Resolution enhancement and realistic speckle recovery with generative adversarial modeling of micro-optical coherence tomography

KAICHENG LIANG,¹,* XINYU LIU,²,³,* SI CHEN,² JUN XIE,² WEI QING LEE,¹,⁴ LINBO LIU,² AND HWE KUAN LEE¹,³,⁴,†

¹Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore
²School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore
³Singapore Eye Research Institute, Singapore
⁴School of Computing, National University of Singapore (NUS), Singapore
⁵Image and Pervasive Access Lab, CNRS, Singapore
*Equal contribution
†leehk@bii.a-star.edu.sg

Abstract: A resolution enhancement technique for optical coherence tomography (OCT), based on Generative Adversarial Networks (GANs), was developed and investigated. GANs have been previously used for resolution enhancement of photography and optical microscopy images. We have adapted and improved this technique for OCT image generation. Conditional GANs (cGANs) were trained on a novel set of ultrahigh resolution spectral domain OCT volumes, termed micro-OCT, as the high-resolution ground truth (~1 μm isotropic resolution). The ground truth was paired with a low-resolution image obtained by synthetically degrading resolution 4x in one of (1-D) or both axial and lateral axes (2-D). Cross-sectional image (B-scan) volumes obtained from in vivo imaging of human labial (lip) tissue and mouse skin were used in separate feasibility experiments. Accuracy of resolution enhancement compared to ground truth was quantified with human perceptual accuracy tests performed by an OCT expert. The GAN loss in the optimization objective, and noise injection in both the generator and discriminator models were found to be important for achieving realistic speckle appearance in the generated OCT images. Qualitative examples applying the models to image data from outside of the training data distribution, namely human retina and mouse bladder, were also demonstrated, suggesting potential for cross-domain transferability. This preliminary study suggests that deep learning generative models trained on OCT images from high-performance prototype systems may have potential in enhancing lower resolution data from mainstream/commercial systems, thereby bringing cutting-edge technology to the masses at low cost.

© 2020 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Optical coherence tomography (OCT) is a 3-dimensional optical imaging technique, and has become part of the standard of care in ophthalmology [1] while growing in importance in other clinical specialties such as gastroenterology [2]. Axial resolution of OCT is governed by the light source bandwidth, while lateral (transverse) resolution is governed by the numerical aperture (NA) of the illumination beam [3]. Hardware efforts to improve axial and lateral resolution can be complex, requiring high performance lasers and imaging objectives, and dispersion matching in the reference and sample paths. Computational techniques have been employed to overcome these constraints. For axial resolution, dispersion mismatch can be corrected by compensation algorithms to restore resolution to the ideal limit set by the light source; recent studies have proposed methods to surpass that limit [4]. For lateral resolution, traditional deconvolution techniques such as the Richardson-Lucy algorithm have been suggested [5], while physics-based
algorithms such as interferometric synthetic aperture microscopy have also been successful [6].

Deep learning has found most success in image classification and feature detection tasks [7], and has also had a growing impact in computational imaging and inverse problems [8] such as resolution enhancement (known as ‘super-resolution’ in the computer vision literature), notably in optical microscopy [9] where deep learning has been used to improve image quality [10]. Generative adversarial networks (GANs) [11], an emerging branch of deep learning, has shown promise in a wide range of imaging applications. GANs use two powerful neural networks competing with each other to greatly enhance the quality and realism of machine-generated images, potentially performing better than a single neural network alone or blind techniques without data priors. Conditional GANs (cGANs) are a flavor of GANs that learn to generate a mapping between two domains by training on image pairs, where a ‘conditional’ image in one domain is co-registered with a ground truth image from another domain [12]. These techniques have been investigated in photography, microscopy [13], as well as OCT studies [14, 15], which focused on denoising and deblurring of commercial ophthalmic OCT images, rather than the learning of higher optical resolution from prototype OCT systems.

