



Surgical scene generation and adversarial networks for physics-based iOCT synthesis

MICHAEL SOMMERSPERGER,^{1,5,6} ALEJANDRO MARTIN-GOMEZ,^{1,2,5,7}  KRISTINA MACH,¹ PETER LOUIS GEHLBACH,³ M. ALI NASSERI,⁴ IULIAN IORDACHITA,² AND NASSIR NAVAB^{1,2}

¹Chair for Computer Aided Medical Procedures and Augmented Reality, Informatics Department, Technical University of Munich, Munich, Bayern, Germany

²Laboratory for Computational Sensing and Robotics, Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland, USA

³Wilmer Eye Institute Research, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

⁴Klinikum Rechts der Isar, Augenklinik, Munich, Bayern, Germany

⁵These authors contributed equally to this work.

⁶michael.sommersperger@tum.de

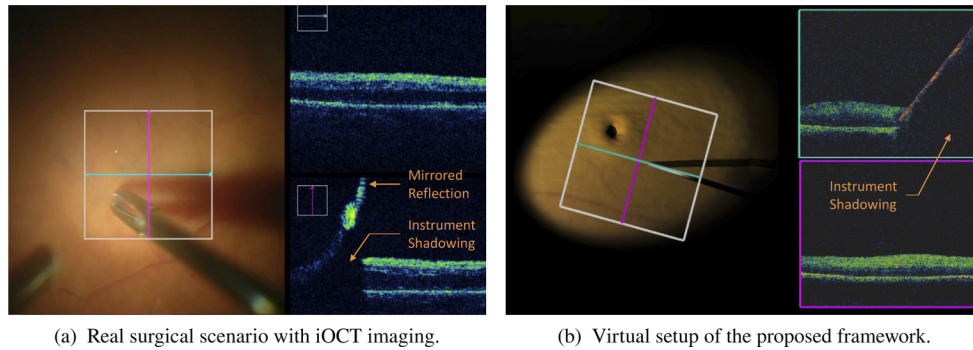
⁷alejandro.martin@jhu.edu

Abstract: The development and integration of intraoperative optical coherence tomography (iOCT) into modern operating rooms has motivated novel procedures directed at improving the outcome of ophthalmic surgeries. Although computer-assisted algorithms could further advance such interventions, the limited availability and accessibility of iOCT systems constrains the generation of dedicated data sets. This paper introduces a novel framework combining a virtual setup and deep learning algorithms to generate synthetic iOCT data in a simulated environment. The virtual setup reproduces the geometry of retinal layers extracted from real data and allows the integration of virtual microsurgical instrument models. Our scene rendering approach extracts information from the environment and considers iOCT typical imaging artifacts to generate cross-sectional label maps, which in turn are used to synthesize iOCT B-scans via a generative adversarial network. In our experiments we investigate the similarity between real and synthetic images, show the relevance of using the generated data for image-guided interventions and demonstrate the potential of 3D iOCT data synthesis.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Optical Coherence Tomography (OCT) is a medical imaging modality originally developed for noninvasive cross-sectional imaging in biological systems [1]. Since then, this imaging modality has become a widely accepted standard for diagnostic purposes in the ophthalmic domain because of its ability to visualize cross-sectional ocular structures at high resolution [2]. Initial time-domain OCT systems rely on mechanical adjustments to acquire single A-scans. Next generation technology utilized spectral-domain systems that measure all echos of a broad bandwidth light source with a spectrometer and a line scan camera without mechanical adjustments, leading to significantly higher acquisition speed [3]. These systems gradually enabled the integration of OCT into the operating room as they provide additional intraoperative visualization of retinal anatomy and surgical instruments to improve surgical outcomes, and support intraoperative diagnosis and decision making (see Fig. 1(a)). Currently, with the emergence of swept-source [4] and spiral scanning [5] technology, and the even higher acquisition speed, intraoperative OCT (iOCT) remains an advancing modality that promotes the creation of novel visualization techniques, the development of computer-assisted algorithms, and the integration of robotic systems to support challenging surgical interventions.



(a) Real surgical scenario with iOCT imaging.

(b) Virtual setup of the proposed framework.

Fig. 1. Comparison of fundus image and cross-sectional iOCT B-scans (a) during a real surgical scenario and (b) generated using our framework. In both images, the cyan and magenta lines superimposed on the microscopic image (left), indicate the position at which two B-scans (right) are acquired. Typical iOCT artifacts such as instrument shadowing and mirroring are highlighted using orange text and arrows.

Subretinal injection is a good example of a surgical procedure that may benefit from iOCT guidance. This procedure requires delivering therapeutic agents into the potential subretinal space by positioning the tip of a micro-surgical cannula at a target area of approximately $20\text{--}30\ \mu\text{m}$ [6] while avoiding irreversible damage to the neighboring tissue structures. In this context, iOCT provides cross-sectional tissue visualization at high resolution to achieve the required surgical precision. Furthermore, iOCT enables image-guided algorithms for the segmentation of retinal layers and instruments, which could facilitate targeted and reproducible interventions [7]. Although research on retinal layer analysis exists, prior work has focused on diagnostic OCT imaging [8–12] with lower noise levels and no surgical instruments or their related artifacts.

Significant differences between diagnostic and intraoperative OCT images are that iOCT often exhibits higher speckle-noise levels, lower contrast between the retinal layers, and contains surgical instruments that produce shadowing and reflection artifacts. Instrument shadowing artifacts result from the emitted OCT scanning laser not reaching the anatomical structures due to the high reflectivity of conventional surgical instruments. This shadow obscures relevant anatomical tissue and results in a loss of information, as depicted in the bottom B-scan of Fig. 1(a) and the upper B-scan of Fig. 1(b). Mirrored reflection artifacts are present in both the commonly employed spectral-domain and emerging swept-source iOCT systems. Although they enable fast scan acquisition, they introduce additional artifacts related to the processing of the detected signal in the Fourier domain [13]. These systems calculate axial scan information based on the echo time delay between a sample and a static reference arm. However, because of the symmetry in the magnitude of the Fourier transform, standard OCT systems cannot distinguish between positive and negative delays of signals returning from the sampled tissue relative to those returning from the reference arm [13]. Therefore, objects above the OCT scanning area appear mirrored in the B-scan images. This artifact is commonly observed in surgical scenarios when the instruments are located above the imaged B-scan area. This scenario is also illustrated in Fig. 1(a) displaying the mirrored instrument reflection in the lower B-scan. Such artifacts introduce additional complexity to iOCT imaging and complicate the transferability of algorithms designed for diagnostic OCT to the intraoperative context.

While state-of-the-art deep learning approaches are capable of segmenting retinal layers, these methods benefit from large databases of annotated B-scans [9]. In this regard, limited access to iOCT systems restricts the availability and generation of dedicated data sets. To overcome the limited availability of medical imaging, existing works have used machine learning approaches for synthesizing multiple modalities including X-Ray [14–16], Computed Tomography

[17,18], Magnetic Resonance [19–21], or Ultrasound [22]. Similar methods have been proposed in ophthalmology to generate microscopic images for vessel segmentation [23]. Although traditional augmentation methods can increase the variations of the data samples used for training learning-based algorithms, they do not always consider specific variations observed in the medical context [24]. These variations include changes in appearance, shape, or location of the observed anatomical structures and instruments. Furthermore, synthetic data sets can be highly effective in promoting the development of new algorithms. These advantages have been shown in applications involving X-ray imaging [25,26]. In the ophthalmic context, generating synthetic iOCT data along with the automatic generation of ground truth label maps would eliminate the time-consuming labeling process of real data and could enable the generation of dedicated iOCT data sets. These data sets could contain specific tool-tissue configurations challenging to generate in model eyes or animal experiments. Potential applications for this include the development of image-guided algorithms for instrument tracking [27–29] or pose estimation [30,31]. These tasks frequently require the acquisition of volumetric scans in controlled environments with known ground truth instrument poses and a known distance between the surgical instruments and retinal layers. Therefore, synthetic data could encourage further research on iOCT guided interventions despite the limitations in the availability or accessibility of iOCT systems.

