Radiation Oncology

RESEARCH

Open Access

Deep learning based automatic segmentation of organs-at-risk for 0.35 T MRgRT of lung tumors

Marvin F. Ribeiro[1], Sebastian Marschner[1], Maria Kawula[1], Moritz Rabe[1], Stefanie Corradini[1], Claus Belka[1,2,3], Marco Riboldi[4], Guillaume Landry[1] and Christopher Kurz[1*]

Abstract

Background and purpose Magnetic resonance imaging guided radiotherapy (MRgRT) offers treatment plan adaptation to the anatomy of the day. In the current MRgRT workflow, this requires the time consuming and repetitive task of manual delineation of organs-at-risk (OARs), which is also prone to inter- and intra-observer variability. Therefore, deep learning autosegmentation (DLAS) is becoming increasingly attractive. No investigation of its application to OARs in thoracic magnetic resonance images (MRIs) from MRgRT has been done so far. This study aimed to fill this gap.

Materials and methods 122 planning MRIs from patients treated at a 0.35 T MR-Linac were retrospectively collected. Using an 80/19/23 (training/validation/test) split, individual 3D U-Nets for segmentation of the left lung, right lung, heart, aorta, spinal canal and esophagus were trained. These were compared to the clinically used contours based on Dice similarity coefficient (DSC) and Hausdorff distance (HD). They were also graded on their clinical usability by a radiation oncologist.

Results Median DSC was 0.96, 0.96, 0.94, 0.90, 0.88 and 0.78 for left lung, right lung, heart, aorta, spinal canal and esophagus, respectively. Median 95th percentile values of the HD were 3.9, 5.3, 5.8, 3.0, 2.6 and 3.5 mm, respectively. The physician preferred the network generated contours over the clinical contours, deeming 85 out of 129 to not require any correction, 25 immediately usable for treatment planning, 15 requiring minor and 4 requiring major corrections.

Conclusions We trained 3D U-Nets on clinical MRI planning data which produced accurate delineations in the thoracic region. DLAS contours were preferred over the clinical contours.

Keywords Deep learning, Auto-segmentation, MR-Linac, MRI-guidance, Thorax

*Correspondence:
Christopher Kurz
christopher.kurz@med.uni-muenchen.de
Full list of author information is available at the end of the article

Introduction

The clinical introduction of magnetic resonance imaging guided radiotherapy (MRgRT) has brought great benefits such as higher dose conformity to the target and more healthy tissue sparing [1]. In contrast to conventional stereotactic body radiotherapy (SBRT) using cone beam computed tomography (CBCT), magnetic resonance (MR) images do not expose the patient to extra dose during image acquisition. The patient's anatomy in treatment position, as depicted on the daily images, is typically used for online plan adaptation. Real-time cine-MR imaging can also be used for breath-hold gated treatment with high geometric gross tumor volume (GTV) coverage [2]. The MRIdian (ViewRay Inc, Cleveland, OH, USA) [3] is a 0.35 T MR-guided linear accelerator (MR-Linac) that enables such treatments. It requires the target and organs-at-risk (OARs) to be accurately segmented not only on the planning image, but also on the image of the day (fraction image) [4]. Although time is a less critical factor for delineating the planning MR image (MRI), it is still a fairly time consuming procedure [5] and prone to inter- and intra-observer variability [6]. These uncertainties may further affect follow-up analyses, such as dose accumulation studies in the scope of MRgRT [7]. While deformable image registration (DIR) is used in the MRgRT current clinical workflow to propagate contours from the planning image to the daily MRI, such contours often necessitate time-consuming manual correction or re-contouring. Treatment time without irradiation ranges from 30 to 70 min for a single fraction [8–10], of which up to 22 min are due to the delineation [11].

Recent publications have shown that deep learning auto-segmentation (DLAS) can produce accurate delineations, which in turn accelerates the workflow, decreasing patient discomfort as well as increasing patient throughput [12]. Network-generated contours are also more consistent, helping to reduce inter- and intra-observer variability [5]. Several groups have demonstrated that artificial neural networks (ANNs) can produce high quality contours in the context of MRgRT. Liang et al. [13] used a support vector machine based model on pancreatic images from a 0.35 T MR-Linac. Fu et al. [14] have used a convolutional neural network (CNN) with two correction networks to achieve promising results in the abdominal region for the same MR-Linac. Eppenhof et al. [15] used a 3D U-Net to segment the clinical target volume (CTV) for prostate cancer patients by generating a deformation field from the planning image to the fraction image of prostate cancer patients from a 1.5 T MR-Linac. Kawula et al. [16] have demonstrated superior accuracy for prostate and bladder segmentation using a patient-specific 3D U-Net on a 0.35 T MR-Linac. Chen et al. [17] have also used a patient-specific CNN model for prostate cancer patients on a 1.5 T MR-Linac. Fransson et al. [18] used a 2D U-Net model trained from scratch on a single 1.5 T prostate patient planning image. Li et al. [19] used a modified version of nnU-Net 2D [20] for a daily updated patient-specific segmentation of pelvic and abdominal fraction images from a 1.5 T MR-Linac.

