Research Article Vol. 15, No. 1 / 1 Jan 2024 / Biomedical Optics Express 222

Biomedical Optics EXPRESS

Optical tomography in a single camera frame using fringe-encoded deep-learning full-field OCT

VIACHESLAV MAZLIN[1,2,*]

[1] Institut Langevin, ESPCI Paris, PSL University, CNRS, 1 rue Jussieu, 75005 Paris, France
[2] Quinze-Vingts National Eye Hospital, 28 Rue de Charenton, 75012 Paris, France
[*] mazlin.slava@gmail.com

Abstract: Optical coherence tomography is a valuable tool for in vivo examination thanks to its superior combination of axial resolution, field-of-view and working distance. OCT images are reconstructed from several phases that are obtained by modulation/multiplexing of light wavelength or optical path. This paper shows that only one phase (and one camera frame) is sufficient for en face tomography. The idea is to encode a high-frequency fringe patterns into the selected layer of the sample using low-coherence interferometry. These patterns can then be efficiently extracted with a high-pass filter enhanced via deep learning networks to create the tomographic full-field OCT view. This brings 10-fold improvement in imaging speed, considerably reducing the phase errors and incoherent light artifacts related to in vivo movements. Moreover, this work opens a path for low-cost tomography with slow consumer cameras. Optically, the device resembles the conventional time-domain full-field OCT without incurring additional costs or a field-of-view/resolution reduction. The approach is validated by imaging in vivo cornea in human subjects. Open-source and easy-to-follow codes for data generation/training/inference with U-Net/Pix2Pix networks are provided to be used in a variety of image-to-image translation tasks.

© 2023 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

Optical sectioning methods play an important role in revealing fine details of biological structures. Compared to wide-field microscopy these methods provide higher axial resolution and improved rejection of out-of-focus light, however they require an extended acquisition time. For example, confocal microscopy (CM) needs to perform prolonged 2D pixel-by-pixel scanning, while structured illumination microscopy (SIM) needs to acquire numerous (e.g. 15) camera frames over time with distinct illumination patterns [1]. Long acquisition time reduces the data throughput and limits the maximal rate of dynamic biological processes that could be followed. In response to this problem research groups have presented a number of fast-sectioning approaches powered by the explosive progress in deep learning (DL). DL methods have proven to be very efficient at suppressing the uncorrelated camera/detector noise, which enabled high-quality SIM reconstruction from just 3 frames (5-fold reduction) [2,3]. Curiously, even the single image-to-single image cross-modality translations were demonstrated, such as wide-field → SIM [4–6], wide-field → CM [7], holography → high-NA sectioning microscopy [8,9], using the generative adversarial network (GAN) [10] architectures.

Optical coherence tomography (OCT) is another optical sectioning method that provides the unique advantage of decoupling axial resolution from the numerical aperture (NA) [11,12]. As a result, high axial sectioning (1 μm) can be achieved with a large field-of-view (2 cm) and long working distance to the sample (6 cm). Moreover, the suppression of out-of-focus light is particularly strong (exponential) in OCT due to coherence gating. These properties have led

#506664 https://doi.org/10.1364/BOE.506664
Journal © 2024 Received 21 Sep 2023; revised 29 Nov 2023; accepted 3 Dec 2023; published 14 Dec 2023

Biomedical Optics EXPRESS

to the wide adoption of OCT as an in vivo diagnostic tool in ophthalmology, dermatology and cardiovascular imaging.

Fundamentally, OCT reconstructs a tomographic image from the multiple optical phases collected with an interferometer. The phases can be obtained in two major ways: by modulating in time the optical path of one interferometric arm (time-domain OCT) and by collecting the interferometric spectral information (spectral-domain OCT and swept-source OCT). Both approaches have seen a significant breakthrough in speed thanks to parallelization of acquisition with 2D cameras. These camera-based methods are referred to as time-domain full-field OCT (TD-FF-OCT) [13,14] and swept-source full-field OCT (SS-FF-OCT) [15,16]. Today, SS-FF-OCT represents the fastest technique for volumetric imaging (7.8 GHz 3D voxel rate [17]), while TD-FF-OCT is the fastest for en face imaging (0.6 GHz en face pixel rate [18]). SS-FF-OCT and TD-FF-OCT were successfully applied for imaging of static structures and dynamic processes in the human eye in vivo (blood flow, tear film evolution, responses to photostimulation) [17–29].

Despite the fast imaging speed, the above methods are still sensitive to ocular motion, because each tomographic image is reconstructed from the multiple camera frames (multiple phases) that are captured sequentially in time. As was shown in [30] even 0.1 ms delay between the frames is sufficient to randomize the phase (between $-\pi$ and π) due to natural eye movements. This leads to phase reconstruction errors as well as doubling, blur and incoherent light artifacts [30]. Some phase errors can be corrected in post-processing in SS-FF-OCT [31]. Several groups proposed solutions to simultaneously capture several optical phases and reconstruct a clear image: via off-axis holographic full-field OCT with spatially coherent [32] and spatially incoherent [33] sources, via dividing the camera into four sectors with 0, $\pi/4$, $\pi/2$, $3\pi/4$ phases using polarization [14,34–37] or using a phase-modulated mirror array [38], via two [39] or four [36] synchronized cameras, by implementing a hyperspectral camera [40], by using line-illumination and diffraction grating [41], by employing multi-depth and multi-angle reference interferometric beams [42]. While being performant, these approaches either significantly reduce the field-of-view (FOV), lower the signal-to-noise ratio (SNR) or increase the cost and complexity.

In this work we present an alternative method, where only a single phase (and a single camera frame) is used for tomographic image reconstruction. The idea is to use fringes in low-coherence interferometry to encode high spatial frequency patterns in a single layer of the sample. High frequencies are caused by the natural roughness of biological tissues on the wavelength scale. This pattern is then efficiently learned and extracted from the background via a deep neural network trained on pairs of filtered raw camera frames and tomographic images. Importantly, this approach uses the full camera FOV, brings even higher SNR and comes at no additional cost, being implemented in the conventional time-domain full-field OCT interferometer. We demonstrate the benefit of the increased speed by imaging in vivo human cornea. The method is implemented in several versions with convolutional (U-Net) and generative (Pix2Pix) neural networks. The open-source codes are prepared to be easily extendable to different input image $\rightarrow$ output image translation problems.

2. Methods

2.1. Optical design

Optically, the device resembles the ophthalmic time-domain full-field OCT [22]. The instrument is composed of a Linnik interferometer with identical microscope objectives in both arms (Air/10×/0.3 NA, LMPLN10XIR, Olympus, Japan). The objectives determine a lateral resolution of 1.7 µm. The light source is a spatially incoherent 850 nm light-emitting diode (LED) with a 30 nm spectral bandwidth (M850LP1, Thorlabs, USA). Such a relatively extended bandwidth results in a short coherence length of 7.7 µm (axial resolution). Thanks to the interferometric optical configuration the light reflected from the different layers of the sample combines on the camera (Q-2A750-CXP, Adimec, Netherlands) with the light reflected from a single plane of

Biomedical Optics EXPRESS

the reference mirror. However, due to the short coherence length of the LED, the interference patterns are observed only for the light reflected from a thin sample layer, which matches with the reference mirror in terms of optical path length (with 7.7 µm precision). It is interesting to note that the observed interference patterns display significant variations and higher frequencies compared to the patterns expected from the interference of the two plane wavefronts (Fig. 1). This is caused by the irregularity of the wavefront reflected from the sample due to inherently diverse axial profiles and angular orientations of the cellular structures. Light reflected from those structures exhibits additional irregular optical phase shift. Thanks to the high resolution of Linnik interferometer we can also notice that the interferences, while being globally uncorrelated, exhibit the local continuity resembling the continuous nature of the underlying cellular meshes. Below, we will exploit these properties of the interference fringes to computationally extract them from the background.

