# Endoscopic ultrasound image synthesis using a cycle-consistent adversarial network

Alexander Grimwood[1], Joao Ramalhinho[1], Zachary M. C. Baum[1], Nina Montaña-Brown[1], Gavin J. Johnson[2], Yipeng Hu[1], Matthew J. Clarkson[1], Stephen P. Pereira[3], Dean Barratt[1], Ester Bonmati[1]

[1]Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK, and the UCL Centre for Medical Image Computing, University College London, London, UK
[2] Department of Gastroenterology, University College London Hospital, London, UK
[3] Institute for Liver and Digestive Health, University College London, London, UK
e.bonmati@ucl.ac.uk

**Abstract.** Endoscopic ultrasound (EUS) is a challenging procedure that requires skill, both in endoscopy and ultrasound image interpretation. Classification of key anatomical landmarks visible on EUS images can assist the gastroenterologist during navigation. Current applications of deep learning have shown the ability to automatically classify ultrasound images with high accuracy. However, these techniques require a large amount of labelled data which is time consuming to obtain, and in the case of EUS, is also a difficult task to perform retrospectively due to the lack of 3D context. In this paper, we propose the use of an image-to-image translation method to create synthetic EUS (sEUS) images from CT data, that can be used as a data augmentation strategy when EUS data is scarce. We train a cycle-consistent adversarial network with unpaired EUS images and CT slices extracted in a manner such that they mimic plausible EUS views, to generate sEUS images from the pancreas, aorta and liver. We quantitatively evaluate the use of sEUS images in a classification sub-task and assess the Fréchet Inception Distance. We show that synthetic data, obtained from CT data, imposes only a minor classification accuracy penalty and may help generalization to new unseen patients. The code and a dataset containing generated sEUS images are available at: https://ebonmati.github.io.

**Keywords:** endoscopic ultrasound, synthesis, classification.

## 1 Introduction

Endoscopic Ultrasound (EUS) is a minimally-invasive procedure to assess the gastrointestinal tract including pancreatobiliary disorders such as pancreatic cancer. It is a complex procedure that combines ultrasound and endoscopy, requiring advanced cognitive and technical skills, such as ultrasound image interpretation [1].

Recent advances in machine learning have made it feasible to automatically classify images and identify standard planes, which can improve ultrasound (US) image interpretation and assist clinicians during navigation with the aim to make the diagnosis more accurate. However, the success of deep-learning based applications relies on the acquisition of a large, well-curated dataset with enough quality to be representative and useful. This is a big challenge in US applications as often training data is limited and models tend to have overfitting problems [2]. Data acquisition during EUS procedures is especially difficult and demanding due to the disruption caused and time required by real-time labelling, as well as the inaccuracies associated with retrospective labelling, because it is difficult to confidently identify landmarks without the 3D spatial and temporal context.

Medical image synthesis using convolutional neural networks (CNN) has been shown to be able to successfully translate Magnetic Resonance Imaging (MRI) to Computed Tomography (CT) [3] and to translate US to MRI [4]. In this work, we evaluate the use of a cycle-consistent adversarial network (CycleGAN) [5] to perform CT-to-EUS image translation to generate synthetic EUS (sEUS) images for the purpose of data augmentation. As an example, CycleGANs have been used before to improve the realism in US simulation from CT in a ray-casting approach, or to generate labelled US images from musculoskeletal US as a data augmentation strategy [6, 7]. The CycleGAN approach is of particular interest for our clinical application as no commercially available endoscopes exist capable of acquiring paired US/CT data, making endoscopy training and patient navigation difficult. The aim of our study is: 1) to assess the similarity between real (EUS) and synthetic (sEUS) images, and 2) to evaluate the use of sEUS images as a data augmentation strategy in a clinically relevant EUS classification task.

## 2 Methods

### 2.1 Data

**CT data.** CT data from five patients was obtained, four were from the MICCAI 2015 workshop and challenge: Multi-Altas Labelling Beyond the Cranial Vault [8]. CT slice dimensions were 512×512 with pixel sizes from 0.59 mm to 0.73 mm. Slice thicknesses were 3 mm, with volume depths from 393 mm to 444 mm. We also included a CT volume of size 512×512×229 with pixel dimensions of 0.55×0.55 mm and a slice thickness of 1 mm. Segmentations of the following structures were available for this study: stomach, pancreas, liver and aorta.