This paper introduces a novel framework capable of generating synthetic iOCT B-scans and volumes based on real retinal layer meshes and virtual instruments in a virtual surgical environment. To illustrate the capabilities of the proposed framework, Fig. 1 depicts a comparison of an image captured during a real retinal surgery under iOCT guidance and our virtual environment integrating synthetic iOCT image generation. To our knowledge, this is the first work towards the generation of fully synthetic iOCT data sets. The fundamental components of the proposed framework used to generate the synthetic data are detailed in Section 2. These components include a virtual environment that replicates a surgical setup using real retinal layer meshes and virtual instrument models. In addition, a specifically designed scene rendering approach is combined with a generative adversarial network (GAN) to synthesize iOCT B-scans using the geometrical information extracted from the virtual environment. This network enables the generation of iOCT-typical instrument shadowing and mirroring artifacts explicitly. At the same time, the GAN models physical properties such as typical iOCT speckle noise and signal attenuation. As presented in Section 3, this particular configuration promotes the generation of data sets to support computer-aided algorithms for image-guided ophthalmic interventions. In this regard, the use of a virtual environment allows not only for the accurate reproduction of actual anatomy but also enables the integration of virtual models of microsurgical instruments that are not present in diagnostic OCT. Virtual environments also facilitate the generation of B-scans from multiple viewpoints and allow for the modification of virtual objects to generate dedicated data sets along with the corresponding ground truth label maps from a controlled environment. The capabilities of this framework and its potential to recreate surgical scenarios a virtual scene are discussed in Sections 4. and 5.

2. Methods

This section details the three main stages of the proposed framework used to generate synthetic iOCT B-scans from virtual scenarios (Fig. 2). First, a data preparation stage creates surface meshes of selected retinal layers segmented from real OCT volumes. In the examples presented in this work, the selected retinal layers correspond to the Inner Limiting Membrane (ILM) and the Retinal Pigment Epithelium (RPE). A *virtual setup* uses these surface meshes to reproduce the anatomy and allows for the introduction of virtual instruments into simulated surgical scenarios. Finally, the *iOCT synthesis* stage uses the geometric information from the scene, including retinal layers and instruments, to produce synthetic iOCT B-scans.

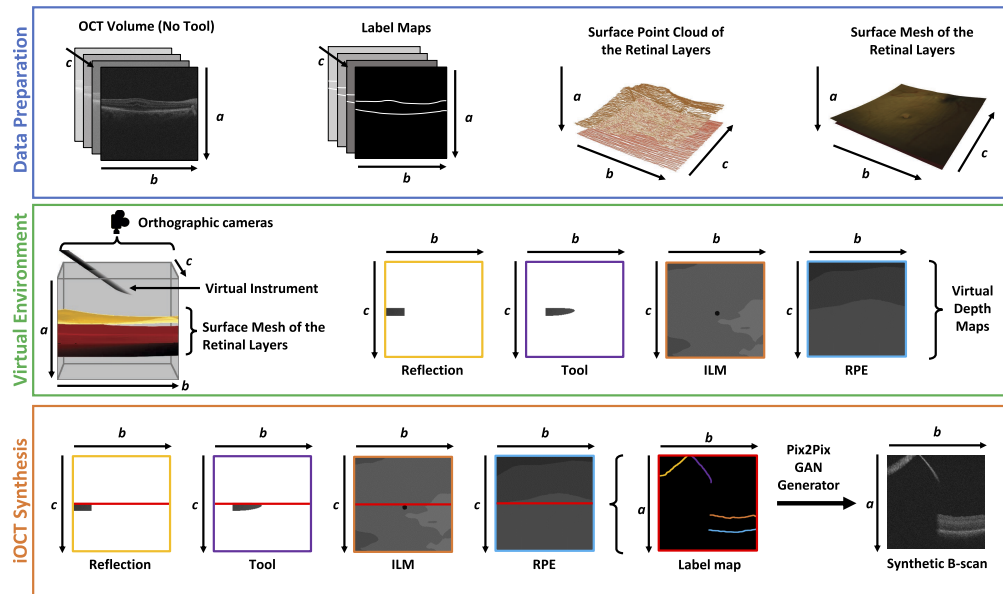


Fig. 2. Our proposed framework uses a *data preparation* stage to extract retinal point clouds from real OCT data. Virtual mesh representations of the layers and surgical tools are combined in a *virtual environment*. The geometric information from the virtual environment is utilized by a generative network to *synthesize iOCT B-scans*.

2.1. Data preparation

The first step toward generating the virtual setup requires the segmentation, and labeling, of L selected retinal layers from the cross-sectional B-scans of a real OCT volume V_{real} (see Fig. 2). The resulting label maps enable the extraction of three-dimensional point clouds using the first occurrence of the retinal layers in the label maps along the OCT A-scan direction. These three-dimensional point clouds enable generating a set of L surface meshes, which are later embedded in a virtual environment. To replicate the enface view during real surgical scenarios (Fig. 1(a)), the retinal surface mesh was colored by manually co-registering the real retinal fundus image with the iOCT volume V_{real} . The result of this process can be seen in Fig. 1(b) and the data preparation box of Fig. 2. This colorization serves only the visual appearance of the virtual environment and does not affect the synthesis of the iOCT B-scans.

2.2. Virtual setup

In a second stage, the generated surface meshes are integrated into a virtual environment respecting their original spatial relationship, generating a virtual reproduction of the retinal anatomy. This property also permits the integration of virtual surgical instruments to simulate interventional scenarios. The following sections describe how a specifically designed scene rendering approach allows generating synthetic data from a sub-area of the virtual environment. The rendering facilitates the generation of cross-sectional label maps, which capture the position of anatomical structures and instruments and facilitate the synthesis of iOCT B-scans.

For this purpose, a set of orthographic cameras is positioned in the virtual environment to simulate the iOCT imaging area. The primary role of each camera is to render a unique *depth map* with dimensions $b \times c$ for a dedicated object in the scene. Each virtual camera is adjusted to capture one specific object to avoid occlusion between the retinal layers, surgical tools and the surface mesh. This additionally facilitates the simultaneous generation of the depth maps for all

objects in the scene. Our setup uses a total of $L + 2$ virtual cameras to generate one depth map per retinal layer, plus two additional depth maps. The first additional depth map captures the surgical instruments, while the second captures the area above the simulated scanning area. This second map is later used to generate the mirroring artifacts typically associated with Fourier-domain OCT systems. The resolution of the depth maps $b \times c$ corresponds to the number of A-scans and B-scans of the synthetic iOCT volume and can be adjusted to match the resolution of existing devices. The images generated by the cameras depict the depth of the objects' surface represented as the ratio between the near and far clipping planes of the orthogonal cameras. The distance between the near and the far clipping planes thus correspond to the imaging depth a of the synthetic iOCT B-scans. To preserve the spatial configuration of the retinal layers and instruments between the depth maps, all cameras, except the camera enabling the mirroring artifacts, are aligned in their position and viewing direction, and are configured to match the position of their far clipping plane distances. The remaining camera, enabling the mirroring artifacts, is aligned in its far plane with the other cameras. However, its clipping distance is twice as large as the others (i.e., it is configured with an imaging depth twice the B-scan depth). This approach allows capturing objects above the B-scan imaging area to create the mirrored reflection artifacts observed in the B-scans. In addition, to replicate the field-of-view of commercially available iOCT systems, the width and height of the orthogonal cameras' view can be regulated. Therefore, as depicted by the white squares in Fig. 1, the virtual cameras can be adjusted to comprise only a part of the large retinal meshes extracted from a wide field-of-view volume. The white rectangle in Fig. 1(b) indicates the simulated iOCT scanning area, which covers a sub-region of the retinal meshes in the virtual environment. The following section describes how this set of depth maps enables the generation of cross-sectional label maps to synthesize iOCT B-scans.