Currently, there are few studies on OAR segmentation of thoracic MR images in general (e.g., Dong et al. [21]), none of which are in the context of MRgRT of lung tumors. The goal of this study was to evaluate the performance of DLAS on important OARs for the treatment of lung tumors (lungs, heart, aorta, esophagus and spinal canal) on a 0.35 T MR-Linac. The generated contours were compared to the clinically used ones in a geometrical analysis. They were also graded by a physician with regards to their clinical usability.

Materials and methods

Database

This study included data from 112 patients with lung tumors treated at the ViewRay MRIdian MR-Linac installed at the Department of Radiation Oncology of the LMU Munich University Hospital. 122 planning MRIs with their corresponding clinical OAR contours created between January 2020 and September 2022 were retrospectively collected. The patients received fractionated treatment in 3–16 fractions. All patients signed an informed consent form. More details about the patient cohort can be found in Table 1. The MR images were acquired using a 3D balanced steady state free-precession sequence. The images had a $1.5 \times 1.5\,mm^2$ in-plane resolution in the axial plane with varying axial size (usually

Table 1 Summary of details about the entire patient cohort and the training, validation and test set subgroups

Information	Training	Validation	Test	Total
Age				
Median	64	60	64	64
Range	19–86	25–75	29–88	19–88
M:F ratio	48:52	53:47	78:22	60:40
Lesion location				
SL l	17%	37%	9%	19%
SL r	21%	11%	26%	20%
ML r	6%	11%	4%	7%
IL l	17%	5%	30%	18%
IL r	21%	16%	17%	20%
Other	17%	21%	13%	17%
Metastasis	71%	79%	57%	70%

Includes median age and age range (min–max), male to female (M:F) ratio, location of the lesion (SL—superior lobe, ML—middle lobe, IL—inferior lobe, r—right, l—left) and the ratio of metastases to primary tumors. Metastasis denotes lesions of any origin in the lung, as opposed to a primary lung tumor

$> 300 \times 300$ pixels) and 3 mm slice thickness, with 144 slices in most cases. Segmentation of the ROIs was performed manually and approved during treatment planning by radiation oncologists. Images were exported from the treatment planning system (TPS) as DICOM files, along with their corresponding contours in DICOM-RT format.

Data pre-processing

Contours were converted to binary masks in ITK Meta Image format (mha), using plastimatch [22] with nearest-neighbour (nn) interpolation on the MR reference grid. Binary masks and images were resampled to 1.5 mm slice thickness with nn interpolation using the SITK [23] Python package, resulting in voxels of isotropic dimension. All images were then center-cropped/zero-padded to a uniform size of $256 \times 256 \times 256$ voxels. Lastly, intensity normalization to values between 0 and 1 while clipping at the image intensity's 99.5th percentile to account for possible high intensity MR artifacts was applied.

Network implementation details

The PyTorch based [24] MONAI [25] implementation of a 3D U-Net was used in this study. The network, inspired by Kerfoot et al. [26], had an identical architecture to the one used in Kawula et al. [16]. It has single input and output channels, which are converted to and from 16 channels in the beginning and end. Each following block then doubles the number of channels, from 16 to 256 in four steps with stride 2 down-convolutions in the encoding arm, and does the same in reverse with up-convolutions in the decoding arm. Each residual block consists of 2 series of a convolution, instance normalization and PReLU activation, of which only the first convolution changes the tensor dimensions as described. A Dice similarity coefficient (DSC) based loss function [27] and the ADAM optimizer [28] were used for training. The Dice loss DL between the network prediction P and the ground truth (GT) Y was computed as

$$\mathrm{DL} = 1 - \frac{2\sum_i^N p_i y_i + \epsilon}{\sum_i^N p_i + \sum_i^N y_i + \epsilon}, \tag{1}$$

summing over all $N = 256 \times 256 \times 256$ voxels $p_i \in P$ and $y_i \in Y$. The term $\epsilon = 10^{-5}$ was used to avoid numerical issues.

Training was performed using an NVidia RTX A6000 (48 GB VRAM) or an NVidia Quadro RTX 8000 (48 GB VRAM) GPU.

Training strategy

Planning MRIs were randomly split into 80 training, 19 validation and 23 test MRIs. A separate model was trained for each OAR. Contours were used according to their availability, so if an OAR had not been delineated for treatment, the planning MRI had to be excluded from the model for that particular OAR. The segmented OARs were chosen based on sufficient data availability, which was also seen as an indicator for the most commonly clinically needed segmentations. This led to a final selection of 6 OARs: right and left lungs, heart, aorta, spinal canal and esophagus. The exact split for each OAR is detailed in Table 2.

The network was trained in a two phase process with successive data augmentation. Data augmentation was used to prevent overfitting. During the first phase, it was trained with few computationally inexpensive augmentations and a mini-batch size of 4 for 75 epochs. The second phase was used to improve the DLAS contours by stronger data augmentations. The training continued with the model parameters of the epoch with the highest DSC on the validation set from the first phase. The mini-batch size was decreased to 1 for the second step, and the network was trained for 500 epochs. To further prevent overfitting, the model at the training epoch with the best performance on the validation set was automatically selected retrospectively as the final model for each OAR individually. The aforementioned two step augmentation process starts with translations and rotations, followed by a random zoom and by random Gaussian noise in the first phase. The probability p_{aug} for translation, rotation, zoom and Gaussian noise being applied was p_1.