**EUS data.** EUS images were obtained from five patients who underwent an EUS-guided examination at University College Hospital London. Data were acquired from a Hitachi Preirus EUS console and a Pentax EG-3270UK or EG-3870UTK US linear video endoscopes with a 7.5 MHz probe. EUS images were collected from video frames of each examination recorded with a resolution of 720×480 pixels at imaging depths from 4 mm to 6 mm and cropped to 522×200 pixels, removing identifiable text and depth-attenuated regions. Anatomical landmarks were identified by an expert and recorded during the procedure. EUS images containing the three clinically relevant anatomical landmarks: pancreas, liver and aorta, were manually identified and collected.

Assorted images outside of these labels were also collected for a background class used in the classification subtask. Images from four patients were used for CycleGAN and classifier training. Data from the remaining patient were used to evaluate classification performance. **Table 1** shows a summary of the number of images and available labels from each patient.

**Table 1.** Summary of EUS images collected for training the CycleGAN and classifier, and for evaluation of the classifier.

| Patient | Task | Images | Labels | | | |
|---|---|---|---|---|---|---|
| | | | Aorta | Liver | Pancreas | Background |
| EUS1 | train | 3299 | 0 | 0 | 1694 | 1605 |
| EUS2 | train | 4758 | 767 | 0 | 3991 | 0 |
| EUS3 | eval | 4451 | 1301 | 729 | 2421 | 0 |
| EUS4 | train | 4996 | 1126 | 1485 | 2385 | 0 |
| EUS5 | test | 1933 | 141 | 503 | 1289 | 0 |

### 2.2 EUS/CT image-to-image translation

Synthetic EUS images were generated using CycleGANs trained to translate 2D CT image planes into sEUS images (**Fig. 1**). The CT plane locations, orientations and bounding dimensions approximated real EUS views. Candidate sEUS locations were automatically identified in CT volumes using the associated CT segmentation labels. Points were randomly sampled along the outer surface of the CT stomach segmentation. Realistic sEUS probe orientations were identified at each point by randomly generating poses within a 30° cone normal to the stomach and retaining only poses where the view intersected an anatomical label of interest (i.e., aorta, liver or pancreas) as shown in **Fig. 2**. These poses were recorded as transformation matrices and saved to file.

During CycleGAN training, 2D CT images were sliced from CT volumes on the fly using a previously reported simulation pipeline [9]. The framework extracted a sEUS field of view, defined by a transformation matrix, in the CT volume. The CT planes were then passed to the CycleGAN with randomly selected EUS images as an unpaired input dataset.

The CycleGAN was based upon a previously described implementation comprising a generator and adversarial discriminator for each imaging modality [5]. Paired EUS and CT plane images were passed to their respective generators, which were trained to map their input modality into synthetic images (i.e., CT to sEUS and EUS to sCT). These synthetic images were subsequently passed to the relevant generator for mapping back to their original modalities. Training was governed by adversarial losses calculated at each discriminator and by cycle consistency losses comparing input images to those mapped to a synthetic modality and then remapped back to their original.

A small pre-trained Gaussian denoising network was added before each discriminator to prevent the generator from embedding information capable of facilitating loss minimization without improving image-to-image translation [10, 11].

### 2.3 Implementation Details

Three CycleGANs were trained, one for each of the three labels: aorta, liver, and pancreas. A batch size of 1 was used for 200 epochs, where the number of iterations per epoch was limited by the relevant EUS dataset size. Other hyperparameters were set to the defaults used in the original implementation [5]. Adam optimization was used on discrimination and generation networks with a learning rate of 0.0002 that decayed linearly after 100 epochs [12]. Image intensity values were normalized between –1 and 1. Data augmentation was applied to all images, incorporating random horizontal flips and random cropping within a 40 pixel margin. The CT plane slicer and CycleGAN frameworks were run simultaneously on a single 16GB NVIDIA Quadro P5000 GPU. CT slices were selectively generated so that EUS images were paired only with CT slices containing >1000 labelled pixels and < 50 pixels with high Hounsfield Units, indicative of bone. For each epoch, a 90/10 training/validation split was randomly applied to the dataset. Losses were plotted against epoch and inspected to ensure convergence was achieved. All models were implemented in TensorFlow 2.2 and CUDA Toolkit 10.1 [13]. Synthetic data for the classifier evaluation sub-task was created using the trained CT-sEUS generator network from each CycleGAN. Open-source code was used where possible and is available at: https://ebonmati.github.io.