2.3. iOCT B-scan synthesis

The final stage of the framework converts the positional information of the layer and instrument surfaces, stored in the *depth maps*, to *cross-sectional label maps* using a ray-casting approach along the A-scans. The method to generate these maps considers OCT-typical shadowing and mirrored reflection artifacts annotated in Fig. 3. This framework stage produces a total of c *label maps* for every set of *depth maps*. The pixels of the *label maps* encode positional surface information of the retinal layers, the instrument, and the mirroring artifact using separate class labels as depicted in Fig. 3. The information obtained from the *depth maps* can be converted to a *label map* as follows:

$$l_{(a,b)}^c = \begin{cases} \textit{tool} & \text{if} & a = \left[a_{res} \cdot d_{(b,c)}^{\textit{Tool}} \right] \\ \textit{mirroring} & \text{if} & a = \left[a_{res} \cdot \left(1 - 2d_{(b,c)}^{\textit{Reflection}} \right) \right] \\ \textit{layer}_x & \text{if} & a = \left[a_{res} \cdot d_{(b,c)}^{\textit{Layer}_x} \right] \text{ and } d_{(b,c)}^{\textit{Tool}} = 1 \text{ and } d_{(b,c)}^{\textit{Reflection}} = 1 \\ \textit{void} & \text{otherwise} & \end{cases} \quad (1)$$

where: $l_{(a,b)}^c$ represents the label map l of the B-scan c at the pixel location (a, b) ; $d_{(b,c)}^X$ is the depth d of the object of interest $X \in \{\textit{tool}, \textit{mirroring}, \textit{layer}, \textit{void}\}$ in the pixel (b, c) encoded as a relative position between the camera's clipping planes with values $[0 \leq d \leq 1]$. Furthermore, a_{res} defines the pixel resolution in axial direction (i.e., the number of pixels along one A-scan). It is important to notice that, if an object of interest X is not present in the axial direction, its

respective depth $d_{(b,c)}^X$ is mapped to 1. Therefore, the class label of a certain $layer_x$ is only included in the label map if the $layer_x$ exists and neither the tool nor the mirroring artifact are present in the A-scan. To generate the iOCT typical mirroring artifacts, the object surfaces' depth, captured above the B-scan area by $d^{Reflection}$, is mirrored about the start of the effective OCT imaging area (see Equation 1). This calculation simulates the mirroring about the zero-delay in Fourier-domain OCT systems [13]. In addition, to create instrument shadowing, the retinal layer classes are only integrated if no instrument or mirroring reflection artifact is identified along the corresponding A-scan. Lastly, to generate the final *label map*, the identified classes are drawn with a thickness of 3 pixels below their calculated surface position to mark the thin boundaries of the retinal layers and surgical instruments. A thin segmentation line is chosen to precisely define the boundaries of anatomical layers and instrument surfaces. The top row of Fig. 4 shows examples of the generated label maps with distinctively colored class labels.

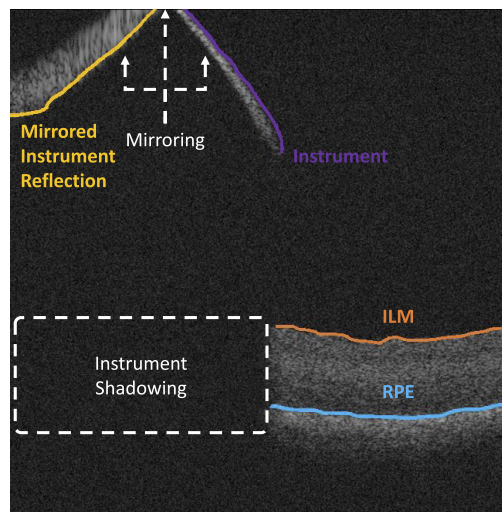


Fig. 3. Annotation of a real ex-vivo porcine B-scan including a surgical cannula, which creates shadowing and mirroring artifacts in the iOCT image. The retinal layer surface classes are marked in orange (ILM) and blue (RPE). The mirroring artifact is highlighted and subdivides the visible instrument.

To synthesize the iOCT B-scans from the segmentation labels, we use a GAN for an image-to-image translation inspired by the Pix2Pix framework [32]. This network has proven capable of synthesizing images from semantic label maps and photos from sketches. This property enables the transformation of label maps containing retinal layers and instruments to synthetic B-scans. Therefore, unlike in unpaired image-to-image translation, we can use image pairs providing direct correspondences between pixels in label maps and B-scans. We adjust the original Pix2Pix network structure to generate high-frequency details and realistic-looking speckle noise of iOCT B-scans. Additionally, contrary to the original Pix2Pix network, our adjusted network structure does not create checkerboard pattern artifacts in the synthetic B-scans (see Fig. 5). These patterns are generated by deconvolution operations, employed by the decoder of the U-net in the original Pix2Pix architecture [32] to construct the output image. The deconvolution layers allow the model to use every point in the low-resolution image to project and create a larger space by filling in the details. The original Pix2Pix architecture uses kernels of size four and a stride of two, which leads to uneven overlaps when these deconvolutions are stacked and produces artifacts on various scales. While the network theoretically could learn weights to avoid this, networks struggle to avoid it altogether, producing images with non-realistic-looking patterns (see Fig. 5).

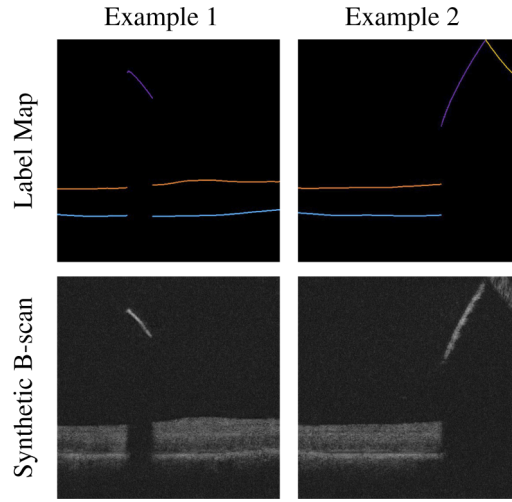


Fig. 4. Examples of paired label maps generated from the virtual setup and the corresponding synthetic iOCT B-scans. The label maps extracted from our virtual environment (top row) are used as the input to the GAN for the generation of synthetic iOCT B-scans (bottom row).

