HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2024 June 15.

Published in final edited form as:

Phys Med Biol. ; 68(12): . doi:10.1088/1361-6560/acda0b.

Automated Deep Learning Auto-segmentation of Air Volumes for MRI-guided Online Adaptive Radiation Therapy of Abdominal Tumors

Ergun Ahunbay[1], **Abdul K. Parchur**[1], **Jiaofeng Xu**[2], **Dan Thill**[2], **Eric S. Paulson**[1], **X Allen Li**[1]

[1]Department of Radiation Oncology, Medical College of Wisconsin, Milwaukee, WI, USA, 53226

[2]Elekta Inc., St. Charles, MO

Abstract

Objective: In the current MR-Linac online adaptive workflow, air regions on the MR images need to be manually delineated for abdominal targets, and then overridden by air density for dose calculation. Auto-delineation of these regions is desirable for speed purposes, but poses a challenge, since unlike CT, they don't occupy all dark regions on the image. The purpose of this study is to develop an automated method to segment the air regions on MRI-guided adaptive radiation therapy (MRgART) of abdominal tumors.

Approach: A modified ResUNet3D deep learning (DL)-based auto air delineation model was trained using 102 patients' MR images. The MR images were acquired by a dedicated in-house sequence named "Air-Scan", which is designed to generate air regions that are especially dark and accentuated. The air volumes generated by the newly developed DL model were compared with the manual air contours using geometric similarity (Dice Similarity Coefficient, DSC), and dosimetric equivalence using Gamma index and dose-volume parameters.

Main results: The average DSC agreement between the DL generated and manual air contours is 99% ± 1%. The gamma index between the dose calculations with overriding the DL vs manual air volumes with density of 0.01 is 97% ± 2% for a local gamma calculation with a tolerance of 2% and 2mm. The dosimetric parameters from planning target volume - PTV and organs at risk - OARs were all within 1% between when DL vs manual contours were overridden by air density. The model runs in less than five seconds on a PC with 28 Core processor and NVIDIA Quadro® P2000 GPU.

Significance: A DL based automated segmentation method was developed to generate air volumes on specialized abdominal MR images and generate results that are practically equivalent to the manual contouring of air volumes.

Keywords

Synthetic CT; MR imaging; Air volume delineation; Adaptive Radiation Therapy

* Eahunbay@mcw.edu .

Disclosure of conflicts of interest

The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

1. Introduction

The MR-Linac treatment systems offer several advantages for daily delineation and motion monitoring during treatment delivery [1, 2]. Unlike computed tomography, the daily MR images do not inherently contain the electron density (ED) information that is necessary for the daily dose calculations. Several methods have been proposed to generate the ED information using the daily MRIs for different tumor sites [3]. Recently, we developed a synthetic computed tomography (CT) solution that utilizes a deformable image registration (DIR) to transfer the ED values from a reference CT to the daily MR image [4]. This approach yielded good accuracy for several sites, but not for abdomen cases where the randomly varying air/gas regions pose a challenge for the DIR [5]. Although the patients are instructed to avoid gassy foods and chewing gum, it is hard to avoid bowel gas during treatment. Sometimes patients are not compliant, or patients have excessive gas secondary to the presence of cancer itself or cancer therapies. In current practice, the air volumes are manually delineated, and their relative ED is overridden with a constant (e.g., 0.01 per ICRU report)[6] before the onset of plan adaptation [7]. It is desirable to automate the delineation of air volumes on the MR images to improve speed, consistency, and potentially to allow for more frequent plan adaptation schemes. Air volumes can also change during the treatment, an automated air segmentation method would be necessary for potential real-time plan adaptation schemes to account for those changes. Unlike CT, air and several other tissue types of the body or artefacts yield similar low signal intensity on MRI, making auto-delineation of air on the MRI challenging, since a simple global thresholding cannot be used.

We previously reported an auto-air delineation method using the DIR between the daily and a reference MR images with delineated GI structures [8]. This method provided air volumes that resulted in dose accuracy practically equivalent to the manual contours. However, there were two drawbacks in the method; (1) it requires the delineation of the air containing GI volumes on a reference image, which must be acquired for each patient separately, and (2) the DIR operation is time consuming (~25 seconds) which can be significant for online adaptation.

In this work, we report on a new and improved method that can quickly auto-segment air volumes on MRI based on deep learning (DL) without reference contours. We trained a DL model using an abdominal MRI dataset of 102 patients with manually delineated air regions. The training dataset is based on MRIs of a special sequence, called air scan — specifically designed in our clinic to better visualize air volumes. The obtained model was evaluated for the air scans using ten new patients and compared with manually delineated air contours.

2. Methods and Materials

2.1. Nomenclature

The following nomenclatures (abbreviations) are used throughout this manuscript:

DL-air: the air volume auto-segmented by the newly developed deep learning models.

DL-dose: the dose calculation generated with the DL-air overridden by rED of 0.01.

Manual-air: the air volumes manually delineated.

Manual-dose: the dose distribution generated by overriding the Manual-air with rED of 0.01.

Repeat-dose: the same as Manual-dose, except that the dose calculation was repeated to assess the statistical difference due to the Monte Carlo algorithm.