### 2.4 Evaluation

Evaluating GANs remains an open challenge, as there is no concrete way to quantify how realistic and diverse the synthetic images are, and no ground truth exists. Often, models are evaluated in a subjective and quantitative manner by asking several observers to rate the images [14]. In this work, we used the Fréchet Inception Distance and a classification sub-task to evaluate our model, as described below.

**Fréchet Inception Distance.** To quantitatively evaluate the quality of the synthetic images, we calculated the Fréchet Inception Distance (FID) [15]. FID is a widely used metric for evaluating the similarity between the generated images (synthetic) and the real images. FID uses the activation distributions of the Inception-v3 model [16] to calculate the distance between real and synthetic images. We used the pre-trained Inception-v3 model available in Keras [16] to obtain the activation distributions for our real and synthetic images, where the FID score was then calculated as follows:

$$FID = \|\mu_X - \mu_Y\|^2 + Tr\left(\Sigma_X + \Sigma_Y - 2\sqrt{\Sigma_X \Sigma_Y}\right), \quad (1)$$

where $\mu_X$ and $\mu_Y$ are the mean of the feature vectors for the real and synthetic images, respectively; $\Sigma_X$ and $\Sigma_Y$ are the covariance matrix for the real and synthetic images, respectively; $\|\mu_X - \mu_Y\|^2$ refers to the sum squared difference between the two mean vectors, and $Tr$ is the trace. A lower FID indicates better-quality synthetic images; conversely, a higher score indicates a lower-quality image. An FID of 0 demonstrates that the activation distribution of the synthetic images is identical to that of the real images. FID is also capable of detecting intra-class mode dropping (i.e., a model that generates

only one type of image for each landmark or class), noise, blurring, and other systematic distortions.

**Classification sub-task.** We evaluate the use of the synthetic EUS images in a classification sub-task. The aim here is to: 1) evaluate the use of synthetic EUS images to classify real EUS images, and 2) to use the synthetic EUS images as a data augmentation strategy. As summarized in **Table 2**, the number of real training images for each class was: 3,194 aorta, 2,214 liver and 10,491 pancreas. A fourth background class was added in training only, incorporating 1,605 EUS images from a mix of indiscernible anatomy and poor quality images. To achieve this, we implemented a simple VGG-16 classification model to classify EUS into the following classes: aorta, liver and pancreas. We used the pre-trained weights from ImageNet, a batch size of 64, a learning rate of $1e^{-7}$ and 100 epochs. As loss function, we used a weighted categorical cross entropy with the weights of 4.23, 4.40, 4.48, and 2.78 for aorta, background, liver and pancreas, respectively. We trained the model using 5 different ratios of synthetic/real images: 0% synthetic + 100% real, 25% synthetic + 75% real, 50% synthetic + 50% real, 75% synthetic and 25% real. For each synthetic ratio, we report the accuracy, precision, recall and F1-measure. Pairs of classifier models were compared using McNemar tests to assess whether differences in accuracy were significant.

**Qualitative evaluation.** We are also interested in the visual explanation and spatial localization of important regions in the EUS and sEUS images that were used to predict the corresponding class. We used the Gradient-weighted Class Activation Mapping (Grad-CAM) to generate the class activation maps for each sample [17]. These maps provide an insight into the model interpretation by backpropagating the gradients from the last convolutional layer.

## 3 Results and Discussion

**Fig. 1** shows a comparison between a real EUS image and a sEUS image for each of the anatomical landmarks selected (aorta, liver, pancreas). Visually inspecting the generated sEUS images, we observed that the sEUS images obtained with CycleGAN look realistic as the main features of the anatomical landmarks are preserved.

