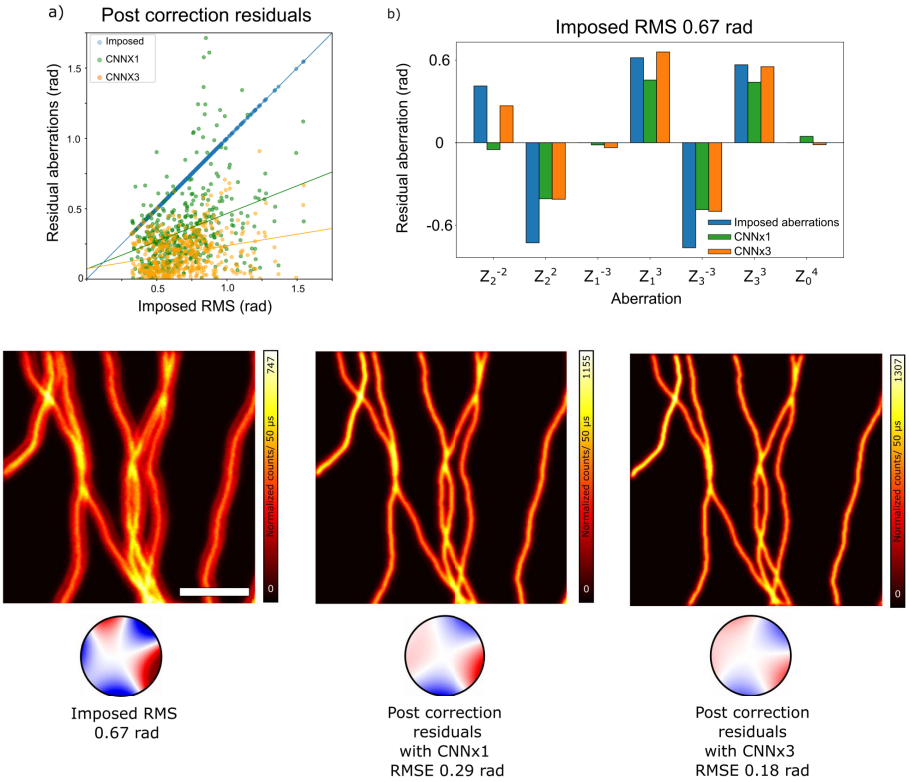

Fig. 5. To assess the error in estimating aberrations in a more complex scenario, we analyzed randomly selected aberrations occurring simultaneously. a) The figure shows the residual errors from 700 simulations after phase retrieval using CNNx1 and CNNx3. In the scatter plot, the imposed aberrations are shown in blue, while the estimated residual aberrations are displayed in green and orange for CNNx1 and CNNx3, respectively. b) A specific set of imposed aberrations with an RMSE value of 0.67 rad is highlighted from the scatter plot. The imposed aberrations and the predicted amplitudes for each aberration are reported in the histogram. This example illustrates how CNNx3 performs better than CNNx1, even in complex scenarios with multiple aberrations occurring simultaneously. Additionally, the bottom row shows the corresponding images along with their respective phase masks. These images demonstrate an enhancement in resolution compared to the aberrated image when using CNNx1, and a further improvement, including an increase in photon counts, when using CNNx3. Scale bar: 5 μm.

To further assess model performance under challenging conditions, we conducted a synthetic experiment (Fig. 5(b)). In this scenario, a total RMS of 0.67 rad was imposed across multiple aberrations. Both CNN models successfully retrieved the imposed aberrations and applied corrections to the same images. Despite the complexity of this scenario, in which five aberrations with varying amplitudes were simultaneously present, both models were able to estimate the imposed aberrations. Notably, predictions from CNNx3 closely aligned with the ground truth, as

Biomedical Optics EXPRESS

shown in the histogram of imposed aberration values. Although CNNx1 showed good overall predictive accuracy, it exhibited limitations in specific cases, such as astigmatism, where its predictions were less precise compared to CNNx3, as expected (Supplement 1, Fig. S7).

Following the application of corrections derived from the predictions, CNNx1 reduced the RMSE to 0.29 rad, while CNNx3 achieved a more precise correction, with an RMSE of 0.18 rad. The corrected images highlight the robustness of the approach, demonstrating its ability to predict and compensate for aberrations in a single acquisition. Additionally, we observed that CNNx1 performed well in less complex scenarios but faced challenges when dealing with aberrations exhibiting significant axial asymmetry, such as the simultaneous presence of astigmatism and spherical aberration (Supplement 1, Fig. S8). These results underscore the superior accuracy of CNNx3 in managing complex aberration profiles, while also demonstrating the efficiency and reliability of CNNx1 in less demanding scenarios, where a single acquisition can provide sufficient precision for aberration estimation.

2.4. ISMAO validation on experimental images

To validate the ISMAO method on real datasets, we implemented a custom ISM system incorporating a 5×5 single-photon avalanche diode (SPAD) array detector and a deformable mirror (DM) positioned within the shared path of the excitation and emission beams. In this work, the DM serves a dual role: it can correct specimen- and system-induced aberrations identified through an AO method, such as our ISMAO approach, but it can also be used to introduce well-defined, controlled optical aberrations into the system. This latter capability enabled the implementation of a transfer learning approach, enriching the CNN training process with conditions not fully captured in purely synthetic datasets. Specifically, we collected a series of ISM datasets by applying specific single-coefficient aberrations. To expand the dataset size, we employed data augmentation strategies and integrated this relatively small real dataset (5–10% of the total training dataset) with the synthetic dataset. The combined synthetic and real dataset was then used to train the CNNs effectively.

