2-2021

Enhancing Reproductive Organ Segmentation in Pediatric CT via Adversarial Learning

Chi Nok Enoch Kan

Taly Gilat-Schmidt

Dong Hye Ye

Enhancing reproductive organ segmentation in pediatric CT via adversarial learning

Kan, Chi Nok Enoch, Gilat-Schmidt, Taly, Ye, Dong Hye

Chi Nok Enoch Kan, Taly Gilat-Schmidt, Dong Hye Ye, "Enhancing reproductive organ segmentation in pediatric CT via adversarial learning," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 1159612 (15 February 2021); doi: 10.1117/12.2582127

Event: SPIE Medical Imaging, 2021, Online Only

Enhancing Reproductive Organ Segmentation in Pediatric CT via Adversarial Learning

Chi Nok Enoch Kan[a], Taly Gilat-Schmidt[a] and Dong Hye Ye[a]

[a]Department of Electrical and Computer Engineering, Marquette University, Milwaukee, USA

ABSTRACT

Accurately segmenting organs in abdominal computed tomography (CT) scans is crucial for clinical applications such as pre-operative planning and dose estimation. With the recent advent of deep learning algorithms, many robust frameworks have been proposed for organ segmentation in abdominal CT images. However, many of these frameworks require large amounts of training data in order to achieve high segmentation accuracy. Pediatric abdominal CT images containing reproductive organs are particularly hard to obtain since these organs are extremely sensitive to ionizing radiation. Hence, it is extremely challenging to train automatic segmentation algorithms on organs such as the uterus and the prostate. To address these issues, we propose a novel segmentation network with a built-in auxiliary classifier generative adversarial network (ACGAN) that conditionally generates additional features during training. The proposed CFG-SegNet (conditional feature generation segmentation network) is trained on a single loss function which combines adversarial loss, reconstruction loss, auxiliary classifier loss and segmentation loss. 2.5D segmentation experiments are performed on a custom data set containing 24 female CT volumes containing the uterus and 40 male CT volumes containing the prostate. CFG-SegNet achieves an average segmentation accuracy of 0.929 DSC (Dice Similarity Coefficient) on the prostate and 0.724 DSC on the uterus with 4-fold cross validation. The results show that our network is high-performing and has the potential to precisely segment difficult organs with few available training images.

Keywords: Medical Image Segmentation, Generative Models, Generative Adversarial Networks, Organ Segmentation

1. INTRODUCTION

Organ segmentation is a popular yet challenging task in medical image analysis. In recent years, researchers have proposed many successful deep learning-based segmentation algorithms such as the U-Net, 3D U-Net, CE-Net and Dense V-Net[1,2] for both pixelwise and volumetric segmentation. However, it remains a challenge to generalize the performances of these algorithms since the amount of available training images varies by organ type. Radiosensitivity and demography have significant impacts on how readily available CT images of a particular organ is. Uterus and prostate are amongst some of the most radiosensitive organs,[3] and hence the availability of CT images containing them is extremely low. Moreover, children are more vulnerable to ionizing radiation exposure in CT scans than adults.[4] Therefore, segmentation performances of most algorithms on these two organs are generally poor. With the recent breakthroughs in generative adversarial networks (GANs),[5] researchers are able to synthesize viable and realistic training images in both conditional and unconditional fashions. However, GANs are usually trained separately from the segmentation algorithms and the training process can be unstable due to gradient flow and mode collapse issues. There exists a need for a segmentation framework that effectively combines the generation of novel training images and the segmentation task itself.

U-Net has gained tremendous popularity for its ability to accurately segment biomedical images. U-Net follows an encoder-decoder architecture, and different networks can be used for its encoder and its decoder. U-Net has inspired many other architectures for 2D CT organ segmentation, such as the dilated residual U-Net (DR U-Net)[6] and organ-attention networks with reverse connections (OAN-RCs).[7] OAN-RC in particular has

Further author information: (Send correspondence to C.N.E.K.)
C.N.E.K.: E-mail: chinokenoch.kan@marquette.edu, Telephone: +1(612)545-7814
T. G.: E-mail: tal.gilat-schmidt@marquette.edu, Telephone: +1(414) 288-4447
D.H.Y.: E-mail: donghye.ye@marquette.edu, Telephone: +1(215)880-7843