No-air-dose: the dose calculation created without overriding rED in the air volumes.

2.2. MR Air Scans

A special MR sequence (Cartesian 3D FLASH or 3D tFE) optimized to acquire air volumes in the abdomen, named Air-Scan, was used. This sequence has a partial echo with a minimum echo time of 1 ms to maximize the air visualization and avoid signal loss due to T2* decay. The Air-Scan images are acquired with a voxel size of $1.88 \times 1.88 \times 3.5\ mm^3$ and a T1-weighted gradient MR acquisition (TE = 1 ms, TR = 2.5 ms, and flip angle = 5°). This image resolution, although coarser than those of daily images used for daily planning, was found to be a good compromise in terms of resolution and speed. A higher spatial resolution increases acquisition times, which decreases the sharpness of the boundary of air and tissue, resulting in inaccuracies in estimation of air volumes. The scan time for this sequence is approximately 9 seconds. The patients are instructed to breathe normally during the acquisition of Air Scans. Partial Fourier and parallel imaging accelerations were used to minimize scanning times. Since 2020, we have used the Air-Scan to manually delineate air volumes in our clinical online adaptation workflow. We use the “2D brush tool” in MIM with an upper range lock value of 2000 to delineate the air volumes on the Air Scans for all patients.

2.3. ResUNet3D Air-segmentation model

A DL automated air segmentation model was developed and implemented in the Admire (Elekta, St Louis MO) environment. A well-established modified 3D Unet architecture was used to build the DL air-segmentation model, called RusUNet3D [9, 10], which is chosen as it is considered as a very good benchmark for general segmentation tasks. The air pockets’ edge/boundaries already have strong features, where the long-skip concatenation of multi-levels between encoder and decoder in ResUnet-style is expected to preserve and represent adequately.

The generation and training of network was implemented in-house by Tensorflow 1.15 platform. The network structure was specified in Figure 1 and a traditional cross entropy function was adopted. The general idea of this network combines the short skip connection in 3D Residual block and long-skip connection between encoder and decoder. Total 100 epochs were performed for this training. To overcome memory issues during training, whole patient volumes were not fed into training, but a slab of consecutive 32 slices (320×320×32) were randomly picked each time. For each epoch, every patient volume was sampled 16 times as the fixed size 320×320×32 from the whole volume, starting slice of which was randomly picked. All kernel sizes are 3×3×3, except for fully connected layers.

The last, i.e., fourth dimension in each block represents the number of filters/channels. To mitigate the potential over-fitting issue, an elastic transform was also applied on-the-fly for the whole volume before sampling. The Adam optimizer was used. The total training epoch was set to 100. The learning rate was fixed at 0.002 for the first half of the total number of epochs and decayed linearly for the second half. Based on our previous similar training experience, the training error curve and five-patient based validation error curve have become flat and stable in the range of 100 epochs.

To learn various representations of MR data, first a new residual block "Res-block", was designed as the basic processing unit. The MR Air-Scan DICOM resolution of the input 3D sub-volume is $1.875 \times 1.875 \times 3.5\ mm^3$, which is the same as its physical resolution. Furthermore, to reduce inhomogeneity and variation, the original Air-Scan images have been standardized and normalized using Z-scores. To reduce potential over-fitting issues, a 3D elastic deformation transform was used as one type of data argumentation for each epoch using a total of ~102 patients to make up the training pool.

The model was trained with 102 Air-Scan sets along with Manual-air. The Air-Scans were acquired on Unity MR-Linac (Elekta, Stockholm, Sweden) during the routine online adaptive treatments of abdominal patients, from January 2019 to March 2022.

2.4. AIR-Segmentation model evaluation

Carefully generated manual air contours from 20 abdominal patients that were treated for their tumors in liver or pancreas were utilized as the ground truth for testing the model. The Table 1 summarizes the demographic information of all 122 patients used in the model, including gender and treatment site specifics. The testing patients' contours were not included in the training dataset.

On these images, DL-air was obtained using the trained DL model in the ADMIRE Research software (v 3.45, Elekta). On these images, the Manual-air contours were generated using threshold 2,000 in MIM software. The small air contour regions with radius less than 1 mm were filtered out. The Manual-air and DL-air were compared geometrically and dosimetrically. All evaluations were performed using specially designed workflows in MIM software (v 7.2).

2.4.1. Geometric comparison—The geometric agreement between the DL-air and Manual-air contours was assessed using the Dice Similarity Coefficient (DSC), a measure of overlap, thus, a relative metric of agreement between the two volumes. We also calculated the absolute disagreement volume (cc) between the DL-air and Manual-air, which should better reflect the dosimetric consequences due to the air volume differences.

2.4.2. Dose calculation—For dosimetric comparison, plans that were generated during online optimization for the treatment of these 20 patients were utilized. The plans are step and shoot Intensity-modulated radiation therapy (IMRT) plans using the Monaco system (version 6.01, Elekta AB) using a Monte Carlo dose engine and considering the dose effect of the 1.5T magnetic field for 20 test cases. The Unity MR-LINAC's 7 MV FFF photon beams were used for all plans. Sequencing parameters of beamlet width, minimum segment

area, minimum segment width, minimum MU/Segment, and maximum segments per plan were set as 2 mm, 2 cm^2, 1 cm, 2, and 70, respectively.