2.2. *In vivo data collection*

The instrument in the hospital was approved by the French health agencies CPP and ANSM (numbers 2019-021B and 2019-A00942-55). The dataset was collected by imaging the in vivo corneas of 3 healthy young subjects (2 females and 1 male) using time-domain full-field OCT. The subject's head was resting on the chin/forehead rests. The imaging was non-contact and comfortable for the subject's thanks to the long working distance of the microscope objective (18 mm). To reduce the effects of eye movements the camera acquisition speed was set to a maximum of 550 frames/second with 1.75 ms exposure time per camera frame. The acquisitions were performed in bursts of about 20 frames followed by a 20 seconds of no-light break to ensure that the light exposure was below the maximal permissible limit. TD-FF-OCT captured en face images from the different depths of central stromal layers and endothelium that constitute 90% of the total corneal thickness (0.55 mm). The tomographic images were reconstructed from the four phase-stepped raw camera frames [14]. The phase step was controlled by the piezo motor moving the reference mirror as well as by the axial movements of the eye [30]. Furthermore, about 5 tomographic images were averaged to increase SNR. Therefore, each averaged tomographic image was reconstructed from 20 frames with an effective 35 ms exposure time. We saved both the averaged 4-phase tomographic images and the corresponding single-phase raw camera frames containing interference patterns.

Overall, we collected 3,000 raw camera frames that formed about 400 averaged 4-phase tomographic images (one averaged tomographic image corresponds to several single-phase raw frames). This dataset was used to create the input (single phase raw frame) ↔ 'ground truth' (averaged 4-phase tomography) image pairs for training the neural networks (Fig. 2). The images within the pair were pre-registered via cross-correlation. The images in the dataset had a higher resolution of 1440 × 1440 pixels compared to 256 ×256 pixels supported by the networks. To adapt to the common network architectures and also to increase the size of the training dataset, we split each image into the non-overlapping 2D mosaics (with each mosaic fragment being 256 × 256 pixels). This augmented the dataset from 3,000 to 48,000 unique image pairs.

2.3. *Single frame tomography processing*

Unfortunately, learning the direct transformation of single-phase raw frame ↔ averaged 4-phase tomography was not feasible, because the interference patterns occupied only a tiny fraction of the total dynamic range of the raw frame. More precisely, for weakly reflecting tissue like the cornea (surface reflectivity of about 2%) the dark-to-bright variation of interference brightness accounted only for 0.5% (0-20) of the full 12-bit range (0-4096). This value is lower compared to the 4-phase tomographic image, where the signal can be bit-sampled with a greater fineness and requires an expanded 7-bit range (0-80 among 0-127). As a result of the above, the small

Biomedical Optics EXPRESS

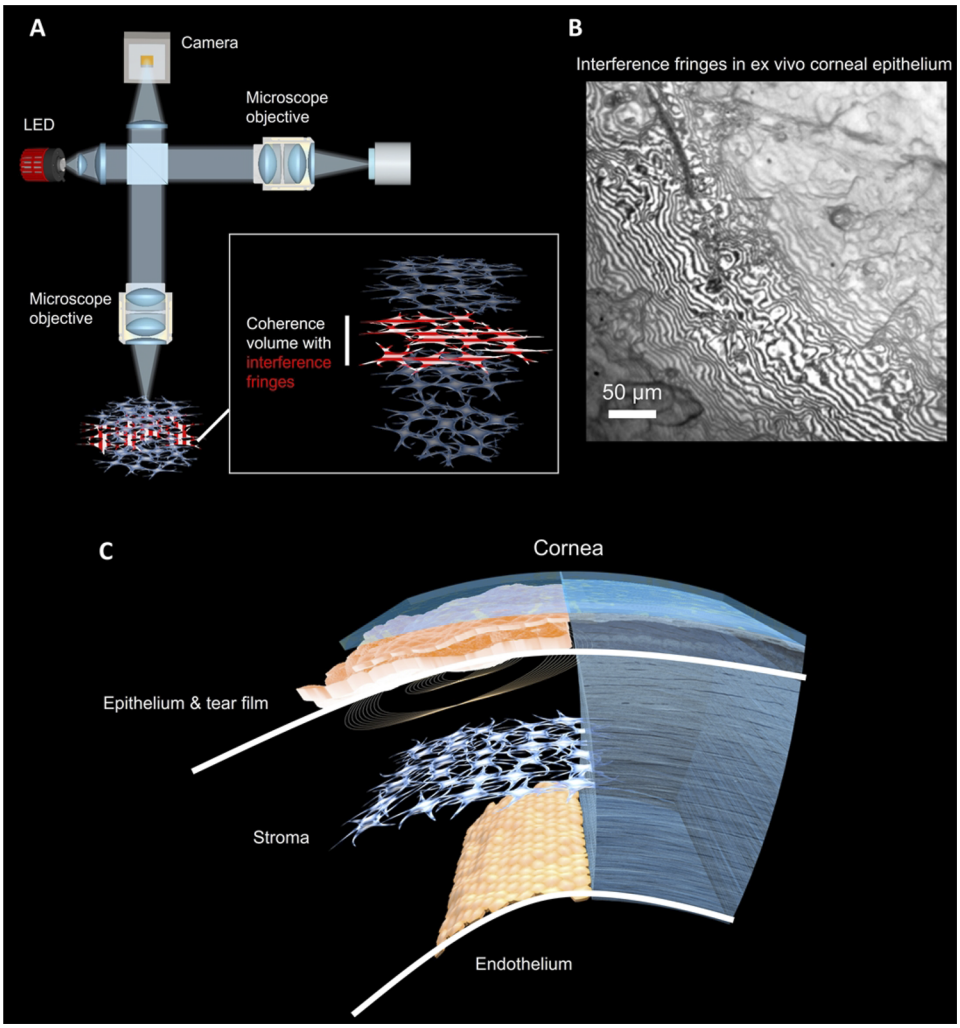

Fig. 1. Full-field OCT and interference fringes. A. Ophthalmic time-domain full-field OCT device. B. The fringes are localized within the thin coherence volume determined by the temporal coherence of the light source. They show irregular shapes with high spatial frequencies, resembling the inherently diverse organization of the cellular structures. Visualization 1 shows the phase-modulated fringes in ex vivo corneal epithelium. The presented fringes are located close to the sample surface and therefore are the most pronounced, while the fringes in the deeper sample layers can be significantly less contrasted. C. Schematic structure of the corneal layers imaged in this study. Full corneal thickness is about 0.55 mm.

interference fluctuations on top of the bright reflection background in raw frames were barely visible and were treated by the NN as a non-essential information.

