Accepted Article Preview: Published ahead of advance online publication

Single-shot 3D incoherent imaging with diffuser endoscopy

Julian Lich, Tom Glosemeyer, Jürgen Czarske and Robert Kuschmierz

Cite this article as: Julian Lich, Tom Glosemeyer, Jürgen Czarske and Robert Kuschmierz. Single-shot 3D incoherent imaging with diffuser endoscopy. *Light: Advanced Manufacturing* accepted article preview 06 March, 2024; doi: 10.37188/lam.2024.015

Received 13 September 2023; revised 01 March 2024; accepted 05 March 2024; Accepted article preview online 06 March 2024

Single-shot 3D incoherent imaging with diffuser endoscopy

Julian Lich[1] *, Tom Glosemeyer[1] † *, Jürgen Czarske[1 2 3 4], and Robert Kuschmierz[1 2 3] †

[1]*Laboratory of Measurement and Sensor System Technique, TU Dresden, Germany*
[2]*Competence Center for Biomedical Computational Laser Systems (BIOLAS), TU Dresden, Germany*
[3]*Else Kröner Fresenius Center for Digital Health (EKFZ), TU Dresden, Germany*
[4]*Cluster of Excellence Physics of Life, TU Dresden, Germany*
*These authors contributed equally to this work.
†Correspondence to: Tom Glosemeyer: tom.glosemeyer@tu-dresden.de or Robert Kuschmierz: robert.kuschmierz@tu-dresden.de

Abstract

Minimally invasive endoscopy offers a high potential for biomedical imaging applications. However, conventional fiberoptic endoscopes require lens systems which are not suitable for real-time 3D imaging. Instead, a diffuser is utilized for passively encoding incoherent 3D objects into 2D speckle patterns. Neural networks are employed for fast computational image reconstruction beyond the optical memory effect. In this paper, we demonstrate single-shot 3D incoherent fiber imaging with keyhole access at video rate. Applying the diffuser fiber endoscope for fluorescence imaging is promising for in vivo deep brain diagnostics with cellular resolution.
Keywords: Endoscopy, Diffuser, Deep Learning, Coherent Fiber Bundle, Multicore Fiber, Neural Networks, Fiber Optics, Speckle Patterns, Fluorescence Imaging.

Introduction

Minimally invasive in vivo imaging at hardly accessible locations is crucial for many biomedical applications inside the human body, such as calcium imaging [1]. Moreover, there is a great research interest in supervising structural and functional processes like brain activity in freely moving animals at cellular resolution [2, 3]. While the observation of neuronal activity in living animals was demonstrated with two-photon microscopy up to a depth of 550 µm [4], imaging deeper inside the brain requires flexible, minimally invasive endoscopes. Therefore, in vivo endoscopic imaging systems need to be robust towards bending and ultra-thin to reduce damage to cells and tissue. Furthermore, three-dimensional (3D) imaging capabilities including high spatial and temporal resolution, chromatic sensitivity and large field of view (FoV) are desired [5].

State-of-the-art lens-based coherent fiber bundle (CFB) endoscopes only allow two-dimensional (2D) imaging in the focal plane of the lens. The CFB enables pixelwise intensity transfer, thus limiting the space-bandwidth product (SBP) by the core number. Therefore, the resolution is determined by the fiber core pitch and the magnification. The use of microlenses allows trading off lateral for axial resolution [6, 7]. Focus tunable lenses and piezo actuators exchange axial for temporal resolution, but are only available at diameters far above 1 mm [8, 9]. 3D imaging can be achieved by using stereo probes, however, the added constructive complexity can be hard to miniaturize effectively [10] . The smallest endoscope diameters can be achieved with single-mode fibers. In combination with Optical Coherence Tomography, 3D imaging of mouse veins was achieved at probe diameters below 0.5 mm [11]. However, coherent light is needed, which prevents applications like fluorescence imaging, and a mechanical focus scanning movement is necessary, which increases the complexity of the sensor head [12]. Multimode fibers (MMF) [13, 14, 15, 16, 17, 18, 19] as well as coherent fiber bundles (CFB) [20, 21, 22, 23, 24, 25, 26] enable lensless endoscopic imaging with passive probe tips and diameters below 500 µm. A high SBP and 3D object reconstruction can be achieved via holography [27, 28, 29, 30, 31] or time-of-flight approaches [32]. However, as the aforementioned lensless endoscopy approaches rely on phase reconstruction, they are prone to fiber bending and require coherent light. The necessary in situ phase recalibration limits the temporal resolution and the measurement of incoherent fluorescent light is challenging. This restricts the transfer to real-world applications.

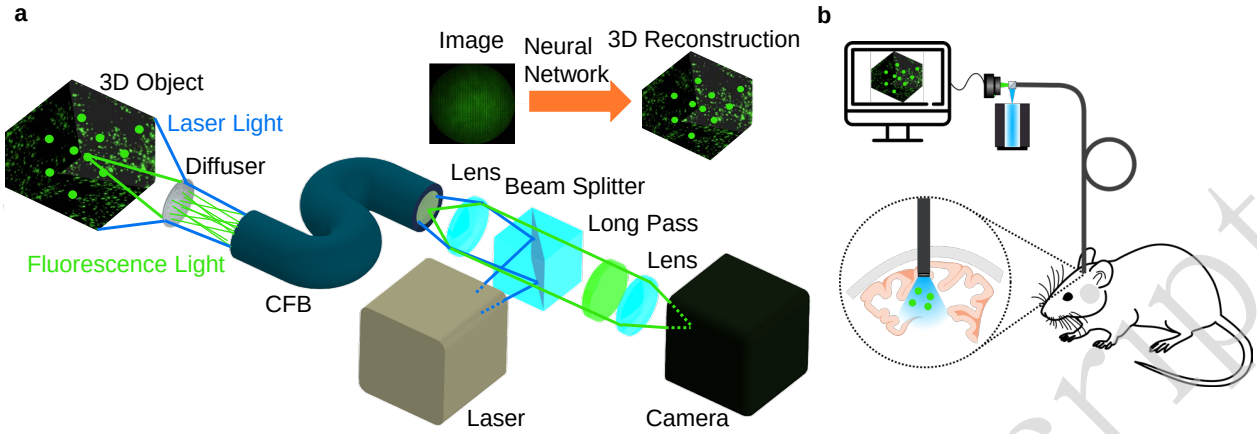

Fig. 1. (**a**) Fluorescence Imaging Setup. A fluorescent object (2D/3D) is illuminated through the system by a laser. Incoherent light emitted by the object is encoded into 2D speckle patterns by a diffuser. The intensity pattern is transmitted by a coherent fiber bundle (CFB) onto a camera. The object image is then computationally reconstructed by a neural network. (**b**) Potential application of the diffuser endoscope for in vivo calcium imaging to observe deep brain activity in living animals.

The use of a CFB and a coded aperture is promising for both 3D imaging with incoherent illumination [33] as well as a higher SBP than equivalent lens-based endoscopes in 2D-imaging scenarios [34, 35]. The light and information loss of coded apertures can be evaded by using diffuse phase masks [36, 37, 38]. In any case, computational image reconstruction is necessary. There are multiple iterative approaches for object retrieval based on convex optimization [39, 40], but they do not offer suitable reconstruction speeds for in vivo real-time imaging. In contrast, neural networks [41, 42, 43, 44, 45] enable operation at video rate after completing the training process [46, 47] and 3D [48] image reconstruction. Moreover, they are robust towards model errors and do not rely on point spread function (PSF) shift invariance [49].