A calculation grid size of 3 mm, and statistical uncertainty of 1% per calculation were used. It should be noted, this 1% uncertainty is only achieved in the high dose regions of the plan, while the lower dose regions typically have higher uncertainty numbers.

2.4.3. Dosimetric evaluation—For each patient, four dose distributions, (1) DL-dose, (2) Manual-dose, and (3) Repeat-dose, and (4) No-air-dose, were calculated using the beam parameters of the clinical plan. The dosimetric effect from the difference between the DL-air and Manual-air was evaluated within the D>80% IDL (iso-dose line), D>50% IDL, and D>30% IDL region using a 3D local gamma analysis using 3%, 3 mm, 2%, 2 mm, and 1%, 1 mm gamma criteria.

Also, the dose volume histogram (DVH) parameters were compared between the dose distributions. The percentage dose deviations for DL-dose vs Manual-dose, Manual-dose vs Repeat-dose, and Manual-dose vs No-air-dose for the DVH parameters PTV_{D95}, $PTV_{0.03cc}$, PTV_{mean}, $Duodenum_{0.03cc}$, $Stomach_{0.03cc}$, Small $bowel_{0.03cc}$, and $Colon_{0.03cc}$ was calculated using a workflow in MIM software.

3. Results

Figure 2 presents a typical output of the air segmentation model, which clearly shows the comparison between the DL model segmented (green contours) and manual contour segmented (pink contours) on representative axial, coronal, and sagittal MR images for a test case. Figure 3 shows the results of the geometric comparisons. This information is also presented in Table 2. The mean DSC agreement is 99%, with 1% standard error, which indicates the DL method can accurately replicate the manual air contouring. The biggest absolute volume disagreement was 5.5cc, out of 70cc of air in patient #2, which as it is shown in the dosimetric analysis below, did not result in a large dosimetric effect.

The dosimetric gamma analysis results are presented in Figure 4 and (see Table S1 in supplementary materials). One can observe that the differences between DL-dose and Manual-dose are rather small, and they are equivalent or even less than the differences between Manual-dose and Repeat-dose Table S1 (see Supplementary Materials). The mean gamma agreement between DL-dose and manual-dose was equal or better than the agreement between manual-dose and repeat-dose for all comparisons. Therefore, it can be concluded that the differences are within the statistical uncertainty of the Monte Carlo calculations, and thus, the dosimetric differences due to air volume discrepancy is almost negligible.

A similar conclusion can be reached for the comparisons of dosimetric criteria (Figures 5 and 6). The DL-dose/Manual-dose differences are all within the ± 1% range, while the No-air-dose/Manual-dose differences tend to exceed this range. In Figure 6, the range between 25 and 75 percentiles is plotted for each DVH point for different comparisons. The DL-dose/Manual-dose differences are all smaller than No-air-dose/Manual-dose differences, and they are mostly smaller or comparable to the Manual-dose/Repeat-dose differences.

The absolute value of Manual-dose/Repeat-dose differences was similar to the DL-dose/manual-dose differences, average 0.32% *vs* 0.39%, p = 0.39. The mean absolute differences between No-air-dose and Manual-dose was 0.64%, which was significantly higher than both DL-dose/Manual-dose differences (p=0.0002) and Manual-dose/Repeat-dose differences (p = 0.00001).

4. Discussion

The results indicate that the DL model is accurate in delineating the abdominal air regions on the Air Scan images. The model can be run in less than five seconds and can be fully automated. Both geometric and dosimetric analyses indicate excellent agreement between DL-air and Manual-air contours. The geometric agreement between DL-air and Manual-air contours is excellent, with average DSC of 0.99. The dosimetric disagreements due to the differences in air volumes are small and are within the statistical uncertainty (±1% per calculation) in the dose calculations. It is important to point out that the DL-dose/manual-dose differences reported here are due to both the statistical uncertainties in the dose calculations, and the dosimetric effects of the air volume differences. On the other hand, the manual-dose/repeat-dose differences are only due to statistical uncertainties. So, the fact that the DL-dose/manual-dose differences are comparable to the manual-dose/repeat-dose differences indicates that any dose effect of the air volume differences is almost indistinguishable in the presence of the dosimetric differences due to statistical uncertainty. Also, it must be noted that the 1% statistical uncertainty used in these calculations is what is being used for the dose calculations in our clinic. Therefore, the dosimetric effects due to DL generated air volumes are well below the level of clinical significance.

Main benefit of the proposed method is in terms of speed, automation, and consistency. The manual delineation currently takes around 1–3 minutes of time, which can be reduced to less than five seconds with the proposed method. Also, automation will allow for improved consistency and reduction in potential human errors. Air segmentation is one of the tasks that constitute the MR guided online adaptive RT and improvements in each task will contribute to the speed of the whole process.

We had previously developed a method (named DIFF) for auto-air delineation, which was also a fully automatic method. However, this DL method is superior in its speed (less than 5 seconds *vs* ~25 seconds), and easier to use. The previously reported DIFF method necessitated delineation of air-containing combined GI volume to be delineated offline [8]. Although this would not be a process that needs to be performed online, it is still an additional burden on the clinic personnel. The current method does not require any special preparation either online or offline and can simply be run as part of the online process.

