ARTICLE

Open Access

Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data

Yijie Zhang^{1,2,3}, Tairan Liu^{1,2,3}, Manmohan Singh⁴, Ege Çetintaş^{1,2,3}, Yilin Luo^{1,2,3}, Yair Rivenson^{1,2,3}, Kirill V. Larin^{4,5} and Aydogan Ozcan^{1,2,3,6}✉

Abstract

Optical coherence tomography (OCT) is a widely used non-invasive biomedical imaging modality that can rapidly provide volumetric images of samples. Here, we present a deep learning-based image reconstruction framework that can generate swept-source OCT (SS-OCT) images using undersampled spectral data, without any spatial aliasing artifacts. This neural network-based image reconstruction does not require any hardware changes to the optical setup and can be easily integrated with existing swept-source or spectral-domain OCT systems to reduce the amount of raw spectral data to be acquired. To show the efficacy of this framework, we trained and blindly tested a deep neural network using mouse embryo samples imaged by an SS-OCT system. Using 2-fold undersampled spectral data (i.e., 640 spectral points per A-line), the trained neural network can blindly reconstruct 512 A-lines in 0.59 ms using multiple graphics-processing units (GPUs), removing spatial aliasing artifacts due to spectral undersampling, also presenting a very good match to the images of the same samples, reconstructed using the full spectral OCT data (i.e., 1280 spectral points per A-line). We also successfully demonstrate that this framework can be further extended to process 3× undersampled spectral data per A-line, with some performance degradation in the reconstructed image quality compared to 2× spectral undersampling. Furthermore, an A-line-optimized undersampling method is presented by jointly optimizing the spectral sampling locations and the corresponding image reconstruction network, which improved the overall imaging performance using less spectral data points per A-line compared to 2× or 3× spectral undersampling results. This deep learning-enabled image reconstruction approach can be broadly used in various forms of spectral-domain OCT systems, helping to increase their imaging speed without sacrificing image resolution and signal-to-noise ratio.

Introduction

Optical coherence tomography (OCT) is a non-invasive imaging modality that can provide three-dimensional (3D) information of optical scattering properties of biological samples. The first generation of OCT systems were based on time-domain (TD) imaging¹, using mechanical

path-length scanning. However, the relatively slow data acquisition speed of the early TDOCT systems partially limited their applicability for in vivo imaging applications. The introduction of the Fourier Domain (FD) OCT techniques^{2,3} with higher sensitivity^{4,5} has contributed to a dramatic increase in imaging speed and quality⁶. Modern FDOCT systems can routinely achieve line rates of 50–400 kHz^{7–12} and there have been recent research efforts to further improve the speed of A-scans to tens of MHz^{13,14}. Some of these advances employed hardware modifications to the optical set-up to improve OCT imaging speed and quality, and focused on, e.g., improving

Correspondence: Aydogan Ozcan (ozcan@ucla.edu)

¹Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA

²Department of Bioengineering, University of California, Los Angeles, CA 90095, USA

Full list of author information is available at the end of the article

These authors contributed equally: Yijie Zhang, Tairan Liu.

© The Author(s) 2021

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

the OCT system design, including improvements in high-speed sources^{13,15,16}, also opening up new applications such as single-shot elastography¹⁷ and others^{18–20}.

Recently, we have experienced the emergence of deep-learning-based image reconstruction and enhancement methods^{21–23} to advance optical microscopy techniques, performing e.g., image super resolution^{23–28}, autofocus^{29–31}, depth of field enhancement^{32–34}, holographic image reconstruction, and phase recovery^{35–38}, among many others^{39–42}. Inspired by these applications of deep learning and neural networks in optical microscopy, here we demonstrate the use of deep learning to reconstruct swept-source OCT (SS-OCT) images using under-sampled spectral data points. Without the need to perform any hardware modifications to an existing SS-OCT system, we show that a trained neural network can rapidly process undersampled spectral data and match, at its output, the image quality of standard SS-OCT reconstructions of the same samples that used 2-fold more spectral data per A-line.

A major challenge in reducing the number of spectral data points in an OCT system without sacrificing resolution is the aliasing artifacts introduced by under-sampling. According to the Nyquist sampling theorem, the maximum axial depth within the tissue that can be imaged without spatial aliasing is proportional to⁴³:

$$z_{\max} \propto \left| \frac{\pi}{2 \cdot \delta_s k} \right| = \left| \frac{\lambda_0^2}{4 \cdot \delta_s \lambda} \right| \quad (1)$$