In **Table 2** we report the classifier performance when trained on varying ratios of sEUS to real EUS images, with the number of sEUS increasing with the ratio. From this table we can observe that classification accuracy is maintained for sEUS ratios up to 75%. We attribute this to the fact sEUS images may provide a consistent representation of patient variation on the selected anatomical landmarks, making it feasible to generalize to new patients. Liver F-measures were consistently low, indicating poor classification performance and degrading overall accuracy scores. Due to the liver's size and position, liver-labelled images can often contain additional anatomical features belonging to the other classes. We speculate this may be a contributing factor to the consistently low F-measures.

Single-sided McNemar tests comparing classifier pairs indicated the small reduction in accuracy, from 0.63 to 0.61 with increasing sEUS ratio, was statistically significant ($p<0.05$).

The FID scores are shown in **Table 3**. Although the FID measure is widely used to evaluate the realism of synthetic images, a significant limitation arises from its reliance on a pre-trained model (ImageNet) that does not comprehensively represent US-specific features. This lack of accurate representation is compounded by the small dataset used in this study (and in medical imaging generally) compared to the originally intended application of FID. As such, we cannot expect the predicted activation distributions to provide authoritative results on our specific clinical application. . An indication of the ideal FID is given by the differences between random subsets within the EUS data (as shown in the EUS vs EUS results). To have a better quantification of a bad FID value, we compared all the EUS images to all the EUS images with added noise using a Gaussian distribution with 0 mean and a standard deviation of 0.1. Our synthetic images achieved lower scores in comparison to that from noisy US images which yielded a FID >300.

Finally, **Fig. 3** shows the Grad-CAM activations for two image examples (one EUS and one sEUS) of the pancreas. Note the model has focused on the area representing the pancreas to make a correct prediction.

Other studies have used GAN-based methods to simulate and augment ultrasound image data: Bargsten and Schlaefer developed SpeckleGAN, which generates intravenous ultrasound speckle simulations from segmentation maps, achieving FID scores < 115 [18]. Peng, et al. generated synthetic ultrasound from MRI images and qualitatively demonstrated their equivalence to numerical simulations [19]. A broader examination of GAN-based approaches in medical imaging was presented by Yi, et al. [20].

In future, this study could be extended to aid EUS navigation by establishing a real-time image labelling and automated landmark recognition framework, for example, by using the Grad-CAM maps to localise salient features in EUS video, as demonstrated in this work. Further potential enhancements include developing a single CycleGAN model capable of generating all three landmark types to enable multi-class object generation and detection.

## 4 Conclusions

The results of this work demonstrate that the generation of synthetic EUS images, from CT data, can support training of a simple classification model when data is scarce as it may better represent the population. It allows generation of a large dataset from specific anatomical landmarks that are relevant for the clinical application of interest, which would not be possible otherwise (as demonstrated by the poor accuracy obtained when using the only real EUS data available). The proposed method is easy to use compared to manual data acquisition and labelling, which is a task that is time consuming and requires the input of a clinical expert.

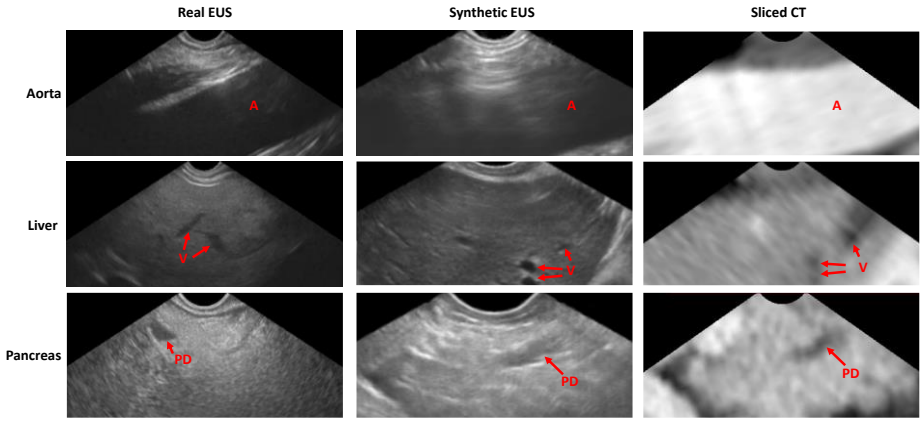


**Fig. 1.** Comparison between a real EUS image, a synthetic EUS image and the sliced CT from which it was generated for each anatomical landmark (aorta, liver and pancreas). Indicative anatomical features are shown in red: A – aorta, V – liver vasculature, PD – pancreatic duct.