The speckle noise distribution of the OCT image signal intensities is modeled in our framework using a log-normal-like (Gamma) function. As numerous prior studies showed [33–35], the Gamma distribution provides an excellent fit to theoretical and experimental results that describe the speckle noise present in this imaging modality. In this context, when the original network fills up the main space during the deconvolutions, it is easy for it to output an average color. This property of the network leads to the necessity of creating an architecture and a training scheme to construct the generator output into a high-frequency image with natural-looking speckle noise. Thus, our framework adopted a U-net [36] architecture as the generator and a patch-based fully convolutional network as the discriminator. As a result of this modification, the original Pix2Pix is adjusted for the purpose of iOCT B-scan synthesis by, instead of using deconvolution blocks, employing a nearest-neighbor up-sampling followed by a padding layer, a standard convolutional layer, and a normalization layer Fig. 6. Additionally, we implemented a noisy-label training scheme for the discriminator. Furthermore, to still create some high-resolution textures, the last layer of the original model remains unchanged. The loss function to train our GAN network is defined as follows:

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G). \quad (2)$$

The generator is denoted as G while the discriminator is defined as D . The adversarial loss is described as:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_x[\log(D(x, G(x)))]. \quad (3)$$

The segmentation is noted as x , while the real OCT is noted as y , and the synthesized OCT is noted as $G(x)$. The weighted L1 loss between real and generated images is defined as:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y} [\|y - G(x)\|_1]. \quad (4)$$

Additionally, we use the same discriminator loss included in the original Pix2Pix GAN [32]. Using the L1 term as part of the loss-function can be helpful when objects appear surrounded by a contrasting background. In iOCT B-scans, retinal layers and tools appear bright and are surrounded by rather dark areas above and below the retina, making the L1 loss a suitable option. In addition, to mitigate the risk of blurring or smoothing the images when using the L1 loss, we

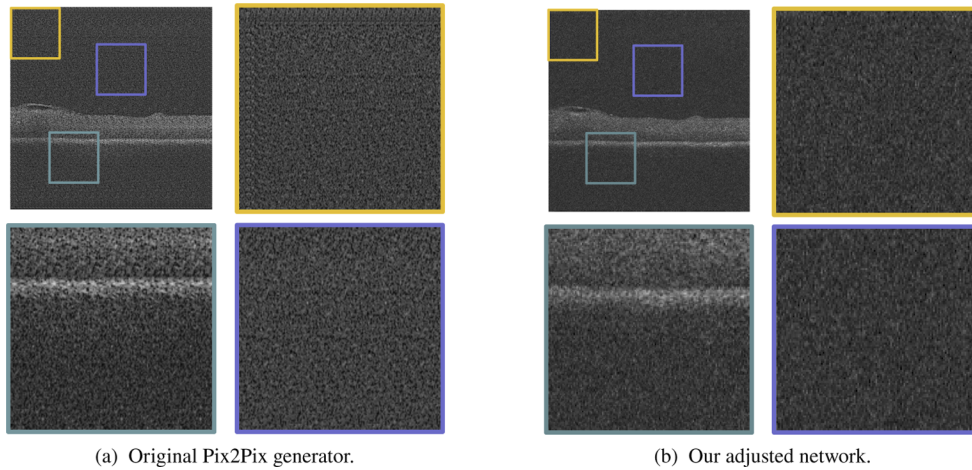


Fig. 5. Comparison of a synthetic B-scan generated from the original Pix2Pix network (a) and our adjusted network (b). Three enlarged patches highlight the artifact pattern created by the original Pix2Pix and justify the need for modifications of the network to generated more realistic looking iOCT B-scans.

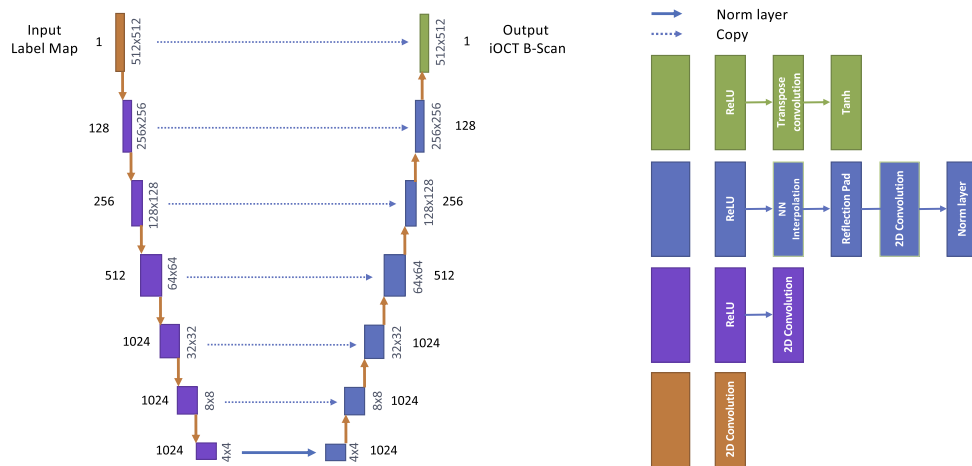


Fig. 6. Adjusted generator network structure of our GAN. We use a new, improved upsampling method including a nearest-neighbor interpolation, a padding layer, a 2D convolution and a normalization layer (marked as the blue components in the network).

introduced errors to the labels. Thus, some generated images, randomly selected, are marked as real, while some real ones are marked as generated. Such a noisy labeling scheme improves the discriminator to distinguish between high-frequency patterns resulting in the generator mimicking the natural-looking speckle noise [37]. To regularize the discriminator when training our GAN we apply this strategy to all images of every 15th batch.

In summary, the proposed generator considers the positional information of the retinal anatomy and the instruments as input images to synthesize iOCT B-scans. Figure 4 shows input *label maps* generated from the virtual environment and their corresponding synthetic iOCT B-scans. The generator learns to model the differences in appearance between the instrument attenuation and the reflection artifacts by specifically separating instrument surfaces and mirroring artifacts in different classes encoded in the generated label maps. The instrument shadowing artifacts, explicitly integrated into the *label maps*, are thus also considered by the network. The network implicitly models the OCT-typical physical phenomena such as signal attenuation and speckle-noise patterns, which occur during the image formation process in real systems. In the next sections we evaluate the similarity of the generated and synthetic iOCT B-scans.

3. Experiments

In this section, we evaluate the effective resolution of our synthetic iOCT and evaluate the quality of the generated images by estimating the similarity between real and synthetic iOCT B-scans. In addition, to demonstrate the relevance of our framework, we illustrate the impact of synthetic B-scans on the performance of retinal layer and instrument segmentation networks. Furthermore, we demonstrate the potential of generating volumetric representations by visually comparing renderings generated from real and synthetic volumes.

3.1. Experimental setup

All volumetric iOCT data used in our experiments was acquired from, evacuated within a time frame of two hours prior to the experiments, using a Zeiss Lumera 700 with integrated Rescan 700 iOCT system (Carl Zeiss AG). To decrease training and inference times, we down-sample the axial resolution of all acquired data by half resulting in a final B-scan resolution of 512×512 pixels. This action results in an axial resolution of $\sim 5 \mu\text{m}$, which is similar to the effective axial scanning resolution of the utilized OCT system and therefore did not decrease imaging accuracy. All iOCT volumes consist of 128 B-scans acquired at a scan depth of 2.8 mm. The volume V_{layer} was generated from an ex-vivo porcine eye and used to extract the retinal meshes and create the virtual scene with a wide field-of-view. Therefore, this volume does not contain surgical instruments and is scanned in an area of 6×6 mm. The data sets DS_{train} and DS_{test} consist of 87 and 9 iOCT volumes, respectively. These volumes, acquired from 6 additional ex-vivo porcine eyes using different locations, are acquired at a scanning area of 3×3 mm and used for training and testing the Pix2Pix network for the B-scan synthesis. For DS_{test} , we reduce the 9 volumes to 100 randomly selected B-scans, which we use to evaluate our methods. Both DS_{train} and DS_{test} include surgical instruments such as forceps or a 27G cannula in different positions within the volumes. The anterior surface of the ILM and RPE, the instrument surface, as well as the mirrored instrument reflection, were automatically pre-segmented using the method described in [7] and the resulting segmentations were manually corrected by an expert to generate the ground truth label maps.