In this work we explore the hypothesis that cGANs can be used to enhance the axial and lateral resolution of OCT images, trained on an ultrahigh resolution OCT ground truth. Using images obtained by micro-OCT [16, 17] with ~1μm resolution, axial and lateral resolutions were synthetically degraded by windowing/averaging the interference spectra, producing an intrinsically co-registered set of paired low-high resolution data for training. Injection of noise in the cGAN architecture was found to substantially improve the quality of image generation. Comparisons were made between our approach and several previously reported techniques - classical blind deconvolution without deep learning (Richardson-Lucy deconvolution), a state of the art non-adversarial deep learning approach, and a vanilla cGAN with no noise injection. Models were separately trained on two datasets - mouse skin and human lip. Three use cases were investigated - the conversion of 1-dimensional low resolution (axial or lateral) to high resolution, and the conversion of 2-dimensional (axial+lateral) low resolution to high resolution. The 2-D case was further investigated for the realism of the speckle reconstruction, using a perceptual quality test performed by an OCT expert, where our GAN approach was found to perform better than previous techniques. We also report training details and hyperparameter heuristics that are specific to OCT image generation. Lastly, we show qualitative examples of our models performing enhancement on OCT data from outside of the training data distribution, namely human retina and mouse bladder, suggesting potential for cross-domain transferability.

Fig. 1. Model architecture of conditional GAN (cGAN) for super-resolution. Gaussian noise is injected during upsampling in the generator and at the input to the discriminator, to stabilize training and produce higher quality results.
2. Materials and methods

2.1. micro-OCT image data and pre-processing

Images were obtained using a prototype micro-OCT system previously reported [18], with axial scan rate of 60 kHz, axial resolution 1.3 μm (tissue) and lateral resolution 1.8 μm. Two datasets were investigated - 10 volumes of mouse skin images, and 8 volumes of human labial (lip) mucosa images, acquired in vivo by a handheld probe and reported in an earlier publication [18]. Each volume had dimensions ~ 800 x 1000 x 500 (~500 B-scans per volume). The pixel size was 0.4 μm (axial) and 0.8 μm (lateral). To generate realistic low axial resolution images, a tight Gaussian window with full width at half maximum (FWHM) set to 25% of the source bandwidth was applied to the raw k-space interference fringe data, degrading the axial resolution to ~5 μm while preserving the depth dimension. To generate low lateral resolution images, the fringes were moving-averaged over 6 A-scan lines, corresponding to ~5 μm in the lateral direction. For 2-D (axial and lateral) low resolution, the fringes were windowed then moving-averaged. Thus the low resolution images were intrinsically co-registered with the high resolution images, with the same pixel dimensions. The images were then cropped to non-overlapping 256 x 256 image patches for model training, with deep low signal regions discarded. The skin volumes were split into 7 volumes for training (28,728 training image patches) and 3 volumes for validation (12,312 validation image patches). The lip mucosa volumes were split into 5 volumes (30,780 training image patches) for training and 3 volumes for validation (18,468 validation image patches). Different models were trained on two versions of data - single frame data, and 3-frame moving averaged data. The 3-frame averaging served as a simple denoising technique that improved the perceptual quality of the images, and is a standard practice in OCT processing when some speckle reduction is preferred.

2.2. cGAN architecture and training

A cGAN architecture was used for the image enhancement deep learning model (Fig. 1). In this architecture, two powerful neural networks learn from each other, thereby improving the quality of the outputs. A ‘generator’ neural network learns from paired training data to produce an enhanced image from a ‘conditional’ input image relative to a ground truth, while a ‘discriminator’ neural network learns to discriminate whether the generator’s output is a generated ‘fake’ or a genuine ground truth, then returning feedback to the generator. As training of the generator and discriminator models is performed alternately, the two models compete till a theoretical limit where the generated images are indistinguishable from the ground truth, although in practice the generated quality does not necessarily converge to an optimum. This previously reported cGAN design is widely known as ‘pix2pix’ [12] and has several open-source skeleton implementations generously made available by the machine learning community [19, 20]. We made specific modifications as described below.

