

RESEARCH

Open Access



A deep image-to-image network organ segmentation algorithm for radiation treatment planning: principles and evaluation

Sebastian Marschner^{1,6*}, Manasi Datar², Aurélie Gaasch¹, Zhoubing Xu³, Sasa Grbic³, Guillaume Chabin³, Bernhard Geiger³, Julian Rosenman⁴, Stefanie Corradini¹, Maximilian Niyazi¹, Tobias Heimann², Christian Möhler⁵, Fernando Vega⁵, Claus Belka¹ and Christian Thieke¹

Abstract

Background: We describe and evaluate a deep network algorithm which automatically contours organs at risk in the thorax and pelvis on computed tomography (CT) images for radiation treatment planning.

Methods: The algorithm identifies the region of interest (ROI) automatically by detecting anatomical landmarks around the specific organs using a deep reinforcement learning technique. The segmentation is restricted to this ROI and performed by a deep image-to-image network (DI2IN) based on a convolutional encoder-decoder architecture combined with multi-level feature concatenation. The algorithm is commercially available in the medical products “syngo.via RT Image Suite VB50” and “AI-Rad Companion Organs RT VA20” (Siemens Healthineers). For evaluation, thoracic CT images of 237 patients and pelvic CT images of 102 patients were manually contoured following the Radiation Therapy Oncology Group (RTOG) guidelines and compared to the DI2IN results using metrics for volume, overlap and distance, e.g., Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD₉₅). The contours were also compared visually slice by slice.

Results: We observed high correlations between automatic and manual contours. The best results were obtained for the lungs (DSC 0.97, HD₉₅ 2.7 mm/2.9 mm for left/right lung), followed by heart (DSC 0.92, HD₉₅ 4.4 mm), bladder (DSC 0.88, HD₉₅ 6.7 mm) and rectum (DSC 0.79, HD₉₅ 10.8 mm). Visual inspection showed excellent agreements with some exceptions for heart and rectum.

Conclusions: The DI2IN algorithm automatically generated contours for organs at risk close to those by a human expert, making the contouring step in radiation treatment planning simpler and faster. Few cases still required manual corrections, mainly for heart and rectum.

Keywords: Radiation, Planning, Algorithm, Deep-image-to-image-network, Contour, Organs at risk, Thorax, Pelvis

Introduction

In radiation treatment planning, delineating the target volumes and organs at risk (OAR) is one of the most important and time-consuming tasks. The dose-volume histogram analysis for plan evaluation, contour-based visual guidance in image-guided radiation therapy, and the dose-response assessment of radiation side effects are depending on the accuracy of the delineation [1, 2].

*Correspondence: sebastian.marschner@med.uni-muenchen.de

⁶ Department of Radiation Oncology, LMU Klinikum, Marchioninistr. 15, 81377 München, Germany
Full list of author information is available at the end of the article



© The Author(s) 2022, corrected publication 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The contouring workload will further increase due to more widely used strategies of adaptive planning where the treatment plan is adapted to anatomical changes such as tumor shrinkage, requiring re-contouring during the fractionated treatment course.

The need for fast and accurate delineation has led to a variety of automated computer-based approaches. Most autosegmentation algorithms used in clinical practice today are atlas-based and required sophisticated atlas creation of self-made contours, with difficulties in case of anatomical variations, low contrast organs or when the anatomy is modified by the presence of a tumor. In addition, atlas-based segmentation is computationally intensive and can take several minutes to complete [3–5]. In the last years, artificial intelligence (AI) and machine learning approaches were developed for autosegmentation aiming to improve accuracy and shorten the time needed for segmentation even in complex anatomical situations [6–8].

In this work, we describe and evaluate an AI algorithm for autosegmentation of organs at risk based on a deep image-to-image-network (DI2IN).