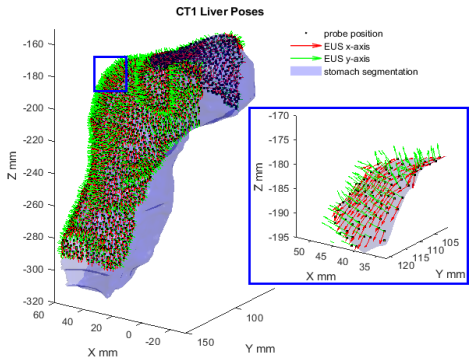


**Fig. 2.** Subset of candidate EUS probe positions and orientations at the stomach surface for liver views within a CT volume.

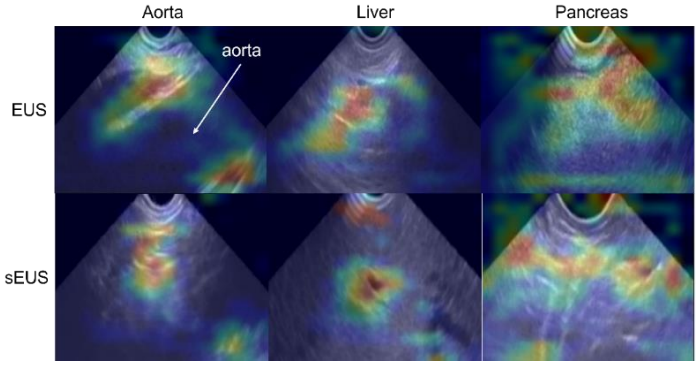


**Fig. 3.** Normalized class activation maps for real EUS images and a synthetic EUS images (sEUS) representing the aorta, liver and pancreas. The red areas represent increased regions of activation used by the model to make a correct prediction.

**Table 2.** Classifier performance when trained on varying ratios of synthetic to real EUS images. The number of sEUS images increases with synthetic ratio.

| Synthetic ratio (%) | Precision | Recall | F1-measure | | | Accuracy |
|---|---|---|---|---|---|---|
| | | | Aorta | Liver | Pancreas | |
| 0 | 0.46 | 0.51 | 0.38 | 0.07 | 0.77 | 0.63 |
| 25 | 0.45 | 0.54 | 0.38 | 0.05 | 0.78 | 0.62 |
| 50 | 0.44 | 0.54 | 0.38 | 0.05 | 0.78 | 0.62 |
| 75 | 0.43 | 0.53 | 0.36 | 0.06 | 0.77 | 0.61 |

**Table 3.** Fréchet inception distance scores when comparing random subsets of real EUS images within the same class, and when comparing real EUS to synthetic EUS (sEUS) images.

| Control (n images) | Compared to (n images) | FID |
|---|---|---|
| EUS pancreas (5250) | EUS pancreas (5241) | 2.00 |
| EUS aorta (1599) | EUS aorta (1595) | 6.96 |
| EUS liver (1110) | EUS liver (1104) | 11.03 |
| EUS all images (7959) | EUS all images (7940) | 1.83 |
| EUS all images (7959) | EUS all images + noise (7940) | 312.56 |
| EUS pancreas (10491) | sEUS pancreas (11763) | 79.88 |
| EUS aorta (3194) | sEUS aorta (2774) | 71.30 |
| EUS liver (2214) | sEUS liver (4365) | 71.68 |
| EUS all images (15899) | sEUS all images (18902) | 55.31 |

## Acknowledgements

This work is supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) (203145/Z/16/Z) and by Cancer Research UK (CRUK) Multidisciplinary Award (C28070/A19985). ZMC Baum is supported by the Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarships-Doctoral Program, and the UCL Overseas and Graduate Research Scholarships. SP Pereira was supported by the UCLH/UCL Comprehensive Biomedical Centre, which receives a proportion of funding from the Department of Health's National Institute for Health Research (NIHR) Biomedical Research Centres funding scheme.