For transfer learning, we used images of lipid droplets [51–53] (Fig. 6(a)) and fluorescent beads (Supplement 1, Fig. S9). Specifically, we captured a series of ISM datasets by introducing single aberrations at varying strengths using the deformable mirror. The Fig. 6 a) illustrates the data acquisition process used for transfer learning. In this case, a specific trefoil aberration (Z_3^{-3})–is applied with varying amplitudes of $[-\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}]$, one for each image. At the bottom right of each image, the corresponding experimental fingerprint from the detector array is displayed. These fingerprints are normalized to the same maximum value to better highlight their differences–specifically, the central image appears sharper compared to the others. Additionally, each image includes the corresponding shift vectors, which exhibit the characteristic "arrow" shape associated with trefoil aberration. These images are subsequently used to train the CNNs, with data augmentation applied to enhance generalization during the learning process. The data acquisition process from the experimental samples has been repeated for all the aberrations under consideration, using the same amplitude interval range.

The newly trained CNNs were subsequently tested on an entirely novel dataset obtained by imaging the cytoskeleton of a cell. This dataset represented a true three-dimensional sample, containing both in-focus and out-of-focus structures, thereby providing a realistic and challenging validation scenario. To validate the ISMAO approach, we introduced three aberrations simultaneously during the imaging experiments, each with random amplitudes, resulting in an RMSE of 1.4 rad, with a prominent contribution from astigmatism. (Fig. 6 b). The resulting image displayed noticeable blurring, making tubulin filaments indistinguishable, particularly those highlighted in the green box. The calculation of the FRC on the raw image estimated a resolution of 302 nm – well above the expected resolution. Subsequently, correlograms were computed from the aberrated image, and the CNN estimated the imposed aberrations, applying corrections

Biomedical Optics EXPRESS

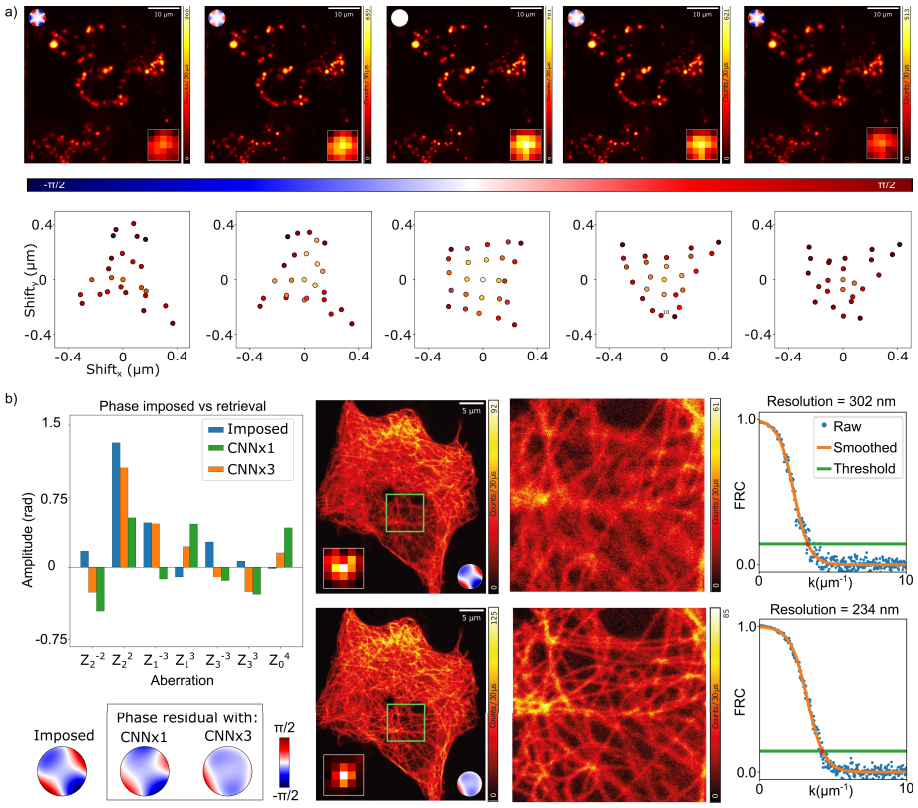

Fig. 6. a)This panel illustrates the data acquisition process used for transfer learning. A specific trefoil aberration (Z_3^{-3}) is introduced at varying amplitudes. The corresponding phase mask for each aberration level–the surface configuration of the deformable mirror–is shown in the top left corner of each image. Each image also includes its associated fingerprint in the bottom right corner and the corresponding shift vectors.b) This panel displays an experimental dataset acquired from fixed cells, which exhibit poorly defined 3D structures. The histogram summarizes the various aberrations introduced simultaneously, resulting in a root mean square error (RMSE) of 1.4 radians. The imposed aberrations are shown in blue, while the predictions from CNNx1 and CNNx3 are represented in green and orange, respectively. The imposed phase mask is also shown, demonstrating CNNx3's ability to correct aberrations–evident from the resulting flattened phase profile. This correction is further supported by the improved experimental images, as indicated by enhanced image quality, sharper fingerprints, and better FRC metrics when compared to the original, uncorrected images.

that yielded the improved image, where the tubulin structures became visible compared to the previous one. This correction also led to an increase in SNR of 24% and resolution of almost 76%. This enhancement demonstrates the ability of the CNN and the ISMAO approach to discern intricate details within complex cellular structures.