In this paper, we demonstrate a minimally invasive diffuser-based fiber endoscope with single-shot 3D live fluorescence imaging capabilities and a probe tip diameter of 700 µm, allowing minimally invasive applications for instance in neurosurgery. The setup is shown in Fig. 1a. Laser light illuminates the tissue through the fiber endoscope. Incoherent light emitted by fluorescent 3D objects is encoded by the diffuser into 2D speckle patterns. The speckle patterns are transmitted by the CFB onto a camera. The 3D object is then computationally reconstructed by neural networks in 20 ms, allowing video-rate imaging at up to 50 fps. Because our approach only processes intensity images, dispersion between fiber cores from optical path length differences as well as phase shifts due to bending can be ignored [34]. In fluorescence settings, the setup can be used for illumination and detection simultaneously, enabling biomedical applications for in vivo imaging at cellular resolution like calcium imaging (Fig. 1b).

Results

To investigate the suitability of the diffuser endoscope for 2D and 3D imaging, different simulation and laboratory experiments were conducted. The approach was first tested in silico in a 2D modality. The system was simulated using Fourier optics to generate pairs of ground truth and camera images as a training dataset. We found a neural network architecture consisting of a combination of Single Layer Perceptron (SLP) and U-Net offered the best reconstruction quality and generalization capabilities. More details on the simulation as well as the neural network training and reconstruction can be found in the Materials and Methods chapter. The reconstruction quality of this and the following experiments was assessed with the metrics of peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) and correlation coefficient (CC) in table 1. For the 2D simulation in a FoV of 416 µm × 416 µm at a distance z of 700 µm between the object plane and the diffuser (Fig. 2a), the evaluation metrics including a mean SSIM of 0.91 indicate a reconstruction with great fidelity.

Table 1: Mean reconstruction quality for the simulation and lab experiments

Metric	2D Sim (n=1000)	2D Exp (n=4)	3D Sim (n=1000)	3D Exp (n=4)
PSNR	28.4 dB	25.3 dB	30.5 dB	29.5 dB
SSIM	0.91	0.89	0.92	0.82
CC	0.93	0.84	0.72	0.43

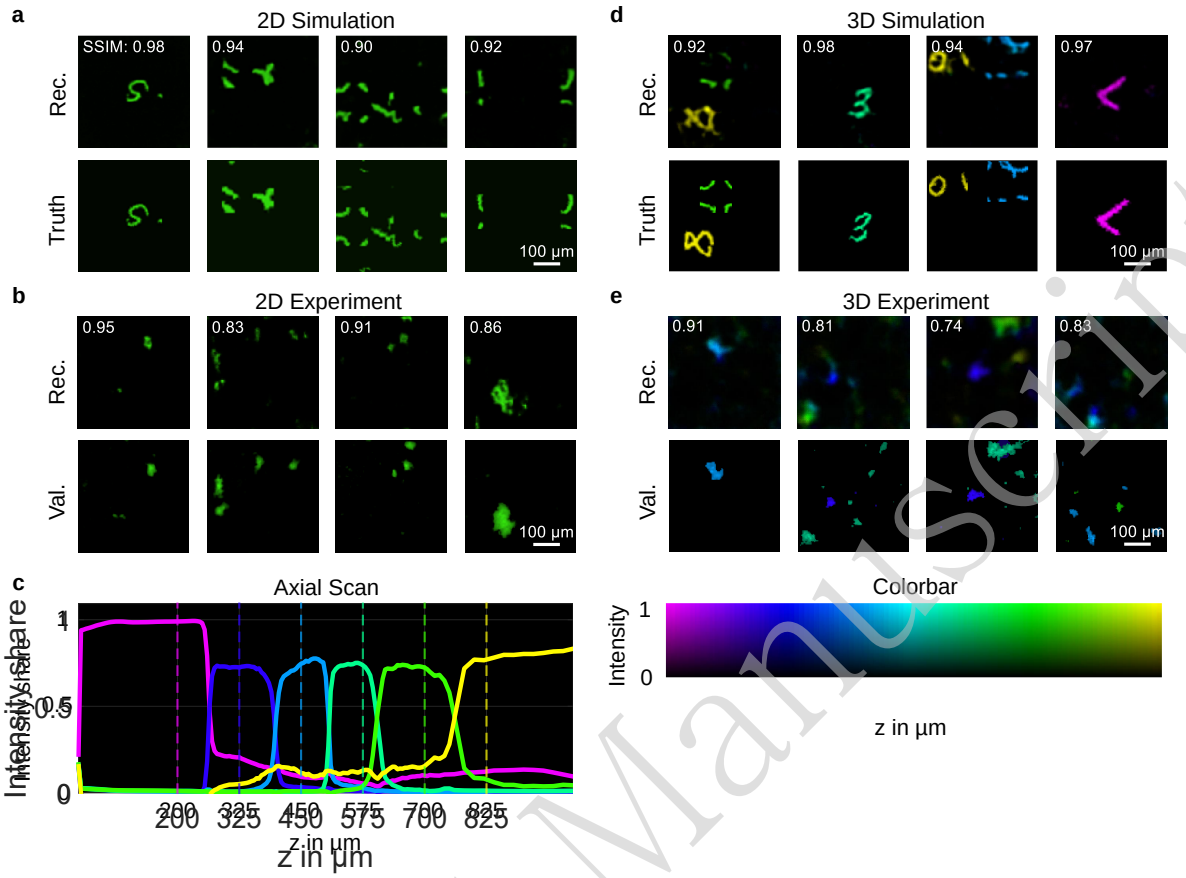

Fig. 2. (**a**) 2D reconstruction of simulated system trained with augmented MNIST digit data. (top) Neural network reconstruction (2D). (bottom) Ground truth images (2D). (**b**) 2D imaging experiment with fluorescence particles at a distance of z = 800 µm from the diffuser. (top) 2D image reconstruction by neural network. (bottom) 2D ground truth captured by validation microscope. (**c**) Intensity share of reconstruction plane for axial scan of projected square over z. The dashed lines mark the positions of the depth planes used for the neural network training. (**d**) 3D reconstruction of simulated system trained with augmented MNIST digit 3D data. The depth position of each voxel in the 3D images is displayed by the color according to the colorbar on the right. (top) Neural network reconstruction (3D). (bottom) Ground truth images (3D). (**e**) 3D imaging experiment with fluorescence particles at two distances z between 200 µm and 800 µm from the diffuser. (top) 3D image reconstruction by neural network. (middle) 2D ground truth for z = 400 µm captured by validation microscope. (bottom) 2D ground truth for z = 575 µm captured by validation microscope. The depth position of each voxel in the 3D images is displayed by the color according to the colorbar on the bottom.

Next, the 3D imaging capability was validated. Therefore, the simulation and the neural network architecture was extended to a FoV of 416 µm × 416 µm × 625 µm consisting of six equally spaced depth layers. Example reconstructions are shown in Fig. 2d where the depth position of each voxel is displayed by a color according to the colorbar on the bottom right. This simulation experiment demonstrates the 3D imaging capabilities of the system. However, the mean PSNR and SSIM values are deceptively high, because the 3D objects are sparsely occupied.

