Deep Rectum Segmentation for Image Guided Radiation Therapy with Synthetic Data

Dimitrios Mallios
School of Electronics & Computer Science
University of Southampton
Southampton SO17 1BJ, United Kingdom
dm1n19@soton.ac.uk

Xiaohao Cai
School of Electronics & Computer Science
University of Southampton
Southampton SO17 1BJ, United Kingdom
x.cai@soton.ac.uk

***Abstract*—Image guidance nowadays is a crucial component for doctors to facilitate the design of the planning radiation therapy dosage. The delineation of soft organs in the planning phase and during the radiation therapy is crucial for the treatment procedure. Deep Learning (DL) flourishes, presenting state-of-the-art results in challenging computer vision tasks; however, the lack of annotated data hardens the research advancements for medical applications. The research in this paper develops DL approaches for the segmentation of organs-at-risk and specifically from images retrieved from a computed tomography system during the radiation treatment of each patient. The proposed approaches are based on convolutional neural architectures trained with only a couple of thousand images, and can also be trained online, showing its learning ability from new patients. The lack of annotated data is also addressed with synthetic data generated by a modified GAN. Experimental results demonstrate the excellent performance of the proposed approaches in rectum segmentation task.**

***Index Terms*—image segmentation, deep learning, medical imaging, image guided radiation therapy**

I. Introduction

One of the most common approaches to treat many cancer types involves Image Guided Radiation Therapy (IGRT) [1], [2]. IGRT involves many complex procedures, such as proper selection of the radiotherapy treatment process and segmentation of the target volumes and organs-at-risk. The treatment process involves mainly two steps, i.e., the planning phase and the treatment phase. In the planning phase the computed tomography (CT) operates at a lower voltage which exposes the patient to high radiation in order to provide better organ imaging to doctors and produce the treatment plan. During the treatment phase, the patient receives the planned radiation dosage while the CT scanner checks organs' positional changes. The organs, and specifically the rectum which is the main focus of this paper, can change shape from day-to-day; therefore, the delineation of organs is crucial during the treatment phase.

The manual segmentation of the organs presents a very time consuming task for the doctors and involves intra and inter-observer variability; thus auto-segmentation algorithms have been proposed in the field to facilitate the annotation process and decrease the annotation time substantially. The simplest method to segment an organ from medical images is by setting threshold boundaries among e.g. the soft tissues and the bones [3], utilizing the information from the pixel intensities. Note that a proper initialization is needed in order the algorithms to work better, and the process can be assisted by edge detection methods to detect the boundaries more effectively. However, the variations in brightness across many images harden the work of finding a proper threshold. Apart from the simple thresholding methods, model-based active contour methods are also popular, i.e., the snakes method [4], its relevant like [5], [6], and the ones exploiting convex optimisations like [7], [8]. These algorithms have been introduced in many works effectively e.g. image guided radiotherapy in [2], [9], [10], [11], within which the registration techniques were also employed to first obtain a valid initialization. One of the main disadvantages of the model-based methods is that the computation time is cost during the treatment which is improper when the segmentation result is in urgent need.

The idea of utilizing deep learning (DL) methods like the U-Net [12] and the Generative Adversarial Networks (GAN) [13] arose recently since both the algorithm and computational resources were greatly improved, and the data along with the annotations were substantially enough to train DL models (e.g. the structured data sets appear in the field of medical imaging in the Grand Challenge[1]). Artificial intelligence (AI) has the most of attention in research while its applications spread almost in every technical and non technical field. One of the most promising applications is in the Healthcare sector where AI aims to revolutionize the diagnosis and the treatment process [14].

In this paper we develop DL methods by utilizing a small set of CT data to segment organs-at-risk during the treatment process. The developed approaches are constructed by DL techniques including convolutional neural networks and GANs, which can be trained using the realistic data sets and newly generated synthetic data sets simultaneously. Moreover, an online training strategy is also proposed by adding new patient's data in the training process to improve the training efficiency and tackle the lack of annotated data. Thorough experiments are being conducted to validate the excellent performance of the proposed approaches. The techniques developed in this paper can expedite the organs segmentation