3. Conclusion

This work introduces ISMAO, a novel sensorless aberration correction strategy for LSM that leverages the unique capabilities of array detector-based imaging (i.e., image scanning microscopy) to provide access to the light (illumination and fluorescence) wavefronts. More generally, we

Biomedical Optics EXPRESS

demonstrate that ISM datasets inherently encode information about optical aberrations, and the CNN serves as an initial attempt to decode this information.

The ISMAO approach employs a custom-designed CNN that processes ISM imaging datasets, and through pre-processing steps grounded in the image formation principles of ISM, achieves robust and efficient aberration prediction. These steps, particularly the calculation of the cross-correlogram, ensure the method is independent of sample structure, making ISMAO broadly applicable. We demonstrate that the CNN trained on synthetic ISM datasets generated through a comprehensive simulation framework can predict aberrations under varied imaging conditions. Transfer learning further enhances the performance, allowing ISMAO to operate effectively even with limited experimental data.

A key innovation of ISMAO is its ability to deliver aberration predictions from a single focal-plane image, with further improvements observed when using a multi-plane acquisition strategy. With just three planes, this approach enriches the network's input with diverse and informative data, significantly improving its capacity to decode complex aberration patterns.

This study highlights the potential of ISMAO for fast aberration estimation and adaptive optics in real-time applications. By leveraging the information encoded in ISM datasets, ISMAO opens the door to new aberration correction possibilities that do not rely on additional dedicated sensors or imaging of point-like structures, making the technique suitable for a wide range of microscopy applications. While the present ISMAO method shows promising results, further validation with more challenging samples, such as tissue or multi-cellular specimens, is needed to fully assess its effectiveness and generalizability. Future work will focus on refining the approach and improving its accuracy, particularly by combining ISMAO with other adaptive optics strategies, such as phase diversity or novel decoding approaches. For instance, more conventional regression analysis could be applied to the aberration "fingerprint," especially as larger detector arrays (e.g., 7×7 elements) become available, offering increased precision and flexibility. Additionally, we plan to expand the use of the ISMAO strategy beyond adaptive optics, supporting advanced post-processing algorithms like multi-image deconvolution, leveraging realistic PSFs directly estimated from the imaging dataset.

By further refining this framework, we aim to unlock new possibilities for next-generation microscopy techniques based on SPAD array detectors. The unique spatiotemporal information provided by SPAD arrays enables seamless integration of ISMAO with other advanced microscopy techniques, such as stimulated-emission depletion microscopy (STED) [41,43,50], fluorescence correlation spectroscopy (FCS) [54,55], and fluorescence lifetime imaging (FLIM) [56–59]. SPAD-based ISM implementations have also been successfully combined with two-photon excitation [60] and, more recently, with single-molecule localization microscopy [61,62]. These results open up exciting opportunities for follow-up research aimed at integrating ISMAO into these advanced microscopy techniques.

4. Methods and materials

4.1. Optical architecture

The experimental setup for ISM comprises a deformable mirror (ALPAO97, Alpao) and a single-photon-avalanche diode (SPAD) detector array [38] (Supplement 1, Fig. S1). A continuous-wave laser with a wavelength of λ = 488 nm serves as the excitation source, coupled with a 4 mm lens (LA4024-A, Thorlabs) into a single-mode polarization-maintaining fiber (P5-405BPM-FC-2,Thorlabs). The light exiting from the fiber is collimated by a 50 mm lens (L2, LA1131-AB-ML, Thorlabs). A dichroic mirror (25 × 36 mm Longpass Dichroic Mirror, 490 nm cut-off) reflects the excitation laser, allowing fluorescence from the sample to pass through. The beam with a diameter of 15 mm reaches the DM (DM97-15 BAX255, pupil diameter 13.5 mm). The DM, positioned in the common path (excitation and fluorescence path), thus correcting aberrations in both illumination and detection paths. The laser beam then passes through lens pairs (L3 and L4

Biomedical Optics EXPRESS

with focal lengths of 600 mm and 75 mm, respectively), producing an 8× de-magnification of the excitation beam. The collimated beam is directed into the sample plane by galvanometric scanner mirrors XY (GVS102, Thorlabs). The pivot point of the scanner is projected by the scan lens (SL, focal length: 50 mm, Leica Microsystems) and a tube lens (TL, focal length: 200 mm, Leica Microsystems) into the back aperture of the objective lens (HCX PL APO $100\times/1.4-0.7$ Oil CS, Leica Microsystems) with a 5.6 mm diameter. To ensure overfilling, the pair of lenses SL and TL produces a 4× magnification of the excitation beam spot onto the objective pupil. Although the pupil of the DM measures 13.5 mm in diameter, after the de-magnification the objective back aperture should be overfilled at least for 10% of its dimension. Otherwise, the contribution of the more lateral actuators can be lost. Then, the pair of lenses (SL and TL) produce a magnification of 4× of the excitation beam spot onto the objective pupil.

The fluorescence signal ($\lambda = 510 - 520$ nm) is collected by the objective lens and de-scanned by the galvanometric mirrors, filtered by DM1, and directed through lenses (L5 with a focal length of 600 mm, L6 with a focal length of 50 mm, LA1422-A-MLd, Thorlabs , and L7 with a focal length of 150 mm, LA1433-A-MLd, Thorlabs) to the detector array. The final lens (L7) images the fluorescent spot onto the detector, with a total magnification of M = 450. Our BrightEyes-MCS software acquires the raw data and controls all parts of the microscope [63].