For addressing that obstacle we collected the so-called background image by averaging around 1,000 frames from the various ex vivo samples and in vivo corneas. This image combined frames of different lateral and axial positions of the sample/cornea relative to the focal plane. There was no specific need for moving the position of the reference mirror outside of the coherence gate during the acquisition of those frames, because the interference patterns of each frame were completely blurred after averaging a thousand of them. This background image thus

Biomedical Optics EXPRESS

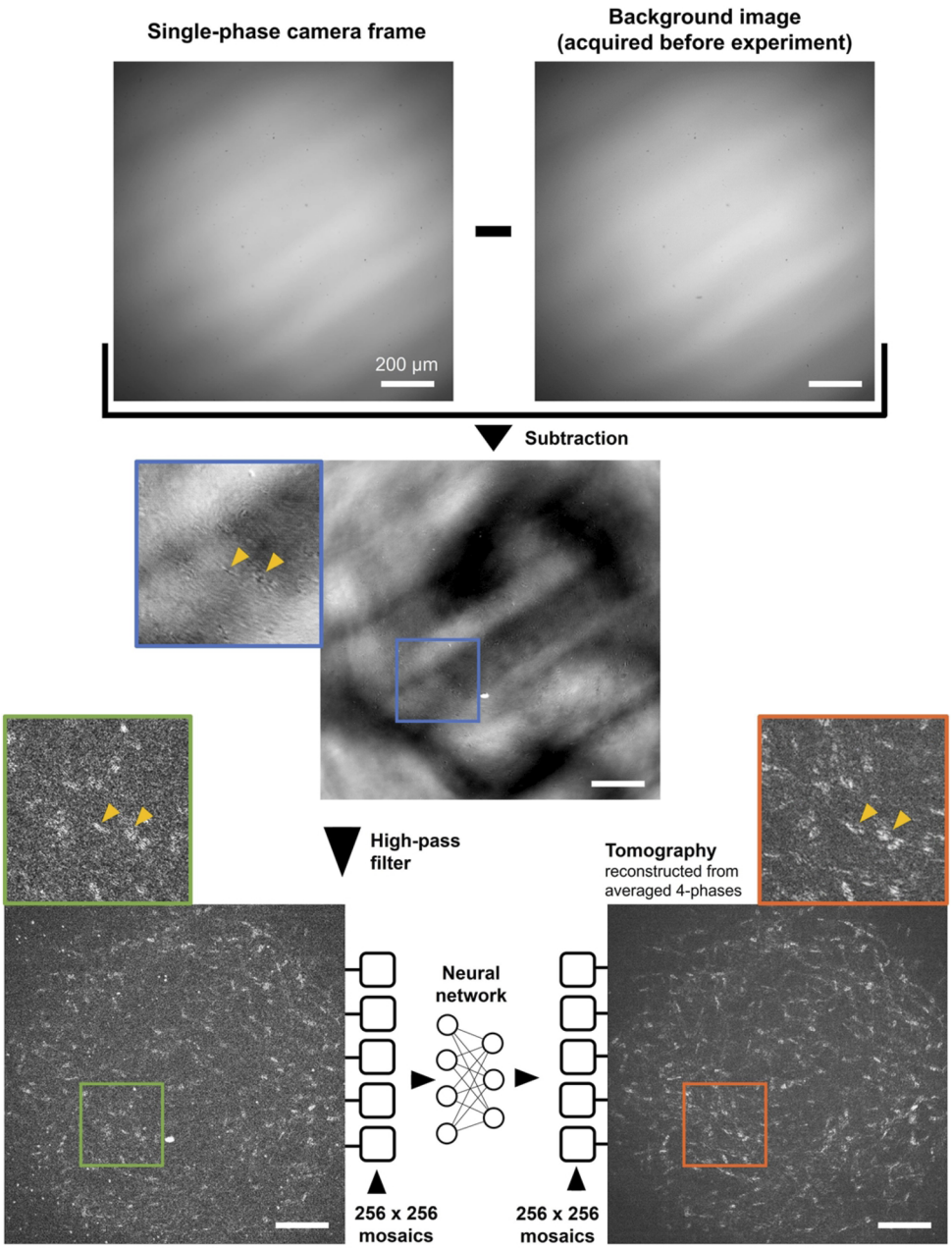

Fig. 2. Method of single-phase tomography. Tomographic image is obtained by: 1) collecting one raw camera frame (one phase), 2) increasing the contrast of interference pattern by subtracting the background image (acquired before the experiment and obtained through averaging around 1,000 frames), 3) high-pass filtering to retrieve the interferences, 4) mosaicking the original image into smaller 256×256 images, adapted for neural networks (NN), 5) NN processing to enhance the interference signals, 6) recombining the mosaics into the full image. The NN is trained on tomographic images obtained using conventional 4-phase reconstruction (in total from 20 camera frames = 4(phases) * 5(averaged)). Yellow arrows point at the nuclei of keratocyte stromal cells of an in vivo human cornea. The diagonal bright patterns in the single-phase camera frame, background image and central subtracted image are caused by the non-uniformity of the LED illumination (stripes of the LED chip). All scale bars are 200 µm.

Biomedical Optics EXPRESS

provided us with a typical background light level containing information about the illumination non-uniformity as well as non-uniformities of reflections from the reference mirror. Importantly, the collection of the background image has to be done only once at the instrumentation assembling stage and before the beginning of the actual imaging experiment.

In order to increase the interference fringe contrast, the background image was subtracted from each frame of the later acquired in vivo corneal dataset (Fig. 2). Then, the high frequency fringes were further highlighted by applying the directional central difference gradient high-pass filters [43]. Unfortunately, the final image contained low signal, comparable to the noise level. In order to enhance the interference signals we trained the NNs.

2.4. Details of neural network architectures

Two common neural network (NN) architectures were tested for recovering interference signals: convolution-based U-Net [44] and generative adversarial-based Pix2Pix [45]. We largely relied on the original implementations of those networks in Fastai/Pytorch [46] and Tensorflow [47] libraries.

The U-Net was composed of one contracting and one expanding ResNet34 backbones with skip connections between them. Additional self-attention [48] and blur layers [49] were added to increase the receptive field of convolution and avoid the checkerboard artifacts, respectively. The network had 41 M parameters and was supporting 8-bit, $256 \times 256 \times 3$ (RGB) input and output images. It was initialized with ImageNet pre-trained weights that were unfrozen during training. The network was trained using the Adam optimizer, weight decay = 1e-3, mean squared error (MSE) loss and a 1cycle policy with 80% / 20% proportion of epochs trained with increasing and decreasing learning rates, respectively (learning rate varied between 3e-3 and 1e-8).

The Pix2Pix was a conditional generative adversarial network composed of a generator and a discriminator. The generator was a U-net with 54 M parameters, while the discriminator was a convolutional PatchGAN classifier with 3 M parameters. The network was supporting 8-bit, $256 \times 256 \times 3$ (RGB) input and output images. The training used the Adam optimizer with a learning rate of 2e-4.

The detailed U-net and Pix2Pix architectures can be found in Supplemental Document 1.

It is important to mention that while the current data preparation routines in Fastai/Pytorch and Tensorflow libraries primarily support 8-bit images, the codes can be adjusted in the future to support 32-bit images as the NN weights are already set for 32-bit precision.

From the 48,000 images of the dataset 90% were used for training and 10% for validation.

U-Net was trained for 40 epochs, which took 10 hours on a single GPU (Titan RTX, Nvidia, USA). Pix2Pix was trained for 100 epochs, which took 20 hours.

3. Results

The tested frames were neither part of the training nor part of the validation datasets. Each raw camera frame containing interference patterns was processed by subtracting the precomputed background image, followed by high-pass filtering. Then the images were split into a 2D mosaic of overlapping sub-images and treated by the U-Net NN. Then, the sub-images were sharpened by the unsharp mask filter (radius 1.0, amount 4.0) [47] to compensate for the well-known problem of blurriness of U-Net output, when trained with MSE loss [48]. Visually, the filtering gave an impression of performing deconvolution as illustrated in Visualization 2. Finally, the sub-images were recombined into the final tomographic view. It is important to note that during the assembling of sub-images their overlapping regions were averaged, which smoothed the contrast mismatch on the sub-image borders and also suppressed the additional noise from the unsharp filter. The resulting single-phase tomographic images were compared with tomographic images reconstructed via the conventional optical 4-phase modulation scheme (computed in total from 20 frames = 5 averaged 4-phase images).