One concern for the current method is that the DL model was specifically developed for the Air-Scan MR sequence and may not be generalized to other MR sequences. Also, if the Air-Scan sequence is to be altered substantially in the future, a retraining of the DL model may be necessary. Air Scan images could possibly be improved by employing non-cartesian methods such as radial or propeller type k-space sampling [11] to minimize blurring from breathing motion, which would increase the accuracy of air volume definitions. Such

improvements in the imaging could be expected to also improve the accuracy of the proposed auto-delineation method.

This method generates the air volumes as contours that can be visually verified once overlaid on the daily MR images quickly during online planning. Correcting random air regions in the abdominal structures would be taking care of the most important dosimetric effect, as air density constitutes the biggest density difference from soft tissue. Although there are several synthetic CT methods in the development phase for the abdomen [12, 13], we are not aware of ones that can accurately address the air regions, as these were not specifically reported. Our method can be combined with other synthetic CT methods as an auxiliary correction on air regions or can be directly used as part of the bulk density correction scheme currently used in the clinic. As part of the clinical scheme, we developed the MIM workflow to run Admire externally and input the generated air contours inside MIM without exiting the current Adapt-to-shape workflows of MIM.

5. Conclusion

Current DL model that can quickly and automatically segment the air regions on abdominal MRI acquired 1.5T MR-LINAC was developed. The model performed well with a DSC of more than 95%, and dosimetric effect within the statistical uncertainty of dose calculations compared with those from manually segmented air volumes. This technique could help speed up and automate the online adaptive planning on MRL for abdominal tumors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding statement

The research was partially supported by the Medical College of Wisconsin (MCW) Fotsch Foundations, and the National Cancer Institute of the National Institutes of Health under award number R01CA247960. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Lagendijk JJ, Raaymakers BW, and van Vulpen M, The magnetic resonance imaging-linac system. Semin Radiat Oncol, 2014. 24(3): p. 207–9. [PubMed: 24931095]
2. Vestergaard A, et al. , The potential of MRI-guided online adaptive re-optimisation in radiotherapy of urinary bladder cancer. Radiother Oncol, 2016. 118(1): p. 154–9. [PubMed: 26631646]
3. Edmund JM and Nyholm T, A review of substitute CT generation for MRI-only radiation therapy. Radiat Oncol, 2017. 12(1): p. 28. [PubMed: 28126030]
4. Ahunbay EE, et al. , A Technique to Rapidly Generate Synthetic Computed Tomography for Magnetic Resonance Imaging-Guided Online Adaptive Replanning: An Exploratory Study. Int J Radiat Oncol Biol Phys, 2019. 103(5): p. 1261–1270. [PubMed: 30550817]
5. Thapa R, et al. , Automated air region delineation on MRI for synthetic CT creation. Phys Med Biol, 2020. 65(2): p. 025009. [PubMed: 31775128]
6. White DR, Griffith RV, and Wilson IJ, ICRU Reports. Reports of the International Commission on Radiation Units and Measurements, 1992. os-24(1): p. 203–205.

7. Paulson ES, et al. , 4D-MRI driven MR-guided online adaptive radiotherapy for abdominal stereotactic body radiation therapy on a high field MR-Linac: Implementation and initial clinical experience. Clin Transl Radiat Oncol, 2020. 23: p. 72–79. [PubMed: 32490218]
8. Ahunbay E, et al. , Development and implementation of an automatic air delineation technique for MRI-guided adaptive radiation therapy. Phys Med Biol, 2022. 67(14).
9. Yu L, et al., Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images Proceedings of the AAAI Conference on Artificial Intelligence, 2017. 31(1).
10. Long J, Shelhamer E, and Darrell T. Fully convolutional networks for semantic segmentation in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
11. Hirokawa Y, et al. , MRI artifact reduction and quality improvement in the upper abdomen with PROPELLER and prospective acquisition correction (PACE) technique. AJR Am J Roentgenol, 2008. 191(4): p. 1154–8. [PubMed: 18806158]
12. Liu L, et al. , Abdominal synthetic CT generation from MR Dixon images using a U-net trained with ‘semi-synthetic’ CT data. Phys Med Biol, 2020. 65(12): p. 125001. [PubMed: 32330923]
13. Kang SK, et al. , Synthetic CT generation from weakly paired MR images using cycle-consistent GAN for MR-guided radiotherapy. Biomed Eng Lett, 2021. 11(3): p. 263–271. [PubMed: 34350052]