4.2. Deformable mirror calibration

After integrating the deformable mirror (DM) into the optical path of the LSM setup, we calibrated it to achieve the desired wavefront modulation. The calibration involved intercepting the excitation beam after the DM using a removable mirror and directing it through a telescope system onto a Shack-Hartmann sensor (SHS, SH-CMOS, Alpao). The telescope was designed to conjugate the DM plane to the microlenses of the SHS and to provide a magnification that aligns the DM size with the SHS active area. The true calibration is finally obtained by using the ALPAO Core Engine (ACE). Additionally, we used the same SHS to correct aberrations potentially introduced by the optical elements between the DM and the objective lens. For this purpose, we introduced a beam splitter before the objective lens to direct the excitation beam onto the SHS. Also in this case, a telescope was used to conjugate the SHS plane and match the beam size effectively.

4.3. Synthetic images for training

The generation of the synthetic dataset for training the CNNs is based on the BrightEyes-ISM package [42]. This package not only offers various reconstruction algorithms for the ISM dataset but also provides tools for generating physics-based simulated ISM scanned images. Specifically, it includes: (i) functions for creating three-dimensional phantom structures that combine point-like features with more complex filamentous structures both parametrized (e.g., size, density); (ii) a point-spread function generator based on vectorial modeling of the focused electromagnetic field, capable of accounting for the detector array geometry, various imaging conditions, and integrating any optical aberrations; and (iii) a simulator for generating scanned ISM datasets by convolving phantom structures with PSFs and incorporating different noise sources (e.g., photon counting noise, thermal dark noise). Since each element of the detector array generates a distinct scanned image, a $2D$ PSF was calculated for each element. To ensure realistic PSFs, the simulations were parameterized using the optical and physical properties of the microscope and detector array. Specifically, the excitation wavelength was set to 488 nm, while the emission wavelength was centered at 515 nm with a bandwidth of ±5 nm. The oil refractive index was assumed to be 1.5 , and the numerical aperture (NA) was randomized between 1.2 and 1.4

Biomedical Optics EXPRESS

A magnification of 450× at the detector array plane were used. The detector was modeled as a 5×5 square array with a pixel pitch of 75 µm and a pixel size of 50 µm. The scanning pixel size was set to 50 nm.

In non-aberrated conditions, we calculated the depth-of-field of our microscope as DOF = $2\lambda n/\mathrm{NA}^2$ = 780 nm. Then, we generated the training set for CNNx3 at defocus values $z = \{-\mathrm{DOF}/2, 0, \mathrm{DoF}/2\}$. Optical aberrations were introduced into the PSFs by modeling the wavefront at the pupil of the objective lens using a Zernike polynomial basis (with polynomials ranging from 5^{th} to 11^{th}).

To randomize the aberrations in the training dataset, the coefficients of the Zernike polynomials were sampled from a double Gaussian distribution with a mean of $\mu = \pm 0.75$ rad and a standard deviation of $\sigma = 0.3$ rad. The coefficients were then constrained in modulus to values between 0.25 rad to 1.5 rad. Additionally, the $2D$ structures generating the images were randomized to increase diversity. A mixture of point-source structures and tubular structures was used, with the paths and sizes of the tubular structures generated randomly. Their concentrations, specifically their emission flux, were also randomized to capture a wide range of potential imaging scenarios. Given the PSFs and the phantoms, the images are generated with a simple 2D convolution. Two sources of noise are successively introduced to the images. The choice of a random pixel-dwell time between $30 - 60$ μs for each simulation, combined with the different photon-flux $10 - 40$ MHz of the phantom allows the generation of images with different photon-counting noise. Furthermore, we also introduce the dark noise intrinsic to each element of the SPAD array detector. A realistic value of dark noise for the array element is measured directly from the SPAD array used in the custom ISM setup. A comprehensive table detailing all parameters used to generate the synthetic images is available in Supplement 1, Table T1. To train the custom CNNs, we generated a dataset comprising a total of 10,000 ISM experiments. This dataset had dimensions of 10,000 (phantoms) × 25 (array elements) × 3 (axial planes), amounting to nearly 110 GB of data. The simulations were performed on the Franklin supercomputer provided by the Istituto Italiano di Tecnologia (IIT) and required approximately one week to complete. After calculating the normalized correlograms, only the central region, typically consisting of 128×128 pixels was used for training the CNN. This choice was motivated by the fact that the central region contains most of the relevant information about the aberration, whereas the peripheral regions are dominated by noise.