Materials and methods

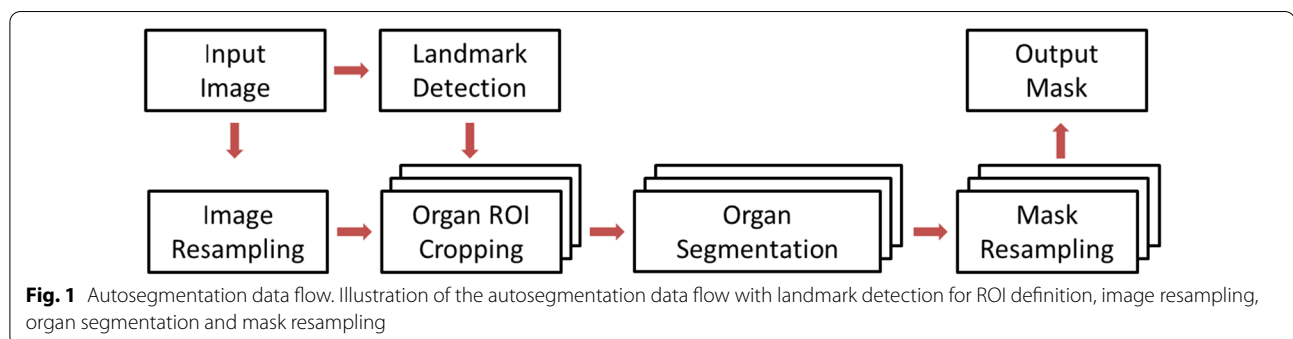
Autosegmentation with a deep image-to-image network (DI2IN)