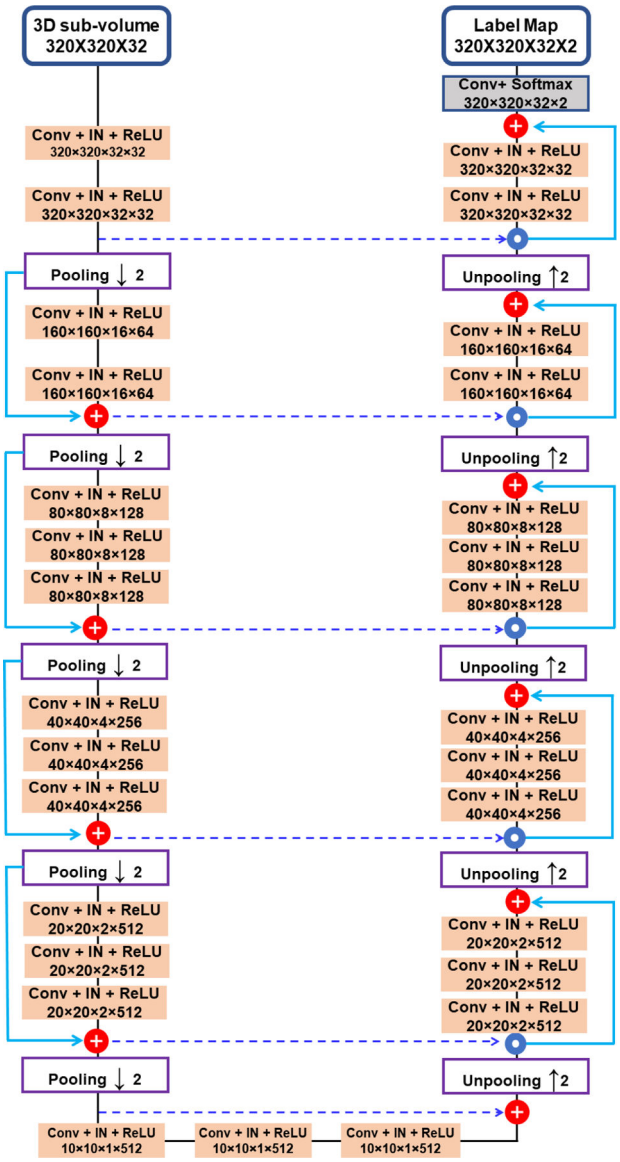

Figure 1.

The schematic illustration shows our in-house 3D ResUnet3D network method used to train our air segmentation model, with clear decoding and encoding structures. The short skip connection with red color "+" represents the addition, while the long blue connection sign "•" represents the concatenation. Conv: Convolution. IN: Instance Normalization. ReLU: Rectifier Linear Unit.