For the validation of these simulation experiments, a setup for imaging of fluorescent Rhodamine B particles was used (Fig. 1a). A digital light projector was employed to project images from six equally spaced axial distances between 200 and 825 µm into the system to train reconstruction networks with a 3D output. The setup and signal processing was then tested by imaging fluorescent particles in 2D and 3D, that were illuminated from the proximal side through the CFB. In both cases, we were able to show that the generalization ability of the networks was sufficient to locate few particles in the FoV. However, the reconstruction quality is reduced with decreasing object sparsity. The reconstruction quality for the fluorescence experiments is worse than it was for the simulations, which could result

from the unoptimized diffuser endoscope setup, noise and the difficult generalization task for the networks.

To validate the depth perception, an object was projected in front of the endoscope at varying axial positions and reconstructed by the neural network. In Fig. 2c, the intensity shares of the reconstructed outputs at the six depths are shown over the projection distance z. With at least 70 % of the intensity attributed to the correct depth plane, an unambiguous discrimination between the axial positions is achieved. Given by the distance between the trained depth planes, an axial resolution of 125 µm results. From the steepness of the slopes in Fig. 2c it follows, that this can be improved further up to 10 µm, e.g. by adding more depth layers to the network architecture, see section 'Materials and methods - Field of view and resolution'.

For both cases of single-shot 2D and 3D imaging, the neural networks are able to reconstruct the image in under 20 ms. The exposure time of around 100 ms was limited by the illumination power of 5 mW in the probe volume, resulting in a total frame rate of 8 fps. No image averaging was performed. Depending on the permitted phototoxicity levels [50], the illumination power and therefore the frame rate can be further increased.

Discussion

The diffuser fiber endoscope we present in this paper was used to demonstrate single-shot 2D and 3D fluorescence imaging with proximal illumination through the CFB, potentially enabling applications like calcium imaging, cancer diagnostics or 3D flow measurement. While this system offers many advantages in comparison to similar approaches, there are also certain limitations that need to be considered.

For many clinical applications, the robustness of the imaging system is an important requirement. In contrast to holographic approaches, no coherent light is required for fluorescence imaging. By using incoherent illumination, only the intensity carries useful information while the phase can be omitted. This makes the imaging system robust towards bending of the fiber bundle which benefits all in vivo applications. Furthermore, the endoscope length should not be a limiting factor for biomedical applications, as long as the signal-to-noise ratio is high enough. The system was illuminated with a laser for the presented experiments, which could be replaced by an LED. This could prevent illumination speckles. Moreover, the diffuser-based endoscope can offer a higher space-bandwidth product than an equivalent lens-based endoscope (Fig. 5), which could also be useful whenever a bigger FoV could make the orientation easier and image data could be acquired faster because less stitching is required. This is a clear advantage over conventional endoscopes where the imaging capabilities are limited by the trade-off between FoV and resolution. Furthermore, the reconstruction approach with neural networks is fast enough to perform live imaging, making real time in vivo applications possible. Moreover, it performs well even if the lateral optical memory effect only covers a fraction of the FoV.

A major limitation of the neural network reconstruction is the sparsity of the object scene to reconstruct. Even if training with data of similar sparsity can counteract this effect, the reconstruction quality suffers from non-sparse objects. Less sparse input to the optical system leads to smaller contrast in the speckle patterns that are transferred to the camera by the CFB and thus, to a decreased reconstruction quality. To enhance the speckle contrast, structured illumination could be implemented, e.g. by focusing the illumination laser onto one fiber core on the proximal side, so that the object on the distal side is illuminated by a speckle pattern. By sequentially focusing through different fiber cores, even the information throughput of the system could be further enhanced at the cost of temporal resolution.

For an application in brain diagnostics, the scattering by the tissue is another limitation. Considering a transport mean free path of 900 µm in brain tissue [51], our endoscope should be capable of imaging up to such a depth. However, because of the increasing noise due to scattering by the tissue for higher penetration depth, a degradation of reconstruction quality is expected. At the moment, the network reconstruction is only able to correct the scattering by a known diffuser and is susceptible to changes. If the reconstruction scheme could be expanded to random diffusers [52], tissue scattering could potentially be decreased too for higher imaging depth.

In the optical setup, the CFB acts as an information bottleneck. For a lens-based endoscope, the space-bandwidth product is equal to the number of fiber cores. For a diffuser endoscope, that limit can be overcome with compressed sensing. The diffuser acts as an encoder to a domain where the information is sparse. Thus, optimizing it to match the problem at hand can potentially improve the imaging capabilities. Currently, a non-optimized random diffuser was mounted on the CFB with a thin glass spacer. This leads to a loss of contrast in the captured images and subsequently to a worse reconstruction by the neural network. Moreover, it has a very small optical memory effect, prohibiting the application of reconstruction approaches that rely on shift invariance. The distal ferrule containing the diffuser and fiber has a diameter of 700 µm, determining the diameter of the endoscope. In principle, the endoscope head could easily be miniaturized to the CFB diameter of 350 µm, but it would also limit the achievable FoV. In the future, optimization of the diffractive optical element on the distal side of the fiber bundle could enhance the imaging capabilites further. Furthermore, the diffuser could be realized as a 3D printed microlens array on top of a gradient-index (GRIN) lens [39] or even a random

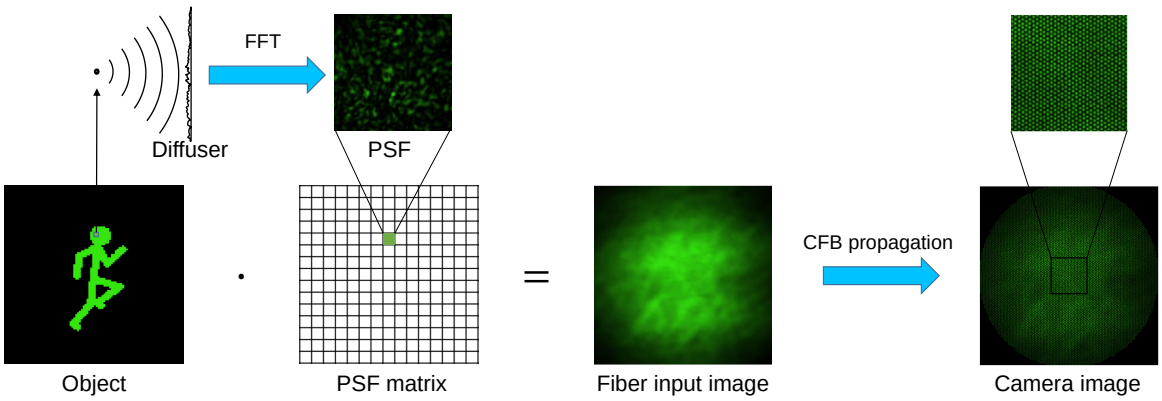

Fig. 3. Simulation of optical system. The object image is encoded with diffuser by superposition of PSFs for illuminated pixels/voxels: For each pixel/voxel in the FoV a PSF is calculated using Fourier Optics. The speckle pattern is then calculated by superposing each PSF weighted by the intensity of the corresponding object pixel/voxel. The transmission through the CFB is simulated with a fiber bundle simulator [53].

phase mask as part of a deep optical system where the optical encoder and the neural network decoder are part of an end-to-end optimization [54, 55]. The neural network architecture could be refined by using a physics-informed network layer modeled after the point spread functions in the FoV for spatially varying deconvolution [56]. This would also allow reconstruction with higher resolution, as less learnable parameters would be required. Deep learning techniques for super-resolution in fluorescence microscopy could be applied as well [57]. Furthermore, a perceptual loss function based on a second network pretrained for image classification could be employed. In comparison to conventional loss functions, it could measure the differences between the reconstructed and the ground truth images in a way that is closer to human perception [58, 59]. After optimization of the optical setup and image reconstruction with the discussed steps, the improved imaging quality could enable previously impossible in vivo applications outside of lab experiments.