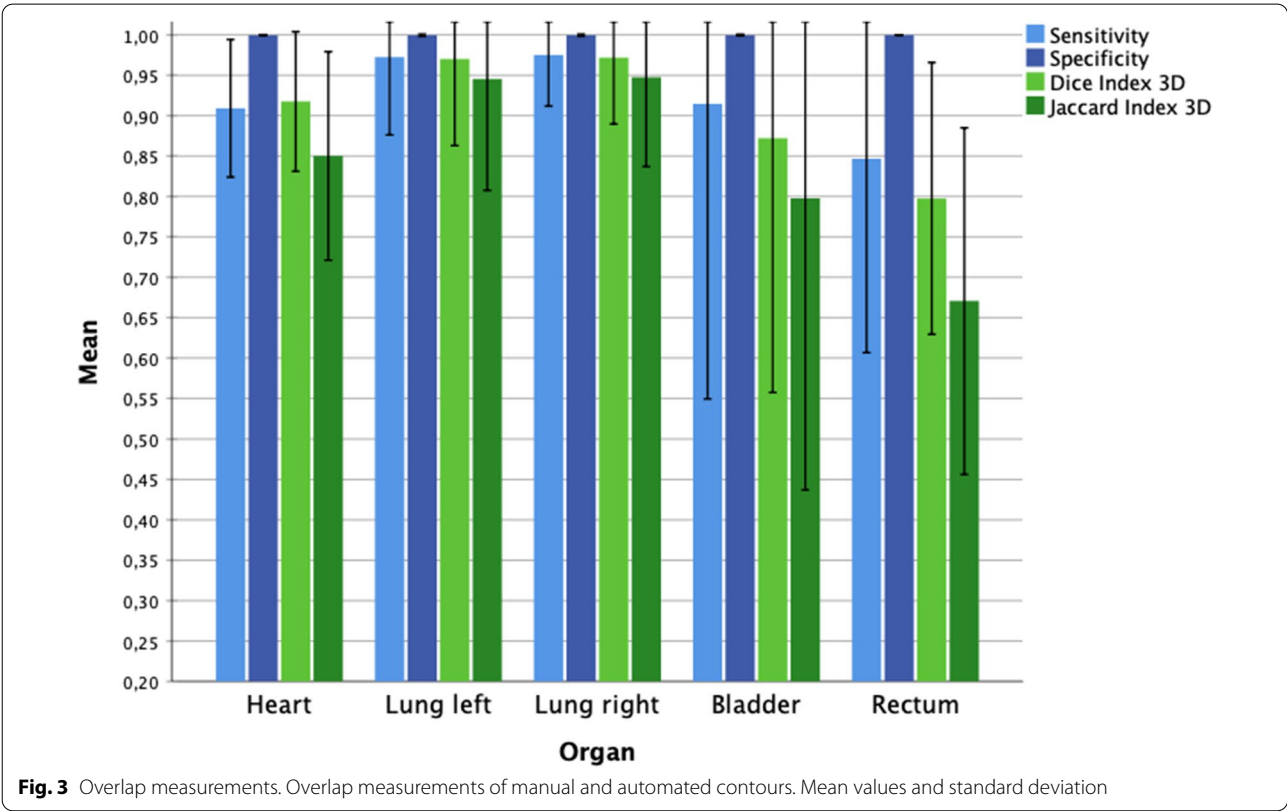
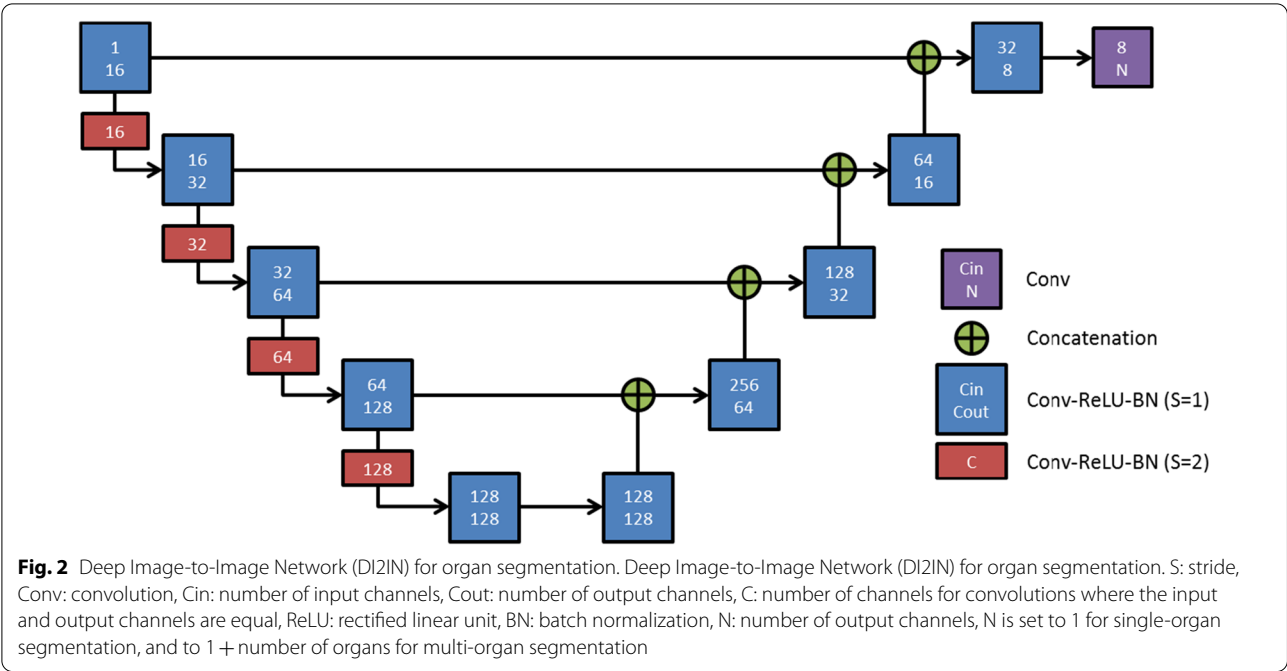
The general data flow of the autosegmentation algorithm provided by SIEMENS Healthineers is illustrated in Fig. 1. The input image is the computed tomography (CT) image of the patient in full resolution. Image resampling is applied as normalization to acquire an isotropic image to present the realistic aspect ratio of the anatomical structures. Due to the high complexity and thus resource requirement from the organ segmentation model, the image is typically downsampled to reduce the computational burden of the algorithm execution.