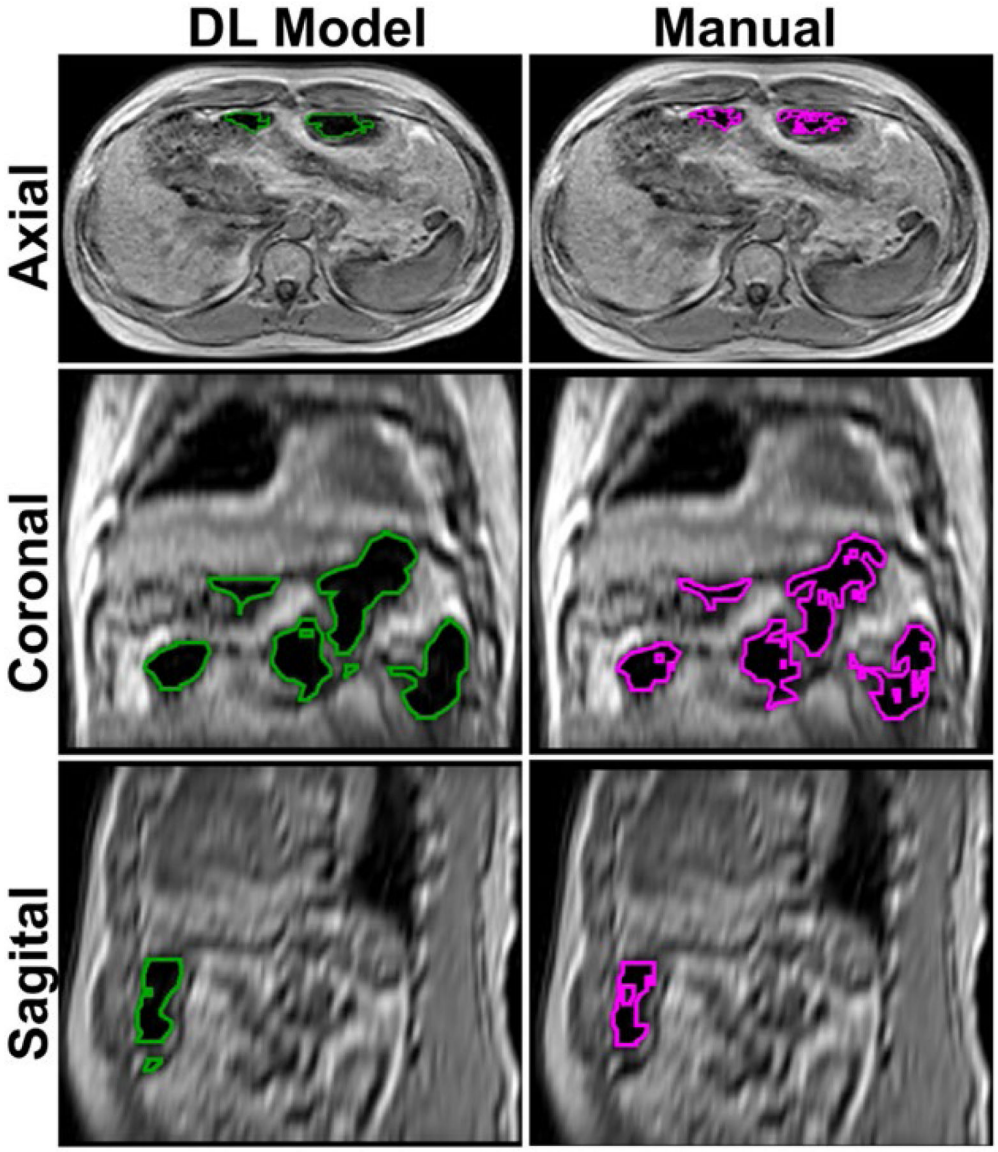

Figure 2.
DL-air (green contours) and Manual-air (pink contours) on representative axial, coronal, and sagittal images for a test case.

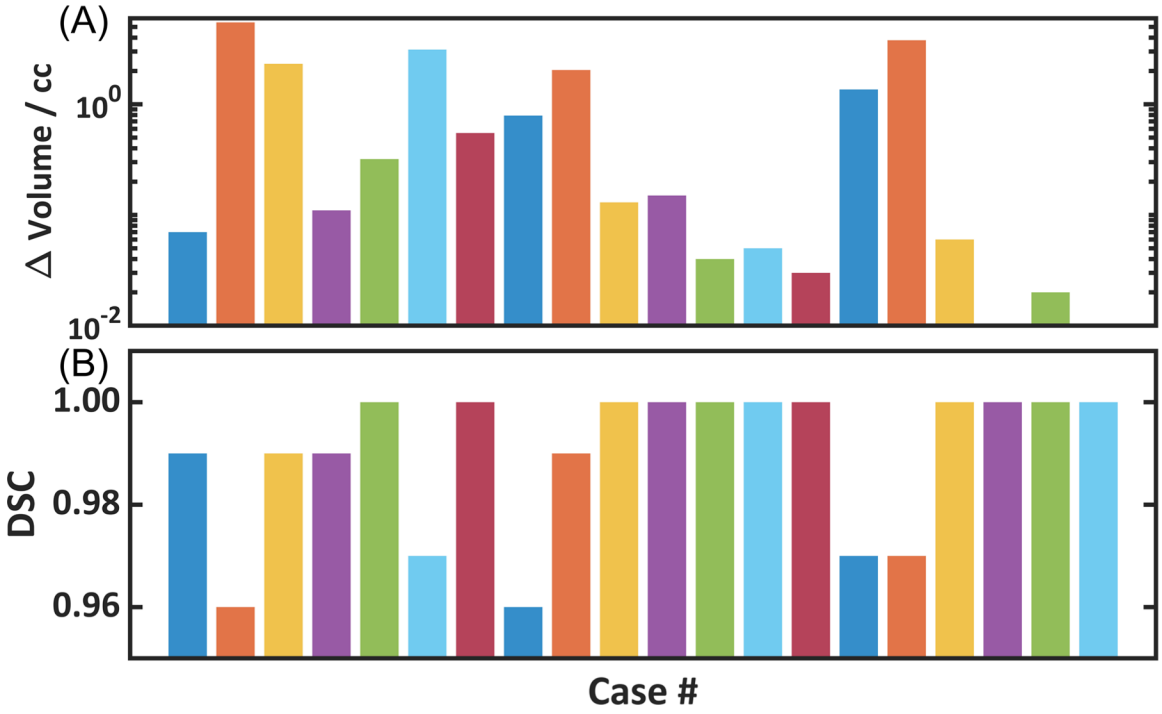

Figure 3.
(A) Difference between DL-air and Manual-air contours and their corresponding (B) DSC comparison.