4.4. CNN architecture

Here, we describe the CNN3 architecture, as CNN1 can be considered its simplified version, reduced to processing only the images of the in-focus plane. CNN is inspired by AlexNet [64,65], and consists of three parallel branches, each designed to process the same input shape of $(24, 128, 128) \times 3$ (Supplement 1, Fig. S10). In each branch, the first step is a convolutional layer that applies 16 filters of size 5×5 to extract initial features from the input. This is followed by batch normalization, which normalizes the outputs to stabilize the learning process and improve convergence. Next, a max pooling layer reduces the spatial dimensions of the feature maps by taking the maximum value in each 2×2 block, helping to decrease the number of parameters and computational load. The branch continues with another convolutional layer, this time applying 32 filters of size 3×3, followed by batch normalization and another max pooling layer. This process is repeated with a third convolutional layer that applies 64 filters of size 2×2, once again followed by batch normalization and max pooling. At the end of each branch, the output is flattened into a one-dimensional vector, preparing it for the subsequent fully connected layers. After all three branches have processed the input, their outputs are concatenated, merging the distinct features learned from each path. This combined representation is then fed into a dense layer with 1024 units, allowing for complex feature combinations. To prevent overfitting, a dropout layer randomly drops 50% of the units during training. This is followed by another dense

Biomedical Optics EXPRESS

layer of 1024 units and another dropout layer. Finally, the network concludes with a dense layer that outputs a vector of size 7, which likely represents class scores for a multi-class classification task. Overall, this architecture effectively combines feature extraction and classification, making it suitable for various image-processing applications. A detailed description of the structure of CNN1 is reported in Table 1. To assess the effectiveness of our CNN models in predicting Zernike polynomial aberrations, it is crucial to employ appropriate evaluation metrics that provide meaningful insights into both accuracy and precision. Given the regression nature of this problem, we focus on metrics that quantify the error or correlation between the predicted and true labels corresponding to the aberration amplitude. The Root Mean Square Error (RMSE) is chosen as the loss function over the amplitude of the Zernike polynomials, with 300 epochs and a batch size of 64. This configuration takes approximately 1.5 hours for CNNx1 and 4 hours for CNNx3. For the transfer learning process, we used only 30 epochs since the model begins with weights that are already close to optimal, and the training process tends to converge more quickly. This faster convergence reduces the need for extensive training over many epochs.

Table 1. Numbers of parameters of CNNx1 model for each layer

Layer	Type	Output Shape	Param #
1	Conv2D	(42, 42, 16)	9616
	BatchNorm	(42, 42, 16)	64
	MaxPool2D	(21, 21, 16)	0
2	Conv2D	(10, 10, 32)	4640
	BatchNorm	(10, 10, 32)	128
	MaxPool2D	(5, 5, 32)	0
3	Conv2D	(4, 4, 64)	8256
	BatchNorm	(4, 4, 64)	256
	MaxPool2D	(2, 2, 64)	0
4	Dense	256	787456
	Dropout	0.5	0
5	Dense	1024	1049600
	Dropout	0.5	0
6	Dense	7	7175

4.5. Transfer learning

Training a CNN from scratch requires large datasets and significant computational resources. By starting with a pre-trained model, which already captures general features from a related domain, transfer learning fine-tunes the network for the specific task, requiring fewer epochs and fewer data in order to improve the generalization capabilities of CNNs. This strategy is particularly beneficial in contexts where labeled data is limited. By utilizing experimental data obtained from observations of fluorescence beads and lipid droplets, we captured features related to real-content conditions, which are critical to replicate in simulations. We collected nearly one hundred experimental images for each focal plane. A subset of this dataset is presented in Supplement 1, Fig. S9, showcasing experimental images of beads under various amplitudes of trefoil Z_3^{-3} aberration. Given the time-intensive process of acquiring thousands of experimental images for CNN training, we employed transfer learning combined with data augmentation techniques on the collected experimental dataset. To further enhance the dataset, we generated an additional 1000 images by cropping the original images around randomly selected central points. To ensure

Biomedical Optics EXPRESS

the quality of the augmented data, we applied a threshold to exclude entirely dark images that might result from cropping. However, due to the nature of the dataset, the data augmentation techniques were limited only to the cropping function; we could not perform all possible geometric transformations on the experimental images, as this would distort the correlograms and affect the classification of the aberrations. Overall, the transfer learning approach aims to enhance the CNNs capability to detect and estimate optical aberrations effectively in practical applications, for this reason, the number of images in the experimental dataset post-augmentation constituted approximately only 10% of the overall data used to train the CNN taking almost only 5 minutes to complete the process.

4.6. Image processing

4.6.1. Confocal image

Given the ISM image dataset (Eq. (1)), we generated the corresponding confocal image by summing all the raw images:

$$i_s(\boldsymbol{x}_s) = \sum_{x_d} i(\boldsymbol{x}_s|\boldsymbol{x}_d) \tag{10}$$

The result is equivalent to a confocal image acquired with a pinhole as large as the detector array.

4.6.2. Shift-vectors and adaptive pixel reassignment

We calculated the shift-vectors of an ISM dataset as:

$$\boldsymbol{\mu}(\boldsymbol{x}_d) = \arg\max_{\boldsymbol{x}_s}\{\mathcal{R}(\boldsymbol{x}_s|\boldsymbol{x}_d)\} \tag{11}$$

The APR reconstruction is calculated as the sum of the aligned images

$$i_{APR}(\boldsymbol{x}_s) = \sum_{x_d} i(\boldsymbol{x}_s + \boldsymbol{\mu}(\boldsymbol{x}_d)|\boldsymbol{x}_d) \tag{12}$$

4.6.3. Correlograms

The correlograms are calculated from the cross-correlation of an ISM dataset as described in Eq. (5), followed by the application of a Fourier transformation with respect to the scan coordinates (Eq. (6)). A normalization step is then applied to eliminate objective contributions (Eq. (7)).

Additionally, note that the denominator in the division of Eq. (7) might contain zeros or very small values.

To avoid divergence, we impose the correlogram to be zero in the positions where the denominator value is smaller than the threshold of $\varepsilon = 1 \times 10^{-7}$.