The automatic segmentation is performed on the region of interest (ROI) around individual organs or a group of multiple organs instead of the entire image volume. This helps the segmentation model to focus on

capturing variations from the organs themselves without disruptions of irrelevant structures and significantly reduces the computational resources. The anatomical landmark detection was trained independently from segmentation algorithms with manually annotated landmark points across the human body as described by Ghesu et al. [9]. To locate the ROIs, anatomical landmarks (including vessel bifurcations, bony structures, and organ center and boundary points) are detected using a deep reinforcement learning technique [9] from the input image. For each landmark, an agent is trained to search for the best path to walk towards the landmark from any location of the image. Specifically, considering the current state as an image patch centred at the current voxel, the agent learns to take one of the actions from the current voxel so that the distance towards the landmark of interest is minimized. During testing time, the agent will move one step at a time and eventually stop at or around the desired landmark where the action estimation converges. To reduce the computational costs, a multi-stage strategy is integrated to search the landmark position at different scales of the image resolution, where the action at the coarser resolution will move the agent close to the landmark position with effectively larger step size. Given the detected landmarks and their heuristic relationships with the organs, ROIs are cropped with the sizes derived from training data distribution with its center based on the associated landmark position and its size being large enough to cover the organs to be segmented.

A deep image-to-image network (DI2IN) [10] based on a convolutional encoder-decoder architecture combined with multi-level feature concatenation is employed for the automatic segmentation step (Fig. 2). Compared to traditional U-Net [6], additional convolutional layers with stride of 2 (red blocks in Fig. 2) are used in the encoder of DI2IN instead of max pooling layers to increase the receptive field while reducing sizes of feature maps. In the decoder of DI2IN, trilinear interpolation is used to upsample the activation maps back to the original input image size. During the training process, the network was





driven by a cross-entropy loss based on a learning rate of 0.001 using the ADAM [11] optimization. The algorithm was trained for the segmentation of left and right

lung (using 10,000 cases), heart (386 cases), and bladder and rectum including anus and rectosigmoid flexure (784 cases). The training CT data were collected from multiple

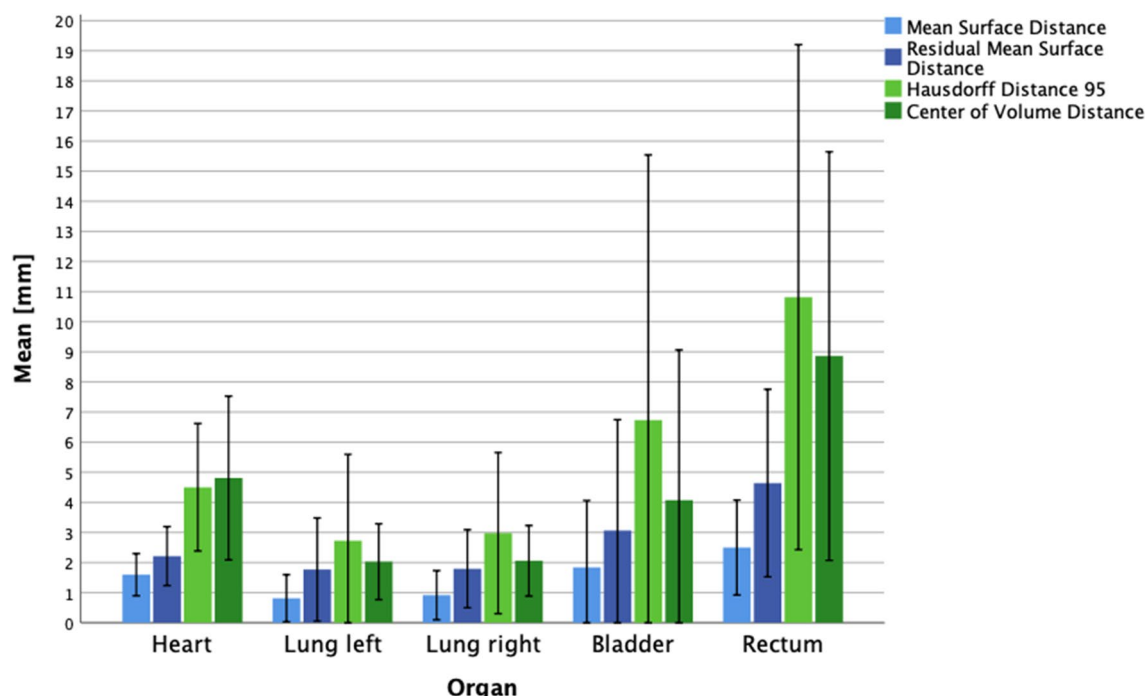


Fig. 4 Distance measurements. Distance measurements between manual and automated contours. Mean values and standard deviation

hospital sites. Data annotation was performed based on RTOG guidelines by a team trained with anatomical knowledge and mentored by radiologists and radiation oncologists.