In addition, for the experiments described in section 3.4, we generated a purely synthetic data set $DS_{\text{synthetic}}$ containing 2,982 B-scans. This data set is generated from 34 different tool positions simulated in our virtual environment. These configurations depict five orientations with the needle tip centered in the volume and aligned to the B-scan direction at 0° , 1° , 2° , 5° , and 90° , two translations of -500 and $-1000 \mu\text{m}$ and 0° in orientation, two translations of 500 and $1000 \mu\text{m}$ and 90° in orientation, and the 25 combinations of five translations (100 , 200 , 300 , 400 , and 500

μm) and five orientations (1, 2, 3, 4, and 5°). For every configuration, three volumes containing the needle above the ILM, in touch with the ILM, and between the ILM and RPE, are acquired. The position of the needle within the volume and the separation between the layer meshes are randomly adjusted during volume generation to ensure diversity in the acquired volumes.

For the implementation and training of the concepts described in Section 2, we use an Intel Core i9-9920X (@3.5GHz) processor and an NVidia GeForce RTX 2080 Ti. The virtual environment is created in Unity 3D (v2019.4.16f1). The virtual models used for the instruments correspond to a 41G cannula and a 23G illuminator frequently used in subretinal injection procedures. Due to the simplicity of their geometry, these models were designed using modeling software, respecting their corresponding dimensions. However, it is worth mentioning that more complex models of surgical tools can be integrated into our virtual setup. The virtual cameras render depth maps at a resolution of 512×512 pixels, and the resolution of the synthetic B-scans corresponds to the sampled resolution of the real B-scans.

3.2. B-scan synthesis evaluation

To generate synthetic iOCT B-scans from tool and layer label maps, we trained our GAN network on DS_{train} for 5 epochs with a batch size of 2, a label flip every 15th batch and a learning rate of 0.0002. The λ parameter in the L1 loss term was set to 100. Afterward, the ground truth segmentations of DS_{test} are used as an input to the GAN to produce synthetic B-scans. In Fig. 7 we show examples of the synthetic B-scans generated by the GAN from the ground truth label maps in DS_{test} and compare them to the corresponding real B-scans in our test set.

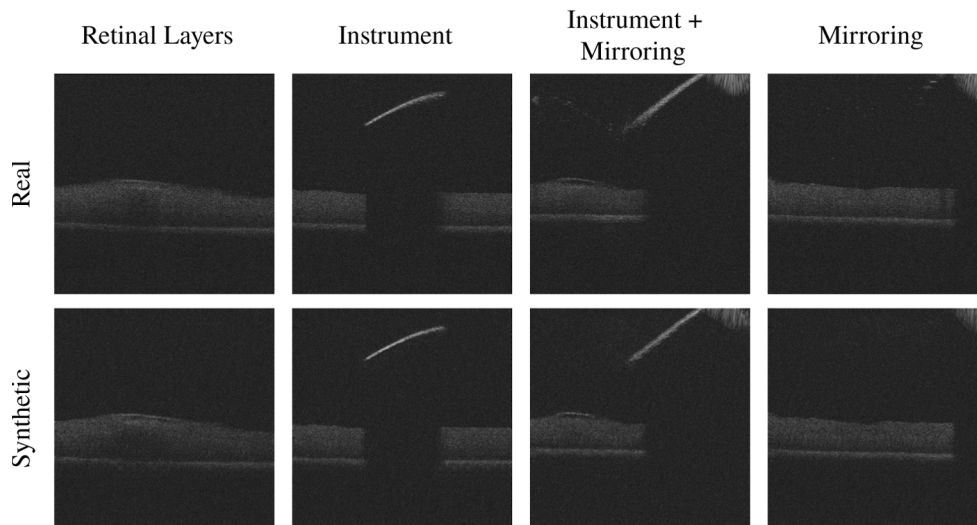


Fig. 7. Comparison of real (top) and corresponding synthetic (bottom) iOCT B-scans. From left to right, these B-scans depict the retinal anatomy and a microsurgical forceps with its associated instrument shadowing and mirrored reflection artifacts.

To evaluate the quality of the generated B-Scans, we use quantitative and qualitative measures. We compute the Structural Similarity Index Measure (SSIM) [38] and the Learned Perceptual Image Patch Similarity (LPIPS) [39] for all pairs of corresponding images in our test set. In this context, pairs are defined as each real B-scan in DS_{test} and the synthetic B-scan that has been created from its ground truth label map using the generator of our GAN. The SSIM score ranges $[0, 1]$ and equals 1 only if both images are identical. In contrast, lower LPIPS scores indicate more similarity between the images. Despite structural similarity scores suggesting that the compared images are distinctive (mean = 0.26, standard deviation = 0.01, min = 0.25,

max = 0.28), perceptual similarity indicates otherwise (mean = 0.13, standard deviation = 0.01, min = 0.11, max = 0.15). Further analysis of our results showed that real and synthetic images showed comparatively high structural similarity in proximity to the retinal anatomy and surgical instruments. However, differences in speckle-noise within the vitreous and the area below the retina containing weak or no image signal decreased the overall SSIM between synthetic and real images.

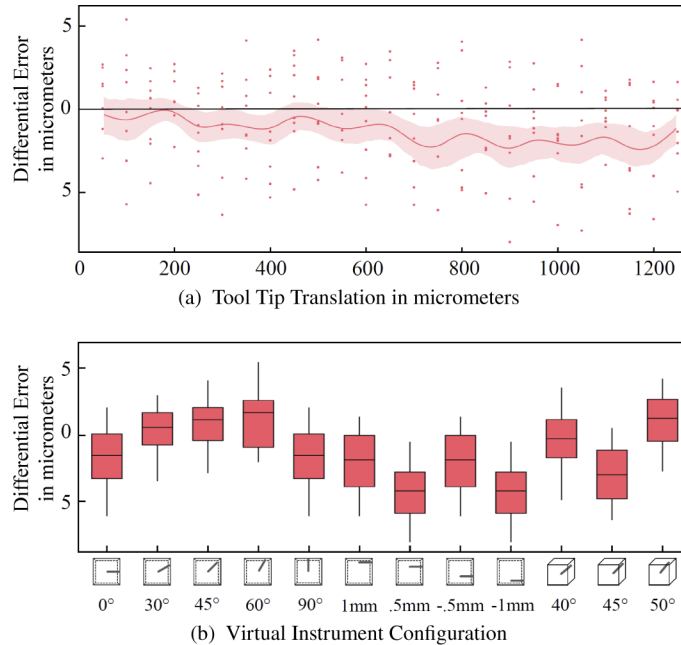


Fig. 8. (a) Differential error in the tooltip position estimation (the red line depicts a smoothed estimate with 95% bootstrapped confidence interval). Each point represents the calculated error between the ground truth location of the tip and its location in the synthetic iOCT volume. (b) Box plots for differential tooltip error for twelve different instrument configurations.

In addition, three clinicians were asked to compare pairs of real and generated B-scans using their own criteria and expertise. We used their comments regarding the similarity of real and synthetic iOCT B-scans as qualitative evaluation. This feedback indicates that, compared to the real B-scans, the synthetic images contain blurrier instrument reflections and layer boundaries, especially at inner retinal tissue. Furthermore, sharp and distinct instrument artifacts and small particles in the vitreous, present in our real data, were not created by the network. However, all the experts confirmed the complexity of distinguishing between real and synthetic B-scans, especially as the differences in the images are predominantly based on subtle structural details.