The generator used a ‘U-Net’ architecture [21] where a series of downsampling and upsampling convolutional paths with skip connections capture patterns at various levels of abstraction. A U-Net is traditionally used for image segmentation, but has also proven effective for the generation task. Recent deep learning papers have suggested that a deeper generator comprising multiple residual network (ResNet) blocks might have superior performance [22], but we did not observe significant differences with this on our training data. The generated image was fed to the discriminator, which was two ‘patchGAN’-style classifiers operating at two image scales [12, 22]. The receptive field of each pixel in the discriminator output was designed to be small (15 and 30 pixels width) relative to the input, such that finer details at the level of speckle might be evaluated by the discriminator. The GAN objective was regularized by an L1 (pixel-wise mean absolute difference) loss as follows: \( L_{GAN} + \lambda L_1 \) where \( \lambda \) was a hyperparameter set to 10. Larger values of \( \lambda \) up to 100 were suggested in prior studies using photographic data [12], but we found these
to be prone to poor speckle generation and blurry images (Fig. 2). We also experimented with using an additional Difference of Structural Similarity (DSSIM) loss term as suggested in the literature [13, 23] but this showed little improvement for OCT data and also produced blurry images; we have observed (Fig. 2) that SSIM may be a poor training objective and evaluation metric for OCT generation. Other training heuristics previously recommended for GANs such as normalization of inputs between -1 and 1, soft and noisy labels, and training the generator for 2-3 iterations for every discriminator iteration were used [24]. Models were trained for 30 epochs. To further improve the quality of the images, Gaussian noise with standard deviation 0.1 was injected at every level of the generative upsampling, as well as at the input to the discriminator, as has been suggested by prior GAN reports [25–27]. The same model architecture and hyperparameters were used for enhancing both human lip and mouse skin data. Even though model training was performed on image patches, the fully convolutional nature of the generator model (with no fully connected layers) enabled the use of image inputs of any size, therefore the full-sized original images could be used in the generator at model prediction time.

2.3. Perceptual accuracy test by human OCT expert

In order to evaluate the quality and realism of the computational reconstructions, a human reader was asked to evaluate the 2D-enhanced images. In typical GAN studies from the deep learning literature, human readers are crowdsourced from online platforms such as Amazon Mechanical Turk to evaluate the quality of generated photographs or artwork, but this is not feasible for specialized imaging data such as OCT. For our study, the reader was an OCT expert (coauthor X.L.) who was involved in planning the study and preparing the data, but was not involved in the machine learning, was blinded to the models, and had not seen the model-generated results beforehand. Image patches (256 x 256 pixels) were shown to the reader one at a time, and the reader given two seconds to evaluate. In a ‘paired’ test, a generated image was shown with its ground truth image side by side and the reader asked to identify the real image. In an ‘unpaired’ test, a single image, either generated or ground truth, was shown one at a time and the reader asked to determine its identity. Two seconds was longer than a typical GAN perceptual test, which gives only one second [26], so as to account for the complexity of a typical OCT image. Before commencing the test, 5 practice examples were showed, each followed by the answer. This was then followed by a test of 50 questions in sequence, with no answers shown. After each test, a ‘confusion score’ was computed as the fraction of incorrectly read images over 50 total images,
giving a percentage. Higher confusion scores closer to 50% would indicate that many images were incorrectly read, suggesting that generated images were realistic and nearly indistinguishable from real high-resolution images (nearly random chance). Lower confusion scores closer to 0% would indicate that most images were correctly read, suggesting that generated images were easily distinguished from real images. The confusion score of the GAN-generated results was compared to separate tests on images produced by a state-of-the-art Unet (non-adversarial training) originally designed for improving signal to noise ratio and image quality of microscopy images [10], as well as a vanilla conditional GAN [12] without the additional injection of noise.

Fig. 3. Enhancement of depth resolution. Scale bar 100 µm.

2.4. Cross-domain validation on real data

As a preliminary qualitative assessment of the models’ performance on real (not simulated) images from a different data distribution from the training data, images of normal (no pathology) human retinal images [28] and mouse bladder tissue [29] were obtained from freely available datasets that accompanied published papers. Retinal images were acquired on a Zeiss Cirrus ophthalmic spectral domain OCT system, with axial and lateral resolution 5 µm and 15 µm respectively, and bladder tissue images were acquired on a Bioptigen Envisu R-class pre-clinical imaging system, with axial and lateral resolution 0.9 µm (tissue) and 8.5 µm respectively. It was necessary to resize the input images such that the size of speckle was approximately similar to that of the training data. The generator model, like most modern neural networks, used convolution operations, for which the convolutional filters had been learned from training data based on the length scales (measured in number of pixels) of image features including speckle noise. Therefore it was necessary for the images entering the generator to have roughly the same length scales of speckle learned by the generator’s convolutional filters (Appendix). For datasets dissimilar to the training data, images were resized to 4x larger in pixel dimensions, using bilinear interpolation, before entering the generator. Low-signal regions deep in the images were cropped, and the
images were marginally resized to have dimensions of a multiple of 256, for ease of input to the trained model. Since higher-resolution ground truths were not available, generated results were qualitatively assessed.