The second phase introduced random elastic deformations, MRI motion artifacts and a random MRI bias fields following translation and rotation transformations. For these, the TorchIO [29] Python library was used. All other transformations were implemented using functions from the MONAI library [25]. There was a second intensity clipping of values below 0 and above 1 after the Gaussian noise was applied. For this phase, p_{aug} was changed to p_2.

Table 2 Summary of number of contours used for each set and ROI

OAR	Training	Validation	Test	Sum
Left lung	75	18	20	113
Right lung	78	18	23	119
Heart	74	15	23	112
Aorta	63	18	19	100
Spinal canal	68	18	23	109
Esophagus	80	18	21	119
Total MR scans	80	19	23	122

Numbers differ due to availability of contours in clinical treatment planning

Data post-processing

The network output probabilities were passed through a sigmoid activation function, followed by a threshold of 0.5 to create a binary mask. Connected components with less than $1/8^{th}$ of the total volume of the mask were removed before evaluation.

Data evaluation

Geometrical analysis

The DSC and the Hausdorff distance (HD) (average HD_{avg} and 95th percentile HD_{95}) were calculated using the clinical contours of the planning MRI as GT. Long, tubular OARs, i. e., the esophagus, spinal canal and aorta were often only delineated in the high dose region near the slices containing the tumor. Thus, these OARs were only evaluated in the axial slices containing the planning target volume (PTV) and the 10 slices above and below.

Physician's grading

A radiation oncologist was presented with the DLAS and the clinical contours of all considered ROIs of the 23 test set MRIs and asked to grade them on a 0–4 scale. These grades represented the following statements: 0—"no clinically relevant correction possible", 1 — "ready to use", 2—"minor corrections required", 3 — "major corrections required", 4—"unusable". To reduce possible evaluation bias, the ROI contours were presented as 46 separate MRIs in random order. The radiation oncologist was thus aware that they were comparing DLAS and clinical segmentation, but not which was which. We used the same scale as Kawula et al., but refined it by adding a category 0 to differentiate between contours with clinically acceptable residual errors from those without.

Results

Network training

The same final set of hyperparameters was used across all OARs. All hyperparameters, aside from p_{aug} and the standard deviation of the Gaussian noise were kept unchanged during both training phases. The range of searched hyperparameters can be found in Table 3. The final set includes a learning rate of 10^{-3}, augmentation probabilities of $p_1 = 0.6$ and $p_2 = 0.85$, a standard deviation for the Gaussian noise of 0.05 and 0.1, a zoom range between 0.9 and 1.1, a rotation range of up to 15° and translation range of up to 22.5 mm in all spatial dimensions. B-Spline elastic deformations used a maximum of 8 control points and a maximum displacement of 24 mm. MR related augmentations were set to 10° and 45 mm for the motion artifacts and an order of 1 and maximum polynomial coefficient magnitude of 0.4 for the bias field. The bias field from TorchIO was modelled as a unit-less quantity that modifies the voxel intensity by multiplying it with a linear combination of polynomial basis functions with randomly chosen coefficients [30]. The average training duration per epoch was around 90 s during the first phase and 5 min in the second phase.

Table 3 Functions used during training and hyperparameters that were manually set

Function	Parameter	Tested range	Final value	
			Phase 1	Phase 2
Probability	p_{aug}	0.25–1	$p_1 = 0.6$	$p_2 = 0.85$
Learning rate	*lr*	10^{-5}–2×10^{-2}	10^{-3}	
Spatial				
Rotation	α_{max} [°]	5–20	15	
Translation	Δ_{max} [mm]	15 — 30	22.5	
Zooming	z_{min}, z_{max}	–	0.9, 1.1	
Deformation	n_{cp}	5–20	–	8
	d [mm]	15–45	–	24
MR				
Motion	m_α[°]	0–15	–	10
	m_Δ [mm]	15–75	–	45
Bias field	order	1–3	–	1
	c_{mag}	0–1	–	0.4
Noise	σ	0.01–0.25	0.05	0.1
	μ	–	0	

Tested range (min–max) and final chosen value for each training phase are given. For more information, refer to documentation of corresponding MONAI or TorchIO functions

Geometric evaluation

Figure 1 shows an exemplary MR image with clinical and DLAS contours. Differences in lung and heart contours can be observed on the coronal slices. Differences in length between the spinal canal contours on both ends and the superior end of the heart contour can be observed on the sagittal slices. The chosen axial slices in the high dose region show good agreement between both sets of contours.

DSC for large OARs (lungs and heart) was high with an averaged median value over the three OARs [interquartile range (IQR)] of 0.95 [0.95–0.96]. For the tubular ROIs (spinal canal, esophagus and aorta), DSC was lower with an average median value of 0.86 [0.78–0.88]. The averaged median value of both HD_{95} (5.0 mm [3.9–6.2 mm] vs 3.0 mm [2.3–5.8 mm]) and HD_{avg} (1.6 mm [1.4–1.9 mm] vs. 1.1mm [0.8–1.6 mm]) was lower for the second group, with the IQR being similar for both groups.

The results per OAR are summarized in Table 4 and visualized in boxplots in Fig. 2.