3.3. Imaging resolution analysis

To evaluate the B-scan resolution achieved with the proposed method, we simulate 12 different volume sequences containing an instrument in the virtual environment and measure the error while tracking the tool's tip throughout the captured sequences. These volumes depict five orientations with the tip of the instrument centered in the volume at 0, 30, 45, 60, and 90°; four offsets of +1000, +500, -500, and -1000 μm and 0° in orientation; and three incident angles of 40, 45 and 50° onto the retina. The virtual tool is placed 1000 μm above the virtual ILM for every configuration. A total of 26 volumes are acquired per configuration with needle motion

of 50 μm in the insertion direction after each acquisition. Obtaining the tool's tip position in the generated volumes allows for estimating the tool motion between volume captures. The differential error between the computed tool's tip motion and the ground truth step size of 50 μm is used to estimate the imaging accuracy of the synthetic iOCT. Results from this experiment show an overall detectable mean error of 1.31 μm and a standard deviation of 2.76 μm , which is comparable to or higher than that of real iOCT systems, such as the one used for acquiring our iOCT data. Figure 8(a) shows the error of the tip tracking of all sequences based on the traveled instrument distance, while Fig. 8(b) shows the errors of the individual volume sequences.

3.4. Relevance to image guided interventions

To offer an example of a relevant application to utilize the synthetic B-scans, we demonstrate that the generated data can support the development of data-driven algorithms. For this purpose, we train a UResNet18 for the joint segmentation of retinal layers and instruments, as proposed in [7]. All segmentation networks employed for these experiments were used without any kind of pre-training. Table 1 depicts the segmentation performance of the network when trained in three variants. First, we trained the network on a subset of 500 B-scans randomly selected from DS_{train} , which we call $DS_{train'}$. In another experiment, we train the segmentation only on the synthetic data in $DS_{synthetic}$. Finally, we use $DS_{synthetic}$ in a pre-training stage before training on an even smaller sub-set of $DS_{train'}$ containing only 80 B-scans, referred to as $DS_{train''}$. The per A-scan detection accuracy and positional error of the layer and tool surfaces of the three training variants were tested on DS_{test} . From Table 1, it can be seen that the network trained exclusively on the synthetic data in $DS_{synthetic}$ achieved good segmentation performance on the real test data in DS_{test} . Additionally, using the synthetic data to pre-train the network before training on a small amount of real data ($DS_{train''}$) drastically improved the segmentation performance compared to the training only on synthetic data. This configuration achieved similar results to training on a big dataset of real B-scans. For this experiment, all models were trained until convergence.

Table 1. Segmentation performance based on different training strategies.
 $DS_{train'}$: training exclusively on real images. $DS_{synthetic}$: training exclusively on synthetic B-scans. $DS_{synthetic} + DS_{train''}$: the synthetic images were used for pre-training, before training on a small set of real B-scans.

Class	$DS_{train'}$			$DS_{synthetic}$			$DS_{synthetic} + DS_{train''}$		
	ACC	Mean	SD	ACC	Mean	SD	ACC	Mean	SD
Tool	99%	0.35	0.16	81%	1.62	0.97	98%	0.35	0.45
ILM	99%	0.27	0.11	91%	3.32	3.88	99%	0.44	1.15
RPE	99%	0.28	0.1	95%	4.29	7.37	99%	0.33	0.11

3.5. Volume rendering

With the increasing acquisition speed of volumetric iOCT data [4,40] over the last years, three-dimensional intraoperative imaging has become a feasible and more popular research area. Demonstrating the potential of our framework to generate 3D data, we follow the volumetric scanning pattern of a commercially available iOCT system, which also has been used for data acquisition in our experiments, by stacking 128 linearly acquired B-scans to generate an iOCT volume and display our rendered results using a direct volume rendering (DVR) approach. To generate the synthetic volumes, B-scans were generated from segmentations of real volumes using our Pix2Pix network. The real volumes were randomly selected from the dataset DS_{test} . Pairs of real and synthetic iOCT volumes are shown in Fig. 9. These volumes show similar intensity values for the instruments in the corresponding pairs. The RPE layer, mainly colored in orange in Fig. 9, also shows similar intensity properties.

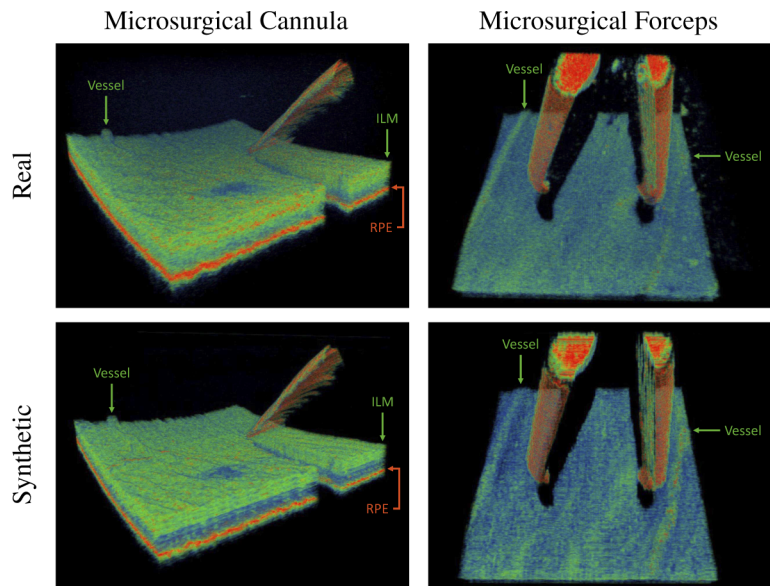


Fig. 9. Renderings of real (top row) and synthetic (bottom row) volumes containing microsurgical instruments such as a microsurgical cannula (left column) and forceps (right column). All visualizations are rendered with the same transfer function.

Furthermore, in Fig. 9, it can be seen that anatomical structures such as single superficial vessels can be inspected in both the real and the synthetic data. However, these structures have been recreated slightly more prominently in the synthetic volume. Additionally, we present visualizations of volumes, purely generated from our virtual scene, containing a virtual model of a 41G cannula and a 23G illuminator. Our illustrations in Fig. 10 show that, in the case of fully virtually generated iOCT volumes, the B-scans remain consistent in the generated intensity values of retinal layers, resulting in uniform colorization of anatomical structures across the synthetic volumes. In Fig. 10, even structures such as the macular hole can be perceived. In this case, the high curvature of the retinal surface results in higher intensities created by the network and are thus visually highlighted in the rendering.

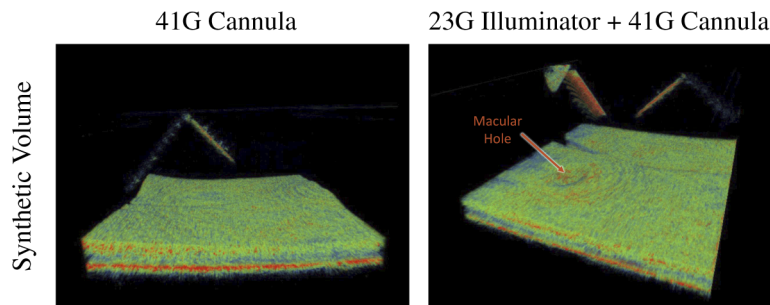


Fig. 10. Synthetic iOCT volumes generated with our framework. The left image shows a 41-gauge injection cannula, while the image on the right additionally contains a model of a 23G illuminator. All visualizations are rendered with the same transfer function.