Medical Imaging 2021: Image Processing, edited by Ivana Išgum, Bennett A. Landman, Proc. of SPIE Vol. 11596, 1159612 · © 2021 SPIE · CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2582127

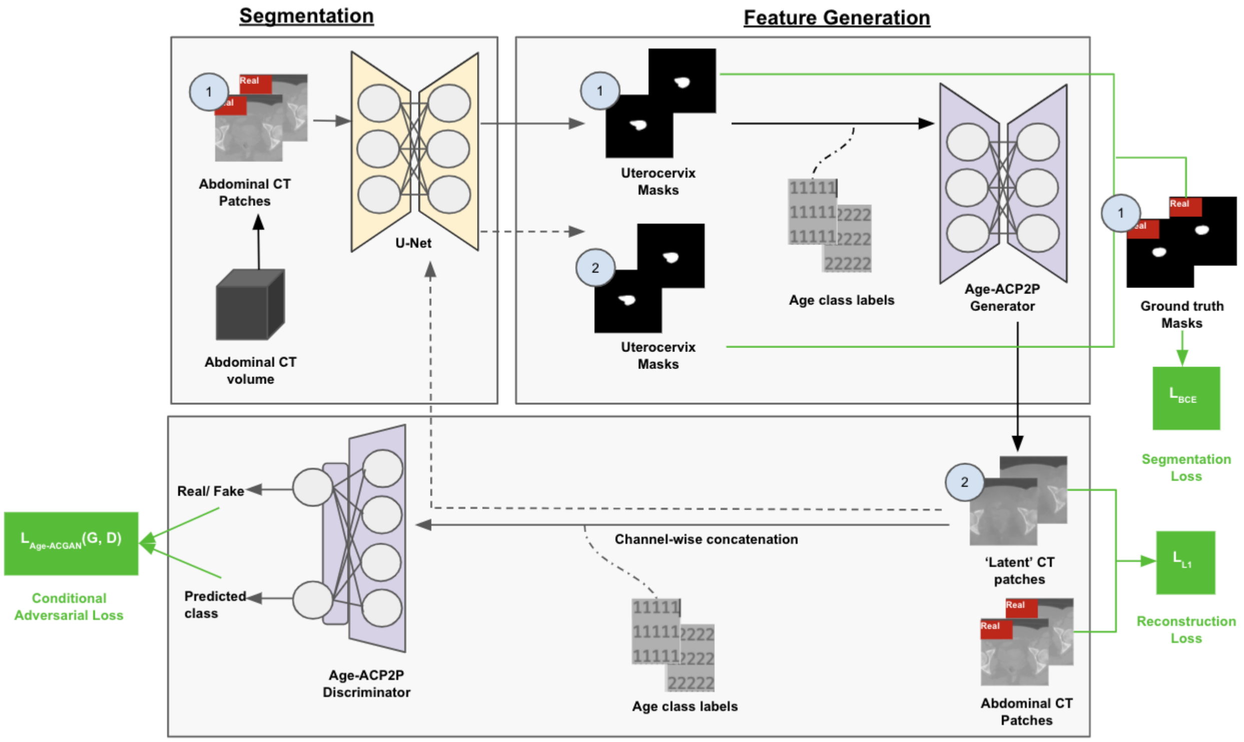

Figure 1. An overview of our proposed CFG-SegNet framework: we first extract patches of abdominal CT images (denoted as 1) to produce their initial segmentation masks. The segmentation masks are subsequently used by the Age-ACP2P generator to reconstruct latent CT patches (denoted as 2). We then retrain the U-Net on the reconstructed patches to obtain a second set of segmentation masks for segmentation loss calculation. By incorporating age information and feedbacks from the Age-ACP2P discriminator, we expect the quality of the latent patches as well as the segmentation masks (1) and (2) to improve over time.

achieved state-of-the-art segmentation performances for 13 organs in abdominal CT scans. However, very few of the reported experiments include reproductive organ segmentation. Shahedi et. al. reported uterus and placenta magenetic resonance (MR) segmentation results using a 3D fully convolutional network (3D-FCN).[8] As for prostate segmentation, Karimi et. al. previously reports a global-local CNN method for T2-weighted prostate MR segmentation.[9] Most of the current work done in reproductive organ segmentation involves MR images, and there is limited literature reporting CT segmentation results for reproductive organs. Therefore, our work also serves as an important baseline for reproductive organ CT segmentation.

