PET-Train: Automatic Ground Truth Generation from PET Acquisitions for Urinary Bladder Segmentation in CT Images using Deep Learning

Christina Gsaxner, Birgit Pfarrkirchner, Lydia Lindner, Antonio Pepe, Peter M. Roth, Jan Egger
Inst. of Computer Graphics and Vision
Graz University of Technology
Graz, Austria

Christina Gsaxner, Jürgen Wallner, Jan Egger
Department of Maxillofacial Surgery
Medical University of Graz
Graz, Austria

Christina Gsaxner, Birgit Pfarrkirchner, Lydia Lindner. Antonio Pepe, Jürgen Wallner, Jan Egger
Computer Algorithms for Medicine Laboratory
Graz, Austria

***Abstract*—In this contribution, we propose an automatic ground truth generation approach that utilizes Positron Emission Tomography (PET) acquisitions to train neural networks for automatic urinary bladder segmentation in Computed Tomography (CT) images. We evaluated different deep learning architectures to segment the urinary bladder. However, deep neural networks require a large amount of training data, which is currently the main bottleneck in the medical field, because ground truth labels have to be created by medical experts on a time-consuming slice-by-slice basis. To overcome this problem, we generate the training data set from the PET data of combined PET/CT acquisitions. This can be achieved by applying simple thresholding to the PET data, where the radiotracer accumulates very distinct in the urinary bladder. However, the ultimate goal is to entirely skip PET imaging and its additional radiation exposure in the future, and only use CT images for segmentation.**

***Index Terms*—Deep Learning, Medical Imaging, Segmentation, PET/CT, Urinary Bladder.**

I. Introduction

Automatic medical image segmentation is known to be one of the more complex problems in image analysis [1]. However, segmentation is often the first step in a computer-aided detection pipeline and therefore incorrect segmentation affects any subsequent steps heavily. To this day, delinearation is often done manually or semi-manually [2] [3], which is a tedious task, since it is time consuming and requires a lot of empirical knowledge. Furthermore, the process of manual segmentation is prone to errors and not reproducible, which emphazises the need for accurate, automatic algorithms. One up-to-date method for automatic image segmentation is the usage of deep neural networks. In the past years, deep learning approaches have made a large impact in the field of image processing and analysis in general, outperforming the state of the art in many visual recognition tasks, e.g., in [4]. Artificial neural networks have also been applied successfully to medical image processing tasks such as segmentation. Therefore, in this paper we propose an approach for fully automatic urinary bladder segmentation in CT images using deep learning and further propose a novel course of action to automatically generate the ground truth labels from PET acquisitions to train neural networks.

II. Related Work

A. Automatic Urinary Bladder Segmentation

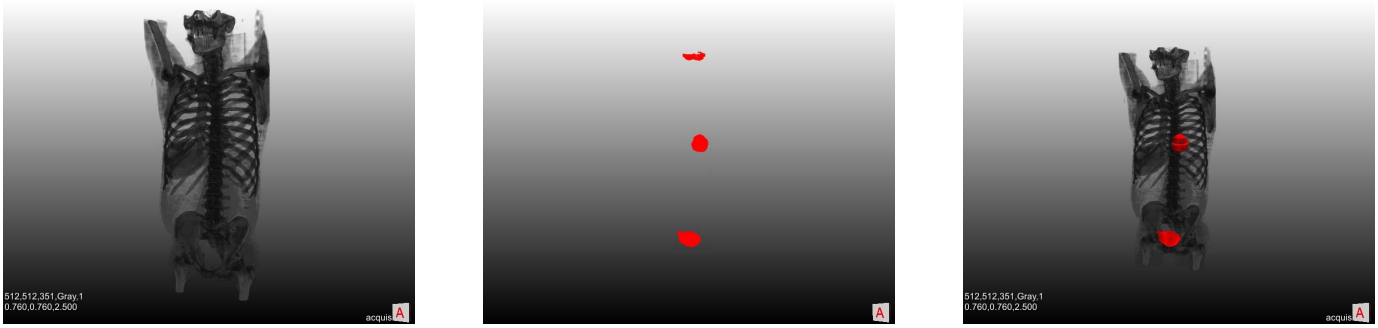
Even though urinary bladder segmentation is still mostly done manually, many different techniques have been attempted in the past years for automatic segmentation of the bladder in CT images. Earlier concepts used deformable models to delineate the urinary bladder wall in 3D [5] [6]. Shi et al. [7] proposed a three step algorithm for segmentation of the urinary bladder. A segmentation attempt using convolutional neural networks for patchwise segmentation was made by Cha et al. in [8] and [9]. The idea of using combined PET/CT scans for generating a ground truth from PET images to train an automatic segmentation algorithm, as introduced in this paper, is, to the best of our knowledge, a new approach. Furthermore, advanced deep learning techniques for semantic segmentation, relying on fully convolutional neural networks and upsampling of coarse feature maps instead of patch-wise classification, have not yet been applied to the problem of urinary bladder segmentation in CT images, as far as we know.