Biomedical Optics EXPRESS

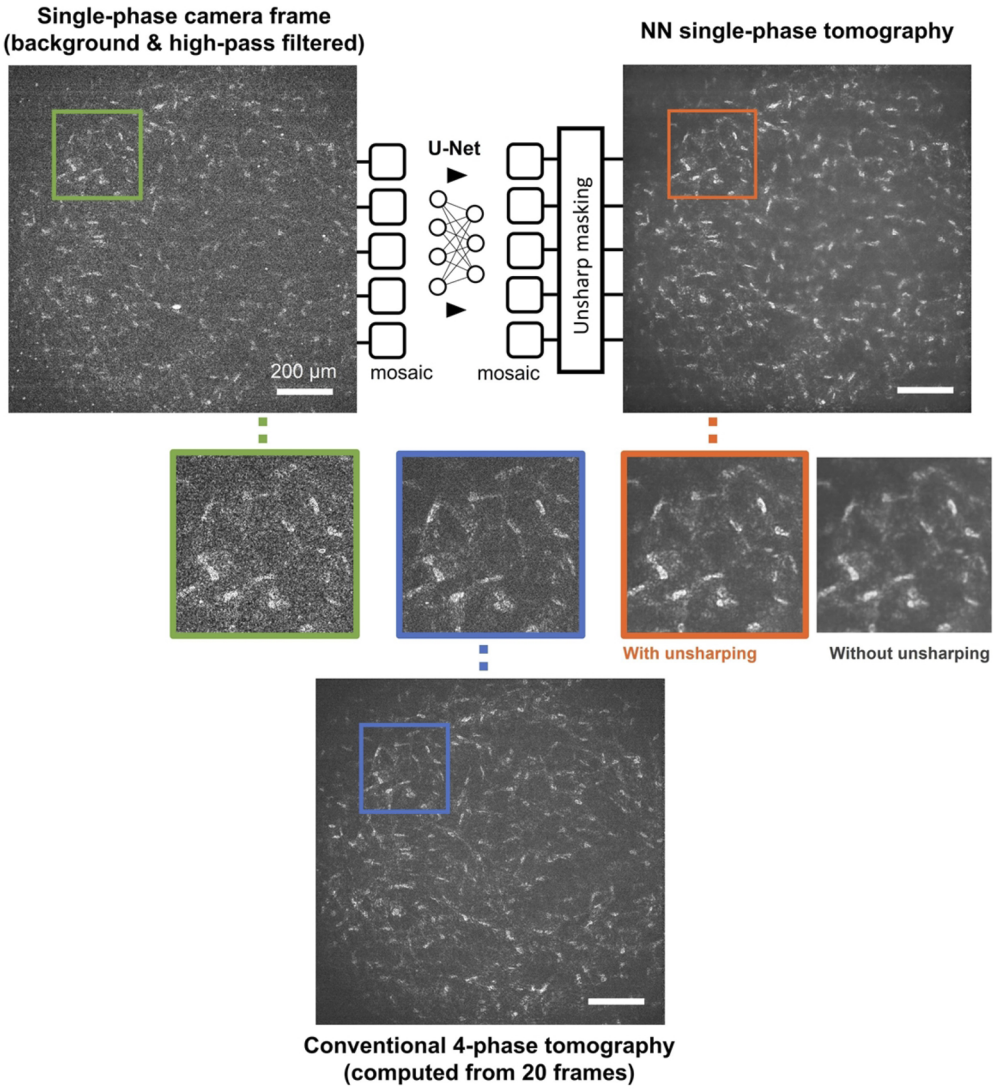

Fig. 3. Single-phase tomography in human corneal stroma in vivo. The method outperforms the conventional averaged 4-phase tomography in terms of speed and SNR. The benefit of unsharp filtering of mosaic images is illustrated in Visualization 2.

Our method successfully reconstructed the tomographic views from the stroma of in vivo human cornea (Fig. 3). The 15 µm oval-shaped nuclei and bodies of keratocyte cells were resolved. One can immediately notice the benefits of U-Net in signal recovery and noise suppression. Indeed, NN brought 25 dB SNR improvement to the original high-pass filtered image (Table 1), and even surpassed the SNR of the ground truth tomographic image reconstructed from 20 frames. This noise-suppression quality of U-Net is well known and is caused by the expected inability of NN to learn and correctly predict the random uncorrelated Gaussian noise. We computed SNR as $10 \cdot \log_{10}[S/\sigma]^2$, where S was the max signal level in the image and σ was the standard deviation of the noise, measured in the corner of the image that was free from sample features. While that SNR method was absolute, the alternative PSNR and SSIM methods were relative and used the conventional 4-phase as a clean reference image.

Biomedical Optics EXPRESS

Table 1. Performance of single-phase tomography relatively to conventional 4-phase imaging

	Single-phase (background & high-pass filtered)	Single-phase (U-Net)	Single-phase (Pix2Pix)	Conventional 4-phase (computed from 5 averaged 4-phase images)
SNR	15 dB	**40 dB**	30 dB	30 dB
PSNR[a] (Higher is better)	12.9	**19.1**	18.2	*Reference*
SSIM[a] (Higher is better)	0.14	0.18	**0.20**	*Reference*
Exposure time	**1.7 ms**	**1.7 ms**	**1.7 ms**	35 ms
Full time (exposure + post-processing)	**2 ms**	1.7 + (>700) ms (2D mosaicking + NN inference on 36 frames)	1.7 + (>800) ms (2D mosaicking + NN inference on 36 frames)	**35 ms**

[a]The marked criteria were computed taking the 4-phase as a reference image

Although the visible structures looked similar in the single-phase and averaged 4-phase tomographies, their structural similarity index (SSIM) was low at 0.18 (Table 1). The reason is that the supposedly clean ground truth 4-phase image actually presents the substantial random variations caused by the eye movements. For example, due to the movements each of the phases is captured from a slightly shifted coherence plane in the sample. Then, the reconstructed 4-phase image integrates all the coherence planes and shows more scattering structures than are visible to each single-phase alone. The axial movements also add a random error to the predetermined optical phases and, as a result, the final image contains the residual interference fringes covering the sample structures. This random error cannot be guessed from a single-phase. Finally, the sharpness of single-phase and 4-phase images can be different, determined by the blur of U-Net processing and the strength of the unsharp filter. Despite this mismatch between the clean image and the trained ground truth 4-phase data, the NN shows the capability to learn a general task of enhancing the interference signal that is present in a single phase.

The high-pass filtering preceding the NN also played a role in the noise reduction by suppressing the low-frequency residual interference fringe artifacts that are known to occur at the flat interfaces such as Descemet's membrane in the cornea [30] (Fig. 4).

We also successfully reconstructed the more regular corneal structures such as endothelial cell mosaics (Fig. 5). Interestingly, the NN reconstructed both the cells and high-frequency residual fringe artifacts. The latter are undesired and can be partially suppressed with a mean filter. NN also learned to remove the dust particle artifacts that were present in the raw images, but were absent in the ground truth tomographic data.

The single-phase method is particularly valuable for in vivo applications. For example, the newly opened possibility of reducing the exposure time by more than 10× (comparing to the averaged 2-phase or 4-phase images with matching SNR) significantly improves the light safety of the device. Alternatively, a more powerful light source can be used at the current safety level to enhance the image quality.