where $\delta_s k$ is the spectral sampling interval in k space, $\delta_s \lambda$ is the wavelength sampling interval, and λ_0 is the central wavelength. When the spectral sampling interval increases, it reduces the maximum depth that can be imaged without spatial aliasing artifacts. In our approach, we first reconstructed each A-line with 2× less spectral data (eliminating every other spectral sample), which resulted in severe spatial aliasing artifacts. We then trained a deep neural network to remove these aliasing artifacts that are introduced by spectral undersampling, matching the image reconstruction results that used all the available spectral data points. To demonstrate the success of this deep learning-based OCT image reconstruction approach, we used an SS-OCT³ system to image murine embryo samples. The trained neural network successfully generalized, and removed the spatial aliasing artifacts in the reconstructed images of new embryo samples that were never seen by the network before. We further extended this framework to process 3× under-sampled spectral data per A-line, and showed that it can be used to remove even more severe aliasing artifacts that are introduced by 3× spectral undersampling, although at the cost of some degradation in the reconstructed image quality compared to 2× spectral undersampling results.

As an alternative approach, we also introduced an A-line-optimized spectral sampling framework to further reduce the acquired spectral data per A-line. The spectral sampling locations and the corresponding OCT image reconstruction network were jointly optimized during the training process, allowing this method to use less spectral data, while achieving better image reconstruction performance compared to 2× or 3× spectral undersampling results.

In addition to overcoming spectral undersampling related image artifacts, the inference time of the deep neural network is also optimized, achieving an average image reconstruction time of 6.73 ms for 512 A-lines, processed all in parallel using a desktop computer; this inference time is further improved to 0.59 ms by simplifying the neural network architecture and using multiple GPUs.

We believe that this deep learning-based OCT image reconstruction method has the potential to be integrated with various swept-source or spectral-domain OCT systems, and can potentially improve the 3D imaging speed without a sacrifice in resolution or signal-to-noise of the reconstructed images.

Results

To demonstrate the efficacy of this deep learning-based OCT image reconstruction framework, which we term DL-OCT, we trained and tested a deep neural network (see “Materials and methods” section) using SS-OCT images acquired on mouse embryo samples. Our 3D image data set consisted of eight different embryo samples, where five of them were used for training and the other three were used for blind testing. For each one of these embryo samples, 1000 B-scans (where each B-scan consists of 5000 A-lines, and each A-line has 1280 spectral data points) were collected by the SS-OCT system shown in Fig. 1a; see “Materials and methods” section for more details. During the network training phase, the original OCT fringes per A-line were first reconstructed using a Fourier transform-based image reconstruction algorithm to form the network’s target (i.e., ground truth) images. Then, the same spectral fringes were 2× down-sampled (by eliminating every other spectral data point), zero interpolated, and reconstructed using the same Fourier transform-based image reconstruction algorithm to form the input images of the network, each of which showed severe aliasing artifacts due to the spectral undersampling (Figs. 1 and 2). Both the real and imaginary parts of these aliased OCT images were used as the network input, where only the amplitude channel of the ground truth was used for the target image during the training phase. After the network training process, which is a one-time effort, taking e.g., ~18 h using a desktop

Fig. 1 Schematic of the DL-OCT image reconstruction framework. **a** Training phase of DL-OCT. Raw OCT fringes were captured by an SS-OCT system. The network target (ground truth) was generated by direct reconstruction of the original OCT fringes as detailed in the “Materials and methods” section. The network input was generated by a 2-fold down-sampling of the spectral data for each A-line, zero interpolation, and reconstruction of the resulting fringes. **b** Testing phase of the DL-OCT. We passed the 2x undersampled OCT image (real and imaginary parts) through a trained network model to create an aliasing-free OCT image, matching the ground-truth reconstruction that used all the spectral data points (see “Materials and methods” section for details).

computer (see “Materials and methods” section), the trained neural network successfully generalized and could reconstruct the images of unknown, new samples that were never seen by the network before, removing the aliasing related artifacts as shown in Fig. 1. Figure 2 further reports a detailed comparison of the network’s input, output, and ground truth images corresponding to different fields of view of mouse embryos, also quantifying the absolute values of the spatial errors made.

The reconstruction results reported in Figs. 1 and 2 clearly reveal that the trained network does not simply keep the connected upper part of the input image as the output. For example, in Fig. 2g, the signal in the ground truth image crosses both the upper and the lower parts of the field-of-view, and in the red circled region, there is an abrupt change, breaking the horizontal connectivity of the image. The DL-OCT network learned to reconstruct the output images by utilizing a combination of the vertical

Fig. 2 Blind testing performance of the DL-OCT framework. The network input, output, and ground-truth images of three mouse embryo samples (never seen by the network before) at different fields-of-view are shown in the first three columns. The error maps of the input and output with respect to the corresponding ground-truth image are provided in the 4th and 5th columns, respectively. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) values were also computed and displayed for each one of these sample fields-of-view.

morphological information exhibited in the target images and the special corrugated patterns caused by aliasing. In an OCT system, the illumination beam naturally forms an axially decaying pattern, where the surfaces or structural discontinuities usually have a stronger signal than the internal structure of the sample⁴³. This characteristic information was effectively captured by the neural network inference, as shown in for example Fig. 2g. This also explains the occasional weak artifacts observed at the network output (see e.g., the yellow circled region in Fig. 2g) for features that lack detectable morphological information along the vertical axis. In general, the trained neural network uses both the vertical and horizontal information at the input image (within its receptive field) to remove various challenging forms of aliasing artifacts such as those emphasized with red color in Fig. 2d.