3. Results and discussion

We have developed a deep learning based algorithm for resolution enhancement of OCT images, based on previously reported techniques in generative adversarial networks. Using very high resolution OCT images as a ground truth, 4x improvement in resolution was demonstrated on images with synthetic resolution degradation. As with typical GAN generation, objective evaluation of the generated outputs was challenging. Given the speckle noise that is inherent to coherent imaging such as OCT or ultrasound, the model was not able nor expected to exactly reproduce the noise content of the ground truth images. Therefore, conventional similarity metrics such as Structural Similarity (SSIM) gave low scores. Excessive regularization produced smoothed, speckle-reduced images with poor resemblance to OCT but still resulted in higher SSIM scores (Fig. 2). Reduced regularization produced speckle noise that appeared qualitatively realistic, suggesting that the noise distribution of the speckle was learned, while the exact details of the generated speckle pattern was different from the ground truth. The generation of realistic yet accurate speckle may be necessary in some specific contexts, and is an interesting possibility for future investigation. Large regularization also seems to suggest itself as a means of speckle noise reduction, although this needs more careful validation and was not the objective of this work.

![Image](image1.png)

Fig. 4. Enhancement of lateral resolution. Scale bar 100 µm.

Image examples showing resolution enhancement in depth, lateral, and both (2-D) axes are shown in Figures 3, 4, 5 and perceptual confusion scores presented in Table 1. The realism of generated speckle appeared to be high. However, it can be observed that the detailed content of the generated speckle pattern can differ from the ground truth, particularly in the 2-D case where
the inference space is larger (Figure 5), even though the larger scale features are preserved and have improved quality. This difference in generated noise pattern, especially in the low-signal (nearly black) background of the images where noise can originate from the laser or other system sources, can lead to poor results when standard quantitative pixel-level similarity metrics such as SSIM are used.

Human perceptual accuracy tests were preferred for evaluating the quality of the enhancement (Table 1). Examples from a range of algorithms are presented in Figures 6 and 7. The Richardson-Lucy technique was generally poor (Appendix) and thus deemed not sufficiently competitive for a human perceptual test. The non-adversarial Unet and vanilla cGAN (no noise injection) produced images that were easily discriminated by a human OCT expert (0% confusion). The noise-injected cGAN confusion scores were substantially higher. The unpaired test results were lower than paired results, which was surprising and opposite to typical GAN studies [26] where readers found single images more confusing. This may be due to our reader being a subject-matter OCT expert, such that in the absence of a confusing alternate image, he was able to tap on pre-existing specialized knowledge of OCT to distinguish realistic speckle. Results from the 3-frame averaged images showed better quality (higher confusion). In practice, the interpretation of OCT images often involves an averaging/denoising process where speckle noise is intended to be suppressed; training a model on denoised data could allow the model to focus on more important image features rather than speckle noise, which is challenging to reproduce.

Noise injection in the architecture was important for quality and realism of the reconstruction. Some examples of low quality images generated by a vanilla cGAN (no noise injection) are shown in Figure 7. The speckle pattern had a repeated grid-like artifact, severely affecting the realism of the images. The images also sometimes showed a noise pattern resembling speckle noise but repeated in most/all generated images (figure insets). This pattern might appear realistic on single images, but was quickly detected by the human reader as a generative artifact when observed over a large number of images from the same generator during a perceptual test.
Data type and model | Single frame | 3-frame average |
|-------------------|-------------|----------------|
|                   | Paired | Unpaired | Paired | Unpaired |
| Mouse skin Unet   | 0     | 0        | 0     | 0        |
| Mouse skin GAN\(_{\text{no noise}}\) | 0     | 6        | 2     | 4        |
| Mouse skin GAN\(_{\text{noise}}\) | 42    | 22       | 52    | 26       |
| Human lip Unet    | 0     | 0        | 0     | 0        |
| Human lip GAN\(_{\text{no noise}}\) | 2     | 2        | 0     | 8        |
| Human lip GAN\(_{\text{noise}}\) | 4     | 6        | 20    | 22       |