B. Neural Networks for Image Segmentation

Fully Convolutional Networks for Semantic Segmentation (FCNs) were introduced by Long et al. in [10]. They adapt classification networks to segmentation networks by transforming fully connected layers into convolution layers. To map the coarse outputs to dense predictions, upsampling layers are proposed. For upsampling, skip connections are used to utilize features from different layers in the network, which have different scales. Since shallower layers preserve more spatial information, finer details from the original image are kept. Their best performing architecture is called FCN-8s.

Atrous Convolution for Semantic Segmentation as proposed in [11] follows a different approach to deal with the problem

978-1-5386-5724-9/17/$31.00



(a) CT data (b) PET data (c) combined PET/CT

Fig. 1: 3D image data obtained from CT, PET and combined PET/CT of the torso and part of the head. While CT data in (a) shows important anatomical structures (mainly bones), the contrast for soft tissue, in example in the abdominal region, is poor. PET data in (b) only shows metabolical active regions, without providing anatomical context, making it impossible to accurately localize lesions. In the co-registered PET/CT scan in (c), it is possible to properly assign active regions anatomically.

of downsampled feature maps resulting after a classification network. They utilize atrous convolution, which is convolution with upsampled filters. By omitting pooling layers and replacing convolutional layers in classification networks with atrous convolution layers, the resolution of feature maps of any layer within a CNN can be controlled. Furthermore, the receptive field of filters can be enlarged without increasing the number of parameters.

III. Methods

A. Dataset and Preprocessing

The implemented methods were trained and tested on the Reference Image Database to Evaluate Therapy Response (RIDER) [12]. The RIDER PET/CT dataset provides serial PET/CT lung cancer patient data and consists of a total of 65 scans. As radiotracer, fluorine-18-labelled fluorodeoxyglucose was used in all PET scans. The public database can be downloaded from the National Cancer Imaging Archive (NCIA) at [13]. After removing patient data with low contrast and high noise from the RIDER PET/CT database, a total of 29 patient datasets were obtained. From them, a total amount of 845 CT image slices around the urinary bladder were extracted for further usage.

B. Generation of Image Data

To increase the amount of available image data, data augmentation was applied to the training images. Affine transformations, specifically rotation and scaling, are applied to CT images as well as the masks. Furthermore, zero-mean Gaussian noise was added to CT images. The generation of segmentations of the urinary bladder as a ground truth for training a deep neural network is performed by using combined positron emission tomography-computed tomography scans. The most commonly used radiotracer, fluorine-18-labelled fluorodeoxyglucose (^{18}F-FDG), accumulates in the urinary bladder, therefore, the bladder always shows up in these PET scans. Contrary to CT, PET images exhibit high contrast and are therefore comparably easy to segment automatically, which is shown in Figure 1. PET images are automatically segmented using a simple thresholding approach. The necessary steps for data generation were implemented in the medical imaging framework MeVisLab (www.mevislab.de) [14] [15]. The corresponding Macro module and Python source code is freely available under: https://github.com/cgsaxner/DataPrep_UBsegmentation [16].