[1]https://grand-challenge.org/challenges/

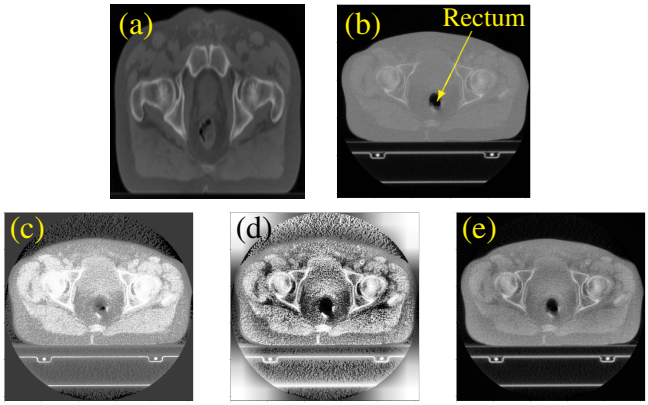

Fig. 1. (a)–(b): a slice sample from the low voltage planning scans and the high voltage treatment scans, respectively; (c)–(e): contrast enhancement of the image in (b) by using the standard HE method, the adaptive HE method, and the CLAHE method, respectively.

in IGRT in medical imaging and many others.

II. Data

The data set used in this paper has been retrieved from ten patients who have been diagnosed with prostate cancer. The data were gathered and annotated for the renowned VoxTox Research project [2]. The data are image slices captured from a CT machine (TomoTherapy HiArt megavoltage CT (MVCT)). The slices present the pelvic area from the patient depicting the organs in the area, see Fig. 1 (a)–(b). The exact delineation of the rectum from the data is extremely challenging. The tissue on the rectum's boundaries along with the air inside the rectum can easily mislead predictions. One more extreme challenge in this problem and in medical imaging is the lack of training data.

The data set contains the planning kilovoltage CT (KVCT) IG scans along with the treatment (delivered) MVCT IG scans throughout treatment for each patient. The planning KVCT scans are 272x272 pixels per slice where the size of the pixel is 1.953 mm with 3 mm thickness. The scanning slices present the entire rectum, from the sigmoid junction of the rectum to the inferior part imaging both the ischial tuberosities. The MVCT scans are 512x512 pixels per slice, and each pixel has a size of 0.754 mm while the thickness of each slice is 6 mm. The MVCT scans are limited to approximately 8–25 slices in order to satisfy the protocols and minimise the additional dose and the time during treatment for the prostate. Thus, during treatment the rectum was partially exposed, imaging only a small part. The rectum in both the planning and treatment scans was manually segmented by an experienced doctor for the training and test purpose. The manually segmented rectum in the planning scans is exceptionally detailed (approx. 35 slices) while that in the treatment is barely half of the planned (approx. 12 slices). In IGRT, the main focus is the segmentation on the treatment scans.

In order to exploit the data set, histogram equalization (HE) algorithms [15], [16], [17] are used as pre-processing to transfer the CT slices in HU to the grayscale format with range [0, 1], with contrast enhancement effect. The standard HE [15] usually works only under certain circumstances, i.e. when the distribution of the pixel intensities of the image is smooth. Otherwise, HE tends to increase the noise in the image. There are some other types of HE which are able to overcome this problem, e.g. adaptive HE [16] and CLAHE [17]. Fig. 1 (c)–(e) presents the HE results using the standard HE, adaptive HE and CLAHE on one slice of the treatment scans, from which we see the CLAHE algorithm achieves the best quality, with kernel size of the neighbourhood pixels set to 32x32. Therefore, the CLAHE algorithm will be used to transfer the whole CT slices in HU to the grayscale format for the subsequent segmentation task.

III. Preliminary

In this section we recall the work in Fully Convolutional Networks (FCNs) and GAN, which will be used in the construction of our approach.

A. Fully Convolutional Networks

FCNs present a basis nowadays for the image segmentation task when dealing with segmentation of tumours and organs from CT or MRI images. The accuracy of the CNN generally relies on the training phase, which requires an enormously massive set of data [18]. The need for using this technology in the healthcare sector is demanding, especially in the field of cancer detection in radiation therapy [19]. In 2015 a fully CNN encoder-decoder architecture called U-Net was proposed outperforming most image segmentation methods by utilizing a smaller amount of trainable parameters [12] compared to other FCN e.g. [20]. The performance of the networks depends on the assumption that the objects of interest were annotated correctly by the annotators, whereas this assumption is not always true in practice. An elegant solution is to annotate the scans using multiple evaluators and include them in the training process. Then a network architecture in [21] can be used to produce a distribution of segmentations for a given input image. Thus, multiple plausible and accurate hypothesis can then be generated to aid the doctor's work.