Table 1. Confusion score in % (0%: zero confusion, \(~50\%:\) maximal confusion) from perceptual accuracy test by a human OCT expert reader on 2-D enhanced images, discriminating between a model output and ground truth. The test consisted of 50 images, such that confusion score was the fraction of incorrectly read images over 50 total images. Higher scores closer to 50% indicate higher quality model outputs nearly indistinguishable from real high-resolution images (random chance).

Fig. 6. Examples of images produced by a range of techniques. Left-right: low-resolution input, Richardson-Lucy deconvolution with Gaussian point spread function of \(\sigma = 2\), non-adversarial Unet, noise-injected cGAN, ground truth.

The adversarial component of the algorithm appeared to be particularly important for OCT generation; the baseline non-adversarial Unet approach has been reported to be successful in microscopy for denoising, increasing signal-to-noise ratio and sharpening, but produced less realistic OCT images than our cGAN approach. Examples of images produced are shown in Figure 6. This was in agreement with most computational super-resolution studies [30,31], which have favored GANs using adversarial learning. Our study had some important limitations. Small numbers of datasets were used in these proof-of-concept experiments, limiting the generalizability of the findings. Low-resolution training data was synthetically created, based on simple operations of spectral cropping and averaging, which may not sufficiently simulate low-resolution images from real-world conventional OCT systems. Our network architecture mimicked a standard end-to-end learning design used in conventional GANs for artwork/photography, and did not incorporate physical or optical models of OCT image formation. Hybrid physics-inspired learning algorithms as suggested in recent computational optics studies [32] may potentially improve performance.

As a preliminary step towards application on real data, publicly available retinal OCT images [28] and mouse bladder OCT images [29] were enhanced (Fig. 8) using the model that performed best on perceptual tests (Table 1), the mouse skin model. Axial resolution was visibly enhanced, but lateral resolution enhancement was marginal. In the future, a more realistic simulation of low lateral resolution in the training data (rather than simply a moving average of spectra) could further improve performance. In the retina, the resolution of layers
particularly the inner/outer segments and retinal pigment epithelium (highlighted by inset in Figure 8) was enhanced, which could have relevance to clinical thickness measurements that have been proposed in previous high-resolution OCT studies [33]. The model performance was found to be sensitive to the input image size; using the original image dimensions produced low quality results. We postulate this to be due to the size and scale of specific image features (e.g. speckle) that are learned from the training data (Appendix). The speckle size of input images should at least roughly match that of the training data; more robust protocols for domain transfer will be developed in future work.

These early proof-of-concept experiments suggest the possibility of packaging the high performance of a prototype imaging system as a low cost software-based image enhancement tool that may be used by scientific/clinical peers who lack access to cutting-edge hardware. As long as the neural network model is trained on a data distribution that is identical or very similar to the intended test usage (e.g. imaging of the same organism and organ under similar conditions), the model can be expected to generate high-quality enhancement results. Even cross-domain OCT applications seem feasible in principle, based on our preliminary experiments, although results will be more variable and will require careful validation. Concerns of super-resolution inference leading to ‘hallucinatory’ artifacts have been reported [34], which should not dampen enthusiasm for this research direction but motivate validation by human readers and comparisons with ground truth images. This concept may also be relevant to high speed swept source OCT systems [35] that typically have lower optical bandwidth and thus worse axial resolution than spectral domain systems. The possibility of having ‘the best of both worlds’ of OCT systems combining high speed and high resolution is an intriguing avenue of future investigation.