We split the available 845 image slices into training and testing data, resulting in 630 images for training and 215 images reserved for testing. The datasets for training were processed with the MeVisLab network to obtain individual, augmented CT slices as well as corresponding ground truth labels. A total amount of 34,020 augmented images and labels were obtained. To additionally be able to analyse the effect of data augmentation, a training dataset containing only the 630 unaugmented images and labels as well as a dataset with only transformed image data (without noise), consisting of 17,010 augmented CT images, was put together.

C. Image Segmentation using Deep Neural Networks

Algorithms for image segmentation using deep neural networks were implemented using TensorFlow 1.3 under Python 3.5. For our segmentation networks, we built on the approach by Pakhomov et al. [17].

We used pre-trained classification models (VGG-16 and ResNet-V2) which are adapted for semantic segmentation for our segmentation task using the approaches by Long et al. in [10] and Chen et al. in [11]. The models were trained on the ILSVRC-2012 dataset for image classification. Our first network definition, called FCN, is based on FCN-8s and uses a pre-trained version VGG-16. The second model definition, called upsampled ResNet, utilizes a pre-trained version of ResNet-V2 with 152 layers and atrous convolution to perform image segmentation. Upsampling of feature maps is performed using transposed convolution. Training was performed using

Image resolution	Network model	Training data	Mean TPR (%)	Mean TNR (%)	Mean DSC (%)	Mean HD (pixel)
256×256	FCN	no augmentation	82.7	99.9	77.6	6.9
		transformed images	**85.0**	**99.9**	**80.4**	**6.1**
		fully augmented images	79.2	99.9	77.6	6.7
	Upsampled ResNet	no augmentation	80.7	99.9	73.5	7.9
		transformed images	82.5	99.9	76.9	6.3
		fully augmented images	79.7	99.9	76.7	7.7
512×512	FCN	no augmentation	80.9	99.9	77.6	13.3
		transformed images	**83.1**	**99.9**	**81.9**	**11.9**
	Upsampled ResNet	no augmentation	68.7	99.9	71.1	23.9
		transformed images	86.5	99.8	67.1	16.9

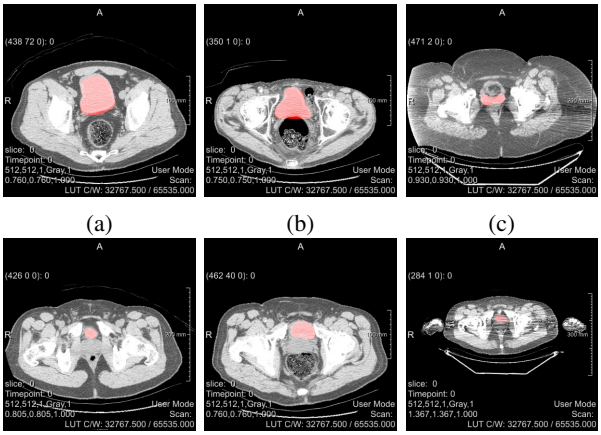
TABLE I: Segmentation evaluation results. This table compares evaluation metrics for FCN and upsampled ResNet architectures trained using unaugmented training data, transformed training data (rotation, scaling) and fully augmented training data (transformations and zero-mean Gaussian noise).

an ADAM optimizer over 34,020 iterations with a batch size of one. We trained on a server equipped with a NVIDIA Tesla K20Xm with 5 GB memory size. Testing was executed using a NVIDIA GeForce GTX 960 with 2 GB of memory. We trained and tested our networks with images of original resolution (512×512) and downsampled images (256×256).