Materials and methods

System model

To describe the optical system, a simple forward model [60]

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{1}$$

can be applied, where $\mathbf{y}$ is the $M \times 1$ flattened measured image vector and $\mathbf{x}$ is the $N \times 1$ flattened object space vector. $\mathbf{H}$ is the $M \times N$ transmission matrix, which can be derived from calibration measurements. For megapixel diffuser camera setups, $\mathbf{H}$ can become very large, so high computational power, memory and calibration effort are necessary. Separable [61, 62] or convolution-based [63, 64] forward models can be applied to reduce the calibration effort and computational complexity by several orders of magnitude [60]. However, separable forward models do not work with arbitrary diffusers, and special phase mask designs with low fabrication tolerances are needed. Convolution-based forward models rely on the shift invariance of the PSF of the diffuser [65], which significantly reduces the effective FoV. By measuring multiple PSF in one object plane, the effective FoV can be enlarged, thus creating a shift variant forward model [66].

Miniaturized CFB only have several thousand fiber cores, so that M can be chosen small enough to directly apply the forward model from Eq. (1). Since usually $M < N$, the equation system in Eq. (1) is underdetermined and therefore has an infinite amount of solutions. A common method for object reconstruction is to solve a convex optimization problem [60]

$$\hat{\mathbf{x}} = \mathrm{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \tag{2}$$

where λ is a Lagrange multiplier. The regularizer R determines the properties of the solution $\hat{\mathbf{x}}$ and must be chosen according to a priori knowledge about the imaged object $\mathbf{x}$. In case of fluorescence imaging, Eq. (2) can be optimized for sparsity by choosing R to be the L_1 norm [60]. Convex optimization is an iterative and time-consuming approach, which impedes video rate imaging. By employing neural networks for the image reconstruction, the imaging rate can be drastically increased.

For the purpose of parameter studies, the imaging part of the system was simulated by angular spectrum propagation [67]. The object image is divided into pixels/voxels which are approximated as point sources (Fig. 3). For each point source, a spherical wave is then propagated through the diffuser, which is modeled as a random phase mask. This mask is then applied to the electrical field distribution before

angular spectrum propagation is employed to propagate the field onto the distal side of the CFB and to retrieve the corresponding PSF of the pixel/voxel. The intensity pattern on the distal fiber facet for a certain object is then retrieved by superposition of the PSFs of all illuminated pixels/voxels of the object weighted by their intensity. To simulate the transmission of the intensity pattern through the fiber, a fiber bundle simulator developed at the University of Kent [53] was used to imitate the Fujikura fiber we used in our lab experiments.

Endoscope setup

In the optical setup for fluorescence imaging, as shown in Fig. 1a, the illumination laser with a wavelength of 488 nm is reflected into the CFB with a dichroic beamsplitter. The illumination light is guided through the CFB and the diffuser to the imaging volume, which contains fluorescent particles emitting incoherent light. The fluorescence light is then scattered by the diffuser. It consists of diffuse tape mounted onto a cylindrical glass spacer with a length of 500 µm which is glued to the distal end of the CFB. The glass spacer and the CFB are enclosed by a metal ferrule determining the total endoscope diameter of 700 µm. The specifications of the diffuser endoscope head are displayed in the following section. The used diffuser acts as a random phase mask and encodes the object image into a speckle pattern. This pattern is then propagated onto the distal facet of the coherent fiber bundle, a Fujikura FIGH-10-350S with 10000 fiber cores on a 325 µm image diameter. The intensity information of the speckle patterns is transmitted onto a Thorlabs Quantalux CS2100M camera on the proximal side of the fiber through the fiber cores, which have a diameter of 1.7 µm and a mean core pitch of 3.1 µm. The shift invariance of the diffuser endoscope was investigated by laterally shifting a point source over the FoV and examining the correlation coefficient between the captured PSF and the PSF at the zero position. The lateral optical memory effect was calculated by using the full width at half maximum (FWHM) of a gaussian curve fit onto the correlation coefficients of the shifted PSFs. For $z = 200$ µm, it was determined at approximately 50 µm, which roughly equals a tenth of the measured FoV.

To gather a large number of training images with the experimental optical setup, a Texas Instruments DLP4710 digital light projector (DLP) including an RGB LED is employed to project different ground truth data with a resolution up to 1920×1080 into the object space (Fig. 4). To achieve the correct demagnification of the object images, the projection optics of the DLP were replaced by an 180 mm Olympus tube lens and an Olympus plan achromat microscope objective with a 40× magnification and an NA of 0.61. For the acquisition of 3D training data, the DLP is moved along the depth axis with a linear stage to introduce images from different distances to the diffuser endoscope.

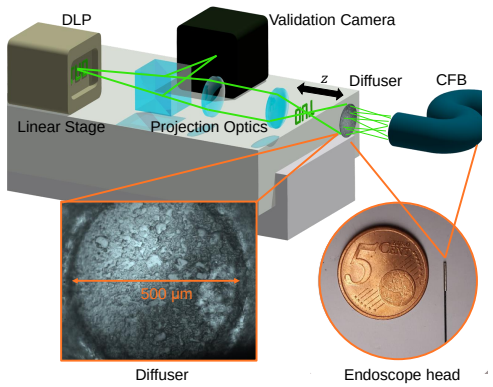

Fig. 4. (top) Training Setup. A digital light projector (DLP) on a stage is employed to project training images from different depths into the optical system and onto a validation camera. (bottom left) Diffuser surface captured by the validation camera. (bottom right) Comparison of the thin probe tip with a diameter of 700 µm to a coin.

The fluorescence images in Figure 2b show agglomerations of Rhodamine B marked Melamin resin particles with a diameter of 10 µm, placed on a cover glass. For the 3D imaging shown in Figure 2e, particles were placed on the front and back surface of the cover glass. The thickness and refractive index of the cover glass were 0.17 mm and 1.53, respectively (ISO 8255-1:1986), resulting in an optical path distance of 0.26 mm between the front and back surface.

Field of view and resolution

The diffuser endoscope's FoV and lateral resolution are compared to an endoscope where the diffuser is replaced with a paraxial lens at the same distance to the CFB. The theoretical lateral FoV of the diffuser endoscope is defined as

$$FoV(z) = \min[w, l + 2 \cdot z_{\mathrm{D}} \cdot \tan\alpha_{\mathrm{C}}] + 2 \cdot z \cdot \tan\alpha \quad (3)$$

where

$$\alpha_{\mathrm{C}} = \sin^{-1}\frac{A_{\mathrm{NC}}}{n} \quad (4)$$

is the light acceptance angle of the fiber cores, l the CFB diameter and w the diffuser diameter (Fig. 6).