We normalized by the fingerprint, defined as follows

$$f(\boldsymbol{x}_d) = \sum_{\boldsymbol{x}_s} i(\boldsymbol{x}_s \mid \boldsymbol{x}_d) \tag{13}$$

Finally, we obtain the correlograms by inverting the Fourier transform and we normalize with the fingerprint Eq. (8).

4.7. Sample preparation

To produce a sample with fixed fluorescent beads we first deposit 150 µL of poly-L-lysine (PLL) on a clean cover slip and incubate it at 37°C for 10 minutes. In the meanwhile, we prepare a dilution of the beads' mother solution in distilled water with a volumetric ratio between

Biomedical Optics EXPRESS

1 : 500 and 1 : 500, which we sonicate for 5 minutes. We dry the coverslip with clean air and deposit150 µL of beads dilution on top of the adhesive film, followed by incubation at 37°C for 10 minutes. We spill the remainder of the solution on the cover glass and dry it with clean air. We add 5 µL of Mowiol mounting medium and seal the cover glass on a microscope slide.

HEPG2, a human hepatocellular carcinoma cell line, and HeLa, a widely known cervical carcinoma cell line, were maintained respectively in Eagle's Minimum Essential Medium (EMEM) and Dulbecco's modified Eagle medium (D-MEM), both containing 10% Foetal Bovine Serum (FBS), at 37°C. For imaging, cells were seeded on poly-L-lysine coated glass slides. When cells reached the proper confluence, slides were rinsed with PBS, fixed with 4% paraformaldehyde/PBS for 15 minutes, then washed again three times with PBS.

Lipid droplets (LDs) in HEPG2 cells were visualized using Nile Red, a fluorescent dye with a high sensitivity to neutral lipids. After fixation and washing, HEPG2 cells were incubated with 0.5 µg/ mL Nile Red (Invitrogen, Thermo Fisher Scientific) in PBS for 30 minutes, washed with PBS, and mounted using ProLong Diamond Antifade Mountant (Thermo Fisher Scientific). Beta-tubulin staining was performed on HeLa cells. To permeabilize cells and block the non-specific binding of the antibodies during the immunostaining, we incubated the coverslides with 3% BSA in 0.5% Triton $X-100$ in PBS for 1 h, at room temperature. Slides were then incubated with 1 : 200 rabbit anti-β tubulin antibody (ab6046, Abcam) overnight at 4°C. The following day, after washing three times with 0.5% Triton-PBS, 1 : 500 Alexa Fluor 488 anti-rabbit antibody (a11008, Abcam) was used as a secondary antibody and incubated at room temperature for 1 h. Then, the slides were rinsed and mounted as above.

Funding. European Research Council "BrightEyes" (CoG N. 818699) .

Acknowledgments. The authors thank all members of the Molecular Microscopy and Spectroscopy labs for the many helpful suggestions: Sabrina Zappone, Dr. Eleonora Perego, Dr. Mattia Donato, Dr. Eli Slenders, Giacomo Garré, Dr. Andrea Bucci, Luca Bega, Sanket Patil, Dr. Sami Walteri Koho, and Dr. Marcus Held. We acknowledge the useful support and advices from Dr. Stefano Bonora. We acknowledge that the research activity herein was carried out using the IIT Franklin HPC infrastructure; we gratefully acknowledge the Data Science and Computation Facility and its Support Team for their support and assistance on the IIT High Performance Computing Infrastructure. This work is part of the "Teaching Science to Computers" Flagship Program of the IIT. **Contributions**. G.V. conceived the idea. F.F., A.Z., and G.V. designed the study. A.Z., A.D., M.J.B, A.D.B, and G.V. supervised the project. F.F. built the microscope with the adaptive optics element. F.F. and F.B. designed the cell experiments. F.F. developed the convolutional neural network with the supervision of P.M. F.F. analyzed the data with the support of all other authors. F.F. and G.V. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Disclosures. G.V. and A.D. have a personal financial interest (co-founder) in Genoa Instruments, Italy.

Data availability. The Python code used in this work is available at [66]. The experimental data generated for this study are available at [67].

Supplemental document. See Supplement 1 for supporting content.

References

1. W. Jiang and H. Li, "Hartmann-shack wavefront sensing and wavefront control algorithm," in *Adaptive Optics and Optical Structures*, vol. 1271 (SPIE, 1990), pp. 82–93.
2. G. E. Artzner, "Microlens arrays for Shack-Hartmann wavefront sensors," Opt. Eng. **31**(6), 1311–1322 (1992).
3. J. Pawley, *Handbook of Biological Confocal Microscopy* (Springer, 2006).
4. L. Schermelleh, A. Ferrand, T. Huser, *et al.*, "Super-resolution microscopy demystified," Nat. Cell Biol. **21**(1), 72–84 (2019).
5. A. Diaspro, eds. *Confocal and Two-photon Microscopy: Foundations, Applications, and Advances* (Wiley-Liss, 2001).
6. J.-A. Conchello and J. W. Lichtman, "Optical sectioning microscopy," Nat. Methods **2**(12), 920–931 (2005).
7. L. Scipioni, A. Rossetta, G. Tedeschi, *et al.*, "Phasor s-FLIM: a new paradigm for fast and robust spectral fluorescence lifetime imaging," Nat. Methods **18**(5), 542–550 (2021).
8. L. Scipioni, G. Tedeschi, M. Navarro, *et al.*, "ESPRESSO: Spatiotemporal omics based on organelle phenotyping," (2024). Pages: 2024.06.13.598932 Section: New Results.
9. J. A. Kubby, *Adaptive Optics for Biological Imaging* (CRC Press, 2013).
10. M. J. Booth, M. A. A. Neil, R. Juškaitis, *et al.*, "Adaptive aberration correction in a confocal microscope," Proc. Natl. Acad. Sci. **99**(9), 5788–5792 (2002).