IV. Results and Evaluation

A. Generation of Training and Testing Data

To illustrate the agreement between the ground truth labels obtained from thresholding the PET data and corresponding CT images, overlays between CT images and generated labels were produced. Some examples of these overlays can be seen in Figure 2. Agreement between the binary masks generated from PET image data and CT image data was overall good, with slices showing a large area surface of the urinary bladder yielding especially accurate results. It can be observed that the shape, size and position within the image of the urinary bladder is highly varying between datasets, which poses a difficulty for automatic segmentation approaches.



(a) (b) (c) (d) (e) (f)

Fig. 2: Examples of overlays between CT data and generated ground truth labels. The underlying CT images are shown in greyscale, while the ground truth labels obtained from PET segmentation are added in red.

B. Segmentation Results

Table I shows several metrics for evaulating our models trained with different training datasets. True positive rate (TPR), true negative rate (TNR), Dice coefficient (DSC) and Hausdorff distance (HD) are reported. It can be seen that we achieved best results with our FCN model trained with images that underwent data augmentation in the form of transformations only. The addition of rotations and scaling to the original data improves the segmentation results significantly. The addition of noise on the other hand does not improve the performance of our proposed networks, but in fact worsens it. A possible explanation for this is that the noisy images are too similar to the original ones, and therefore, the models start to overfit. Another reason might be that the added zero mean Gaussian noise is not meaningful in the presented context, and therefore the network learns spurious patterns.

Figure 3 shows several representative examples of the obtained segmentation results for images downsampled to a resolution of 256×256. For better illustration, original image data was overlaid with the contour of the ground truth in green as well as the prediction made by our deep networks in red. Sample (a) shows that for input images with good soft tissue contrast, a large, homogeneous area surface and a regular shape of the urinary bladder, all models perform reasonably well. For images with lower contrast, as seen in sample (b), FCN architecture produces significantly better results than ResNet architecture. Sample (c), however, shows that the upsampled ResNet adapts better to very irregular shapes of the urinary bladder. Sample (d) presents one drawback

of our approach: The ground truth label does not fit the underlying structure of the urinary bladder perfectly in this case. Even though some of our models produce reasonalble qualitative results in this case, evaluation scores will be rather low. Therefore, the presented evaluation metrics might not always represent the actual performance of the models that well and qualitative results should also be considered when assessing segmentation performance. Structures with a small area surface often pose a difficulty for automatic segmentation approaches. However, sample (e) shows that the proposed models are able to identify those small structures.

V. Discussion and Conclusion

Small dataset sizes and the lack of annotations due to the complexity of manual segmentation are big limitations to deep learning applications in medical image processing. There have been some attempts to overcome this obstacle, including the usage of existing tools to create labels for pre-training [18], the usage of sparse annotations [19] or the generation of artificial data with Generative Adversarial Networks (GANs) [20]. We introduce a new solution to this problem by generating training and testing datasets for deep learning algorithms by exploiting ^{18}F-FDG accumulation in the urinary bladder to produce ground truth labels, and by the application of data augmentation to enlarge a small dataset. We showed that when making use of combined PET/CT data, an automatic low-level segmentation of PET image data can be used to attain a fully automatic, high-level segmentation of corresponding CT data. We achieved satisfying segmentation results with a very small image database and completely without the usage of manually segmented image data. Since combined scanners are becoming increasingly more widespread, it can be expected that more, larger PET/CT image databases will be available in the future. Our approach presents a promising tool for automatically processing such databases and can be generalized to all applications of combined PET/CT or combined PET/MRI, such as cancerous tumours in the lung or in the head and neck area, just to name a few.

We used the generated data to train and test two different well-known deep learning models for semantic image segmentation. Our qualitative results show that the proposed segmentation methods can accurately segment the urinary bladder in CT images and are in many cases more accurate than the ground truth labels obtained from PET image data. It is shown that the used FCN architecture generally performs better in terms of evaluation metrics than the proposed ResNet architecture. We achieved the best segmentation performance with our FCN network which was trained with transformed image data. Future work could include the implementation of post-processing algorithms. In many publications, including Che et al. [11], fully-connected conditional random fields are used to accurately recover object boundaries that are smoothed within the deep neural network. In our case, this might especially improve performance in cases were the urinary bladder has irregular, distinct shapes.