The maximum angle of light acceptance α is defined as

$$\alpha = \beta + \sin^{-1}(n \cdot \sin\alpha_{\min}) \quad (5)$$

with

$$\alpha_{\min} = \min\left[\alpha_{\mathrm{C}}, \tan^{-1}\left(\frac{l+w}{2 \cdot z_{\mathrm{D}}}\right)\right] \quad (6)$$

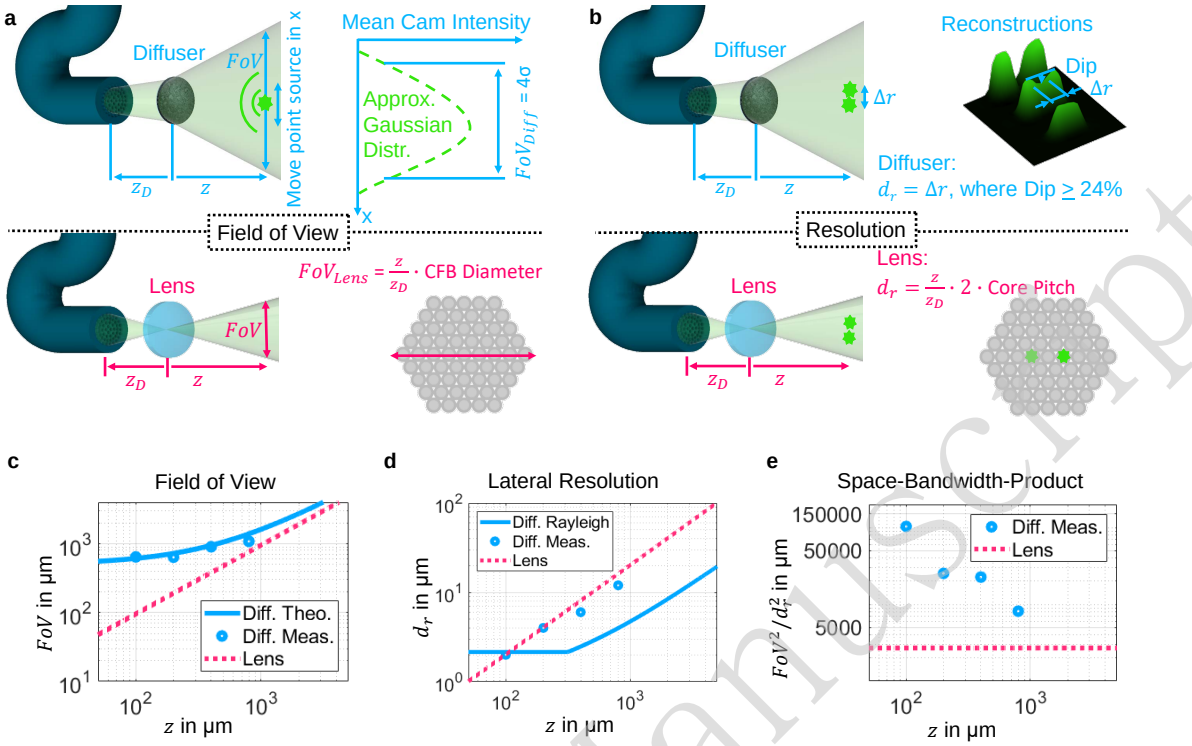

Fig. 5. (a) Determination of FoV for diffuser endoscope (top) and equivalent lens system (bottom). The refractive power of the lens is assumed to vary in such a way that the conjugate plane of the facet is always at the object distance z. (b) Determination of lateral resolution for diffuser endoscope (top) and equivalent lens system (bottom). The NA of the lens system is assumed to be the core NA. The lateral resolution of the lens system is determined by the magnified core pitch and not the diffraction limit. (c) Theoretical and measured FoV for the diffuser endoscope and theoretical FoV of equivalent lens system. (d) Theoretical and measured lateral resolution for the diffuser endoscope and theoretical lateral resolution of equivalent lens system. (e) Theoretical and measured SBP for the diffuser endoscope and theoretical FoV of equivalent lens system.

The stray angle β was defined as twice the standard deviation of a gaussian curve fitted to the stray angle spectrum. The stray angle spectrum was measured by illuminating the diffuser tape with a collimated beam and capturing the far field onto a camera. The FoV of the diffuser endoscope was additionally determined experimentally as the lateral object distance from which light is captured into the CFB (Fig. 5a, top).

The theoretical and experimental results of the diffuser endoscope's FoV are in good agreement. An offset of the diffuser diameter magnitude at zero axial distance is visible (Fig. 5c). In contrast, the FoV of the lens endoscope is proportional to the axial distance and goes down to zero at $z = 0$. Hence, the diffuser endoscope has an inherently larger FoV in the near field.

A theoretical limit for the lateral and axial resolution is given by the object sided numerical aperture of the system in air

$$A_N = \sin u_{max} \tag{7}$$

It is defined by the maximum angle u_{max}, from which light from an object on the optical axis can enter the CFB

$$u_{max} = \min\left[\tan^{-1}\left(\frac{\min[w_D, w]}{2\,z}\right), \alpha_C + \beta\right] \tag{8}$$

Here w_d is the ray intersection diameter at the diffuser for the case of an infinitely large core NA is defined as

$$w_D = l - \frac{2\,z_D}{n} \cdot \tan\left(\tan^{-1}\left[\frac{l/2}{z + z_D/n}\right] - \beta\right) \tag{9}$$

For the values given in table 2, the NA is limited by the fiber core aperture angle α_C at $z < 300$ µm and by the stray angle β at larger distances. For the fluorescence wavelength $\lambda = 0.65$ µm, a lateral resolution $1.2{\cdot}\lambda/A_N$ of (2...4) µm and an axial resolution $2{\cdot}\lambda/A_N^2$ of (10...32) µm can be achieved within the measurement volume according to the Rayleigh criterion.

Analogously, the diffuser endoscope's lateral resolution was experimentally determined by the distance between two points at which the intensity between the points drops

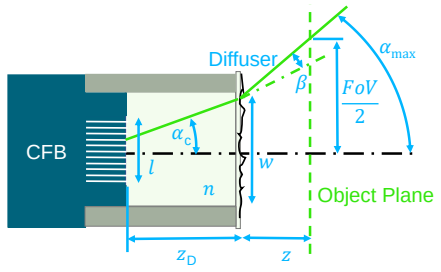

Fig. 6. Diffuser endoscope head. The FoV is determined by the object distance z and the maximum angle $\alpha_{\max}$ from which light can enter the cores of the CFB. $\alpha_{\max}$ is increased by the diffuser stray angle β and is either limited by the light acceptance angle of the fiber cores α_C or by the properties of the endoscope head, including the CFB diameter l, the glass spacer length z_D and refractive index n as well as the diffuser diameter w.

Table 2: Specifications of the diffuser endoscope head.

Quantity	Symbol	Magnitude
Metal ferrule diameter		700 µm
CFB diameter	l	350 µm
Core NA [68, 69]	A_{NC}	0.39
Core diameter [69]		1.7 µm
Core open area ratio [69]		0.21
Core pitch [69]		3.4 µm
Refractive index of glass spacer	n	1.5
Diffuser-CFB distance	z_D	500 µm
Diffuser diameter	w	500 µm
Diffuser stray angle (2σ)	β	6.2°

by 24 % (Fig. 5b). In case of the lens endoscope, two points are distinguishable in the CFB plane if there is at least one non-illuminated core between them. The lateral resolution of the diffuser endoscope and the lens endoscope have similar magnitudes and are both proportional to the axial distance of the system (Fig. 5d). This implies that the lateral resolution of the diffuser endoscope is also limited by the core pitch and the magnification z/z_D. The diffuser endoscope is diffraction limited at $z = 100$ µm (Fig. 5d). At larger distances, the magnified core pitch limits the two point resolution. The space-bandwidth-product (SBP) is defined as the squared quotient of the FoV and the lateral resolution. While the SBP of the lens endoscope is bound to the core number and therefore constant, the SBP of the diffuser endoscope increases with decreasing object distance, leading to an inherent imaging advantage in the near field (Fig. 5e). At an axial distance of 100 µm it exceeds the constant lens endoscope SBP by a factor of 38.