Next, to quantify the performance of DL-OCT image reconstructions, two quantitative metrics were calculated for 13,131 different test image patches: peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) (see “Materials and methods” section for details). PSNR is a non-normalized metric that represents an estimation of the human perception of the image reconstruction quality. For images with pixels ranging from 0 to 1 with double-precision (such as the test images in our framework), a 20–30 dB PSNR value is generally acceptable for noisy target images⁴⁴. The SSIM, on the other hand, is a normalized metric that focuses more on image structure similarity between two images. This metric can take a value between 0 and 1 (where 1 represents an image that is identical to the target)⁴⁴. Overall, compared to the target (ground truth) images that used all the spectral data points, the spectrally undersampled input images with aliasing artifacts achieved a PSNR and an SSIM of 18.3320 dB and 0.2279, respectively, averaged over 13,131 test image patches. Both of these metrics were significantly improved at the network’s output images, achieving 24.6580 dB and 0.4391, respectively, also averaged over 13,131 test image patches. Some examples of these image comparisons with the resulting PSNR and SSIM values are also reported in Fig. 2.

To further test the robustness of the DL-OCT approach, it was also tested on other types of samples (i.e., human finger, human nail, human palm, human wrist, the limbus of human eye, anterior chamber of the human eye, and mouse eye). In total, 7 different samples for each type of tissue were imaged (except for mouse eye, where only 4 samples were imaged) by another SS-OCT imaging system (see Supplementary Methods for details). A single image reconstruction network was trained with all these types of tissue, where one sample for each type was reserved for blind testing. During the testing phase, the network consistently achieved high-quality image reconstructions (Supplementary Fig. S4) and obtained an

average PSNR of 28.7683 dB and an SSIM of 0.7239 on all the testing image patches (see Supplementary Methods for details).

We also used spatial frequency analysis to further quantify our network inference results against the ground truth images. To perform this comparison, we converted the network input, output, and ground truth images into the spatial frequency domain by performing a 1D Fourier transform along the vertical axis (for each A-line). The results of this spatial frequency comparison for each A-line are shown in Fig. 3d–f, which further reveal the success of the network’s output inference, closely matching the spatial frequencies of the corresponding ground truth image. The quantitative comparison in Fig. 3g–i also demonstrates that the network output very well matches the ground truth images for both the low and high-frequency parts of a sample.

Discussion

In our results reported so far, we used zero interpolation to pre-process the $2\times$ undersampled spectral data per A-line, before generating the network’s input image with severe spatial aliasing. Alternatively, zero-padding is another method that can be used to pre-process the undersampled spectral data for each axial line. However, other spectral interpolation methods such as the nearest neighbor, linear, or cubic interpolation may result in various additional artifacts due to the non-smooth structure of each spectral line. We performed a comparison of these different interpolation methods used to pre-process the same undersampled spectral data, the results of which are summarized in Fig. 4; in these results, each DL-OCT network was separately trained using the same undersampled spectral data, pre-processed using a different interpolation method. Among these interpolation methods, cubic interpolation was found to generate the most severe spatial artifacts at the network output. Both zero padding and zero interpolation methods shown in Fig. 4 consistently resulted in successful image reconstructions at the network output, removing aliasing artifacts observed at the input images, providing a decent match to the ground truth. On the contrary, other interpolation methods, such as cubic interpolation, introduced additional artifacts at the network output image (see, e.g., the red circled region in Fig. 4c) due to the inconsistent interpolation of missing spectral data points at the input. To further quantify this comparison, we also calculated the SSIM and PSNR values between the network output images and the corresponding ground truth SS-OCT images for five different pre-processing methods (Table 1). This quantitative analysis reported in Table 1 reveals that the zero interpolation method (presented in the “Results” section) achieves the highest PSNR and SSIM values for reconstructing SS-OCT images using

Fig. 3 Frequency spectrum analysis of DL-OCT. **a–c** SS-OCT images of a sample field-of-view, forming the network input, network output and ground truth, respectively. **d–f** Log-scaled spatial frequency spectra of **a–c** represented in spectral-spatial domain using 1D Fourier transform along the A-line direction of each image. **g, h** Averaged intensity of the spectral profile over two specific spatial regions (① and ② shown in **f**). **i** is the same as in **g, h**, except that it is averaged over the entire spatial axis, shown in **d–f**.

a 2-fold undersampled spectrum per A-line. It is also worth noting that the zero interpolation and zero padding methods achieve very close quantitative results, and significantly outperform the other spectral interpolation methods, including cubic, linear and nearest-neighbor interpolation, as summarized in Table 1.