4. Discussion

The method presented in this work facilitates the generation of entirely new iOCT B-scans using only the positional information of anatomical structures and instruments extracted from a virtual scene. Because the retinal meshes only define the layer positions, the pixel intensity values of the generated B-scans are not defined by a pre-existing OCT volume, but rather fully modeled by the GAN. Therefore, the GAN can be trained on iOCT volumes that show a specific signal quality or that originate from a specific OCT device, allowing for much higher flexibility in the quality and the properties of the generated images. In this context, the GAN facilitates the modeling of physical phenomena such as signal attenuation and speckle noise. However, using a GAN to create other dominant iOCT typical artifacts, such as instrument shadowing and mirroring, would require extremely large amounts of data or may not be possible. Therefore, these artifacts are explicitly modeled during the label map generation using the virtual environment. The experiments presented in Section 3.2 demonstrate the potential of the proposed framework for the generation of synthetic iOCT data. Qualitative feedback provided by medical experts highlighted the high similarity perceived from observing real and generated images. These qualitative results were supported by quantitative metrics used to estimate the similarity between images as perceived by humans. In this regard, LPIPS results revealed high similarity between the real and synthetic iOCT data sets. Interestingly, despite the promising results that provided evidence of high similarity between the images, the evaluation of the generated B-scans revealed low SSIM values. This behavior could result from the very sparse label maps that only outline the surface of the retinal layers and surgical instruments. In this context, as the network has no additional information, the difference between real and synthetic images is higher in areas further from the labeled structures. This premise could also explain why the network cannot recreate small structures in the vitreous, which were present in some of the real B-scans. Furthermore, previous studies have shown that many similarity metrics, including SSIM, can be confused by noise patterns [39]. As a result of this consideration, further analysis of the images showed that the SSIM was higher in patches close to or including retinal layer structures and surgical tools.

Furthermore, the experiments presented in Section 3.3 showed that the synthetic images generated with the proposed framework depict similar properties to those of commercially available iOCT systems. Furthermore, Section 3.4 showed that the generation of synthetic data could support the development of learning based algorithms. In this context, the generation of synthetic iOCT data sets can be useful for training these algorithms in the case of lacking real data, or to generate data sets with known ground truth or with desired tool-tissue configurations. Although our results support the idea of using this framework for the training of learning based algorithms, it is important to mention that the experiments conducted here used porcine eyes. Therefore, their transferability to human anatomy must be investigated in further studies.

We have also shown the potential of the proposed framework to generate synthetic volumetric iOCT data and to integrate virtual models of surgical tools. The renderings presented in Section 3.5 show that our method can generate synthetic volumes with visible anatomical features and realistic surgical instrument reflections, along with their associated artifacts such as mirroring and shadowing. These results could support the future investigation of algorithms to synthesize volumetric iOCT data from virtual scenes, in order to generate even more uniform intensities across the volume and shared features between neighboring B-scans.

Our method explicitly models typical iOCT instrument shadowing and mirrored reflection artifacts during the label map generation. Besides resulting in synthetic B-scans, which are more similar to real iOCT B-scans, expert feedback suggests that these artifacts also provide valuable cues to aid orientation in the surgical environment and navigate the instruments during surgery. Therefore, we consider it important to model such artifacts in our synthetic data.

Despite the design consideration of the proposed framework, the data synthesis could further benefit from modeling realistic tissue deformations during tool-tissue interactions. In addition,

due to the varying quality of the data acquired from ex-vivo porcine eyes, the results presented in this work only considered the two most distinctive retinal layers, which are also the most relevant for many interventions. However, incorporating additional anatomical information could further improve the quality of the generated images.

Although multiple studies have shown the anatomical similarity of human and porcine retinas [41–44] the image quality from ex-vivo animal experiments is lower compared to iOCT scans acquired during surgery on real patients. In this regard, we consider synthesizing iOCT data sets that more closely resemble human retinas as future work. However, collecting the necessary data under this conditions involves additional technical and ethical challenges that must be considered. These challenges include unforeseen patient motion (including motion due to breathing) as well as hand tremor of the surgeon that can lead to unwanted motion artifacts when acquiring iOCT volumes. The possibility of observing motion during the collection of the volume becomes particularly relevant when considering the update rate of commercially available iOCT systems, such as the one we used for our experiments (≈ 0.4 volumes/second). Furthermore, collecting data when the tool is positioned very close to the retina or interacting with the retina may pose a risk for the patient's health. It is important to notice that this work represents a first step towards the generation of synthetic iOCT data sets, for which our experiments have shown promising results.

In addition, we consider modeling pathological retinas in the synthesis of B-scans as part of our future work. This could be achieved for example by generating meshes from the segmentation of a pathology in real OCT and adding them to the virtual environment. The proposed network would then also take into account the position of the pathological anatomy. Incorporating more detailed information in our virtual scene would in turn allow us to focus more on the visualization of delicate anatomical structures, such as the macular hole illustrated in Fig. 10. These results motivate us to envision the application of the proposed framework not only to support the development of learning-based algorithms. In this context, exploring visualizations on synthesized iOCT B-scans could enable a fast adaptation of the technology for their use in interventional scenarios and could further support the ideation of new surgical procedures.

5. Conclusion

In this paper, we introduce a novel framework composed of a virtual scene for the generation of surgical environments and a GAN for synthesizing iOCT B-scans. The virtual scene enables the combination of retinal layer meshes extracted from real anatomy with models of surgical instruments. This scene allows for the generation of label maps that model typical iOCT-related imaging artifacts such as mirrored instrument reflection and instrument shadowing. A generator network then uses these maps to synthesize iOCT B-scans, recreating the properties of the iOCT image formation, such as the typical speckle-noise pattern and signal attenuation. Our experiments, including mathematical models and expert feedback, have shown that the real and synthetic images are perceptually similar. Furthermore, we showcase the usability of the synthetic data for the development of image-guided algorithms and demonstrate the potential for synthetic 3D iOCT data generation.

Funding. National Institutes of Health (1R01EB025883-01A1).