To determine the axial resolution experimentally, the intensity share curves in Fig. 2c were differentiated. The axial resolution was then defined as 2σ of the resulting peaks in approximation to the Rayleigh criterion. Within the measurement volume, the axial resolution ranges from 13 µm to 32 µm, depending on the object distance, which is in good agreement with the theory above.

Further improvement of the spatial resolution are possible, e.g. by increasing the stray angle of the diffuser. Furthermore, the fiber-facet-diffuser distance z_D can be adapted to match the magnified core pitch to the diffraction limit at other distances.

Neural network reconstruction

For the image reconstruction, two neural networks are trained sequentially. The first network is a Single Layer Perceptron (SLP), which consists of a single fully connected layer with weights and biases between all pixels of the input and the output layer. Moreover, the activation function was omitted, so that the network represents the pseudo-inversion of the intensity transmission matrix $\mathbf{H}$ from the linear forward model in (1). However, some artifacts remain in the background due to non-linearities in the optical system. Thus, the trained SLP is used to generate training data for a second network, a slightly modified U-Net, which serves as denoising network to clean up the artifacts arising from the intermediate SLP reconstruction. While the U-Net was originally developed for image segmentation [70], it can easily be adapted to be applicable for image-to-image regression by replacing the output layer. The employed U-Net was comprised of three encoding and decoding stages for the 2D and 3D reconstruction. For both networks, the loss function used for training is the mean squared error (MSE) between the reconstructed image and the ground truth. The reconstruction scheme is displayed in Fig. 7a. For all network trainings, an Adam optimizer [71] with a learning rate of 10^{-4} and a mini-batch size of 32 were used. As a convergence criteria, validation patience was employed, where the training is stopped after a certain number of validation iterations have passed without an improvement in the validation loss. For this paper, validations were calculated every 20 iterations and the training was stopped after 200 validations without improvement for the SLP and 100 validations for the U-Nets.

The performance of the combination of SLP and U-Net as well as each network alone is compared to the iterative ADMM algorithm introduced for a diffuser-based camera [40] in Fig. 7b for ”Running Man” test data which differs from the ”Augmented Digits” training data. The reconstruction quality is also examined with different metrics in table 3 as a cross-validation for ”Augmented Digits” training and test data. While the

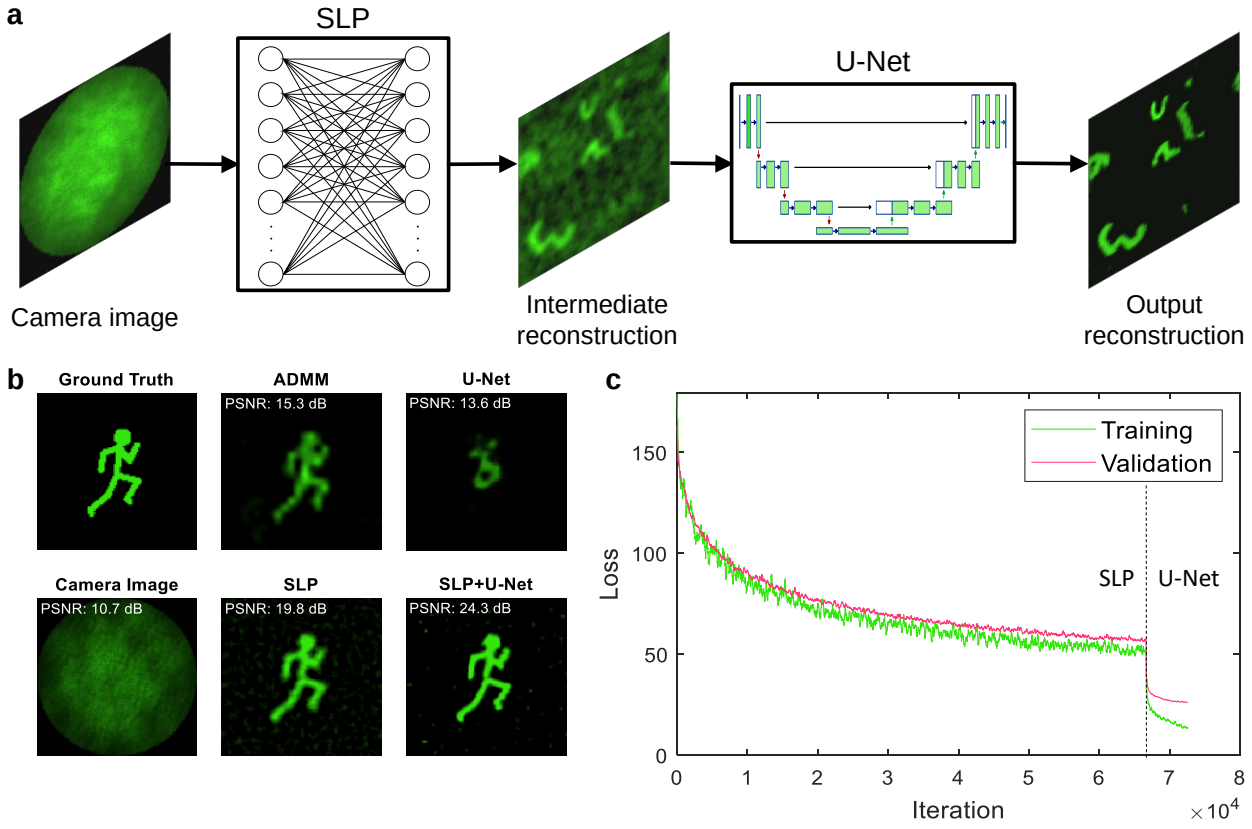

Fig. 7. (**a**) Neural Network Architecture for image reconstruction. The input measurement is fed into a Single Layer Perceptron (SLP) which is trained with Augmented MNIST digits so that the learned weights represent the inverse of the optical transfer function. The intermediate reconstruction is then denoised by a U-Net to retrieve the final output. The networks are trained separately. The training input of the U-Net are the predictions of the training data by the SLP. (**b**) 2D Reconstruction of "Running Man" test data with different approaches. Top, from left to right: Ground Truth Image. Reconstruction by iterative ADMM algorithm [40]. Reconstruction by single U-Net. Bottom, from left to right: Simulated Camera Image. Reconstruction by Single Layer Perceptron (SLP). Reconstruction by combination of SLP and U-Net. (**c**) Training and validation loss progress for SLP and U-Net.

Table 3: Comparison of reconstruction approaches for "Augmented Digits" data with cross-validation results for neural networks. Values are given with 1σ standard deviation.