We demonstrated that training data augmentation in the form of transformations, like rotation and scaling, can significantly improve the performance of segmentation networks, however, the addition of zero-mean Gaussian noise to the training data did not result in an enhanced performance in our case. Subsequent work could go into further exploring the effects of data augmentation on the segmentation results, by generating even bigger augmented datasets and by applying different noise types to the original image data.

ACKNOWLEDGMENT

This work received funding from the Austrian Science Fund (FWF) KLI 678-B31: “enFaced: Virtual and Augmented Reality Training and Navigation Module for 3D-Printed Facial Defect Reconstructions” (Principal Investigators: Jürgen Wallner and Jan Egger), the TU Graz Lead Project (Mechanics, Modeling and Simulation of Aortic Dissection) and CAMed (COMET K-Project 871132) which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Federal Ministry for Digital and Economic Affairs (BMDW) and the Styrian Business Promotion Agency (SFG).

References

[1] Isaac Bankman. *Handbook of Medical Image Processing and Analysis*. Academic Press, 2008.

[2] Jan Egger et al. GBM volumetry using the 3D slicer medical image computing platform. *Scientific reports*, 3:1364, 2013.

[3] Jan Egger. Refinement-cut: user-guided segmentation algorithm for translational science. *Scientific reports*, 4:5164, 2014.

[4] Alex Krizhevsky et al. Imagenet: Classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[5] Benjamin Haas et al. Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies. *Physics in Medicine and Biology*, 53(6):1751, 2008.

[6] Maria Jimena Costa et al. Automatic segmentation of the bladder using deformable models. In *Biomedical Imaging: IEEE International Symposium*. IEEE, 2007.

[7] Feng Shi et al. Automatic segmentation of bladder in ct images. *Journal of Zhejiang University-Science A*, 10(2):239–246, 2009.

[8] Kenny H Cha et al. Comparison of bladder segmentation using deep-learning convolutional neural network with and without level sets. In *SPIE Medical Imaging*, 2016.

[9] Kenny H Cha et al. Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Medical Physics*, 43(4):1882–1896, 2016.

[10] Jonathan Long et al. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[11] Liang-Chieh Chen et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *Transactions on Pattern Analysis and Machine Intelligence: 40(4), 834–848*, 2018.

[12] Samuel G Armato et al. The reference image database to evaluate response to therapy in lung cancer (RIDER) project: A resource for the development of change-analysis software. *Clinical Pharmacology & Therapeutics*, 84(4):448–456, 2008.

[13] National cancer imaging archive. http://www.cancerimagingarchive.net/. Accessed: 2017-10-09.