2. METHOD

Our main goal is to construct an effective segmentation framework that generates new training images at each iteration. A new loss function is proposed to simultaneously train a conditional GAN (cGAN) and a segmentation network. The quality of the synthesized features and the segmentation masks is expected to improve over time.

2.1 Adversarial Learning Framework for Reproductive Organ Segmentation

The proposed framework consists of two networks, namely an Age Auxiliary Classifier Pix2Pix (Age-ACP2P) network and a U-Net. While the U-Net is responsible for segmentation of the training images, the Age-ACP2P network generates additional 'latent' training features at each iteration. Fig. 1 provides a graphical overview of our proposed method.

In the first iteration of training, segmentation masks are generated from the training images by the U-Net. These segmentation masks are translated back into latent features using an Age-ACP2P network. An Age-ACP2P network is simply a Pix2Pix variant of Age-ACGAN (Age Auxiliary Classifier GAN), which was previously used to conditionally synthesize CT scans containing the pancreas.[10] Similar to Age-ACGAN, we incorporate age information by adding an additional auxiliary classifier to Age-ACP2P's discriminator and by channel-wise concatenation of age class labels.

2.2 Loss Function

The original GAN formulation consists of two competing neural networks, a generator network G and a discriminator network D.[5] G takes in a random noise vector $\mathbf{z}$, and transforms it into an image $G(\mathbf{z})$. The training objective of D is to maximize $log(D(\mathbf{z})) + log(1 - D(G(\mathbf{z}))$, the probability of assigning correct labels to both training images and synthetic (or 'fake') images generated by G . The G network is trained to minimize $log(1-D(G(\mathbf{z}))$, the log of the inverted probability of D's prediction of fake images. In practice, the minimization of the inverted probability is not an easy task. Therefore we seek to maximize $D(G(\mathbf{z}))$ instead. In summary, the objective function of GAN can be formulated as a minimax loss:

$$\arg\min_{\mathbf{G}}\max_{\mathbf{D}} L_{GAN}(G, D) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[logD(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[log(1 - D(G(\mathbf{z})))] \tag{1}$$

Age-ACP2P, the feature generation portion of our proposed CFG-SegNet uses a variant of GAN called conditional GANs (cGANs). In particular, we choose to use the Pix2Pix network, a type of cGAN commonly used in image-to-image translation tasks.[11] The loss function of Pix2Pix is a combination of conditional adversarial ($L_{cGAN}(G, D)$) and reconstruction (L_{L1}) losses:

$$G^* = \arg\min_{\mathbf{G}}\max_{\mathbf{D}} L_{cGAN}(G, D) + \lambda L_{L1} \tag{2}$$

We can replace $L_{cGAN}(G, D)$ with the adversarial loss proposed in auxiliary classifier GANs (ACGANs) to incorporate class information from the training images.[12] To specifically include age information in our proposed network, we use a variant of ACGAN architecture called Age-ACGAN.[10,12] Age-ACGAN modifies the original objective function in ACGAN to compute the log-likelihoods of the correct source (L_s) and the correct age class (L_a) as following:

$$L_s = E[\log P(S_{CT} = real|X_{real})] + E[\log P(S_{CT} = fake|X_{fake})] \tag{3}$$

$$L_a = E[\log P(C_{age} = age|X_{real})] + E[\log P(C_{age} = age|X_{fake})] \tag{4}$$

The training objective of discriminator D is to maximize $L_a + L_s$. This ensures the discriminator always maximizes the log likelihood it assigns to the correct source of CT image CT_{source} and the correct age class CT_{age}. By denoting Age-ACGAN's objective functions (3) and (4) as a single minimax loss $L_{age-ACGAN}$ and substituting it into (2), we obtain the objective function of Age-ACP2P:

$$G^* = \arg\min_{\mathbf{G}}\max_{\mathbf{D}} L_{Age-ACGAN}(G, D) + \lambda_{L1} L_{L1} \tag{5}$$

Finally, we incorporate binary cross-entropy (BCE) loss into our combined loss function (5) to get the final objective function:

$$G^* = \arg\min_{\mathbf{G}}\max_{\mathbf{D}} L_{age-ACGAN}(G, D) + \lambda_{L1} L_{L1} + \lambda_{BCE} L_{BCE} \tag{6}$$

The two λs in our final objective function control the weighting of reconstruction and segmentation losses. If λ_{BCE} is 0, we get Age-ACP2P's objective function.