After organ segmentation, the estimated organs mask is resampled back to the original image resolution, where each organ-specific mask is aggregated in a single multi-organ mask.

Evaluation patient cohort

An independent evaluation patient cohort was established from CT images of patients treated at LMU (Ludwig-Maximilians-Universität) university hospital. The scans were acquired for treatment planning without contrast medium on a Toshiba CT scanner with 3 mm slice thickness. For the thoracic region, 237 female patients treated for breast cancer were included, resulting in 237 usable heart contours and 233/234 usable left/right lung contours. For the pelvic region, 102 male and female patients treated for various tumors (e.g. cervical and prostate cancer) were included resulting in 98 usable bladder and 102 usable rectum contours. OARs with gross tumor volume or tumor infiltration were excluded.

The CT data was anonymized for the scientific purpose of this work. This study complies with the declaration of Helsinki, Good Clinical Practice (GCP) and Good Epidemiological Practice (GEP). The data acquisition and

analysis were in accordance with Bavarian hospital law (Art.27 Abs. 4 BayKrG).

Manual segmentation

All manual contours were drawn by an experienced radiation oncologist following the guidelines of the RTOG [12, 13] using Oncentra Masterplan by Elekta AB, Sweden. *Lungs*: All inflated and collapsed, fibrotic and emphysematic lungs were contoured including small vessels extending beyond the hilar regions; hilars and trachea/main bronchus were not included. *Heart*: Contoured along the pericardial sac. The superior aspect (base) began at the level of the inferior aspect of the pulmonary artery passing the midline and extend inferiorly to the apex of the heart. *Bladder*: Contoured inferiorly from its base, and superiorly to the dome. *Rectum*: Contouring ended inferiorly from the lowest level of the ischial tuberosities (right or left), and superiorly before the rectum lost its round shape in the axial plane and connected anteriorly with the sigmoid.

Comparison of manual and automatic segmentation

Manual contours (MAN) were considered as ground truth and were compared to the automatic contours (AUTO) generated by a software prototype (provided by Siemens Healthineers) of the DI2IN algorithm. We used several quantitative geometric measures in the

categories volume (absolute and ratio), overlap (Sensitivity, Specificity [14], Jaccard Conformity Index [15, 16], Dice Similarity Coefficient [17], Discordance Index [18, 19], Geographical Miss Index) and distance (Mean Surface Distance [14], Center of Volume Distance, Residual Mean Surface Distance [20, 21], Hausdorff Distance HD_{95} [22], and difference of the superior, inferior, right, left, anterior and posterior boundaries defined by the furthest reaching voxel belonging to the contour in the respective direction). All formulas used are summarized in the appendix. All results were imported into IBM SPSS Statistics version 25.0.0 and subsequently processed and analysed.

Additionally, MAN and AUTO were inspected visually for identification of regions which are still challenging for the algorithm.

Results

General algorithm performance

In all cases, the DI2IN algorithm was able to generate the automatic contours without any user interaction. The computation with the prototype took roughly 30 s per organ.

Volume comparison

The volumes of manual (MAN) and automatic (AUTO) contours are summarized in Table 1. The highest variations in absolute volume from patient to patient were observed for the left and right lung as the largest structure types, however, the mean volume difference between manual and automatic contours e.g. of the left lung were only 17 ml. Also, for all other organs the absolute volumes of manual and automatic contours were similar, with the volume ratio either 0.9 or 1.0.

Overlap and distance measurements

The results for all metrics of overlap and distance comparisons are summarized in Table 2.