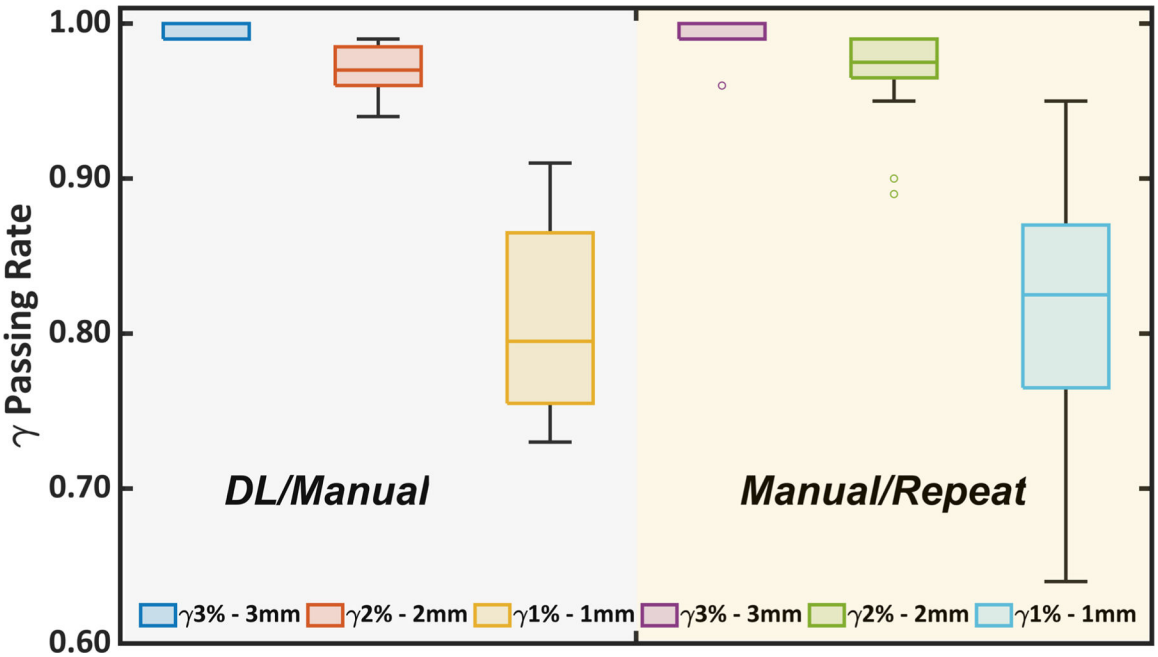

Figure 4.
The gamma passing rates for DL/manual and manual/repeat comparisons for D>30%IDL.

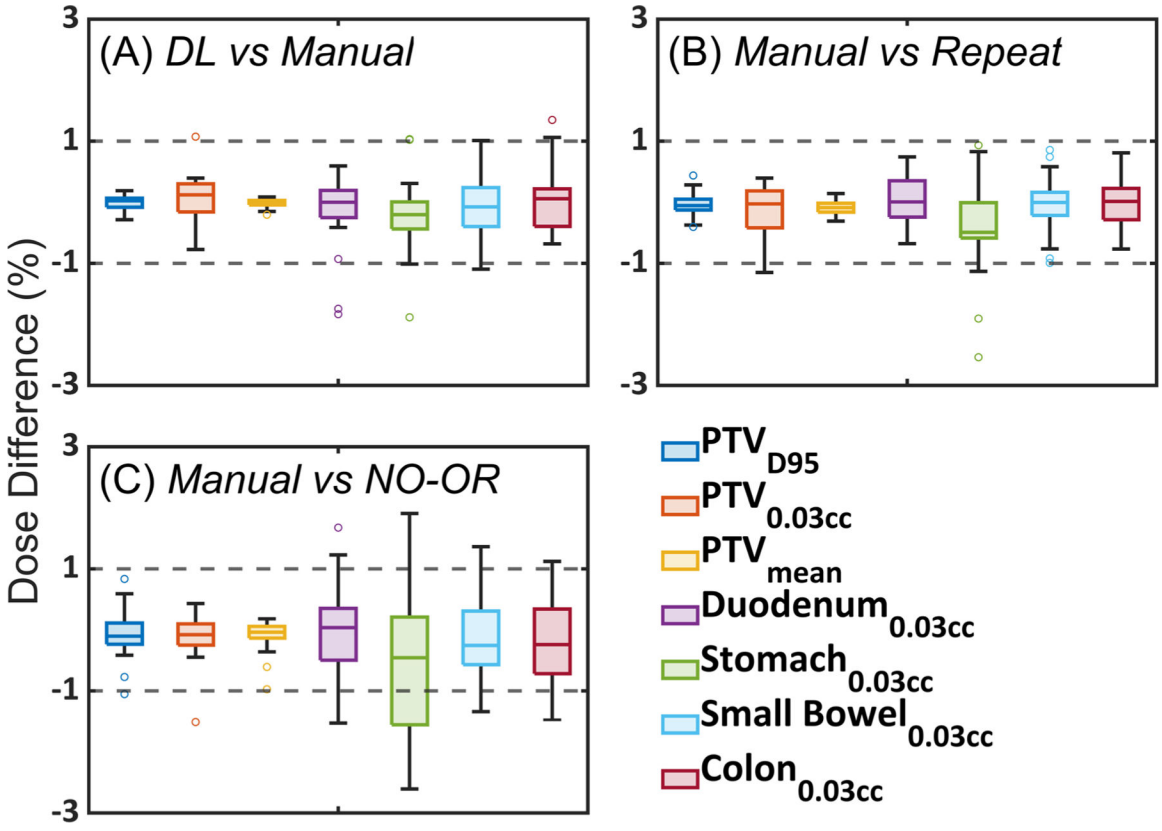

Figure 5.

Percentage dose deviation in (A) DL-dose vs Manual-dose, (B) Manual-dose vs Repeat-dose, and (C) Manual-dose vs No-air-dose for PTV_{D95}, $PTV_{0.03cc}$, PTV_{mean}, $Duodenum_{0.03cc}$, $Stomach_{0.03cc}$, Small $bowel_{0.03cc}$, and $Colon_{0.03cc}$. Dashed lines show 1% deviation line on dose difference % axis.