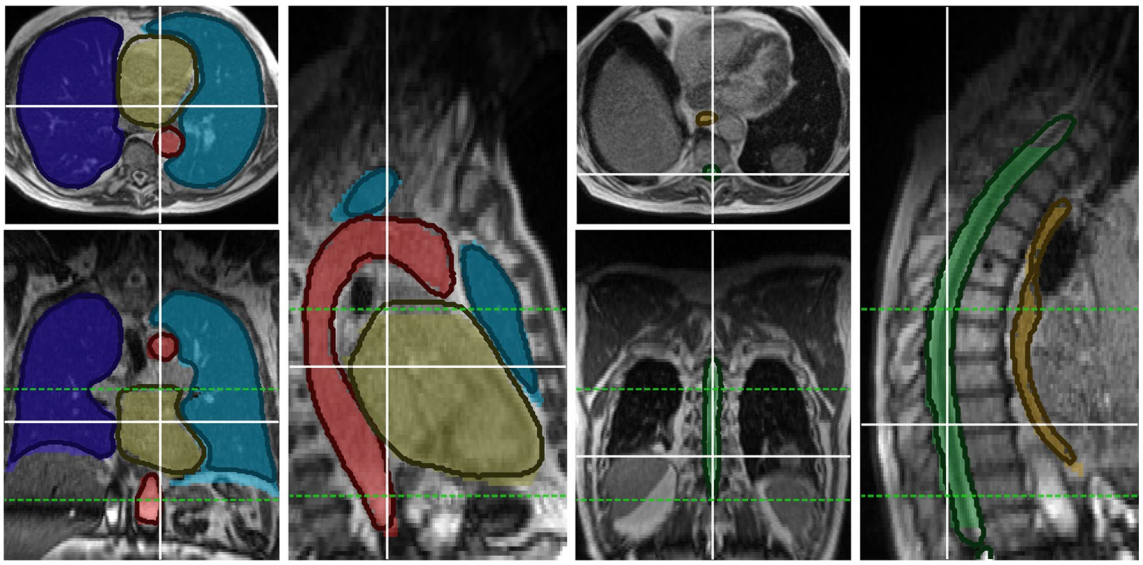

Fig. 1 Example DLAS contours (solid outline) and GT contour (colored overlay) in coronal, axial and sagittal view, subdivided into images containing lungs, heart and aorta for the left three panels, and esophagus and spinal canal for the right three panels. The dotted, green lines represent slices containing the PTV ± 10 slices in superior and inferior direction, where the geometrical evaluation of the esophagus, spinal canal and aorta was performed. MR images from a single patient shown

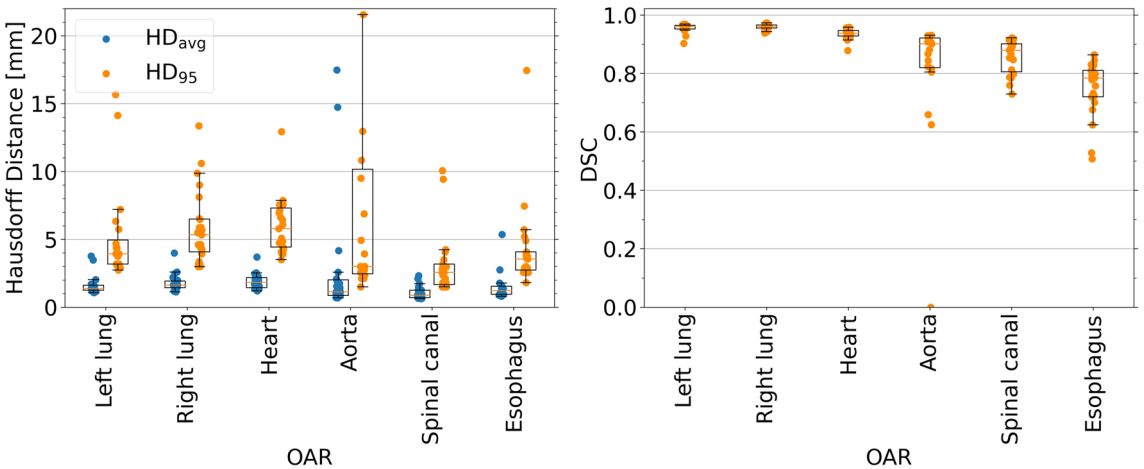

Fig. 2 Box plots of Hausdorff distance (left) and Dice similarity coefficient (right) for all test set contours per ROI. Not on display in the left plot are two data points for the HD_{95} of the aorta at 29.8 mm and 37.9 mm

Table 4 Summary of geometrical analysis, showing DSC and HD (95th percentile and average value) for each segmented OAR, median [IQR]

	DSC	HD_{95} (mm)	HD_{avg}
Left lung	0.96 [0.95–0.96]	3.9 [3.2–4.9]	1.4 [1.2–1.6]
Right lung	0.96 [0.96–0.97]	5.3 [4.1–6.5]	1.6 [1.4–1.9]
Heart	0.94 [0.93–0.95]	5.8 [4.4–7.3]	1.8 [1.4–2.2]
Aorta	0.90 [0.82–0.92]	3.0 [2.5–10.2]	1.1 [0.9–2.0]
Spinal canal	0.88 [0.81–0.90]	2.6 [1.7–3.2]	0.9 [0.7–1.3]
Esophagus	0.78 [0.72–0.81]	3.5 [2.7–4.1]	1.2 [1.0–1.5]

Physician's grading

The physician's grading favored the DLAS contours, which were given the rating "0—no clinically relevant correction possible" 85 times, compared to 18 times for clinical contours. This was most apparent for the lungs, the spinal canal and the aorta. The DLAS as well as the clinical contours of heart and esophagus received lower scores than the other OARs on average.