Metric	ADMM	U-Net		SLP		**SLP+U-Net**	
	Test	Train	Test	Train	Test	**Train**	**Test**
PSNR	17.3±2.5 dB	23.0±4.3 dB	22.2±4.7 dB	23.1±2.8 dB	22.5±3.1 dB	**28.8±3.3 dB**	**26.4±4.3 dB**
SSIM	0.73±0.07	0.67±0.15	0.65±0.16	0.30±0.06	0.28±0.06	**0.92±0.04**	**0.89±0.06**
CC	0.09±0.14	0.79±0.08	0.75±0.09	0.78±0.10	0.76±0.11	**0.94±0.04**	**0.90±0.05**
Time	1541±28 ms	4.5±2.5 ms		2.9±2.5 ms		**7.4±4.8 ms**	

ADMM algorithm is capable of reconstructing images scattered by a diffuser with high fidelity given a sufficiently large memory effect, the addition of the CFB disturbs its reconstruction capabilities significantly. Due to the implemented regularization, the reconstruction is sparse, especially in the outer image are where the memory effect is not high enough. This leads to a rather high SSIM even if the correlation coefficient indicates a very bad result. The U-Net alone suffers from bad generalization and fails to extract features from the speckle patterns. The dark parts of the image are reconstructed accurately, resulting again in a misleadingly high SSIM. The SLP, on the other hand, is generalizing very well and is capable of reconstructing most of the features from the ground truth image, however, the

output image is blurry and contains some background noise leading to a low SSIM. With the help of the U-Net applied to the SLP output, this effect can be reduced significantly to retrieve a reconstruction with high fidelity. Furthermore, the reconstruction by the two consecutively applied neural networks is about 200 times faster than the iterative ADMM reconstruction approach. The combination of SLP and U-Net also outperforms the single networks as well as the ADMM algorithm in all metrics.

The dataset "Augmented Digits" is comprised of 10000 pairs of ground truth and speckle images for training as well as 1000 image pairs for validation and testing. The ground truth images are created by randomly selecting up to five images from the MNIST Digit dataset (28 × 28 pixels), randomly shifting, scaling and rotating them, and placing them randomly in the final image of 104 × 104 pixels. A reconstructed example can be seen in Fig. 7a (right). For 3D ground truth images, the samples from the MNIST digits dataset are placed at a random depth chosen from a set of layers (Fig. 2d). All network training and testing was performed on an NVIDIA RTX A6000 GPU of our workstation.

Acknowledgement

This work was supported by the German Research Foundation (DFG) under grant (CZ 55/48-1).

Author contributions

Conceptualization: JL, TG, RK; Supervision: RK, JC; Funding Aquisition: RK, JC; Project Administration: RK, JC; Methodology: JL, TG, RK; RK and JC are the principal investigators. JL performed the lab experiments, TG performed the simulation experiments and designed and evaluated the neural network reconstruction. JL and TG wrote the manuscript. All authors read and commented on the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Prevedel, R. et al. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature Methods* **11**, 727–730 (2014).

[2] Resendez, S. L. & Stuber, G. D. *In vivo* calcium imaging to illuminate neurocircuit activity dynamics underlying naturalistic behavior. *Neuropsychopharmacology* **40**, 238 (2015).

[3] Jacob, A. D. et al. A compact head-mounted endoscope for *in vivo* calcium imaging in freely behaving mice. *Current Protocols in Neuroscience* **84**, e51 (2018).

[4] Stirman, J. N. et al. Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nature biotechnology* **34**, 857–862 (2016).

[5] Bouchard, M. B. et al. Swept confocally-aligned planar excitation (SCAPE) microscopy for high-speed volumetric imaging of behaving organisms. *Nature Photonics* **9**, 113–119 (2015).

[6] Liu, J. D. et al. Light field endoscopy and its parametric description. *Optics Letters* **42**, 1804–1807 (2017).

[7] Zhou, P. et al. Light field endoscope calibration based on virtual objective lens and virtual feature points. *Optical Engineering* **59**, 104101 (2020).

[8] Aljasem, K. et al. Scanning and tunable micro-optics for endoscopic optical coherence tomography. *Journal of Microelectromechanical Systems* **20**, 1462–1472 (2011).

[9] Wurster, L. M. et al. Endoscopic optical coherence tomography angiography using a forward imaging piezo scanner probe. *Journal of Biophotonics* **12**, e201800382 (2019).

[10] Geng, J. & Xie, J. Review of 3-d endoscopic surface imaging techniques. *IEEE Sensors Journal* **14**, 945–960 (2014).

[11] Li, J. W. et al. Ultrathin monolithic 3D printed optical coherence tomography endoscopy for preclinical and clinical use. *Light: Science & Applications* **9**, 1–10 (2020).

[12] Kirsten, L. et al. Endoscopic optical coherence tomography with wide field-of-view for the morphological and functional assessment of the human tympanic membrane. *Journal of Biomedical Optics* **24**, 031017 (2018).

[13] Čižmár, T. & Dholakia, K. Exploiting multimode waveguides for pure fibre-based imaging. *Nature Communications* **3**, 1027 (2012).

[14] Turtaev, S. et al. High-fidelity multimode fibre-based endoscopy for deep brain *in vivo* imaging. *Light: Science & Applications* **7**, 92 (2018).

[15] Caramazza, P. et al. Transmission of natural scene images through a multimode fibre. *Nature Communications* **10**, 1–6 (2019).

[16] Zhu, C. Y. et al. Image reconstruction through a multimode fiber with a simple neural network architecture. *Scientific Reports* **11**, 896 (2021).

[17] Zhang, Q. et al. Learning the matrix of few-mode fibers for high-fidelity spatial mode transmission. *APL Photonics* **7**, 066104 (2022).

[18] Rothe, S. et al. Intensity-only mode decomposition on multimode fibers using a densely connected convolutional network. *Journal of Lightwave Technology* **39**, 1672–1679 (2021).

[19] Wen, Z. et al. Single multimode fibre for *in vivo* light-field-encoded endoscopic imaging. *Nature Photonics* **17**, 679–687 (2023).

[20] Porat, A. et al. Widefield lensless imaging through a fiber bundle via speckle correlations. *Optics Express* **24**, 16835–16855 (2016).

[21] Weiss, U. & Katz, O. Two-photon lensless micro-endoscopy with *in-situ* wavefront correction. *Optics Express* **26**, 28808–28817 (2018).

[22] Tsvirkun, V. et al. Flexible lensless endoscope with a conformationally invariant multi-core fiber. *Optica* **6**, 1185–1189 (2019).

[23] Sun, J. W. et al. Real-time complex light field generation through a multi-core fiber with deep learning. *Scientific Reports* **12**, 7732 (2022).

[24] Stasio, N., Moser, C. & Psaltis, D. Calibration-free imaging through a multicore fiber using speckle scanning microscopy. *Optics Letters* **41**, 3078–3081 (2016).

[25] Sun, J. W. et al. Quantitative phase imaging through an ultra-thin lensless fiber endoscope. *Light: Science & Applications* **11**, 204 (2022).

[26] Zhou, Y. et al. Light-field micro-endoscopy using a fiber bundle: a snapshot 3d epi-fluorescence endoscope. *Photon. Res.* **10**, 2247–2260 (2022).

[27] Kang, M. et al. Fourier holographic endoscopy for imaging continuously moving objects. *Optics Express* **31**, 11705–11716 (2023).

[28] Badt, N. & Katz, O. Real-time holographic lensless micro-endoscopy through flexible fibers via fiber bundle distal holography. *Nature Communications* **13**, 6055 (2022).

[29] Choi, W. et al. Flexible-type ultrathin holographic endoscope for microscopic imaging of unstained biological tissues. *Nature Communications* **13**, 4469 (2022).

[30] Di Leonardo, R. & Bianchi, S. Hologram transmission through multi-mode optical fibers. *Optics Express* **19**, 247–254 (2011).

[31] Scharf, E. et al. Video-rate lensless endoscope with self-calibration using wavefront shaping. *Optics Letters* **45**, 3629–3632 (2020).