3. EXPERIMENTAL RESULTS

To display our proposed framework's ability to improve reproductive organ segmentation performance, we design an experiment with a data set containing 64 CT volumes collected by the Medical College of Wisconsin. Out of the 64 CT volumes, 24 of them are from female patients containing the uterus and the rest are from male patients containing the prostate. Each image volume is zero-mean whitened ($\mu = 0, \sigma = 1$) and center-cropped

Method	Uterus	
	DSC	HD
CFG-SegNet (Ours)	**0.724±0.0413**	**0.709±1.56**
CE-Net	0.706±0.0415	1.08±1.67
U-Net	0.697±0.0419	1.09±1.92

Table 1. Mean uterus segmentation results with our proposed CFG-SegNet, CE-Net and U-Net. The values shown are the average results of the 4-fold cross-validation experiment. Best results are highlighted in bold.

Method	Prostate	
	DSC	HD
CFG-SegNet (Ours)	**0.929±0.200**	**0.338±0.905**
Attention U-Net	0.925±0.195	0.414±1.21
U-Net	0.923±0.167	0.390±1.01

Table 2. Mean prostate segmentation results with our proposed CFG-SegNet, Attention U-Net and U-Net. The values shown are the average results of the 4-fold cross-validation experiment. Best results are highlighted in bold.

around the region containing the organ. Image volumes are uniform in size ($200 \times 256 \times 256$), and each slice is either a positive (uterus or prostate present) or a negative (no uterus and no prostate) sample. We perform 2.5D segmentation with CFG-SegNet, in which each training batch contains 2D slices. The produced segmentation mask from each slice is then grouped by patient to produce a volumetric segmentation mask for evaluation. We compute the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) averaged across 4-fold cross validation.

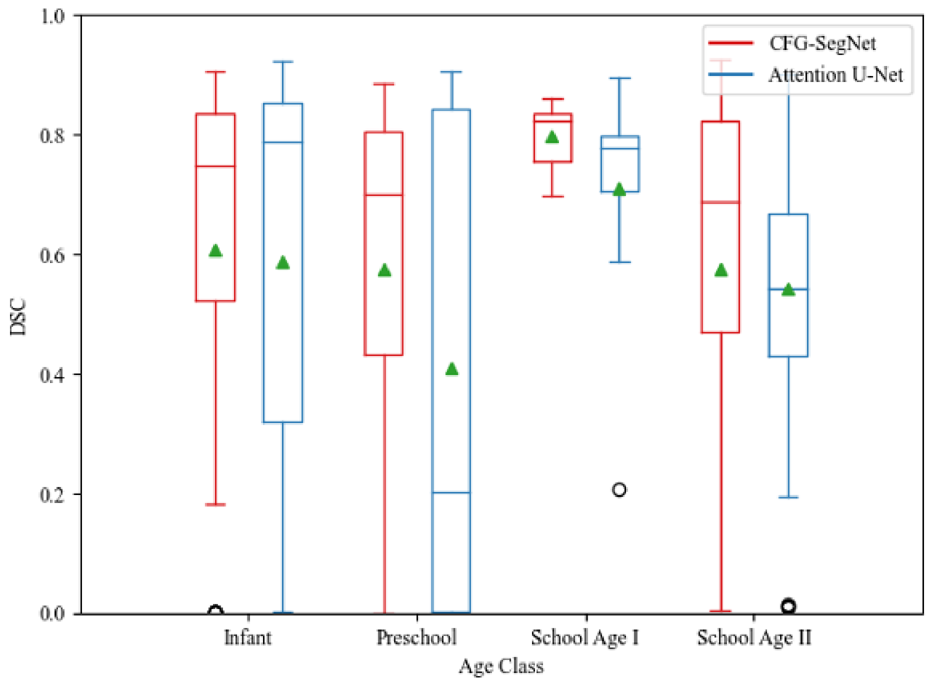

Figure 2. Paired class-wise boxplot of CFG-SegNet and Attention U-Net prostate segmentation results (center cropping). CFG-SegNet is better than Attention U-Net in all four pediatric age classes.