85% of DLAS and 65% of clinical contours were deemed at least "ready to use" (grades 0 or 1). In more detail, 70% and 61% of heart and esophagus DLAS contours and 48% and 67% of the clinical contours respectively

received these ratings. Of the remaining OARs (lungs, aorta and spinal canal), 95% of DLAS contours received a grading of "no clinically relevant correction possible" or 1—"ready to use", versus 69% of clinical contours. A more detailed breakdown can be seen in Fig. 3.

Discussion

In this study, DLAS for 0.35 T MR-Linac planning images of lung tumor patients was evaluated via geometric analysis by comparing it to the clinically used contours, and via clinical grading by a radiation oncologist. The geometric analysis showed that the DLAS contours were close to the clinically used ones. We achieved reasonably high DSCs for all OARs, which were in line with or better than average (where curated challenge datasets were not used) CT-based DLAS results [31, Table 1] [32]. As mentioned previously, due to the lack of studies on MRI-based DLAS, this was our only basis for comparison. In the majority of cases, the DLAS contours were preferred over the clinical contours by the radiation oncologist.

For both lungs, a DSC of 0.96 was achieved and all DLAS contours received the best grade (0). While high DSC values are more easily achieved by large, high contrast organs, manual lung segmentation is time consuming in the clinical TPS version currently in use due to a lack of automation tools. Our model therefore may lead to substantial time savings. The main differences were attributed to differing contouring styles near the bronchial tree or the tumor, which were included in a few cases, but usually excluded for clinical contours.

The tubular shape of the spinal canal and the aorta result in an increased surface area to volume ratio, and therefore DSCs were lower. Nonetheless, HD was comparable to the lungs. Only 2 out of 23 DLAS contours for the spinal canal and 2 out of 19 for the aorta were deemed to require any kind of correction. These were mostly due to small holes in the structures (grade 3) or some slices only having the OAR partially contoured on one side (grade 2), e.g. due to MRI artifacts. Examples can be seen in Fig. 4.

The average DSC and HD achieved for the heart were comparable to the lungs, which is partly due to a large volume with comparatively small surface area. Grading resulted in 6 out of 23 contours requiring minor corrections, while 1 required major corrections. These were mainly due to the heart wall not being included in the DLAS contour.

As with previous studies on CT images by other groups [31, Table 1] [32], average DSC scores of the esophagus contours were worse than for all other OARs. The poor contrast to the surrounding tissue and possibly decreased sharpness due to abdominal motion, make it difficult to segment for both physicians and the network. The much larger variety in possible shapes in the axial slices and

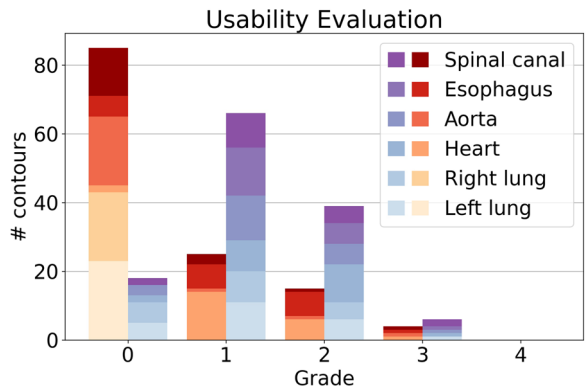

Fig. 3 Physician's evaluation of test set contours (total of 129, see Table 2), subdivided by OARs. DLAS contours on the left in (red hues), clinical contours on the right (blue hues). Grade descriptions: 0—"no clinically relevant correction possible", 1—"ready to use", 2—"minor corrections required", 3—"major corrections required", 4—"unusable"

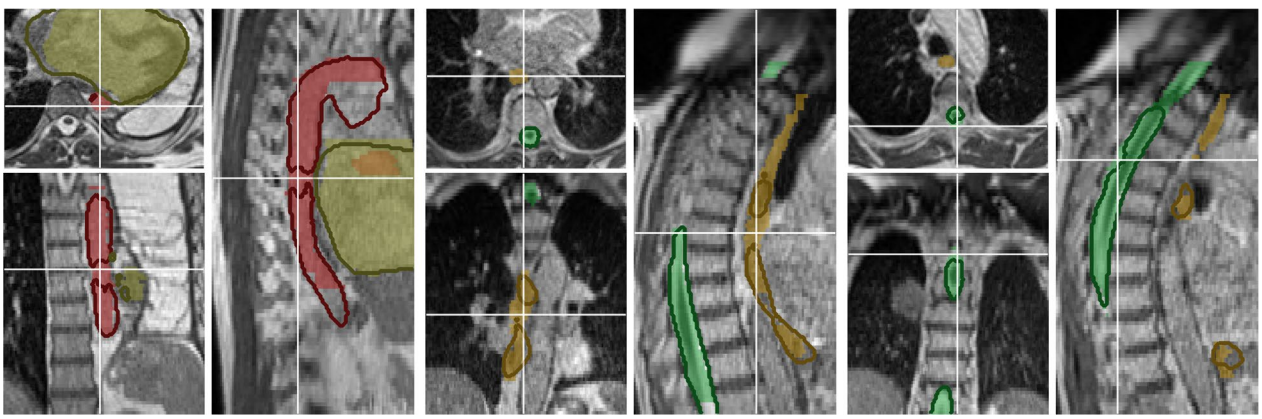

Fig. 4 Examples of poor DLAS contours (solid outline) compared to GT contour (colored overlay) in coronal, axial and sagittal view of the aorta (red, left) for P23 and esophagus (orange, middle) and the spinal canal (green, right) for P14

differing lengths of the contour in the training data are also possible reasons for the worse performance for this OAR. Other groups such as Fu et al. [14] reported similar problems with the duodenum, which behaves similarly to the esophagus in this context.