However, all these interpolation/padding methods require a similar amount of time to generate the network input images compared to reconstructing the conventional OCT

images without undersampling, which might partially limit the adaptability of DL-OCT to high-speed imaging applications. An alternative pre-processing method that requires approximately m -fold less reconstruction time for $m\times$ spectral undersampling is reported in Supplementary Information. This method squeezes the spectral data by m -fold compared to its original size after undersampling, and applies a Fast Fourier Transform (FFT) directly onto the squeezed spectral data. Then, through simple copy/flip and

Fig. 4 Comparison of different pre-processing methods for DL-OCT. **a** Raw SS-OCT spectral fringes and the corresponding reconstructed OCT image (ground truth), where f_c indicates the cut-off frequency of the spectral data. **b** 2-fold undersampled OCT fringes. **c** Undersampled OCT fringes that are pre-processed using different interpolation methods. Three separate neural networks were trained for each one of the pre-processing methods to generate the network outputs. PSNR and SSIM values are also displayed for each one of these fields-of-view.

concatenation processes, a network input that is equivalent to the zero interpolation method can be obtained (Supplementary Methods). Visual inspection and quantitative results also suggest that this method can achieve identical performance to the zero interpolation method (Supplementary Fig. S2 and Supplementary Table S1).

We also analyzed the inference speed of the trained DL-OCT network to reconstruct SS-OCT images with

undersampled spectral measurements. For a batch size of 128 B-Scans, where each B-scan consists of 512 A-lines (with 640 spectral data points per A-line), the neural network is able to output a new OCT image in ~ 6.73 ms per B-scan using a desktop computer (Fig. 5). This inference time can be further reduced with some simplifications made in the neural network architecture; for example, a reduction of the number of channels from

Table. 1 Comparison of PSNR and SSIM values between the network output images and the corresponding ground truth SS-OCT images for five different pre-processing methods (also see Fig. 4).

Processing method	PSNR (dB)		SSIM	
	Average	Standard deviation	Average	Standard deviation
Zero interpolation ("Results" section)	24.6580	1.3225	0.4391	0.0320
Zero-padding	24.3447	1.8167	0.4378	0.0338
Cubic interpolation	24.3610	1.2398	0.4018	0.0291
Linear interpolation	24.4303	1.3493	0.4169	0.0343
Nearest neighbor interpolation	24.4763	1.3144	0.4135	0.0288

The highest average PSNR and SSIM values are shown in bold.

48 to 16 at the first layer of the neural network (Fig. 6) helped us reduce the average inference time down to ~1.69 ms per B-scan (512 A-lines). Through visual inspection, one can see that the 16-channel network can reconstruct decent OCT images compared with the 48-channel network results (shown in Fig. 5). Quantitatively compared using 13,131 image patches, the average SSIM and PSNR values downgraded, due to the reduced number of channels, from 0.4391 to 0.4122 and from 24.6580 dB to 24.2523 dB, respectively. Furthermore, with additional parallelization through the use of a larger number of GPUs, the inference speed per B-scan can be further improved. For example, with the use of 8 NVIDIA Tesla A100 GPUs (Nvidia Corp., Santa Clara, CA, USA) in parallel, the inference time was further reduced to ~1.42 ms and ~0.59 ms per B-scan for 48-channel and 16-channel networks, respectively (shown in Fig. 5). This can be used to better serve various applications that demand rapid reconstruction of 3D samples.

Finally, we explored to see whether DL-OCT can be extended to use an even smaller number of spectral data points (N_{spec}) per A-line to perform an image reconstruction. First, we investigated the case for 3× undersampled spectral data per A-line. For this, we used the same neural network architecture as before, which was this time trained with input SS-OCT images that exhibited even more extensive spatial aliasing since for every spectral measurement data point that is kept, 2 neighboring wavelengths were dropped out, resulting in $N_{\text{spec}} = 427$ spectral data points contributing to an A-line, whereas the ground truth images of the same samples had 1280 spectral measurements per A-line. In addition to this, we implemented an A-line-optimized undersampling method, where the number of spectral data points per

A-line was further reduced to $N_{\text{spec}} = 407$ (see “Materials and methods” section). The image reconstruction results for the 3× undersampling method ($N_{\text{spec}} = 427$) and A-line-optimized undersampling method ($N_{\text{spec}} = 407$) are reported in Fig. 7, in comparison with the 2× undersampling method ($N_{\text{spec}} = 640$). This comparison in Fig. 7 reveals that, while DL-OCT can successfully process 3× undersampled spectral data with decent image reconstructions at its output, it also starts to exhibit some spatial artifacts in its inference when compared with the ground truth images of the same samples (see, e.g., the red marks in Fig. 7). Furthermore, we observe that the A-line-optimized undersampling method can visually achieve almost identical performance to the 2× undersampling results. A quantitative comparison of these three methods is reported in Table 2. It is worth mentioning that the A-line-optimized undersampling method achieved the best quantitative reconstruction performance among the three methods (Table 2) because this framework can learn and optimize both the A-line spectral undersampling grid and the OCT image reconstruction neural network, which makes it easier for this framework to better fit to the target data and imaging task.