Shorter exposure time enables tomographic frame-rates at full camera speed of 550 Hz (Visualization 3). That being said, the single-phase approach comes at a cost of additional 700 ms delay per captured frame in post-processing time. Although the single NN inference takes only 20 ms on Titan RTX GPU, one has to perform the processing 36 times for each mosaicked 256×256 sub-image of a full 1440×1440 camera frame. The inference takes even longer, if one uses overlapped mosaicking to avoid the edge artifacts. We expect that orders of magnitude improvement in the inference speed can be reached in the future by training models adapted for

Biomedical Optics EXPRESS

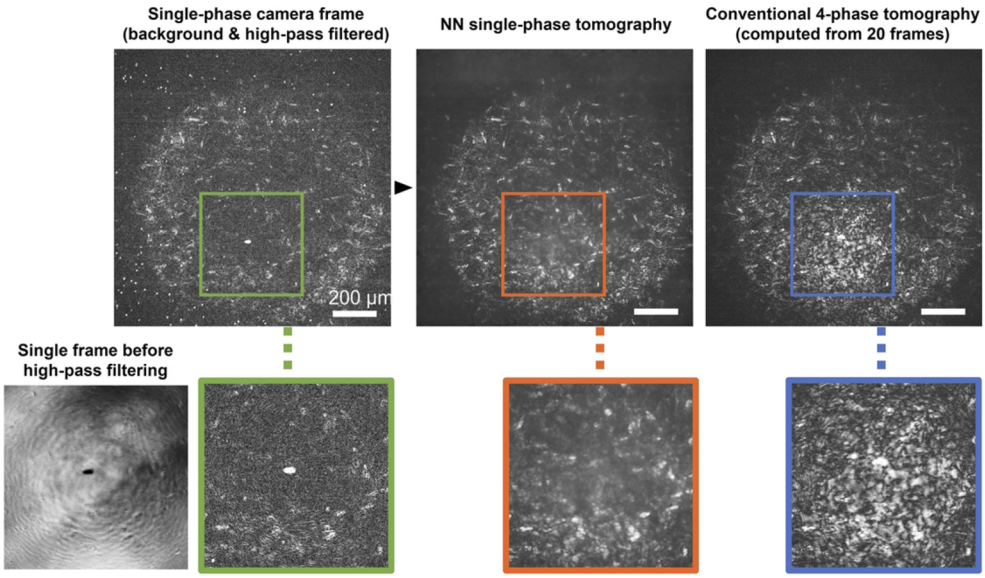

Fig. 4. Suppression of low-frequency residual interference artifacts in single-phase tomography. The fringe artifacts originate from the regular and flat interface of Descemet's corneal membrane. The suppression occurs at the high-pass filtering step that is absent in the conventional 4-phase tomographic method.

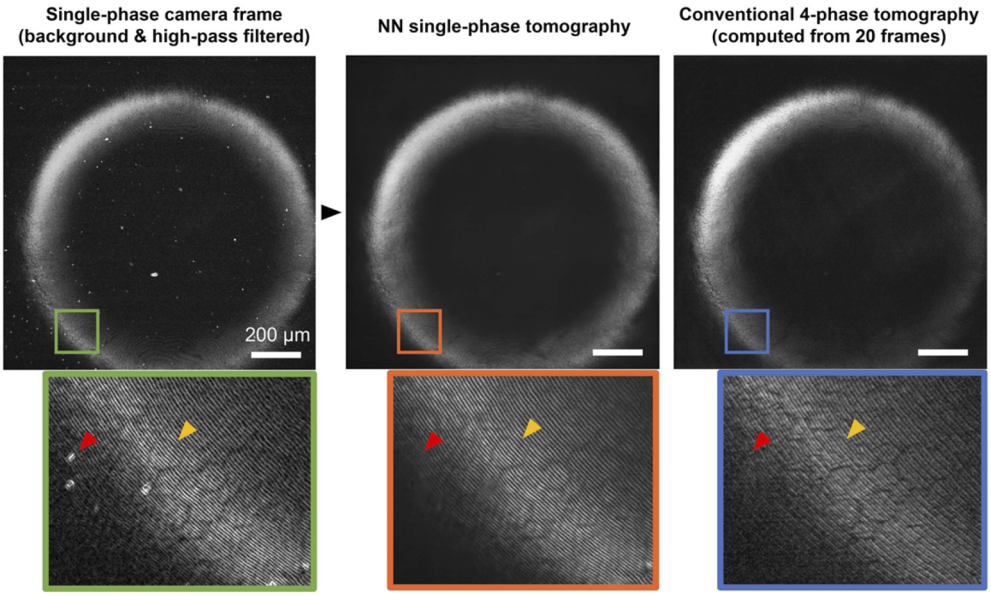

Fig. 5. Endothelial cell mosaic imaging with single-phase tomography. The regular cell structures are well reconstructed (e.g., as shown with a yellow arrow). However, the NN also learns to produce the high-frequency interference artifacts that are inherent to both input and ground truth training data. On the other hand, NN efficiently removes the dust particles (e.g., as indicated by a red arrows).

Biomedical Optics EXPRESS

larger data. For example, a U-Net model adapted for 2048×2048 images can process the full 1440×1440 camera frame in a single pass without mosaicking. Although making such a model is within reach even on consumer GPU (12 GB), it would require collecting considerably more training data as the mosaicking cannot be used for data augmentation.

One important advantage of the single-phase imaging is that it waives the demand to have at least two well-aligned frames for tomographic reconstruction. This is important for addressing several known problems. For example, it was shown before that the lateral shift between the consecutive camera frames leads to out-of-focus artifacts in tomographic images, while the axial shift results in signal intensity blinking due to random phase fluctuations [30]. As is shown in Fig. 6 and Visualization 4 the single-phase tomography produces images that are clean from the out-of-focus artifacts. It also makes the signal level consistent over time and immune to phase modulation between the consecutive camera frames (Fig. 7 and Visualization 5). The only movements that can still affect the signal level are the fast ones that make a significant shift during the exposure time of one phase leading to fringe washout. The above decoupling of the signal level from the movements is particularly important as the majority of clinical diagnostic methods use tissue reflectivity to discriminate between the healthy and pathological cases.

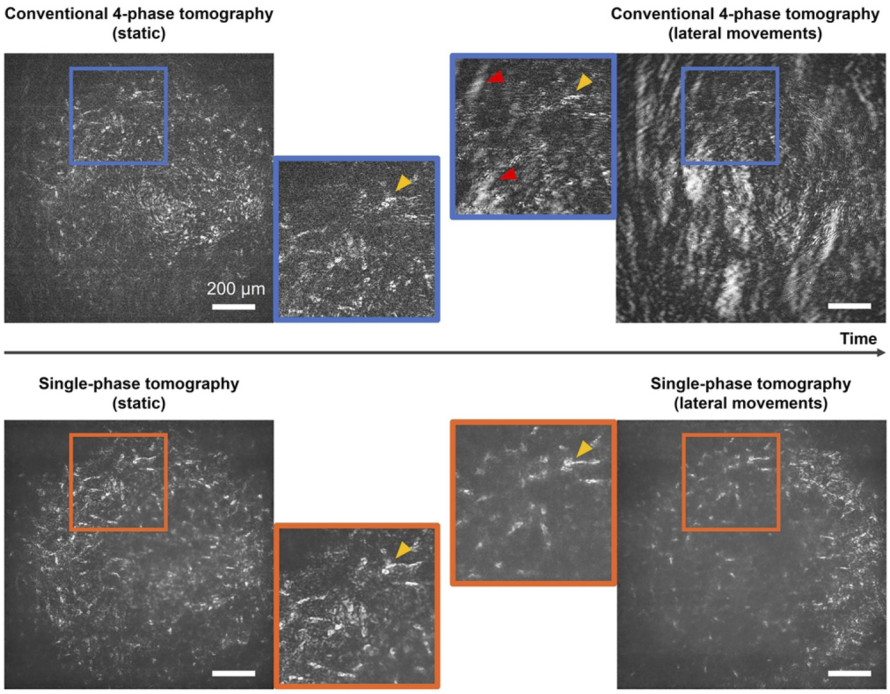

Fig. 6. Suppression of the lateral movement artifacts in single-phase tomography. The lateral movements cause misalignment between the consecutive camera frames leading to imperfect rejection of the out-of-focus light in the final 4-phase tomographic images. The artifacts (red arrows) show a defocused view from the air/sample reflecting interface and hinder the keratocyte cell nuclei (yellow arrows). Single-phase tomography reconstructs the tomographic image without artifacts albeit the signal may be reduced, if the movement occurs during the exposure time of one frame. The comparison is illustrated in Visualization 4.