Segmentation on patients with an uncommon anatomy, such as a collapsed or removed lung, were also included (1 in training, 1 in validation, 2 in the test set, P18 and P23). DLAS contours of test case P18 received a near perfect score (grade 0 for right lung, spinal canal, esophagus and aorta, grade 1 for heart), whereas test case P23's DLAS contours received the worst grading overall (grade 0 for right lung and spinal canal, grade 2 for heart and esophagus, grade 3 for aorta).

In Fig. 2, five cases stand out (aorta segmentations with DSC of 0.00, 0.63 and 0.66, P3, P21 and P23 respectively, and esophagus segmentations with DSC of 0.51 and 0.53, P13 and P14 respectively). The largest deviation in DSC was the aorta segmentation for P3 with a DSC of 0 and a HD_{95} of 30 mm. In this case, the ground truth contour was only done in the inferior part of the MRI, despite the target region being located in the superior part. The physician's grading of this patient was 0 for the DLAS and 3 for the ground truth. Similarly, P21 had a DSC of 0.63 and a HD_{95} of 38 mm for the aorta, but was graded with 0 for DLAS, and the ground truth was graded with 1. The poor DSC and HD were due to the network fully segmenting the aorta, while the clinical contour did not include the ascending aorta and the arch of the aorta, since they were farther away from the tumor. Excluding these cases would lead to median [IQR] values of 0.91 [0.84–0.92], 2.9 [2.4–6.9] mm, 1.0 [0.9–1.7] mm for DSC, HD_{95} and HD_{avg} respectively for the aorta. The aorta DLAS contour of P23 had a DSC of 0.66 and a a HD_{95} of 22 mm and received a grade of 3. In this case, the DLAS failed by including a sizable part of the heart in the segmentation, exhibiting a minor hole and not continuing the contour far enough into the heart. The esophagus DLAS contours had two cases with large deviations. P13's DLAS contour was however graded with 0, compared to the ground truth's 1, despite the DSC of 0.51 and HD_{95} of 17 mm. For P14, DSC and HD_{95} were 0.53 and 7.4 mm respectively. The DLAS contours received a grading of 3 and the ground truth a grading of 1. Here, the DLAS contour exhibited a hole in the middle of the contour.

In these cases, it can be concluded that the contour quality is not always well reflected by the DSC and HD. A geometric analysis using clinical data as the ground truth has its limitations, as evidenced by Vaassen et al. [33]. For example, a segmentation with a good geometric evaluation can still lead to low gamma pass rates when used for treatment planning, as indicated by Kawula et al. [34]. The grading system was found to be more helpful at gauging the quality of DLAS contours for the purpose of this study. We also acknowledge that the physician might look at the contours differently in a review compared to a treatment setting. Clinical contours are generated during treatment workflows, and the physicians tend to focus on the high dose region around the lesion, which is most relevant for treatment planning. Clinical contour quality farther away from the treatment region may thus be lower.

This appears to not have noticeably hindered training, as most deviations appear to even out in DLAS models, given enough training data. Lustberg et al. [35] have also found that using models trained on non-curated local data could still save 50% time compared to manual contouring. Nonetheless, the variations in length of the esophagus, spinal canal and aorta contours remain a challenge. We used masks to only consider the tumor region when selecting the best model during validation, but did not take any measures to alter the training process in this regard. This was intentional, as any clipping of the masks would result in DLAS contours being shorter in general. We judged that, in a scenario where these contours are presented to a physician as a starting point, it would be more time efficient for them to remove or ignore distal, inaccurate parts, as opposed to having to expand a contour that is too short. However, we suspect that the inconsistent length of these contours in the training data might be a contributing factor for the occasional holes in these OARs. A more uniform training set, acquired by re-contouring the training data, could lead to some further segmentation improvements for the underperforming OARs (esophagus and heart). Similarly, patient specific model fine-tuning with a single training patient would likely create more consistent contours (as demonstrated by Kawula et al. [16]). Albeit that would only apply to the fraction images, as opposed to new planning images.

The goal of automatic segmentation is to help physicians with delineating structures. This means reducing the time spent manually delineating structures and decreasing observer variability. Evaluation methods need to be chosen with these aspects in mind. The DLAS contours should therefore not perfectly fit the existing ground truth, but rather require as few corrections as possible. In a next step, we will quantify the time saved in the MRgRT workflow by prospectively providing physicians with these contours for OAR delineation in treatment planning [36, 37].

Conclusion

In conclusion, we trained U-Nets for contouring the lungs, heart, aorta, spinal canal and esophagus on thoracic images from an 0.35 T MR-Linac. They were able

to produce contours that were most of the time preferred to the clinical contours by a radiation oncologist.