Since CFG-SegNet uses age information to enhance the quality of segmentation, we divide our patients into 6 age classes. These age classes are: infant (ages 1 to 3), preschool (ages 4 to 6), school age I (ages 7 to 9), school age II (ages 10 to 12), adolescent I (ages 13 to 15) and adolescent II (ages 16 to 17). For our uterus segmentation experiment, we compare our proposed method to the state-of-the-art Context Encoder Network (CE-Net) and the U-Net. CE-Net has achieved state-of-the-art segmentation results on multiple medical imaging datasets including lung CT segmentation. Its success can be attributed to the addition of dense atrous convolution (DAC) and residual multi-kernel pooling (RMP) blocks, which aim to better preserve spatial information than the traditional U-Net. We use Attention U-net,[13] a self-attention variant of the U-Net which uses attention gates (AG) in place of CE-Net in our prostate segmentation experiment. Attention U-Net is previously used to segment the pancreas in 150 abdominal CTs, where it outperforms several state-of-the-art pancreas segmentation algorithms such as the Multi-Model 2D FCN and Hierarchical 3D FCN. Each of the 3 networks in our uterus and prostate segmentation experiments is trained for a total of 50 epochs, and the best validation weights are saved for testing. As shown in both Table 1 and Table 2, the resulting mean DSC and HD values with 4-fold cross-validation demonstrate the superiority in uterus and prostate segmentation performance of the proposed method.

Figure 2 is a paired class-wise boxplot which summarizes the segmentation results (DSC) of our proposed CFG-SegNet and Attention U-Net. DSC values of background slices (negative samples that do not contain the prostate label) are excluded from the plot since both networks perform well on identifying true negatives. As shown in figure 3, CFG-SegNet consistently performs better than Attention U-Net across all four pediatric age classes.

Qualitative evaluation of our experiment also shows that the proposed method is able to generate uterus and prostate segmentation masks with above-average accuracy, particularly for younger patients that fall within the preschool class. As visible in the results shown in Figure 2, CFG-SegNet is able to generate reasonably-shaped segmentation masks for two 5-year-old patients under 50 epochs. The segmentation quality of both the CE-Net and U-Net are comparatively lower. This clearly shows the advantage of conditional feature generation as CFG-SegNet is able to produce high-quality reproductive organ segmentation masks for classes that are generally underrepresented in the training data set.