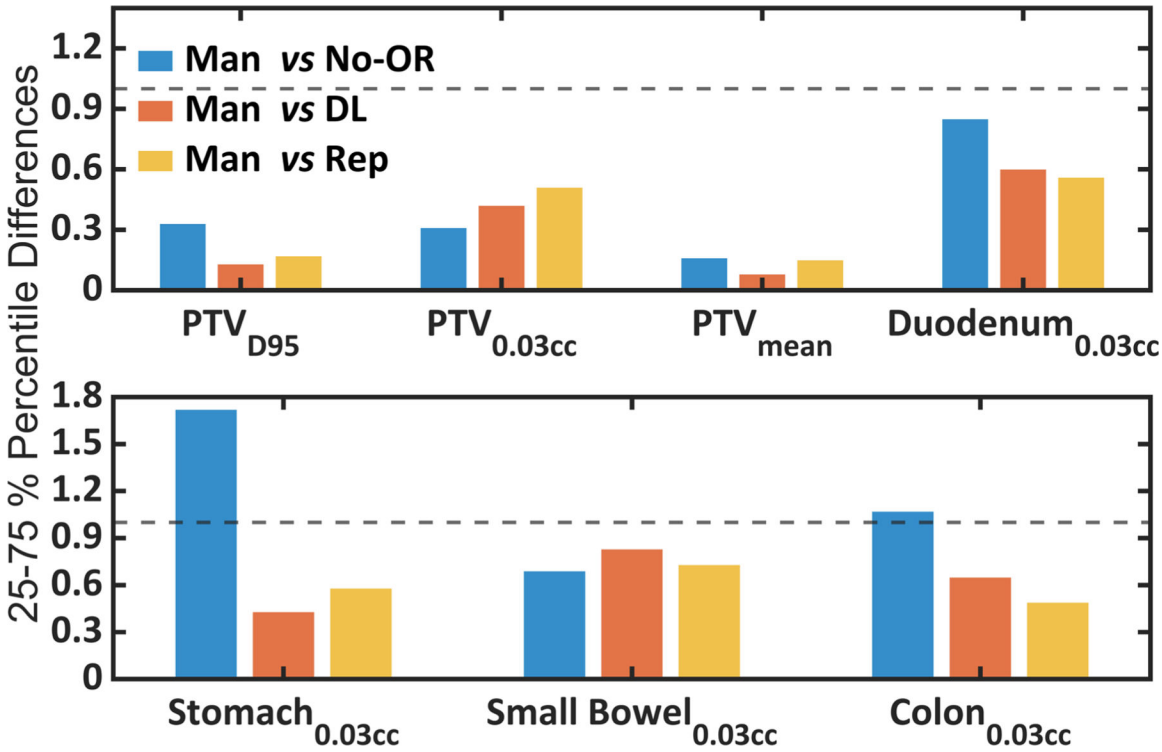

Figure 6.

The 25–75% percentile dose difference of PTVD95, $PTV_{0.03cc}$, PTV_{mean}, $Duodenum_{0.03cc}$, $Stomach_{0.03cc}$, Small $bowel_{0.03cc}$, and $Colon_{0.03cc}$ for manual *vs* No-OR, manual *vs* DL, and manual *vs* Rep.

Table 1:

Demographic information for all 122 patients used in training and testing of the Air Segmentation model.

Gender Statistics					
Gender	**Number**	**Percentage (%)**	**Mean Age**	**Median Age**	**Max Age**
Female	47	38.5	66.6	67.0	83
Male	75	61.5	66.3	67.0	90
Treatment Site Statistics					
Site	**Number**	**Percentage (%)**	**Mean Age**	**Median Age**	**Max Age**
Pancreas	56	45.9	66.9	67.0	90
Liver	46	37.7	65.9	68.5	82
Abdominal/Other	20	16.4	66.3	66.0	81

Table 2:

Results of the geometric comparisons for 10 patients. The DL-air and Manual-air volumes in cc, and the agreement (intersection) and disagreement volumes (XOR) are also listed in cc.

Patient [#]	[*]V2 / cc	[#]V1 /cc	Intersection (V1, V2) /cc	XOR (V1, V2) /cc	DSC (V1, V2)
1	4.55	4.54	4.5	0.07	0.99
2	69.21	74.03	68.90	5.51	0.96
3	94.40	96.13	94.10	2.32	0.99
4	6.79	6.90	6.80	0.11	0.99
5	38.99	39.28	39.00	0.32	1.00
6	56.12	58.65	55.80	3.12	0.97
7	72.63	73.17	72.60	0.55	1.00
8	8.52	9.30	8.50	0.79	0.96
9	91.91	92.74	91.30	2.04	0.99
10	17.64	17.52	17.50	0.13	1.00
11	49.80	49.67	49.66	0.15	1.00
12	18.28	18.32	18.28	0.04	1.00
13	26.58	26.63	26.58	0.05	1.00
14	7.12	7.13	7.11	0.03	1.00
15	21.81	21.87	21.16	1.36	0.97
16	53.44	56.98	53.31	3.79	0.97
17	14.64	14.64	14.61	0.06	1.00
18	10.77	10.78	10.77	0.01	1.00
19	10.02	10.02	10.01	0.02	1.00
20	5.70	5.69	5.69	0.00	1.00
Average				**1.02 ± 1.55**	**0.99 ± 0.01**

[*]V2 – (DL-air volume)

[#]V1 – Manual-air volume)

Execution of the model takes less than five seconds on an Intel® Xeon® Gold 6312 Processor (28-Core) PC with NVIDIA Quadro® P2000 GPU (5GB GPU memory).