In biomedical imaging, the most common architecture used in segmentation tasks is the U-Net architecture. It is composed of two parts, namely the encoder and the decoder. The encoder part captures and encodes the information from the image by utilizing successively convolutional operations followed by max pooling. U-Net can also handle more channels in the input instead of the unit channel. Usually, this refers to the RGB colour image. This type of network is usually mentioned as 2.5D in the literature because it utilises more contextual information from the image slices [22].

B. Generative Adversarial Networks

Another variation in segmentation methods has been based on GAN [13]. GANs present a new concept in the training of neural networks by putting two individual networks – namely the generator and the discriminator – to compete each other in a minimax game. The generator tries to learn the distribution of

the real data so as to reproduce them and fool the discriminator which acts as a critic. GANs are useful to generate synthetic data to overcome the lack of training data set dilemma [23].

The conditional GAN (cGAN) [24], a variation of the original GAN, was firstly introduced in feed forward networks. An architecture was designed in [25] based on cGAN and can be adapted to segmentation tasks. The generator network is simply a U-Net based network while the discriminator is a 'patchGAN' network [26] which evaluates the generator's output by evaluating KxK local patches if they are fake or real. The 'patchGAN' discriminator according to [26] seems more effective as it helps the generator to produce higher quality samples and penalizes the structure of the generator's output locally. The cycle-GAN model was introduced in [27] for multi-organ nuclei segmentation in histopathology images, which also inspired us with the construction of the approach in this paper.

IV. Proposed Methods

In this section we present our realistic-synthetic online approach for soft organ (e.g. rectum) segmentation in IGRT. The main goal is to utilize as few annotated CT scans from patients as possible to achieve the best satisfying segmentation results in new patient's scans.

Fig. 2 presents the main workflow of the approach, which contains two branches after the DICOM data are pre-processed to retrieve the CT slices. The first branch (top) shows the data being fed into the FCN network (where the U-Net is used in our experiments) in order to be trained with a predefined loss function. The second branch presents the synthetic data generation module (where cGAN is used in our experiments), which takes a manual segmentation of the rectum as input and generates a synthetic image. The generated synthetic image is also fed into the FCN network in conjunction with the realistic samples. Note, importantly, that the second branch can address the issue of the lack of training data. The FCN network is then trained and outputs a segmentation mask of the rectum, followed by a post-processing step in order to remove outliers from the output segmentation mask and just keep the largest contour which presents the rectum position. It is worth emphasising that the post-processing is not utilized for training as it does not present a differentiable module.

This proposed realistic-synthetic pipeline can be trained with single slice (1-channel) images straightforwardly. It can also be trained firstly with only the first branch and then, along with the second branch and the data generation. With respect to training with three adjacent slices, only the first branch is utilized. Training with three adjacent slices and predicting the middle segmentation mask can exploit additional spatial information from the previous and the next slice leveraging the relevant position of the rectum which resides closely among the slices in the spatial axis. If the last part of the rectum appears in the last indexed slice of the data, in other words, the last slice of the rectum does not have a slice after it, in order to form three slices for training, the slice before the first slice of the rectum will be used.

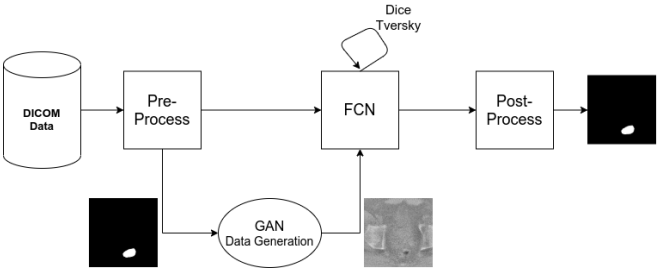

Fig. 2. Realistic-synthetic approach for rectum segmentation in IGRT. Dice and Tversky represent two different loss functions. The binary image is the manual segmentation of the rectum, and the grayscale image on the right hand side of the GAN network is a generated synthetic image.