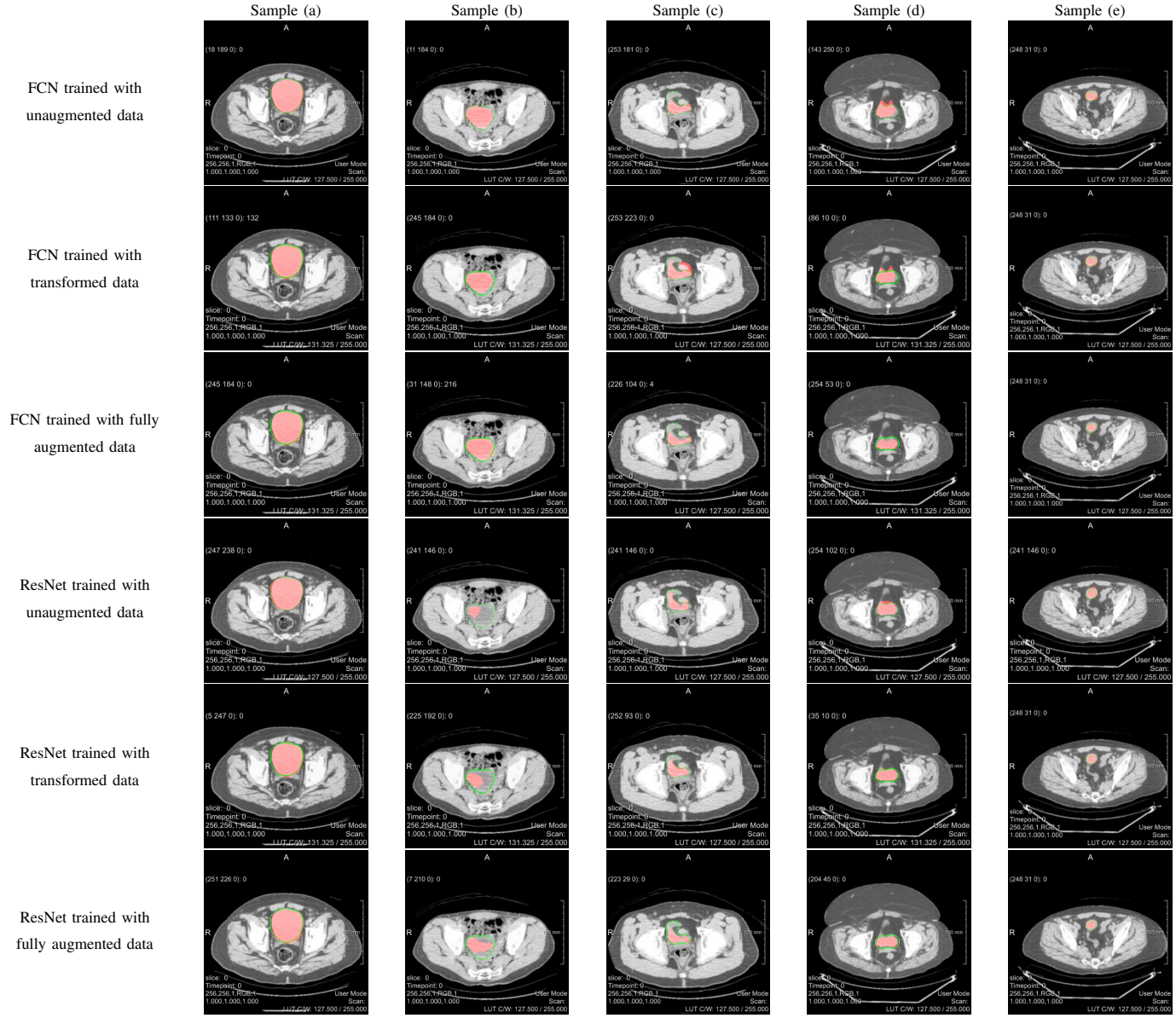


Fig. 3: Qualitative segmentation result overlays for images scaled to 256×256. Ground truth labels are shown by the contours in green, the predictions made by the deep learning models are overlaid in red.

[14] Jan Egger et al. HTC Vive MeVisLab integration via OpenVR for medical applications. *PloS one*, 12(3):e0173972, 2017.

[15] Jan Egger et al. Preoperative measurement of aneurysms and stenosis and stentsimulation for endovascular treatment. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 392–395. IEEE, 2007.

[16] Christina Gsaxner et al. Exploit 18 F-FDG enhanced urinary bladder in PET data for deep learning ground truth generation in CT scans. In *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10578, page 105781Z. International Society for Optics and Photonics, 2018.

[17] Daniil Pakhomov et al. Deep residual learning for instrument segmentation in robotic surgery. *arXiv preprint arXiv:1703.08580*, 2017.

[18] Abhijit Guha Roy et al. Error corrective boosting for learning fully convolutional networks with limited data. *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2017.

[19] Özgün Çiçek et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.

[20] Thomas Neff et al. Generative adversarial network based synthesis for supervised medical image segmentation. *Joint OAGM and ARW Workshop*, 2017.