Another benefit of the single-phase method is that the tomographic imaging speed is becoming controlled primarily by exposure time and not the frame rate. This means that the expensive specialized 500–10,000 FPS cameras can be substituted by using cheap conventional 20 - 30 FPS

Biomedical Optics EXPRESS

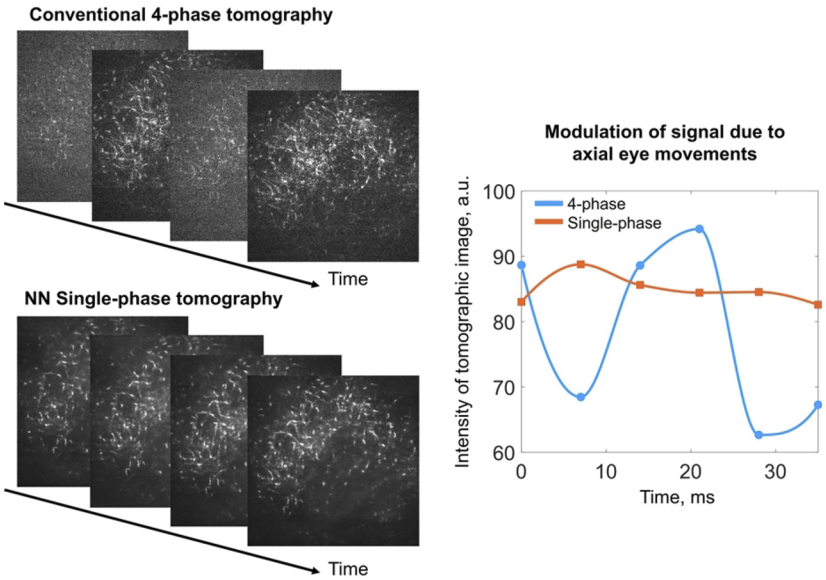

Fig. 7. Suppression of the axial movement artifacts in single-phase tomography. The axial movements cause randomization of optical phase between the consecutive camera frames leading to uncontrollable signal blinking in the 4-phase tomographic images. On the other hand, the signal is consistent in the tomographic images reconstructed from the single phase. The comparison is further illustrated in Visualization 5.

ones and setting a low exposure time. As a result, the fast tomographic imaging becomes broadly accessible.

Lastly, we compared the performances of the convolutional U-Net with that of the alternative neural network architecture - generative adversarial Pix2Pix (Fig. 8). Contrary to U-Net, Pix2Pix network could reconstruct the sharp images matching the target 4-phase tomography without using the unsharp filter. This was reflected in the slightly higher SSIM (Table 1) of Pix2Pix. However, Pix2Pix partially replicated the noise in the target images resulting in the considerably lower SNR. Pix2Pix also struggled to reconstruct the weaker signals as seen in the corners of the image leading to the additional hazy noise. One significant drawback of Pix2Pix is the unstable training. For example, the undertrained Pix2Pix network produced pattern artifacts that were repeating across the field and that could be mistakenly confused for being the sample structures. On the contrary, U-Net did not show signs of that behavior and the undertraining resulted in just the more blurry image. As such the U-Net would be more appropriate in the medical setting.

We should mention one major limitation of the single-phase tomographic method – it successfully produced images from the middle and deep corneal structures (100 µm – 550 µm) but not close to the surface (above 100 µm depth). When imaging the deeper structures the surface becomes defocused and does not show the high-frequency details. Thus, the only high-frequency patterns in the sample are generated by the optical interferences from the tomographic layer of interest. In this case the tomographic view can be easily extracted. On the other hand, when imaging at the small depths the surface structures are insufficiently defocused and have high frequencies overlapping with those of interferences. As a result, the surface structures hinder the underlying tomographic layer.

Biomedical Optics EXPRESS

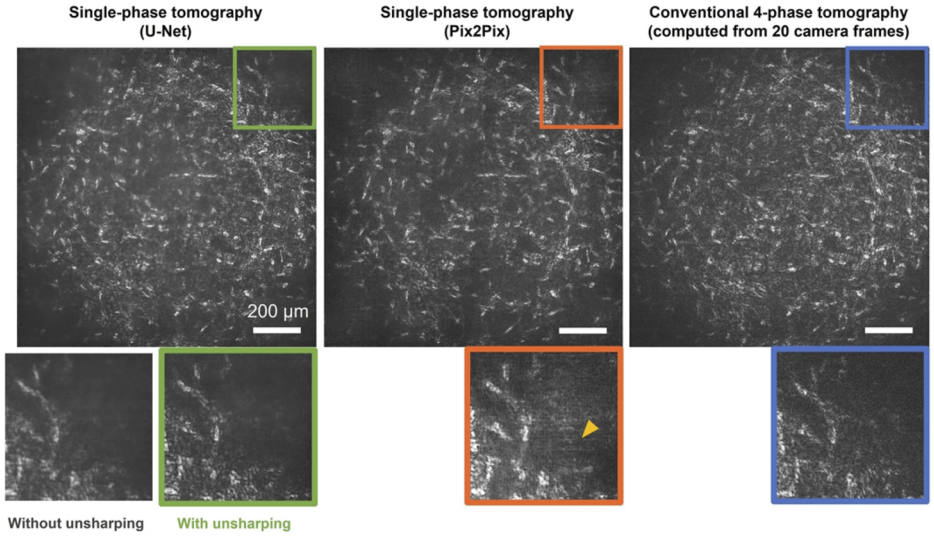

Fig. 8. Comparison of U-Net and Pix2Pix networks for tomographic image reconstruction. Pix2Pix produces sharp images even without the unsharp filter thanks to the discriminator in the architecture that easily labels the blurred images as fake. However, the same architecture forces the network to create structures even in complicated noisy parts of the images leading to haze artifacts (yellow arrow).

In order to quantitatively estimate this effect we used the Gaussian beam model. The PSF beam diameter in the focus is:

$$D_0 = \frac{2 \cdot \lambda \cdot \sqrt{n^2 - NA^2}}{(\pi \cdot n \cdot NA)} \tag{1}$$

While the PSF diameter at axial depth z is:

$$D_z = D_0 \cdot \sqrt{1 + \left(\frac{z}{z_0}\right)^2} \tag{2}$$

where the $z_0 = \pi \cdot {D_0}^2 \cdot n/(4 \cdot \lambda)$.

Considering $\lambda = 0.85\,\mu m$, $NA = 0.3$, $n = 1$ (*in air*), we measure the PSF diameter $D_0 = 1.7\,\mu m$ in the focus, which matches the lateral resolution of the optical system. At the smallest depth that we can probe (around 100 µm) the PSF diameter becomes $D_{100\,\mu m} \approx 120\,\mu m$. The extension of the PSF by 70× is sufficient to completely blur the background and suppress the high frequencies overlapping with interference fringes.

One can achieve the faster background blurring and larger accessible depth range by using optics with higher NA. For example, for NA = 0.5 the 70× blurring will be reached already at a depth of $z = z_0 \cdot \sqrt{70^2 - 1} = 14\,\mu m$. Given that 14 µm almost matches the coherence gate thickness, such configuration would enable full range corneal imaging.