Acknowledgements. We also want to extend our appreciation to Prof. Dr. med. Mathias Maier and Dr. med. Daniel Zapp, Klinik und Poliklinik für Augenheilkunde, Klinikum rechts der Isar of the Technical University of Munich for all their valuable feedback during the development of this work.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science* **254**(5035), 1178–1181 (1991).
2. M. L. Gabriele, G. Wollstein, H. Ishikawa, L. Kagemann, J. Xu, L. S. Folio, and J. S. Schuman, "Optical coherence tomography: history, current status, and laboratory work," *Invest. Ophthalmol. Visual Sci.* **52**(5), 2425–2436 (2011).
3. A. G. Podoleanu, "Optical coherence tomography," *J. Microsc.* **247**, 209–219 (2012).
4. O. Carrasco-Zevallos, B. Keller, C. Viehland, P. Hahn, A. N. Kuo, P. J. DeSouza, C. A. Toth, and J. A. Izatt, "Real-time 4D visualization of surgical maneuvers with 100kHz swept-source Microscope Integrated Optical Coherence Tomography (MIOCT) in model eyes," *Invest. Ophthalmol. Visual Sci.* **55**, 1633 (2014).
5. O. M. Carrasco-Zevallos, C. Viehland, B. Keller, R. P. McNabb, A. N. Kuo, and J. A. Izatt, "Constant linear velocity spiral scanning for near video rate 4D OCT ophthalmic and surgical imaging with isotropic transverse sampling," *Biomed. Opt. Express* **9**(10), 5052–5070 (2018).
6. M. Karampelas, D. A. Sim, P. A. Keane, V. P. Papastefanou, S. R. Sadda, A. Tufail, and J. Dowler, "Evaluation of retinal pigment epithelium-bruch's membrane complex thickness in dry age-related macular degeneration using optical coherence tomography," *Br J Ophthalmol* **97**(10), 1256–1261 (2013).
7. M. Sommersperger, J. Weiss, M. A. Nasser, P. Gehlbach, I. Iordachita, and N. Navab, "Real-time tool to layer distance estimation for robotic subretinal injection using intraoperative 4d oct," *Biomed. Opt. Express* **12**(2), 1085–1104 (2021).
8. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express* **8**(8), 3627–3642 (2017).
9. A. Shah, L. Zhou, M. D. Abrámoff, and X. Wu, "Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images," *Biomed. Opt. Express* **9**(9), 4509–4526 (2018).
10. J. I. Orlando, P. Seeböck, H. Bogunovic, S. Klimscha, C. Grechenig, S. Waldstein, B. S. Gerendas, and U. Schmidt-Erfurth, "U2-Net: A Bayesian U-Net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological OCT scans," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (2019), pp. 1441–1445.
11. D. Ma, D. Lu, M. Heisler, S. Dabiri, S. Lee, G. W. Ding, M. V. Sarunic, and M. F. Beg, "Cascade dual-branch deep neural networks for retinal layer and fluid segmentation of optical coherence tomography incorporating relative positional map," in *Proceedings of Machine Learning Research*, vol. 121, T. Arbel, I. B. Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, and C. Pal, eds. (PMLR, Montreal, 2020), pp. 493–502.
12. A. Tran, J. Weiss, S. Albarqouni, S. Faghi Roohi, and N. Navab, "Retinal layer segmentation reformulated as OCT language processing," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, eds. (Springer International Publishing, Cham, 2020), pp. 694–703.
13. J. G. F. Wolfgang Drexler, *Optical Coherence Tomography Technology and Applications* (Springer International Publishing, 2015), 2nd ed.
14. H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2018), pp. 990–994.
15. A. Madani, M. Moradi, A. Karargyris, and A. Syeda-Mahmood, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," in *Medical Imaging 2018: Image Processing*, vol. 10574 (International Society for Optics and Photonics, 2018), p. 105741M.
16. B. Teixeira, V. Singh, T. Chen, K. Ma, B. Tamersoy, Y. Wu, E. Balashova, and D. Comaniciu, "Generating synthetic x-ray images of a person from the surface geometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 9059–9067.
17. M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing* **321**, 321–331 (2018).
18. M. J. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis," in *15th International Symposium On Biomedical Imaging* (IEEE, 2018), pp. 240–244.
19. F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *International Conference on Artificial Neural Networks* (Springer, 2017), pp. 626–634.
20. C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "Gan-based synthetic brain mr image generation," in *15th International Symposium on Biomedical Imaging* (IEEE, 2018), pp. 734–738.
21. C. Bermudez, A. J. Plassard, L. T. Davis, A. T. Newton, S. M. Resnick, and B. A. Landman, "Learning implicit brain MRI manifolds with deep learning," in *Medical Imaging 2018: Image Processing*, vol. 10574 (International Society for Optics and Photonics, 2018), p. 105741L.
22. A. Zaman, S. H. Park, H. Bang, C.-w. Park, I. Park, and S. Joung, "Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images," *Int. journal computer assisted radiology surgery* **15**(6), 931–941 (2020).

23. A. Lahiri, K. Ayush, P. Kumar Biswas, and P. Mitra, "Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale microscopy images: Automated vessel segmentation in retinal fundus image as test case," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), pp. 42–48.
24. X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.* **58**, 101552 (2019).
25. M. Unberath, J.-N. Zaech, C. Gao, B. Bier, F. Goldmann, S. C. Lee, J. Fotouhi, R. Taylor, M. Armand, and N. Navab, "Enabling machine learning in x-ray-based procedures via realistic simulation of image formation," *Int. journal computer assisted radiology surgery* **14**(9), 1517–1528 (2019).
26. M. Grimm, J. Esteban, M. Unberath, and N. Navab, "Pose-dependent weights and domain randomization for fully automatic x-ray to ct registration," *IEEE Trans. Med. Imaging* **40**(9), 2221–2232 (2021).
27. M. Zhou, H. Roodaki, A. Eslami, G. Chen, K. Huang, M. Maier, C. Lohmann, A. Knoll, and M. A. Nasseri, "Needle segmentation in volumetric optical coherence tomography images for ophthalmic microsurgery," *Appl. Sci.* **7**(8), 748 (2017).
28. M. Zhou, X. Wang, J. Weiss, A. Eslami, K. Huang, M. Maier, C. P. Lohmann, N. Navab, A. Knoll, and M. A. Nasseri, "Needle localization for robot-assisted subretinal injection based on deep learning," in *2019 International Conference on Robotics and Automation (ICRA)* (2019), pp. 8727–8732.
29. M. Zhou, K. Huang, A. Eslami, H. Roodaki, D. Zapp, M. Maier, C. P. Lohmann, A. Knoll, and M. A. Nasseri, "Precision needle tip localization using optical coherence tomography images for subretinal injection," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (2018), pp. 4033–4040.
30. M. Zhou, K. Huang, A. Eslami, H. Roodaki, H. Lin, C. Lohmann, A. Knoll, and M. A. Nasseri, "Beveled needle position and pose estimation based on optical coherence tomography in ophthalmic microsurgery," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (2017), pp. 308–313.
31. M. Zhou, M. Hamad, J. Weiss, A. Eslami, K. Huang, M. Maier, C. P. Lohmann, N. Navab, A. Knoll, and M. A. Nasseri, "Towards robotic eye surgery: marker-free, online hand-eye calibration using optical coherence tomography images," *IEEE Robot. Autom. Lett.* **3**(4), 3944–3951 (2018).
32. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 1125–1134.
33. N. M. Grzywacz, J. de Juan, C. Ferrone, D. Giannini, D. Huang, G. Koch, V. Russo, O. Tan, and C. Bruni, "Statistics of optical coherence tomography data from human retina," *IEEE Trans. Med. Imaging* **29**(6), 1224–1237 (2010).
34. G. Farhat, G. J. Czarnota, M. C. Kolios, and V. X. D. Yang, "Detecting cell death with optical coherence tomography and envelope statistics," *J. Biomed. Opt.* **16**(2), 1–7 (2011).
35. D. A. Jesus and D. R. Iskander, "Assessment of corneal properties based on statistical modeling of oct speckle," *Biomed. Opt. Express* **8**(1), 162–176 (2016).
36. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015), pp. 234–241.
37. J. Brownlee, *Generative Adversarial Networks with Python: Deep Learning Generative Models for Image Synthesis and Image Translation* (Machine Learning Mastery, 2019).
38. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Process.* **13**(4), 600–612 (2004).
39. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR* (2018).
40. J. P. Kolb, W. Draxinger, J. Klee, T. Pfeiffer, M. Eibl, T. Klein, W. Wieser, and R. Huber, "Live video rate volumetric OCT imaging of the retina with multi-MHz A-scan rates," *PLoS One* **14**(3), 1–20 (2019).
41. L. V. Del priore, H. J. Kaplan, R. Hornbeck, Z. Jones, and M. Swinn, "Retinal pigment epithelial debridement as a model for the pathogenesis and treatment of macular degeneration," *Am. J. Ophthalmol.* **122**(5), 629–643 (1996).
42. M. Chandler, P. Smith, D. Samuelson, and E. MacKay, "Photoreceptor density of the domestic pig retina," *Vet. Ophthalmol.* **2**(3), 179–184 (1999).
43. A. Hendrickson and D. Hicks, "Distribution and density of medium-and short-wavelength selective cones in the domestic pig retina," *Exp. Eye Res.* **74**(4), 435–444 (2002).
44. A. Hendrickson, K. Bumsted-O'Brien, R. Natoli, V. Ramamurthy, D. Possin, and J. Provis, "Rod photoreceptor differentiation in fetal and infant human retina," *Exp. Eye Res.* **87**(5), 415–426 (2008).