4. Conclusion

In this proof-of-concept study, the feasibility of axial and lateral resolution enhancement of OCT images using a generative adversarial network was investigated. A high resolution ground truth acquired with micro-OCT, paired with simulated low resolution image inputs were used to train a neural network to generate resolution-enhanced outputs. Results were evaluated by a human OCT expert for perceptual realism. Preliminary cross-domain experiments were performed on image data from outside of the training data distribution. Future work will involve the acquisition of more realistic training data, such as true low lateral resolution images taken with low numerical
Fig. 8. Preliminary experiments with cross-domain application, applying a model trained on mouse skin microOCT to human retinal images (courtesy of [28]) and mouse bladder images (courtesy of [29]) from commercial OCT systems.

aperture optics and low axial resolution images taken with reduced source bandwidth, larger amounts of data with more variety of quality including typical imaging artifacts, and studies of cross-domain transferability and robustness.

**Funding**

This research is supported by the Ministry of Education Singapore under its Academic Research Fund Tier 1 (2018-T1-001-144), National Medical Research Council Singapore under its Open Fund - Individual Research Grant (MOH-OFIRG19may-0009), and Agency for Science, Technology and Research (A*STAR) under its Industrial Alignment Fund (Pre-positioning) (H17/01/a0/008).

**Acknowledgments**

We thank Dr. Chen-Hsin Sun for helpful conversations on ophthalmic OCT.

**Disclosures**

The authors declare no conflicts of interest.

**References**

1. J. Fujimoto and E. Swanson, “The Development, Commercialization, and Impact of Optical Coherence Tomography,” Invest. ophthalmology & visual science 57, OCT1–OCT13 (2016).
2. M. J. Gora, M. J. Suter, G. J. Tearney, and X. Li, “Endoscopic optical coherence tomography: technologies and clinical applications [Invited],” Biomed. Opt. Express 8, 2405–2444 (2017).

3. Y.-Z. Liu, F. A. South, Y. Xu, P. S. Carney, and S. A. Boppart, “Computational optical coherence tomography [Invited],” Biomed. Opt. Express 8, 1549–1574 (2017).

4. X. Liu, S. Chen, D. Cui, X. Yu, and L. Liu, “Spectral estimation optical coherence tomography for axial super-resolution,” Opt. Express 23, 26521–26532 (2015).

5. S. A. Hajiotosslami, M. R. N. Avanaki, and A. G. Podoleanu, “Image quality improvement in optical coherence tomography using Lucy-Richardson deconvolution algorithm,” Appl. Opt. 52, 5663–5670 (2013).

6. T. S. Ralston, D. L. Marks, P. Scott Carney, and S. A. Boppart, “Interferometric synthetic aperture microscopy,” Nat. Phys. 3, 129–134 (2007).

7. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521, 436–444 (2015).

8. G. Barbashitis, A. Ozcan, and G. Sittu, “On the use of deep learning for computational imaging,” Optica 6, 921–943 (2019).

9. C. Belthangady and L. A. Royer, “Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction,” Nat. Methods 16, 1215–1225 (2019).

10. M. Weigert, U. Schmidt, T. Boothe, A. Mäijller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, “Content-aware image restoration: pushing the limits of fluorescence microscopy,” Nat. Methods 15, 1090 (2018).

11. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Networks,” arXiv:1406.2661 [cs, stat] (2014). ArXiv: 1406.2661.

12. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” arXiv:1611.07004 [cs] (2016). ArXiv: 1611.07004.

13. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. GâijnaydÄśn, L. A. Bentolilla, C. Kural, and A. Ozcan, “Deep learning enables cross-modality super-resolution in fluorescence microscopy,” Nat. Methods 16, 103 (2019).

14. Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, “Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN,” Biomed. Opt. Express 9, 5129–5146 (2018).

15. Y. Huang, Z. Lu, Z. Shao, M. Ran, J. Zhou, L. Fang, and Y. Zhang, “Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network,” Opt. Express 27, 12289–12307 (2019).

16. E. Liu, J. A. Gardecki, S. K. Nadkarni, J. D. Toussaint, Y. Yagi, B. E. Bouma, and G. J. Tearney, “Imaging the subcellular structure of human coronary atherosclerosis using microÅ²ptical coherence tomography,” Nat. Medicine 17, 1010–1014 (2011).

17. D. Cui, K. K. Chu, B. Yin, T. N. Ford, C. Hyun, H. M. Leung, J. A. Gardecki, G. M. Solomon, S. E. Birket, L. Liu, S. M. Rowe, and G. J. Tearney, “Flexible, high-resolution micro-optical coherence tomography endobronchial probe toward in vivo imaging of cilia,” Opt. Lett. 42, 867–870 (2017).