[32] Stellinga, D. et al. Time-of-flight 3D imaging through multimode optical fibers. *Science* **374**, 1395–1399 (2021).

[33] Wu, J. C. et al. Single-shot lensless imaging with fresnel zone aperture and incoherent illumination. *Light: Science & Applications* **9**, 53 (2020).

[34] Shin, J. et al. A minimally invasive lens-free computational microendoscope. *Science Advances* **5**, eaaw5595 (2019).

[35] Anand, V., Rosen, J. & Juodkazis, S. Review of engineering techniques in chaotic coded aperture imagers. *Light: Advanced Manufacturing* **3**, 24 (2022).

[36] Vellekoop, I. M., Lagendijk, A. & Mosk, A. P. Exploiting disorder for perfect focusing. *Nature Photonics* **4**, 320–322 (2010).

[37] van Putten, E. G. et al. Scattering lens resolves sub-100 nm structures with visible light. *Phys. Rev. Lett.* **106**, 193905 (2011).

[38] Li, D. & Yao, Y. Scattering-lens based quantum imaging beyond shot noise. *Scientific Reports* **11**, 7785 (2021).

[39] Yanny, K. et al. Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy. *Light: Science & Applications* **9**, 171 (2020).

[40] Antipa, N. et al. Diffusercam: lensless single-exposure 3D imaging. *Optica* **5**, 1–9 (2018).

[41] Lyu, M. et al. Deep-learning-based ghost imaging. *Scientific Reports* **7**, 17865 (2017).

[42] Li, S. et al. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).

[43] Li, Y. Z., Xue, Y. J. & Tian, L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).

[44] Cheng, Q. Q. et al. De-noising imaging through diffusers with autocorrelation. *Applied Optics* **60**, 7686–7695 (2021).

[45] Khan, S. S. et al. Flatnet: Towards photorealistic scene reconstruction from lensless measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**, 1934–1948 (2020).

[46] Monakhova, K. et al. Learned reconstructions for practical mask-based lensless imaging. *Optics Express* **27**, 28075–28090 (2019).

[47] Rivenson, Y., Wu, Y. C. & Ozcan, A. Deep learning in holography and coherent imaging. *Light: Science & Applications* **8**, 85 (2019).

[48] Kuschmierz, R. et al. Ultra-thin 3D lensless fiber endoscopy using diffractive optical elements and deep neural networks. *Light: Advanced Manufacturing* **2**, 30 (2021).

[49] Borhani, N. et al. Learning to see through multimode fibers. *Optica* **5**, 960–966 (2018).

[50] Icha, J. et al. Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* **39**, 1700003 (2017).

[51] Ali, J. H. Spectral optical properties of gray matter in human male brain tissue measured at 400-1100 nm. *Optics* **4**, 1–10 (2023).

[52] Luo, Y. et al. Computational imaging without a computer: seeing through random diffusers at the speed of light. *eLight* **2**, 4 (2022).

[53] Hughes, M. Fibre bundle simulator (2022). At URL https://www.mathworks.com/matlabcentral/fileexchange/75157-fibre-bundle-simulator.

[54] Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).

[55] Wang, T. Y. et al. Image sensing with multilayer nonlinear optical neural networks. *Nature Photonics* **17**, 408–415 (2023).

[56] Yanny, K. et al. Deep learning for fast spatially varying deconvolution. *Optica* **9**, 96–99 (2022).

[57] Wang, H. D. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103–110 (2019).

[58] Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016, 694–711.

[59] Barbastathis, G., Ozcan, A. & Situ, G. H. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).

[60] Boominathan, V. et al. Recent advances in lensless imaging. *Optica* **9**, 1–16 (2022).

[61] Asif, M. S. et al. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging* **3**, 384–397 (2017).

[62] DeWeert, M. J. & Farm, B. P. Lensless coded-aperture imaging with separable doubly-toeplitz masks. *Optical Engineering* **54**, 023102 (2015).

[63] Singh, A. K. et al. Scatter-plate microscope for lensless microscopy with diffraction limited resolution. *Scientific Reports* **7**, 10687 (2017).

[64] Ludwig, S. et al. Scatter-plate microscopy with spatially coherent illumination and temporal scatter modulation. *Optics Express* **29**, 4530–4546 (2021).

[65] Li, S. H. et al. Memory effect assisted imaging through multimode optical fibres. *Nature Communications* **12**, 3751 (2021).

[66] Kuo, G. et al. On-chip fluorescence microscopy with a random microlens diffuser. *Optics Express* **28**, 8384–8399 (2020).

[67] Schmidt, J. D. Numerical Simulation of Optical Wave Propagation with Examples in MATLAB (Bellingham: SPIE, 2010).

[68] Han, J.-H. & Kang, J. U. Effect of multimodal coupling in imaging micro-endoscopic fiber bundle on optical coherence tomography. *Applied Physics B* **106**, 635–643 (2012).

[69] Steelman, Z. A. et al. Comparison of imaging fiber bundles for coherence-domain imaging. *Applied Optics* **57**, 1455–1462 (2018).

[70] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer–Assisted Intervention. Munich, Germany: Springer, 234–241.

[71] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations. San Diego, CA, USA: ICLR, 2014.

[72] Mahalati, R. N., Gu, R. Y. & Kahn, J. M. Resolution limits for imaging through multi-mode fiber. *Optics Express* **21**, 1656–1668 (2013).

[73] Scharf, E., Kuschmierz, R. & Czarske, J. Holographic lensless fiber endoscope with needle size using self-calibration. *tm-Technisches Messen* **86**, 144–150 (2019).

[74] Ferrer-Roca, O. Telepathology and optical biopsy. *International Journal of Telemedicine and Applications* **2009**, 740712 (2009).

[75] Allen, E. & Triantaphillidou, S. The manual of photography. 10th edn. (Amsterdam: Elsevier, 2011).

[76] Pratt, V. Direct least-squares fitting of algebraic surfaces. *ACM SIGGRAPH computer graphics* **21**, 145–152 (1987).

[77] Wu, J. C. et al. Learned end-to-end high-resolution lensless fiber imaging toward intra-operative real-time cancer diagnosis. *Print at https://arxiv.org/abs/2203.00008* (2022).

[78] Rothe, S. et al. Transmission matrix measurement of multimode optical fibers by mode-selective excitation using one spatial light modulator. *Applied Sciences* **9**, 195 (2019).

[79] Morales-Delgado, E. E., Psaltis, D. & Moser, C. Focusing and scanning of femtosecond pulses through a multimode fiber: applications in two-photon imaging and polymerization. Australian Conference on Optical Fibre Technology 2016. Sydney, Australia: Optica Publishing Group, 2016, AT5C–1 (2016).

[80] Czarske, J. W. et al. Transmission of independent signals through a multimode fiber using digital optical phase conjugation. *Optics Express* **24**, 15128–15136 (2016).

[81] Dorfman, R. A formula for the gini coefficient. *The Review of Economics and Statistics* **61**, 146–149 (1979).

[82] Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Computer Vision – ECCV 2016 (eds Leibe, B. et al.) Lecture Notes in Computer Science, 694–711 (Springer International Publishing, Cham, 2016).

[83] Shaw, P. J. Comparison of widefield/deconvolution and confocal microscopy for three-dimensional imaging. in Handbook of biological confocal microscopy (ed Pawley, J. B.) (New York: Springer, 2006), 453–467.