Abbreviations

MR	Magnetic resonance
MRI	Magnetic resonance image
MRgRT	Magnetic resonance (imaging) guided radiotherapy
DLAS	Deep learning autosegmentation
DSC	Dice similarity coefficient
HD	Hausdorff distance
SBRT	Stereotactic body radiotherapy
CBCT	Cone beam computed tomography
GTV	Gross tumor volume
MR-Licac	MR-guided linear accelerator
OAR	Organ-at-risk
DIR	Deformable image registration
ANN	Artificial neural network
CNN	Convolutional neural network
CTV	Clinical target volume
TPS	Treatment planning system
mha	Meta image format
nn	Nearest neighbour
HD_{avg}	Average Hausdorff distance
HD_{95}	95th percentile Hausdorff distance
GT	Ground truth
PTV	Planning target volume
IQR	Interquartile range

Acknowledgements
Special thanks to Antonio Carrasco and Samira Hosseini for exporting a majority of the patient data.

Author contributions
MFR ran the study, trained the neural networks, performed the data analysis and wrote the manuscript. SM graded the contours, provided clinical insights and reviewed the manuscript. MK provided the core of the 3D-Unet, participated in exporting and augmenting the data and in hyperparameter tuning. MRa was involved in discussions during all stages of the study and reviewed the paper. SC provided clinical insights and reviewed the manuscript. CB and MRi reviewed the manuscript. GL and CK designed the study, participated in all stages of this work, especially data preparation, network training, data analysis and writing the manuscript. All authors read and approved of the manuscript.

Funding
M.K. was funded by the Wilhelm Sander-Stiftung (2019.162.1 and 2019.162.2). M.Ra. was supported by the Munich Medical & Clinician Scientist Program (MCSP) by the Medical Faculty of the LMU Munich.

Availability of data and materials
The data used in this study cannot be made available due to data protection regulations.

Declarations

Ethics approval and consent to participate
All patients provided informed written consent within the scope of an ethically approved study protocol in place at the Department of Radiation Oncology of the LMU Munich University Hospital (ethics project number 20-291).

Consent for publication
Not applicable.

Competing interests
The Department of Radiation Oncology of the LMU University Hospital has received research grants from ViewRay Inc. (Oakwood Village, OH, USA). ViewRay did not fund this study, was not involved and had no influence on the study design, the collection or analysis of data, on the writing of the manuscript, or the decision to submit the manuscript for publication.

Author details
[1]Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany. [2]German Cancer Consortium (DKTK), Partner Site Munich, Munich, Germany. [3]Bavarian Cancer Research Center (BZKF), Munich, Germany. [4]Department of Medical Physics, Ludwig-Maximilians-Universität München, Garching, Germany.

Received: 21 April 2023 Accepted: 3 August 2023
Published online: 14 August 2023

References

1. Finazzi T, Palacios MA, Haasbeek CJ, Admiraal MA, Spoelstra FO, Bruynzeel AM, Slotman BJ, Lagerwaard FJ, Senan S. Stereotactic MR-guided adaptive radiation therapy for peripheral lung tumors. Radiother Oncol. 2020;144:46–52.
2. de Koste JRVS, Palacios MA, Bruynzeel AM, Slotman BJ, Senan S, Lagerwaard FJ. MR-guided gated stereotactic radiation therapy delivery for lung, adrenal, and pancreatic tumors: a geometric analysis. Int J Radiat Oncol Biol Phys. 2018;102(4):858–66.
3. Klüter S. Technical design and concept of a 0.35 t mr-linac. Clin Transl Radiat Oncol. 2019;18:98–101.
4. Corradini S, Alongi F, Andratschke N, Belka C, Boldrini L, Cellini F, Debus J, Guckenberger M, Hörner-Rieber J, Lagerwaard F, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. Radiat Oncol. 2019;14(1):1–12.
5. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. Sem Radiat Oncol. 2019;29:185–97.
6. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. Radiother Oncol. 1998;47(3):285–92.
7. Rabe M, Palacios MA, van Sörnsen de Koste JR, Eze C, Hillbrand M, Belka C, Landry G, Senan S, Kurz C. Comparison of MR-guided radiotherapy accumulated doses for central lung tumors with non-adaptive and online adaptive proton therapy. Med Phys. 2023. https://doi.org/10.1002/mp.16319.
8. Sahin B, Mustafayev TZ, Gungor G, Aydin G, Yapici B, Atalar B, Ozyar E. First 500 fractions delivered with a magnetic resonance-guided radiotherapy system: initial experience. Cureus. 2019. https://doi.org/10.7759/cureus.6457.
9. Hadi I, Eze C, Schönecker S, von Bestenbostel R, Rogowski P, Nierer L, Bodensohn R, Reiner M, Landry G, Belka C, et al. MR-guided SBRT boost for patients with locally advanced or recurrent gynecological cancers ineligible for brachytherapy: feasibility and early clinical experience. Radiat Oncol. 2022;17(1):1–9.
10. Rogowski P, von Bestenbostel R, Walter F, Straub K, Nierer L, Kurz C, Landry G, Reiner M, Auernhammer CJ, Belka C, et al. Feasibility and early clinical experience of online adaptive MR-guided radiotherapy of liver tumors. Cancers. 2021;13(7):1523.
11. Lamb J, Cao M, Kishan A, Agazaryan N, Thomas DH, Shaverdian N, Yang Y, Ray S, Low DA, Raldow A, et al. Online adaptive radiation therapy: implementation of a new process of care. Cureus. 2017. https://doi.org/10.7759/cureus.1618.
12. Cusumano D, Boldrini L, Dhont J, Fiorino C, Green O, Güngör G, Jornet N, Klüter S, Landry G, Mattiucci GC, et al. Artificial intelligence in magnetic resonance guided radiotherapy: medical and physical considerations on state of art and future perspectives. Phys Med. 2021;85:175–91.
13. Liang F, Qian P, Su K-H, Baydoun A, Leisser A, Van Hedent S, Kuo J-W, Zhao K, Parikh P, Lu Y, et al. Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: An intelligent, multi-level fusion approach. Artif Intell Med. 2018;90:34–41.
14. Fu Y, Mazur TR, Wu X, Liu S, Chang X, Lu Y, Li HH, Kim H, Roach MC, Henke L, et al. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. Med Phys. 2018;45(11):5129–37.