18. S. Chen, X. Liu, N. Wang, Q. Ding, X. Wang, X. Ge, E. Bo, X. Yu, H. Yu, C. Xu, and L. Liu, “Contrast of nuclei in stratified squamous epithelium in optical coherence tomography images at 800 nm,” J. Biophotonics 12, e201900073 (2019).

19. P. Isola, “Github repository philippix2pix2pix,” (2020). Original-date: 2016-11-16T22:48:50Z.

20. E. Linder-Noren, “Github repository eriklindernoren/Keras-GAN,” (2020). Original-date: 2017-07-11T16:24:53Z.

21. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv:1505.04597 [cs] (2015). ArXiv: 1505.04597.

22. T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs,” arXiv:1711.11585 [cs] (2017). ArXiv: 1711.11585.

23. H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss Functions for Image Restoration With Neural Networks,” IEEE Transactions on Comput. Imaging 3, 47–57 (2017).

24. S. Chintala, “Github repository soumith/gan-hacks,” (2020). Original-date: 2016-12-09T16:09:27Z.

25. T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” arXiv:1812.04948 [cs, stat] (2018). ArXiv: 1812.04948.

26. T. R. Shaham, T. Dekel, and T. Michaeli, “SimGAN: Learning a Generative Model from a Single Natural Image,” arXiv:1905.01164 [cs] (2019). ArXiv: 1905.01164.

27. F. Huszar, “Instance Noise: A trick for stabilising GAN training,” (2016).

28. P. Gholami, P. Roy, M. K. Parthasarathy, and V. Lakshminarayanan, “OCTID: Optical Coherence Tomography Image Database,” arXiv:1812.07056 [cs] (2019). ArXiv: 1812.07056.

29. K. C. Zhou, R. Qian, S. Degan, S. Farsiu, and J. A. Izzat, “Optical coherence refraction tomography,” Nat. Photonics 13, 794–802 (2019).

30. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” arXiv:1609.04802 [cs, stat] (2016). ArXiv: 1609.04802.
mask-based lensless imaging,” Opt. Express **27**, 28075–28090 (2019).

33. C. D. Lu, B. Lee, J. Schottenhamml, A. Maier, E. N. Pugh, and J. G. Fujimoto, “Photoreceptor Layer Thickness Changes During Dark Adaptation Observed With Ultrahigh-Resolution Optical Coherence Tomography,” Invest. Ophthalmol. & Vis. Sci. **58**, 4632–4643 (2017).

34. J. P. Cohen, M. Lack, and S. Honari, “Distribution Matching Losses Can Hallucinate Features in Medical Image Translation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer International Publishing, Cham, 2018), Lecture Notes in Computer Science, pp. 529–536.

35. K. Liang, O. O. Ahsen, Z. Wang, H.-C. Lee, W. Liang, B. M. Potsaid, T.-H. Tsai, M. G. Giacomelli, V. Jayaraman, H. Mashimo, X. Li, and J. G. Fujimoto, “Endoscopic forward-viewing optical coherence tomography and angiography with MHz swept source,” Opt. Lett. **42**, 3193–3196 (2017).
5. Appendix

5.1. Image resizing for cross-domain transfer of generative models

As described in Methods, the input image to the generator should have length scale of features approximately matched to that of the generator’s convolutional filters (learned from training data). Figure 9 shows the effect of image scaling. Beyond 4x on input scaling, the generator appeared to produce unrealistic artifacts. In future studies of cross-domain generative transfer, the scale factor can be a hyperparameter to be optimized, although generated images will still require careful inspection by human experts for quality and realism.

![Fig. 9. Generated images from dataset (retina) outside of the training data distribution (mouse skin), with various scale factors on the input image.](image)

5.2. Richardson-Lucy deconvolution parameter selection

Richardson-Lucy deconvolution was generally unable to produce good quality or realistic OCT images of higher resolution. The point spread function was estimated by a Gaussian. The generated images were sensitive to the parameters of \( \sigma \), the standard deviation of the Gaussian PSF, and number of iterations. Some examples are shown in Figure 10.

![Fig. 10. Images from Richardson-Lucy deconvolution with various parameters.](image)