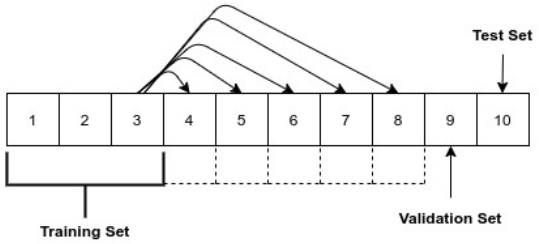

Fig. 3. Online training diagram. The dotted lines present the training data set that is augmented after inference on each patient. The numbers in the blocks represent each patient.

Together with the realistic-synthetic pipeline shown in Fig. 2, an online training manner is also constructed to enhance the approach's learning capability from new CT images retrieved from different patients, see Fig. 3 using ten patients' data we have as an example. In detail, firstly, the first three patients' data are used to train the realistic-synthetic network shown in Fig. 2 for a specific amount of epochs. Then the trained network is used to predict the rectum segmentation masks of the forth patient, and the predicted segmentation masks are added back to the training set. Repeat this training process on each upcoming patient until it reaches the data limit, say to the eighth patient in our case. The trained network is then evaluated on the validation data set (i.e., the ninth patient in our case) to select the best hyper-parameters and tested on the test data set (i.e., the tenth patient).

Before closing this section, we emphasise that the approaches developed above can also be adopted for other segmentation purposes.

V. Experimental Results

The planning scans and treatment scans can be linked via some proper projections observed by e.g. registration techniques. The slices received during the treatment stage which contain the rectum can also be easily identified using the prior knowledge in the planning scans. Therefore, for simplification, the CT scans received during the treatment stage which contain the rectum are utilised to train the proposed approach.

The ten patients' data are separated as follows: five patients containing 2062 samples are used for training, three patients

containing 1080 samples are used only for synthetic image generation in order to avoid biases, one patient containing 379 samples is used for validation, and one patient containing 376 samples is used for test. Moreover, the training was conducted on cropped patches with size of 256x256 instead of the original 512x512 to ignore the irrelevant information and facilitate the learning efficiency. The cropping coordinates were obtained by a thorough inspection of the data set and the position of the rectum in the slices. Data augmentation techniques such as random horizontal flipping and rotation of ± 10 degrees are also involved for the comparison purpose. The augmentation techniques are applied randomly with a probability of 0.5.

The accuracy of the approach is measured quantitatively by utilising the Dice Coefficient, say S_{DC}, defined as

$$S_{\mathrm{DC}} = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \epsilon}, \tag{1}$$

where p_i represents i-th pixel predicted as rectum or not, g_i represents the ground truth of the i-th pixel, and ϵ is close to 0 added to avoid a zero denominator. The Dice similarity coefficient loss [28] say L_{DSC} is used as the loss function to train our approach, i.e., $L_{\mathrm{DSC}} = 1 - S_{\mathrm{DC}}$.

For easy of discussion, we name our realistic-synthetic approach RS-FCN, its first branch R-FCN, and R-$\mathrm{FCN}_{\mathrm{DA}}$ when the data augmentation technique is used with R-FCN. The models in our approaches are trained with a starting learning rate of 10^{-4} and a weight decay of 10^{-4}. See [27] for the hyperparameters used to generate our synthetic images. The code of our approaches is available on GitHub[2].

A. Results

The rectum segmentation quantitative results using the developed methods R-FCN, R-$\mathrm{FCN}_{\mathrm{DA}}$ and RS-FCN are shown in Table I. In particular, these methods are also tested using one and three channels (slices) as the training ways. Note that the RS-FCN method only works with 1 channel setting (the way of extending it for multi-channel is left for future investigation). From Table I, we see that for 1 channel setting, the RS-FCN method achieves the best accuracy, which shows the effectiveness of the synthetic data generated in its second branch using cGAN. In particular, the accuracy it achieved is higher than that of the method R-$\mathrm{FCN}_{\mathrm{DA}}$, which shows the synthetic data is better than the ones generated using data augmentation techniques. The results using the 3 channels setting can generally achieve higher accuracy than using 1 channel setting, which is reasonable because of the usefulness of the spacial information. It is also interesting to notice that the data augmentation technique is not that useful for the 3 channels setting, which might be because the spacial information provides similar/equivalent contribution.

The segmentation performance of the developed methods via visual cross-validation by comparing with the ground truth was also conducted. Fig. 4 presents the results of the method R-FCN with 3 channels setting on three real unseen slices.