15. Eppenhof KA, Maspero M, Savenije M, de Boer J, Van der Voort van Zyp J, Raaymakers BW, Raaijmakers A, Veta M, van den Berg C, Pluim JP. Fast contour propagation for MR-guided prostate radiotherapy using convolutional neural networks. Med Phys. 2020;47(3):1238–48.
16. Kawula M, Hadi I, Nierer L, Vagni M, Cusumano D, Boldrini L, Placidi L, Corradini S, Belka C, Landry G, et al. Patient-specific transfer learning for auto-segmentation in adaptive 0.35 T MRgRT of prostate cancer: a bi-centric evaluation. Med Phys. 2023;50(3):1573–85.
17. Chen X, Ma X, Yan X, Luo F, Yang S, Wang Z, Wu R, Wang J, Lu N, Bi N, et al. Personalized auto-segmentation for magnetic resonance imaging-guided adaptive radiotherapy of prostate cancer. Med Phys. 2022;49(8):4971–9.
18. Fransson S, Tilly D, Strand R. Patient specific deep learning based segmentation for magnetic resonance guided prostate radiotherapy. Phys Imag Radiat Oncol. 2022;23:38–42.
19. Li Z, Zhang W, Li B, Zhu J, Peng Y, Li C, Zhu J, Zhou Q, Yin Y. Patient-specific daily updated deep learning auto-segmentation for MRI-guided adaptive radiotherapy. Radiother Oncol. 2022;177:222–30.
20. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11.
21. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, Liu T, Yang X. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys. 2019;46(5):2157–68.
22. Sharp, G.C., Li, R., Wolfgang, J., Chen, G., Peroni, M., Spadea, M.F., Mori, S., Zhang, J., Shackleford, J., Kandasamy, N. Plastimatch: an open source software suite for radiotherapy image processing. In: Proceedings of the XVI'th International conference on the use of computers in radiotherapy (ICCR), Amsterdam, Netherlands 2010.
23. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of SimpleITK. Front Neuroinform. 2013;7:45.
24. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32 2019.
25. Ma N, Li W, Brown R, et al. Project MONAI. CERN: Zenodo; 2021.
26. Kerfoot E, Clough J, Oksuz I, Lee J, King AP, Schnabel JA. Left-ventricle quantification using residual u-net. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young A, Rhode K, Mansi T, editors. Statistical Atlases and Computational Models of the Heart: Atrial Segmentation and LV Quantification Challenges. Cham: Springer; 2019. p. 371–80. https://doi.org/10.1007/978-3-030-12029-0_40.
27. Milletari, F., Navab, N., Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 2016;. IEEE.
28. Kingma, D.P., Ba, J. 2014; Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.
29. Pérez-García F, Sparks R, Ourselin S. TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput Methods Programs Biomed. 2021;208: 106236.
30. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imag. 1999;18(10):897–908.
31. Zhang T, Yang Y, Wang J, Men K, Wang X, Deng L, Bi N. Comparison between atlas and convolutional neural network based automatic segmentation of multiple organs at risk in non-small cell lung cancer. Medicine. 2020;99(34):e21800.
32. Liu X, Li K-W, Yang R, Geng L-S. Review of deep learning based automatic segmentation for lung cancer radiotherapy. Front Oncol. 2021;11: 717039.
33. Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, van Elmpt W. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. Phys Imag Radiat Oncol. 2020;13:1–6.
34. Kawula M, Purice D, Li M, Vivar G, Ahmadi S-A, Parodi K, Belka C, Landry G, Kurz C. Dosimetric impact of deep learning-based CT auto-segmentation on radiation therapy treatment planning for prostate cancer. Radiat Oncol. 2022;17(1):21.
35. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, van Elmpt W, Dekker A. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol. 2018;126(2):312–7.
36. Vaassen F, Boukerroui D, Looney P, Canters R, Verhoeven K, Peeters S, Lubken I, Mannens J, Gooding MJ, van Elmpt W. Real-world analysis of manual editing of deep learning contouring in the thorax region. Phys Imag Radiat Oncol. 2022;22:104–10.
37. Brouwer CL, Boukerroui D, Oliveira J, Looney P, Steenbakkers RJ, Langendijk JA, Both S, Gooding MJ. Assessment of manual adjustment performed in clinical practice following deep learning contouring for head and neck organs at risk in radiotherapy. Phys Imag Radiat Oncol. 2020;16:54–60.

Publisher's Note