In summary, we demonstrated the ability to rapidly reconstruct SS-OCT images using a deep neural network that is fed with undersampled spectral data. This DL-OCT framework, with its rapid and parallelizable inference capability, has the potential to speed up the image acquisition process for various SS-OCT systems without the need for any hardware modifications to the optical setup. Although the efficacy of this presented framework was demonstrated using an SS-OCT system, DL-OCT can also be used in various spectral-domain OCT systems that acquire spectral interferometry data for 3D imaging of samples.

Materials and methods

Data acquisition

All the animal handling and related procedures were approved by the Baylor College of Medicine (University of Houston, USA) Institutional Animal Care and Use Committee and adhered to its animal manipulation policies. The animal protocol for mouse embryo imaging was the University of Houston (UH) 16-026. The mouse eye imaging reported in the Supplementary Information was under animal protocol UH: PROTO202000028. All human skin and human eye samples in the Supplementary Information were obtained under IRB UT Health (University of Texas Health Science Center) HSC-MS-16-0383 and UH: STUDY00001723, respectively. Timed matings of CD-1 mice were set up overnight. The presence of a vaginal plug was considered 0.5 days post coitum (DPC). At 13.5 DPC, embryos ($N = 8$) were dissected out of the mother and immediately prepared for OCT imaging.

Fig. 5 DL-OCT inference time as a function of the B-Scan batch size for blind testing. **a** With increasing batch size, the average inference time per B-Scan (512 A-lines) rapidly decreases owing to the parallelizable nature of the neural network computation. The average inference time converged to ~6.73 ms per B-Scan for a batch size of 128. If the number of channels in the neural network's first layer is reduced from 48 down to 16, the average inference time further improved to ~1.69 ms per B-Scan. Our GPU memory size limited further reduction of the average inference time of DL-OCT. By using 8 NVIDIA Tesla A100 GPUs in parallel, the inference time was further reduced to ~1.42 ms and ~0.59 ms per B-scan for the 48-channel and 16-channel networks, respectively. All inference times were obtained by averaging 1000 independent runs, computed on a desktop computer (see "Materials and methods" section). **b** Sample fields-of-view are shown for network input, network output (using 48 channels vs. 16 channels in the first layer), and ground truth images. PSNR and SSIM values are also displayed for each one of these fields-of-view.

Special care was taken to ensure that the yolk sac was not damaged during dissection. The embryos were immersed in Dulbecco's Modified Eagle Media (DMEM) in a standard culture dish and imaged with the SS-OCT system (OCS1310V2, Thorlabs Inc., NJ, USA). The OCT system had a central wavelength of ~1300 nm, a sweep range of ~100 nm, and an incident power of ~12 mW. The axial and transverse resolutions of the system have been

characterized as ~12 μ m and ~10 μ m, respectively, in air. More details on the performance of the OCT system can be found in the previous work⁴⁵. In this work, a sample area of 12 mm $\times$ 12 mm $\times$ 6 mm (X, Y, Z) was imaged. Each raw A-scan consisted of 1280 spectral data points that were sampled linearly in the wavenumber domain by a k-clock on the OCT system. 3D imaging was performed by raster scanning the OCT beam across the sample with

a pair of galvanometer-mounted mirrors. Each B-scan consisted of 5000 A-scans, and each sample volume consisted of 1000 B-scans.

Image processing

After the data acquisition, the raw OCT fringes were processed using $2\times$ down-sampling (by eliminating every other spectral data point), followed by zero interpolation to generate the $2\times$ spectrally undersampled SS-OCT reconstruction (which is used as the network input). Reconstruction of the target SS-OCT image (ground truth) from the raw spectral data was performed using multiple steps. First, to decrease the effect of sharp transitions and spectral leakage, each raw A-scan was windowed with a Hanning window. Next, the filtered fringes were processed by an FFT to get complex OCT data. Then, the norm of the complex vector was converted to dB scale, and the complex conjugate was discarded. A background subtraction step was performed by subtracting the mean of all the A-scans in each OCT volume from each A-scan. The resulting B-scans (after the background subtraction and windowing) was utilized as the network training targets (ground truth).