[2]https://github.com/dimimal/deepRectumSegmentation

1 channel			3 channels	
R-FCN	R-$\mathrm{FCN}_{\mathrm{DA}}$	RS-FCN	R-FCN	R-$\mathrm{FCN}_{\mathrm{DA}}$
$0.76^{\pm 0.13}$	$0.77^{\pm 0.11}$	$\mathbf{0.79}^{\pm 0.11}$	$\mathbf{0.85}^{\pm 0.10}$	$0.82^{\pm 0.15}$

TABLE I
SEGMENTATION ACCURACY OF R-FCN, R-$\mathrm{FCN}_{\mathrm{DA}}$ AND RS-FCN.

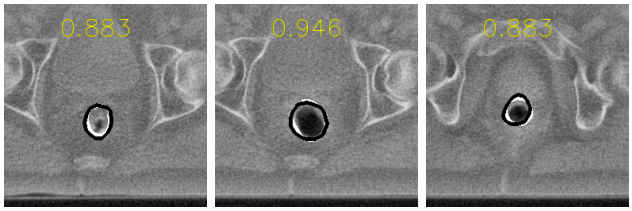

Fig. 4. Rectum segmentation using the developed method R-FCN with 3 channels setting. The boundary in white colour presents the ground truth while the black colour presents the results of our method. The number in each image is the Dice Coefficient.

These three slices possess different features which all lead to segmentation difficulty. For example, the first one in Fig. 4 has a tiny black spot inside while the following two have larger black spots which usually are the gas that resides inside the rectum; note that the rectum intensity (except for the black spot) is very close to its adjacent regions. In Fig. 4, the segmentation results of our method are visualized with the black colour while the ground truth with white. Also, the Dice Coefficient is projected on top of each sample. It clearly shows that our method is able to predict the rectum's position with immense precision.

Finally, we test the developed method RS-FCN equipped with the online training diagram given in Fig. 3. The training follows the way shown in Fig. 3, i.e., start with the data sets of three patients, and then involving the following patients one by one where their rectums are segmented using the pre-trained network. The training started with 20 epochs initially for the three patients, and then the online learning with two epochs for each upcoming patient. The results regarding the Dice Coefficient are presented in Fig. 5. The horizontal axis presents the data sets of each patient, while the numbers on the line depict the number of training samples used at each step. It

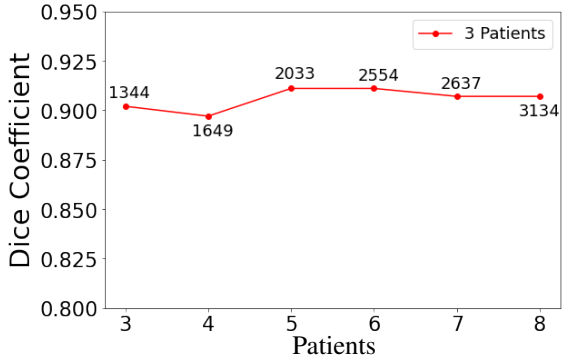

Fig. 5. Rectum segmentation using the developed method RS-FCN equipped with the online training diagram. The numbers on the line depict the number of training samples used at each step.

shows the excellent performance of the developed method RS-FCN equipped with the online training diagram, with achieved Dice Coefficient 0.9 on average, which is comparable with the segmented result delivered by experts manually.

VI. Conclusion

In this paper we developed DL approaches for rectum segmentation in IGRT, presenting the potential of using synthetic images to address the lack of labelled data in reality in order to train very deep models. Experiments were conducted to showcase the excellent performance of the developed approaches, including the online learning strategy which demonstrated the feasibility of learning from new patients on-the-fly. The CT images presented in this paper were gathered explicitly from patients with prostate cancer and the experiments were conducted from the CT images obtained during the treatment phase. The methods developed in the paper however are highly likely to be adopted for other segmentation purposes.

One of the key findings in this paper is the feasibility of generating synthetic data and then introducing them during training large models to obtain satisfactory results. Synthetic data generation shows the potential of closing the gap of learning and the limitation of annotated data in medical applications, which remains a significant issue. In this work, the models with synthetic data were trained with a similar amount of realistic data, and quite promising results were achieved. The online learning at the beginning was crucial to showcase the stability of the model when learning from new patients. This is important due to the variability of the rectum whose shape is deformable.