4. Discussion

Single-phase tomography employs three key ideas: 1) encoding structural patterns in the sample layer using low-coherence interference fringes, typically considered as undesired artifacts, 2) relying on the inherent roughness of the biological tissues on the wavelength scale to increase the spatial frequency of the patterns, 3) training a neural network to efficiently extract these structured patterns from the noisy background.

Biomedical Optics EXPRESS

In comparison to the conventional reconstruction methods requiring several phases such as time-domain full-field OCT [22], optical transmission tomography (OTT) [50,51], the single-phase approach is indifferent to the phase shift between the consecutive camera frames. The phase errors induced by the sample movements do not propagate to create the artifacts and the tomographic signal stays consistent over time. This decoupling of signal from the movements is an important milestone for bringing full-field OCT techniques to clinics, as the majority of diagnostic biomarkers rely on the comparison of tissue reflectivities. The new approach also allows one to suppress the fastest movements occurring during the single frame by reducing the camera exposure time. The sub-millisecond exposure time is currently available in every global shutter CMOS camera including the consumer ones, which opens a path for a low-cost tomography in the future. However, the chosen sensor should still possess the full-well capacity and dynamic range, sufficient to detect the small interference-induced contrast variations within the bright background. The requirement for full-well capacity is less strict for more reflective samples as their interference signals can occupy larger proportion of the full dynamic range of the sensor.

The proposed single-phase tomography is a general method and is applicable to any samples that exhibit structural roughness on the micrometer scale. The latter requirement makes the biological tissues and in vivo organs naturally suited candidates for imaging. The key challenge is to acquire a significantly large image database for the desired organ. Although, we showed the feasibility of collecting such a database from one of the most dynamic organs – in vivo human eye (cornea), the target images were not completely free from the motion-related interference fringe artifacts. As a result, the neural network also learned to reproduce those artifacts in the final tomographic views. On the other hand networks appeared to be very efficient at suppressing the uncorrelated Gaussian noise, producing images with higher SNR than the ground truth averaged 4-phase data, the latter requiring 10× - 20× longer imaging time.

It is interesting to mention that the spatial frequency of interferences can also be increased optically by controlling the wavelength, by tilting the reference arm relatively to the sample arm, by using a glass plate in one optical arm [52] or by using the curved reference mirror [27]. These solutions open future possibilities of applying the single-phase FF-OCT method for inspection of flat samples with inherently low fringe frequency, such as semiconductor display panels.

One limitation of the proposed single-phase approach is the long post-processing time ranging from 1 second up to several minutes for a single image, which introduces a considerable delay in real-time imaging applications. Orders of magnitude faster speeds are anticipated to become possible by using larger versions of the networks adapted to process the full image at once without mosaicking. However, such networks would require more training data.

Another limitation of the method is the reduced imaging range. The imaging plane should be sufficiently deep inside the sample to enable defocused blurring of the sample surface. Although with the 0.3 NA the usable imaging range begins at 100 µm depth, the range can be efficiently extended with the use of higher NA optics. For example, 0.5 NA is expected to blur the surface faster, starting already at 14 µm depth. Some existing 0.5 NA objectives with the long working distance (10 mm) can be suited for non-contact in vivo examination. Alternatively, any NA is expected to work with immersion objectives which remove the strongly reflective air/tissue interface or for the sample layers that are naturally located deeper than the surface, such as retina.

From our experience the convolutional U-Net architectures were considerably more robust to underfitting and lack of data compared to their generative counterparts, such as Pix2Pix. The inherent blurring of U-Net can be efficiently corrected with an unmask filter applied before assembling the full FOV out of mosaicked sub-images.

We expect the increasingly broader integration of the neural networks solutions into the medical imaging devices in the future. As of today FDA has authorized more than 500 AI-enabled medical devices [53], including those that use deep learning for image reconstruction [54] as

Biomedical Optics EXPRESS

well as those capable of autonomous diagnosis in ophthalmology without the doctor [55]. The deep learning methods hold promise to decrease the patient exposure to radiation, improve the diagnostic accuracy through image enhancement and through integration of the different clinical data in a single-text based analysis. They are also expected to tackle the known shortcomings of the medical system such as lack of trained personnel and increasing costs.

Funding. Agence Nationale de la Recherche (ANR-10-IDEX-0001-02 PSL, ANR-22-CE19-0018, ANR-22-CE45-0005).

Acknowledgement. The author would like to warmly thank Albert Claude Boccara and Samer Alhaddad for many fruitful discussions.

Disclosures. VM: SharpEye SAS (I).

Data availability. Full Jupyter notebooks with U-Net/Pix2Pix codes and trained weights are provided in [56] . The codes include the data preparation routines for convenient training and inference on an arbitrary image-to-image translation tasks. The full-scale image processing examples (raw camera frame -> background subtraction -> high pass filtering -> NN processing) can also be found in the link above.

Supplemental document. See Supplement 1 for supporting content.

References

1. A. Markwirth, M. Lachetta, V. Mönkemöller, *et al.*, "Video-rate multi-color structured illumination microscopy with simultaneous real-time reconstruction," Nat. Commun. **10**(1), 4315 (2019).
2. L. Jin, B. Liu, F. Zhao, *et al.*, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," Nat. Commun. **11**(1), 1934 (2020).
3. C. Ling, C. Zhang, M. Wang, *et al.*, "Fast structured illumination microscopy via deep learning," Photonics Res. **8**(8), 1350 (2020).
4. X. Zhang, Y. Chen, K. Ning, *et al.*, "Deep learning optical-sectioning method," Opt. Express **26**(23), 30762 (2018).
5. H. Zhuge, B. Summa, J. Hamm, *et al.*, "Deep learning 2D and 3D optical sectioning microscopy using cross-modality Pix2Pix cGAN image translation," Biomed. Opt. Express **12**(12), 7526 (2021).
6. C. Bai, J. Qian, S. Dang, *et al.*, "Full-color optically-sectioned imaging by wide-field microscopy via deep-learning," Biomed. Opt. Express **11**(5), 2619 (2020).
7. B. Li, S. Tan, J. Dong, *et al.*, "Deep-3D microscope: 3D volumetric microscopy of thick scattering samples using a wide-field microscope and machine learning," Biomed. Opt. Express **13**(1), 284 (2022).
8. Y. Wu, Y. Luo, G. Chaudhari, *et al.*, "Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," Light: Sci. Appl. **8**(1), 25 (2019).
9. Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," Light: Sci. Appl. **8**(1), 85 (2019).
10. I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2014), Vol. 27.
11. D. Huang, E. A. Swanson, C. P. Lin, *et al.*, "Optical Coherence Tomography," Science **254**(5035), 1178–1181 (1991).
12. W. Drexler and J. G. Fujimoto, eds., *Optical Coherence Tomography* (Springer International Publishing, 2015).
13. E. Beaurepaire, A. C. Boccara, M. Lebec, *et al.*, "Full-field optical coherence microscopy," Opt. Lett. **23**(4), 244–246 (1998).
14. A. Dubois, *Handbook of Optical Coherence Microscopy: Technology and Applications* (Pan Stanford, 2015).
15. B. Považay, A. Unterhuber, B. Hermann, *et al.*, "Full-field time-encoded frequency-domain optical coherence tomography," Opt. Express **14**(17), 7661 (2006).
16. T. Bonin, G. Franke, M. Hagen-Eggert, *et al.*, "In vivo Fourier-domain full-field OCT of the human retina with 15 million A-lines/s," Opt. Lett. **35**(20), 3432 (2010).
17. E. Auksorius, D. Borycki, P. Stremplewski, *et al.*, "In vivo imaging of the human cornea with high-speed and high-resolution Fourier-domain full-field optical coherence tomography," Biomed. Opt. Express **11**(5), 2849 (2020).
18. V. Mazlin, P. Xiao, J. Scholler, *et al.*, "Real-time non-contact cellular imaging and angiography of human cornea and limbus with common-path full-field/SD OCT," Nat. Commun. **11**(1), 1868 (2020).
19. J. Scholler, V. Mazlin, O. Thouvenin, *et al.*, "Probing dynamic processes in the eye at multiple spatial and temporal scales with multimodal full field OCT," Biomed. Opt. Express **10**(2), 731 (2019).
20. C. Pfäffle, H. Spahr, L. Kutzner, *et al.*, "Simultaneous functional imaging of neuronal and photoreceptor layers in living human retina," Opt. Lett. **44**(23), 5671 (2019).
21. D. Hillmann, H. Spahr, C. Pfäffle, *et al.*, "In vivo optical imaging of physiological responses to photostimulation in human photoreceptors," Proc. Natl. Acad. Sci. U.S.A. **113**(46), 13138–13143 (2016).
22. V. Mazlin, P. Xiao, E. Dalimier, *et al.*, "In vivo high resolution human corneal imaging using full-field optical coherence tomography," Biomed. Opt. Express **9**(2), 557–568 (2018).
23. L. Puyo, H. Spahr, C. Pfäffle, *et al.*, "Retinal blood flow imaging with combined full-field swept-source optical coherence tomography and laser Doppler holography," Opt. Lett. **47**(5), 1198 (2022).