Biomedical Optics EXPRESS

11. M. Booth, D. Andrade, D. Burke, *et al.*, "Aberrations and adaptive optics in super-resolution microscopy," Microscopy **64**(4), 251–261 (2015).
12. K. M. Hampson, R. Turcotte, D. T. Miller, *et al.*, "Adaptive optics for high-resolution imaging," Nat. Rev. Methods Primers **1**(1), 68 (2021).
13. M. J. Booth, "Adaptive optics in microscopy," Philos. Trans. R. Soc., A **365**(1861), 2829–2843 (2007).
14. J. Wang and Y. Zhang, "Adaptive optics in super-resolution microscopy," Biophys. Rep. **7**(4), 267–279 (2021).
15. T. Furieri, D. Ancora, G. Calisesi, *et al.*, "Aberration measurement and correction on a large field of view in fluorescence microscopy," Biomed. Opt. Express **13**(1), 262–273 (2022).
16. D. Ancora, T. Furieri, S. Bonora, *et al.*, "Spinning pupil aberration measurement for anisoplanatic deconvolution," Opt. Lett. **46**(12), 2884–2887 (2021).
17. Q. Zhang, Q. Hu, C. Berlage, *et al.*, "Adaptive optics for optical microscopy," Biomed. Opt. Express **14**(4), 1732–1756 (2023).
18. B. Platt and R. Shack, "Production and use of a lenticular hartmann screen," Spring Meeting of Optical Society of America, 1971 (Vol. 656) (1971).
19. B. C. Platt and R. Shack, "History and principles of Shack-Hartmann wavefront sensing," Journal of Refractive Surgery **17**, S573–S577 (2021).
20. A. Kudryashov, V. Samarkin, A. Alexandrov, *et al.*, "Shack-Hartmann wavefront sensor - advantages and disadvantages," in *2010 International Conference on Advanced Optoelectronics and Lasersin*, (2010), pp. 76–77. ISSN: 2160–1534.
21. M. J. Booth, "Wavefront sensorless adaptive optics for large aberrations," Opt. Lett. **32**(1), 5–7 (2007).
22. J. Antonello, A. Barbotin, E. Z. Chong, *et al.*, "Multi-scale sensorless adaptive optics: application to stimulated emission depletion microscopy," Opt. Express **28**(11), 16749–16763 (2020).
23. K. M. Hampson, J. Antonello, R. Lane, *et al.*, *Sensorless Adaptive Optics* (Zenodo, 2020).
24. Q. Hu, J. Wang, J. Antonello, *et al.*, "A universal framework for microscope sensorless adaptive optics: Generalized aberration representations," APL Photonics **5**(10), 100801 (2020).
25. C. Johnson, M. Guo, M. C. Schneider, *et al.*, "Phase-diversity-based wavefront sensing for fluorescence microscopy," Optica **11**(6), 806–820 (2024).
26. L. Möckl, P. N. Petrov, and W. E. Moerner, "Accurate phase retrieval of complex 3d point spread functions with deep residual neural networks," Appl. Phys. Lett. **115**(25), 251106 (2019).
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. P. A. Khorin, A. P. Dzyuba, P. G. Serafimovich, *et al.*, "Neural networks application to determine the types and magnitude of aberrations from the pattern of the point spread function out of the focal plane," J. Phys.: Conf. Ser. **2086**(1), 012148 (2021).
29. B. P. Cumming and M. Gu, "Direct determination of aberration functions in microscopy by an artificial neural network," Opt. Express **28**(10), 14511–14521 (2020).
30. E. Durech, W. Newberry, J. Franke, *et al.*, "Wavefront sensor-less adaptive optics using deep reinforcement learning," Biomed. Opt. Express **12**(9), 5423–5438 (2021).
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. M. Bertero, P. Brianzi, and E. R. Pike, "Super-resolution in confocal scanning microscopy," Inverse Problems **3**(2), 195–212 (1987).
33. C. Sheppard, "Super-resolution by confocal fluorescent microscopy," Optik International Journal For Light and Electron Optics (1982).
34. C. B. Müller and J. Enderlein, "Image scanning microscopy," Phys. Rev. Lett. **104**(19), 198101 (2010).
35. C. J. R. Sheppard, "Resolution and super-resolution," Microsc. Res. Tech. **80**(6), 590–598 (2017).
36. J. Huff, "The airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution," Nat. Methods **12**(12), i–ii (2015).
37. I. M. Antolovic, C. Bruschini, and E. Charbon, "Dynamic range extension for photon counting arrays," Opt. Express **26**(17), 22234–22248 (2018).
38. M. Buttafava, F. Villa, M. Castello, *et al.*, "SPAD-based asynchronous-readout array detectors for image-scanning microscopy," Optica **7**(7), 755–765 (2020).
39. C. J. R. Sheppard, M. Castello, G. Tortarolo, *et al.*, "Pixel reassignment in image scanning microscopy: a re-evaluation," J. Opt. Soc. Am. A **37**(1), 154–162 (2020).
40. C. J. R. Sheppard, M. Castello, G. Tortarolo, *et al.*, "Pixel reassignment in image scanning microscopy with a doughnut beam: example of maximum likelihood restoration," J. Opt. Soc. Am. A **38**(7), 1075–1084 (2021).
41. A. Zunino, M. Castello, and G. Vicidomini, "Reconstructing the image scanning microscopy dataset: an inverse problem," Inverse Problems **39**(6), 064004 (2023).
42. A. Zunino, E. Slenders, F. Fersini, *et al.*, "Open-source tools enable accessible and advanced image scanning microscopy data analysis," Nat. Photonics **17**(6), 457–458 (2023).
43. G. Tortarolo, A. Zunino, F. Fersini, *et al.*, "Focus image scanning microscopy for sharp and gentle super-resolved microscopy," Nat. Commun. **13**(1), 7723 (2022).