## References

1. Bonmati, E., Hu, Y., Gibson, E., Uribarri, L., Keane, G., Gurusami, K., Davidson, B., Pereira, S.P.S.P., Clarkson, M.J.M.J., Barratt, D.C.D.C.: Determination of optimal ultrasound planes for the initialisation of image

registration during endoscopic ultrasound-guided procedures. Int. J. Comput. Assist. Radiol. Surg. 13, 875–883 (2018). https://doi.org/10.1007/s11548-018-1762-2.

2. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T.: Deep Learning in Medical Ultrasound Analysis: A Review, (2019). https://doi.org/10.1016/j.eng.2018.11.020.
3. Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with context-aware generative adversarial networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 417–425. Springer Verlag (2017). https://doi.org/10.1007/978-3-319-66179-7_48.
4. Jiao, J., Namburete, A.I.L., Papageorghiou, A.T., Noble, J.A.: Self-Supervised Ultrasound to MRI Fetal Brain Image Synthesis. IEEE Trans. Med. Imaging. 39, 4413–4424 (2020). https://doi.org/10.1109/TMI.2020.3018560.
5. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob, 2242–2251 (2017).
6. Zhang, L., Portenier, T., Goksel, O.: Learning ultrasound rendering from cross-sectional model slices for simulated training. Int. J. Comput. Assist. Radiol. Surg. 16, 721–730 (2021). https://doi.org/10.1007/s11548-021-02349-6.
7. Cronin, N.J., Finni, T., Seynnes, O.: Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images. Comput. Methods Programs Biomed. 196, 105583 (2020). https://doi.org/https://doi.org/10.1016/j.cmpb.2020.105583.
8. Landman, B., Xu, Z., Igelsias, J.E., Styner, M., Langerak, T.R., Klein, A.: Multi-Atlas Labeling Beyond the Cranial Vault. https://doi.org/10.7303/syn3193805.
9. Ramalhinho, J., Tregidgo, H.F.J., Gurusamy, K., Hawkes, D.J., Davidson, B., Clarkson, M.J.: Registration of Untracked 2D Laparoscopic Ultrasound to CT Images of the Liver Using Multi-Labelled Content-Based Image Retrieval. IEEE Trans. Med. Imaging. 40, 1042–1054 (2021). https://doi.org/10.1109/TMI.2020.3045348.
10. Porav, H., Musat, V., Newman, P.: Reducing Steganography In Cycle-consistency GANs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recogni-tion (CVPR) Workshops. pp. 78–82 (2019).
11. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Trans. Image Process. 26, 3142–3155 (2017). https://doi.org/10.1109/TIP.2017.2662206.
12. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization, (2017).
13. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Y.J., Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, M.S., Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, J.S., Benoit Steiner, Ilya Sutskever, Kunal

Talwar, P.T., Vincent Vanhoucke, Vijay Vasudevan, F.V., Oriol Vinyals, Pete Warden, Martin Wattenberg, M.W., Yuan Yu, and X.Z.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/, (2015). https://doi.org/10.5281/zenodo.4724125.

14. Lucic, M., Kurach, K., Michalski, M., Gelly, S., Bousquet, O.: Are GANs Created Equal? A Large-Scale Study. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.) Advances in Neural Information Processing Systems. Curran Associates, Inc. (2018).
15. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 6629–6640 (2017). https://doi.org/10.5555/3295222.3295408.
16. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 2818–2826. IEEE Computer Society (2016). https://doi.org/10.1109/CVPR.2016.308.
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int. J. Comput. Vis. 128, 336–359 (2020). https://doi.org/10.1007/s11263-019-01228-7.
18. Bargsten, L., Schlaefer, A.: SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. Int. J. Comput. Assist. Radiol. Surg. 15, 1427–1436 (2020). https://doi.org/10.1007/s11548-020-02203-1.
19. Peng, B., Huang, X., Wang, S., Jiang, J.: A Real-Time Medical Ultrasound Simulator Based on a Generative Adversarial Network Model. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 4629–4633 (2019). https://doi.org/10.1109/ICIP.2019.8803570.
20. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Med. Image Anal. 58, 101552 (2019). https://doi.org/https://doi.org/10.1016/j.media.2019.101552.