For $2\times$ down-sampling of the measured spectral data points, the even elements of the acquired spectrum for each A-line were removed. For $3\times$ down-sampling results reported in Fig. 7, two successive spectral measurements were eliminated, in a repeating manner, for each spectral data point that was kept. Next, zeros were interpolated in the exact same positions, where the spectral data points were removed. Then, the mean of the zero interpolated spectral data was subtracted out before applying the FFT function. Both the real and imaginary parts of the down-sampled OCT complex data, resulting from the FFT, were kept as input data for the network. Each pair of input and

ground truth images were normalized such that they have zero mean and unit variance before they were fed into the DL-OCT network.

DL-OCT network architecture, training, and validation

For DL-OCT, we used a modified U-net architecture⁴⁶ as shown in Fig. 6. Following the processing of the down-sampled OCT reconstructions and regular OCT images (ground truth images, using all the spectral data points), the resulting volumetric images were partitioned into patches of 640×640 pixels, forming training image pairs (B-scans); all blank image pairs (without sample features) were removed from training. The training loss function was defined as:

$$l = L_1 \{z_{\text{label}}, G(x_{\text{input}})\} \quad (2)$$

where $G(\cdot)$ refers to the output of the neural network, z_{label} denotes the ground truth SS-OCT image without undersampling, and x_{input} represents the network input. The mean absolute error, L_1 norm, was used to regularize the output of the network and ensure its accuracy.

The modified version of the U-net architecture is shown in Fig. 6, which has five down-blocks followed by five up-blocks. Each one of the down-blocks consists of two convolution layers and their activation functions, which together double the number of channels. A max-pooling layer with a stride and kernel size of two is added after the two convolution layers to downsample the features. The up-blocks first upscale the output of the center layer using bilinear interpolation by a factor of two. And then two convolution layers and their activation functions, which decrease the number of channels by a factor of two, are added after the upscaling. Between each one of the up- and down-sampling blocks of the same level, a skip

(see figure on previous page)

Fig. 7 Comparison of DL-OCT blind testing results using 3× undersampled, 2× undersampled, and A-line-optimized input spectral data. Three mouse embryo samples (never seen by either of these DL-OCT networks) are imaged for blind testing. PSNR and SSIM values are also displayed for each one of these fields-of-view. Ground truth images used 1280 spectral data points per A-line, whereas 2× and 3× DL-OCT networks used $N_{\text{spec}} = 640$ and $N_{\text{spec}} = 427$ spectral data points per A-line, respectively. The A-line-optimized network used $N_{\text{spec}} = 407$ spectral data points per A-line; also see Supplementary Fig. S5.

Table. 2 Comparison of PSNR and SSIM values between the network output images and the corresponding ground truth SS-OCT images for three different undersampling methods using zero interpolation (also see Fig. 7).

Undersampling method	PSNR (dB)		SSIM	
	Average	Standard deviation	Average	Standard deviation
2× undersampling	24.6580	1.3225	0.4391	0.0320
3× undersampling	23.4012	1.2536	0.3080	0.0259
A-line-optimized undersampling	25.0618	1.2807	0.4638	0.0295

connection concatenates the output of the down-blocks with the up-sampled images, enabling the features to be directly passed at each level. After these down- and up-blocks, a convolution layer is used to reduce the number of channels to one, which corresponds to the reconstructed output image, approximating the ground truth OCT image.

Throughout the U-net structure, the convolution filter size is set to be 3×3; the output of these filters is followed by a Leaky ReLU (Rectified Linear Unit) activation function, defined as:

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \quad (3)$$

The learnable variables were updated using the adaptive moment estimation (Adam⁴⁷) optimizer with a learning rate of 10^{-4} . The batch size for the training was set to be 3.

Quantitative metrics

PSNR is defined as:

$$\text{PSNR} = 10 \times \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (4)$$

where MAX_I is the maximum possible pixel value of the ground truth image. MSE is the mean squared error between the two images being compared, which is defined as:

$$\text{MSE} = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} [\mathbf{I}(i, j) - \mathbf{K}(i, j)]^2 \quad (5)$$

where $\mathbf{I}$ is the target image, and $\mathbf{K}$ is the image that is compared with the target.

SSIM is defined as:

$$\text{SSIM}(a, b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{a,b} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)} \quad (6)$$

where μ_a and μ_b are the mean values of a and b , which represent the two images being compared, σ_a and σ_b are the standard deviations of a and b , $\sigma_{a,b}$ is the cross-covariance of a and b , respectively, and C_1 and C_2 are constants that are used to avoid division by zero. Note that both PSNR and SSIM metrics can be affected by background noise in an OCT image. Therefore, to compute these two metrics we used the network output and target (ground truth) images that are over the noise level (70 dB in our SS-OCT system) and then converted them into grayscale with a range from 0 to 1, using double precision.