Biomedical Optics EXPRESS

44. A. Zunino, G. Garré, E. Perego, *et al.*, "Structured detection for simultaneous super-resolution and optical sectioning in laser scanning microscopy," (2024).
45. J. Goodman, *Introduction To Fourier Optics* (W.H. Freeman, 2017), 4th ed.
46. Y. Liu, V. Stergiopoulou, J. Chuah, *et al.*, "Revisiting psf models: unifying framework and high-performance implementation," (2025).
47. M. Siemons, C. N. Hulleman, R. Ø. Thorsen, *et al.*, "High precision wavefront control in point spread function engineering for single emitter localization," Opt. Express **26**(7), 8397–8416 (2018).
48. D. Ancora, A. Zunino, G. Vicidomini, *et al.*, "Image scanning microscopy reconstruction by autocorrelation inversion," J. Phys. Photonics **6**(4), 045003 (2024).
49. M. Castello, C. J. R. Sheppard, A. Diaspro, *et al.*, "Image scanning microscopy with a quadrant detector," Opt. Lett. **40**(22), 5355–5358 (2015).
50. M. Castello, G. Tortarolo, M. Buttafava, *et al.*, "A robust and versatile platform for image scanning microscopy enabling super-resolution FLIM," Nat. Methods **16**(2), 175–178 (2019).
51. F. Baldini, L. Zeaiter, F. Diab, *et al.*, "Nuclear and chromatin rearrangement associate to epigenome and gene expression changes in a model of in vitro adipogenesis and hypertrophy," Biochim. Biophys. Acta, Mol. Cell Biol. Lipids **1868**(10), 159368 (2023).
52. F. Baldini, M. Calderoni, L. Vergani, *et al.*, "An overview of long non-coding (lnc)rnas in neuroblastoma," Int. J. Mol. Sci. **22**(8), 4234 (2021).
53. G. Vecchione, E. Grasselli, A. Voci, *et al.*, "Silybin counteracts lipid excess and oxidative stress in cultured steatotic hepatic cells," World J. Gastroenterol. **22**(26), 6016–6026 (2016).
54. E. Slenders, M. Castello, M. Buttafava, *et al.*, "Confocal-based fluorescence fluctuation spectroscopy with a SPAD array detector," Light:Sci. Appl. **10**(1), 31 (2021).
55. E. Slenders, E. Perego, M. Buttafava, *et al.*, "Cooled SPAD array detector for low light-dose fluorescence laser scanning microscopy," Biophys. Rep. **1**(2), 100025 (2021).
56. A. Rossetta, E. Slenders, M. Donato, *et al.*, "The BrightEyes-TTM as an open-source time-tagging module for democratising single-photon microscopy," Nat. Commun. **13**(1), 7406 (2022).
57. E. Perego, S. Zappone, F. Castagnetti, *et al.*, "Single-photon microscopy to study biomolecular condensates," Nat. Commun. **14**(1), 8224 (2023).
58. G. Tortarolo, A. Zunino, S. Piazza, *et al.*, "Compact and effective photon-resolved image scanning microscope," Adv. Photonics **6**(01), 016003 (2024).
59. A. Bucci, G. Tortarolo, M. O. Held, *et al.*, "4d single-particle tracking with asynchronous read-out single-photon avalanche diode array detector," Nat. Commun. **15**(1), 6188 (2024).
60. S. V. Koho, E. Slenders, G. Tortarolo, *et al.*, "Two-photon image-scanning microscopy with SPAD array and blind image reconstruction," Biomed. Opt. Express **11**(6), 2905–2924 (2020).
61. N. Radmacher, O. Nevskyi, J. I. Gallea, *et al.*, "Doubling the resolution of fluorescence-lifetime single-molecule localization microscopy with image scanning microscopy," Nat. Photonics **18**(10), 1059–1066 (2024).
62. E. Slenders, S. Patil, M. O. Held, *et al.*, "Array detection enables large localization range for simple and robust MINFLUX," (2024). Pages: 2024.07.08.602588 Section: New Results.
63. M. Donato, E. Slenders, A. Zunino, *et al.*, "Brighteyes-mcs: a control software for multichannel scanning microscopy," Journal of Open Source Software **9**(103), 7125 (2024).
64. C. Lu, Q. Tian, L. Zhu, *et al.*, "Mitigating the ambiguity problem in the CNN-based wavefront correction," Opt. Lett. **47**(13), 3251–3254 (2022).
65. 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).
66. F. Fersini, "ISMAO," Github, 2025, https://github.com/VicidominiLab/ISMAO.
67. F. Fersini, "Data for ISMAO," Zenodo, 2025, https://doi.org/10.5281/zenodo.13789465