Biomedical Optics EXPRESS

24. P. Xiao, V. Mazlin, K. Grieve, *et al.*, "In vivo high-resolution human retinal imaging with wavefront-correctionless full-field OCT," Optica **5**(4), 409 (2018).
25. E. Auksorius, D. Borycki, and M. Wojtkowski, "Crosstalk-free volumetric in vivo imaging of a human retina with Fourier-domain full-field optical coherence tomography," Biomed. Opt. Express **10**(12), 6390 (2019).
26. P. Mecê, J. Scholler, K. Groux, *et al.*, "High-resolution in-vivo human retinal imaging using full-field OCT with optical stabilization of axial motion," Biomed. Opt. Express **11**(1), 492 (2020).
27. V. Mazlin, K. Irsch, K. Irsch, *et al.*, "Curved-field optical coherence tomography: large-field imaging of human corneal cells and nerves," Optica **7**(8), 872–880 (2020).
28. J. Zhang, V. Mazlin, K. Fei, *et al.*, "Time-domain full-field optical coherence tomography (TD-FF-OCT) in ophthalmic imaging," Ther. Adv. Chronic Dis. **14**, 204062232311701 (2023).
29. V. Mazlin, K. Irsch, V. Borderie, *et al.*, "Compact orthogonal view OCT: clinical exploration of human trabecular meshwork and cornea at cell resolution," (2022).
30. V. Mazlin, P. Xiao, K. Irsch, *et al.*, "Optical phase modulation by natural eye movements: application to time-domain FF-OCT image retrieval," Biomed. Opt. Express **13**(2), 902 (2022).
31. C. Pfäffle, H. Spahr, D. Hillmann, *et al.*, "Reduction of frame rate in full-field swept-source optical coherence tomography by numerical motion correction [Invited]," Biomed. Opt. Express **8**(3), 1499 (2017).
32. H. Sudkamp, P. Koch, H. Spahr, *et al.*, "In-vivo retinal imaging with off-axis full-field time-domain optical coherence tomography," Opt. Lett. **41**(21), 4987 (2016).
33. E. M. Seromenho, A. Marmin, S. Facca, *et al.*, "Single-shot off-axis full-field optical coherence tomography," Appl. Phys. Lett. **121**(11), 113702 (2022).
34. M. Žurauskas, R. R. Iyer, and S. A. Boppart, "Simultaneous 4-phase-shifted full-field optical coherence microscopy," Biomed. Opt. Express **12**(2), 981 (2021).
35. C. Dunsby, Y. Gu, and P. M. W. French, "Single-shot phase-stepped wide-field coherence-gated imaging," Opt. Express **11**(2), 105 (2003).
36. H. M. Subhash, "Full-Field and Single-Shot Full-Field Optical Coherence Tomography: A Novel Technique for Biomedical Imaging Applications," Adv. Opt. Technol. **2012**, 1–26 (2012).
37. M. S. Hrebesh, R. Dabu, and M. Sato, "In vivo imaging of dynamic biological specimen by real-time single-shot full-field optical coherence tomography," Opt. Commun. **282**(4), 674–683 (2009).
38. W. Nugroho, Y. Ito, M. S. Hrebesh, *et al.*, "Basic characteristics of interference image obtained using spatially phase-modulated mirror array," Opt. Rev. **18**(2), 247–252 (2011).
39. D. Sacchet, M. Brzezinski, J. Moreau, *et al.*, "Motion artifact suppression in full-field optical coherence tomography," Appl. Opt. **49**(9), 1480 (2010).
40. R. R. Iyer, M. Žurauskas, Q. Cui, *et al.*, "Full-field spectral-domain optical interferometry for snapshot three-dimensional microscopy," Biomed. Opt. Express **11**(10), 5903 (2020).
41. Y. Watanabe and M. Sato, "Quasi-single shot axial-lateral parallel time domain optical coherence tomography with Hilbert transformation," Opt. Express **16**(2), 524 (2008).
42. J. Moon, Y.-S. Lim, S. Yoon, *et al.*, "Single-shot multi-depth full-field optical coherence tomography using spatial frequency division multiplexing," Opt. Express **29**(5), 7060 (2021).
43. "Find directional gradients of 2-D image - MATLAB imgradientxy," https://www.mathworks.com/help/images/ref/imgradientxy.html.
44. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," (2015).
45. P. Isola, J.-Y. Zhu, T. Zhou, *et al.*, "Image-to-Image Translation with Conditional Adversarial Networks," (2016).
46. J. Howard and S. Gugger, "Fastai: A Layered API for Deep Learning," Information **11**(2), 108 (2020).
47. "pix2pix: Image-to-image translation with a conditional GAN | TensorFlow Core," https://www.tensorflow.org/tutorials/generative/pix2pix.
48. H. Zhang, I. Goodfellow, D. Metaxas, *et al.*, "Self-Attention Generative Adversarial Networks," (2018).
49. Y. Sugawara, S. Shiota, and H. Kiya, "Super-Resolution using Convolutional Neural Networks without Any Checkerboard Artifacts," (2018).
50. V. Mazlin, O. Thouvenin, S. Alhaddad, *et al.*, "Label free optical transmission tomography for biosystems: intracellular structures and dynamics," Biomed. Opt. Express **13**(8), 4190 (2022).
51. S. Alhaddad, O. Thouvenin, M. Boccara, *et al.*, "Comparative analysis of full-field OCT and optical transmission tomography," Biomed. Opt. Express **14**(9), 4845 (2023).
52. P. Mecê, K. Groux, J. Scholler, *et al.*, "Coherence gate shaping for wide field high-resolution in vivo retinal imaging with full-field OCT," Biomed. Opt. Express **11**(9), 4928 (2020).
53. C. for D., and R. Health, "Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices," FDA (2022).
54. "MR image reconstruction with AIR^{TM} Recon DL," https://www.gehealthcare.com/products/magnetic-resonance-imaging/air-technology/air-recon-dl.
55. "IDx-DR (EU)," https://www.digitaldiagnostics.com/products/eye-disease/idx-dr-eu/.
56. V. Mazlin,"Image-to-image translation for large images," Github, 2023, https://github.com/vmazlin/i2i.