A-line-optimized spectral undersampling method

The workflow of the A-line-optimized undersampling method is shown in Fig. 8. The 2× undersampling method was used as the baseline, and further optimization/learning was applied upon it to be able to use even less spectral data points for OCT image reconstruction. A continuous trainable vector was firstly generated, and it was binarized by thresholding (with a threshold of $T = 0.5$, shown by the red dashed line in Fig. 8) to form a binary grid. Then, this binary grid was applied to the regular 2× undersampling grid to generate the final optimized undersampling grid with a total number of spectral data points less than 640. After the optimized undersampling grid was obtained, the same pre-processing and U-net training protocol was adopted as in the regular 2× undersampling method. During the network training process, the continuous trainable vector

(for spectral sampling) and the variables of the U-net were jointly optimized by the backpropagated gradient of the training loss.

Implementation details

The network was implemented using Python version 3.6.0, with TensorFlow framework version 1.11.0. Network training was performed using a single NVIDIA GeForce RTX 2080Ti GPU (Nvidia Corp., Santa Clara, CA, USA) and testing was performed using a desktop computer with 4 GPUs (NVIDIA GeForce RTX 2080Ti). The data set used for our training contained ~20,000 image pairs (640 A-lines in each image), which was split into training and validation sets with a ratio of 9:1. The training process took about 18 h for 22 epochs. DL-OCT inference times as a function of the batch size are reported in Fig. 5.

Acknowledgements

The Ozcan Lab at UCLA acknowledges the support of NSF and HHMI. The Larin Lab at UH acknowledges the support of NIH (R01AA028406, R01HD096335, R01EB027099, and R01HL146745).

Author details

¹Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA. ²Department of Bioengineering, University of California, Los Angeles, CA 90095, USA. ³California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA. ⁴Department of Biomedical Engineering, University of Houston, Houston, TX 77204, USA. ⁵Department of Molecular Physiology and Biophysics, Baylor College of Medicine, University of Houston, Houston, TX 77204, USA. ⁶Department of

Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

Author contributions

Y.Z., T.L., E.Ç., Y.L., and Y.R. contributed to the algorithms and analysis. M.S. performed the OCT experiments. A.O., Y.Z., T.L., and M.S. prepared the manuscript; all the authors contributed to the manuscript editing. A.O. and K.L. supervised the research. A.O. initiated and came up with the presented concept.

Data availability

The deep-learning models reported in this work used standard libraries and scripts that are publicly available in TensorFlow. All the data and methods needed to evaluate the conclusions of this work are present in the main text. Additional data can be requested from the corresponding author (A.O.).

Conflict of interest

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41377-021-00594-7>.

Received: 20 March 2021 Revised: 2 July 2021 Accepted: 6 July 2021

Published online: 29 July 2021

References

- Huang, D. et al. Optical coherence tomography. *Science* **254**, 1178–1181 (1991).
- Fercher, A. F. et al. Measurement of intraocular distances by backscattering spectral interferometry. *Opt. Commun.* **117**, 43–48 (1995).
- Chinn, S. R., Swanson, E. A. & Fujimoto, J. G. Optical coherence tomography using a frequency-tunable optical source. *Opt. Lett.* **22**, 340–342 (1997).
- Choma, M. A. et al. Sensitivity advantage of swept source and Fourier domain optical coherence tomography. *Opt. Express* **11**, 2183–2189 (2003).