Based on the above conclusions, the research in the future should be focused on more sophisticated online learning schemes and expand the work on synthetic data generation for 3D slices (or generate multi adjacent slices). This could be achieved with further fine-tuning of GANs. Moreover, there is room for improvement with image pre-processing techniques as the original images usually contain too much noise. As to adapting the proposed approaches in the industrial section for medical purposes, a thorough clinical evaluation is compulsory so that it could be in line with the government's relevant policy.

Acknowledgment

The authors would like to thank the VoxTox research group from Cambridge University for fruitful discussions and providing the practical data used in this paper.

References

[1] C. F. G. Delaney, S. Jacob and M. Barton, "The role of radiotherapy in cancer treatment," *Cancer*, vol. 104, no. 6, pp. 1129–1137, 2005.

[2] N. Burnet, J. Scaife, M. Romanchikova, and S. Thomas, "Applying physical science techniques and cern technology to an unsolved problem in radiation treatment for cancer: the multidisciplinary 'voxtox' research programme," *CERN IdeaSquare Journal of Experimental Innovation*, vol. 1, no. 1, p. 3, 2017.

[3] K. Brock, *Image Processing in Radiation Therapy*. CRC Press, 2016.

[4] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[5] V. Caselles and et al., "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, pp. 61–79, 1997.

[6] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[7] X. Cai, R. Chan, and T. Zeng, "A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 368–390, 2013.

[8] X. Cai, R. H. Chan, C.-B. Schönlieb, G. Steidl, and T. Zeng, "Linkage between piecewise constant Mumford–Shah model and Rudin–Osher–Fatemi model and its virtue in image segmentation," *SIAM J. Sci. Comput.*, vol. 41, no. 6, pp. B1310–B1340, 2019.

[9] L. Shelley, M. Sutcliffe, K. Harrison, and J. Scaife, "Autosegmentation of the rectum on megavoltage image guidance scans," *Biomedical Physics & Engineering Express*, vol. 5, no. 2, p. 025006, 2019.

[10] X. Cai, C.-B. Schönlieb, J. Lee, J. Scaife, H. Karl, M. Sutcliffe, M. Parker, and N. Burnet, "Automatic contouring of soft organs for image-guided prostate radiotherapy," *Radiotherapy and Oncology*, vol. 119, pp. S895–S896, 2016.

[11] J. Scaife, K. Harrison, A. Drew, X. Cai, and et al., "Accuracy of manual and automated rectal contours using helical tomotherapy image guidance scans during prostate radiotherapy," *J Clin Oncol*, vol. 33, p. 94, 2015.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[13] I. Goodfellow, J. Pouget-Abadie, and et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, and et al., Eds., 2014, pp. 2672–2680.

[14] I. Boon, T. Au Yong, and C. Boon, "Assessing the role of artificial intelligence (ai) in clinical oncology: Utility of machine learning in radiotherapy target volume delineation," *Medicines*, vol. 5, no. 4, p. 131, 2018.

[15] P. E. Trahanias and A. N. Venetsanopoulos, "Color image enhancement through 3-d histogram equalization," in *Proceedings., 11th IAPR ICPR*, 1992, pp. 545–548.

[16] S. M. Pizer, E. P. Amburn, and et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355 – 368, 1987.

[17] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. Morgan Kaufmann, 1994, pp. 474–485.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012, pp. 1097–1105.

[19] R. F. Christopher R.Deig, Aasheesh Kanwar, "Artificial intelligence in radiation oncology," *Hematology/Oncology Clinics of North America*, vol. 33, no. 6, pp. 1095–1104, 2019.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038, 2014.

[21] S. Kohl, B. Romera-Paredes, and et al., "A probabilistic u-net for segmentation of ambiguous images," *CoRR*, vol. abs/1806.05034, 2018.

[22] K. Kushibar, S. Valverde, and et al., "Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features," *CoRR*, vol. abs/1709.09075, 2017.

[23] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks," *Scientific Reports*, vol. 9, no. 1, p. 16884, Nov 2019.

[24] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014.

[25] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016.

[26] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017.

[27] F. Mahmood, D. Borders, and et al., "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *CoRR*, vol. abs/1810.00236, 2018.

[28] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, 2016.