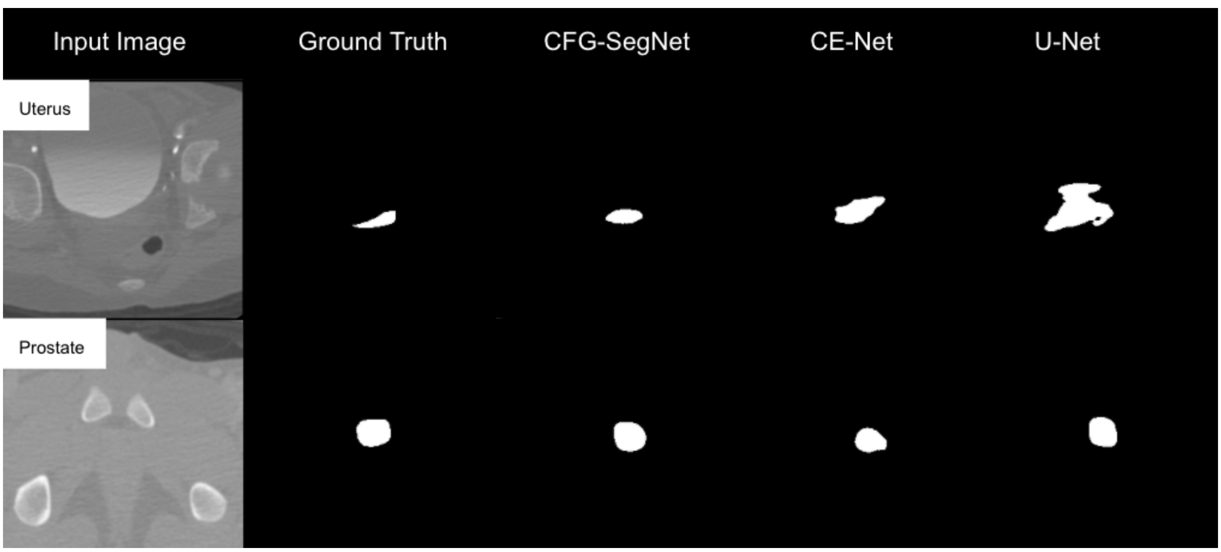

Figure 3. Qualitative results of uterus (top) and prostate (bottom) segmentation of two 5-year-old patients from our 2.5D segmentation experiment. Despite having fewer training samples available for patients of younger age, our proposed method is capable of generating better segmentation masks for all patients than existing methods. U-Net seems to outperform CE-Net in prostate segmentation on our 5-year-old subject. However, out of all three networks CFG-SegNet still produces the best segmentation masks for both organs.

4. CONCLUSION

In this paper, we present a novel framework for reproductive organ segmentation which uses an Age-ACP2P network to conditionally generate training features to improve segmentation performance. Our method effectively tackles class imbalance issues in training data by combining segmentation with age-conditioned feature generation. Segmentation results of our pediatric abdominal CT data set with few training images for younger patients show the proposed method is capable of segmenting the uterus and the prostate with high accuracy across all age classes when compared to the current state-of-the-art. A possible extension of this work is to combine CFG-SegNet with self-supervised learning. If CFG-SegNet can be further improved to self-supervise, then the empirically-set age class definition would not be necessary. Moreover, Age-ACP2P can also be adapted for volumetric synthesis of 3D latent features. Overall, our work serves as an important baseline study of 2.5D reproductive organ CT segmentation and contains a novel approach to continuously generate image features in order to improve segmentation performance.

REFERENCES

[1] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., and Liu, J., "CE-Net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging* **38**(10), 2281–2292 (2019).

[2] Hesamian, M. H., Jia, W., He, X., and Kennedy, P., "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging* **32**(4), 582–596 (2019).

[3] Rubin, P. and Casarett, G. W., "Clinical radiation pathology as applied to curative radiotherapy," *Cancer* **22**(4), 767–778 (1968).

[4] Pearce, M., Salotti, J., and Little, M., "Radiation Exposure from CT Scans in Childhood and Subsequent Risk of Leukaemia and Brain Tumours: A Retrospective Cohort Study," *Lancet* **380**, 499–505 (01 2012).

[5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*], *NIPS'14*, 2672–2680, MIT Press, Cambridge, MA, USA (2014).

[6] Vesal, S., Ravikumar, N., and Maier, A. K., "A 2d dilated residual u-net for multi-organ segmentation in thoracic CT," *CoRR* **abs/1905.07710** (2019).

[7] Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E., and Yuille, A., "Abdominal multi-organ segmentation with organ-attention networks and statistical fusion," *Medical image analysis* **55**, 88–102 (2019).

[8] Shahedi, M., Dormer, J. D., T T, A. D., Do, Q. N., Xi, Y., Lewis, M. A., Madhuranthakam, A. J., Twickler, D. M., and Fei, B., "Segmentation of uterus and placenta in MR images using a fully convolutional neural network," *Proceedings of SPIE–the International Society for Optical Engineering* **11314**, 113141R (Feb. 2020). Edition: 2020/03/16.

[9] Karimi, D., Samei, G., Shao, Y., and Salcudean, S., "A deep learning-based method for prostate segmentation in t2-weighted magnetic resonance imaging," (2019).

[10] Kan, C. N. E., Maheenaboobacker, N., and Ye, D. H., "Age-conditioned synthesis of pediatric computed tomography with auxiliary classifier generative adversarial networks," *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (2020).

[11] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-Image Translation with Conditional Adversarial Networks," *CoRR* **abs/1611.07004** (2016).

[12] Odena, A., Olah, C., and Shlens, J., "Conditional Image Synthesis With Auxiliary Classifier GANs," in [*ICML*], (2016).

[13] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. C. H., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D., "Attention u-net: Learning where to look for the pancreas," *CoRR* **abs/1804.03999** (2018).