5. De Boer, J. F. et al. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Opt. Lett.* **28**, 2067–2069 (2003).
6. De Boer, J. F., Leitgeb, R. & Wojtkowski, M. Twenty-five years of optical coherence tomography: the paradigm shift in sensitivity and speed provided by Fourier domain OCT [Invited]. *Biomed. Opt. Express* **8**, 3248–3280 (2017).
7. Oh, W. Y. et al. Ultrahigh-speed optical frequency domain imaging and application to laser ablation monitoring. *Appl. Phys. Lett.* **88**, 103902 (2006).
8. Huber, R., Wojtkowski, M. & Fujimoto, J. G. Fourier Domain Mode Locking (FDML): a new laser operating regime and applications for optical coherence tomography. *Opt. Express* **14**, 3225–3237 (2006).
9. Huber, R., Adler, D. C. & Fujimoto, J. G. Buffered Fourier domain mode locking: unidirectional swept laser sources for optical coherence tomography imaging at 370,000 lines/s. *Opt. Lett.* **31**, 2975–2977 (2006).
10. Yun, S. H. et al. Comprehensive volumetric optical microscopy in vivo. *Nat. Med.* **12**, 1429–1433 (2006).
11. Adler, D. C. et al. Three-dimensional endomicroscopy using optical coherence tomography. *Nat. Photonics* **1**, 709–716 (2007).
12. Potsaid, B. et al. Ultrahigh speed spectral/Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second. *Opt. Express* **16**, 15149–15169 (2008).
13. Klein, T. & Huber, R. High-speed OCT light sources and systems [Invited]. *Biomed. Opt. Express* **8**, 828–859 (2017).
14. Wei, X. M. et al. 28 MHz swept source at 1.0 μm for ultrafast quantitative phase imaging. *Biomed. Opt. Express* **6**, 3855–3864 (2015).
15. Oh, W. Y. et al. 400 kHz repetition rate wavelength-swept laser and application to high-speed optical frequency domain imaging. *Optics Lett.* **35**, 2919–2921 (2010).
16. Tsai, T. H. et al. Ultrahigh speed endoscopic optical coherence tomography using micromotor imaging catheter and VCSEL technology. *Biomed. Opt. Express* **4**, 1119–1132 (2013).
17. Singh, M. et al. Phase-sensitive optical coherence elastography at 1.5 million A-lines per second. *Opt. Lett.* **40**, 2588–2591 (2015).
18. Wieser, W. et al. High definition live 3D-OCT in vivo: design and evaluation of a 4D OCT engine with 1 GVoxel/s. *Biomed. Opt. Express* **5**, 2963–2977 (2014).
19. Blatter, C. et al. Ultrahigh-speed non-invasive widefield angiography. *J. Biomed. Opt.* **17**, 070505 (2012).
20. Baumann, B. et al. Total retinal blood flow measurement with ultrahigh speed swept source/Fourier domain OCT. *Biomed. Opt. Express* **2**, 1539–1552 (2011).
21. de Haan, K. et al. Deep-learning-based image reconstruction and enhancement in optical microscopy. *Proc. IEEE* **108**, 30–50 (2020).
22. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
23. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
24. Wang, H. D. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
25. De Haan, K. et al. Resolution enhancement in scanning electron microscopy using deep learning. *Sci. Rep.* **9**, 12050 (2019).
26. Boyd, N. et al. DeepLoco: fast 3D localization microscopy using neural networks. Preprint at <https://www.biorxiv.org/content/10.1101/267096v1> (2018).
27. Ouyang, W. et al. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).
28. Nehme, E. et al. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018).
29. Luo, Y. L. et al. Single-shot autofocusing of microscopy images using deep learning. *ACS Photonics* **8**, 625–638 (2021).
30. Pinkard, H. et al. Deep learning for single-shot autofocus microscopy. *Optica* **6**, 794–797 (2019).
31. Pitkäaho, T., Manninen, A. & Naughton, T. J. Performance of autofocus capability of deep convolutional neural networks in digital holographic microscopy. In *Proceedings of Digital Holography and Three-Dimensional Imaging, Jeju Island, Korea, Optical Society of America, 2017, W2A.5* (2017).
32. Luo, Y. C. et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nat. Methods* **16**, 1323–1331 (2019).
33. Yang, X. L. et al. Deep learning-based virtual refocusing of images using an engineered point-spread function. *ACS Photonics* **8**, 2174–2182, <https://doi.org/10.1021/acsp Photonics.1c00660> (2021).
34. Huang, L. Z. et al. Recurrent neural network-based volumetric fluorescence microscopy. *Light Sci. Appl.* **10**, 62 (2021).
35. Rivenson, Y. et al. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).
36. Wu, Y. C. et al. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* **5**, 704–710 (2018).
37. Liu, T. R. et al. Deep learning-based color holographic microscopy. *J. Biophotonics* **12**, e201900107 (2019).
38. Liu, T. R. et al. Deep learning-based holographic polarization microscopy. *ACS Photonics* **7**, 3023–3034 (2020).
39. Nguyen, T. et al. Deep learning approach for Fourier ptychography microscopy. *Opt. Express* **26**, 26470–26484 (2018).
40. Helgadottir, S., Argun, A. & Volpe, G. Digital video microscopy enhanced by deep learning. *Optica* **6**, 506–513 (2019).
41. Nguyen, T. et al. Automatic phase aberration compensation for digital holographic microscopy based on deep learning background detection. *Opt. Express* **25**, 15043–15057 (2017).
42. Hershko, E. et al. Multicolor localization microscopy and point-spread-function engineering by deep learning. *Opt. Express* **27**, 6158–6183 (2019).
43. Drexler, W. & Fujimoto, J. G. *Optical Coherence Tomography: Technology and Applications* (Springer, Berlin, 2008).
44. Sara, U., Akter, M. & Uddin, M. S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *J. Comput. Commun.* **7**, 8–18 (2019).
45. Singh, M. et al. Applicability, usability, and limitations of murine embryonic imaging with optical coherence tomography and optical projection tomography. *Biomed. Opt. Express* **7**, 2295–2310 (2016).
46. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany: Springer, 2015* 234–241 (2015).
47. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).