The overlap measurements sensitivity, specificity, Jaccard Conformity Index (JCI) and Dice Similarity Coefficient (DSC) are illustrated in Fig. 3. The sensitivity showed values of 0.98 for the left/right lungs, 0.93 for the bladder and 0.91 for the heart, and the lowest value for

the rectum with 0.84. The specificity was excellent with 0.99 for all structure types. The JCI with mean values of 0.95 for the right and left lung showed nearly complete overlap between manual and automatic contours. Again, the poorest result was obtained for the rectum with a JCI of 0.67. The same ranking was seen for the DSC, with best values for left/right lung (DSC 0.97), then heart (DSC 0.92), bladder (DSC 0.88) and rectum (DSC 0.79).

The Discordance Index (DisI) performed best for left and right lung with only 3% of AUTO being outside MAN. Heart, bladder and rectum reach 6%, 13% and 22%, respectively. Comparable results were seen for the Geographical Miss Index (GMI), where for the left/right lungs 2/3% of MAN were outside AUTO, 8% for the bladder, 7% for the heart, and 16% for the rectum.

The distance measurements in terms of center of volume distance, MSD, RMSD and HD_{95} are illustrated in Fig. 4. The MSD and RMSD showed mean values between 0.8 and 4.6 mm for all organs. The Hausdorff distance HD_{95} was best for left /right lung with of 2.7/2.9 mm, and worst for the rectum with 10.8 mm

Most boundary differences were around or below 3 mm (the thickness of one CT slice). Bigger discrepancies were seen for the superior, inferior and anterior boundaries of the rectum with -8.7 mm, 7.0 mm and 5.2 mm respectively, and the inferior and superior boundary of the heart with a mean deviation of -8.5 mm and -4.9 mm respectively.

The visual inspection showed an overall excellent agreement between manual and automated contours, with the most challenging organs, as already identified by the boundary analysis, being the rectum and the heart. Two exemplary cases are shown in Fig. 5.

Discussion

The quantitative comparison between MAN and AUTO contours showed an excellent agreement in most cases for all geometric metrics in terms of volume, overlap and distance, with some exceptions especially for the heart and the rectum. The discrepancies for the rectum can be explained by the fact that the training data set did include anus and the rectosigmoid flexure, but the manual segmentation did not. Some cases would require

Table 1 Volume comparison

	Bladder mean \pm SD	Rectum mean \pm SD	Heart mean \pm SD	Lung left mean \pm SD	Lung right mean \pm SD
Volume MAN [ml]	331 \pm 178	89 \pm 39	569 \pm 115	1,897 \pm 604	2,220 \pm 647
Volume AUTO [ml]	362 \pm 209	97 \pm 44	557 \pm 108	1,914 \pm 623	2,240 \pm 668
Difference (MAN-AUTO) [ml]	32 \pm 74	8.1 \pm 22	11 \pm 54	17 \pm 158	19 \pm 151
Ratio (MAN/AUTO)	0.9 \pm 0.2	0.9 \pm 0.2	1.0 \pm 0.1	1.0 \pm 0.1	1.0 \pm 0.1

Volume comparison between manual (MAN) and automatic contours (AUTO), mean values \pm standard deviation

Table 2 Overlap and distance measurements

	Bladder mean \pm SD	Rectum mean \pm SD	Heart mean \pm SD	Lung left mean \pm SD	Lung right mean \pm SD
Sensitivity	0.93 \pm 0.15	0.84 \pm 0.1	0.91 \pm 0.04	0.98 \pm 0.03	0.98 \pm 0.03
Specificity	0.99 \pm 0.01	0.99 \pm 0.01	0.99 \pm 0.01	0.99 \pm 0.01	0.99 \pm 0.01
Jaccard Conf. Index (JCI) 3D	0.81 \pm 0.15	0.67 \pm 0.10	0.85 \pm 0.06	0.95 \pm 0.06	0.95 \pm 0.05
Dice Sim. Coeff. (DSC) 3D	0.88 \pm 0.13	0.79 \pm 0.08	0.92 \pm 0.04	0.97 \pm 0.05	0.97 \pm 0.04
Discordance Index (DisI) 3D	0.13 \pm 0.1	0.22 \pm 0.10	0.06 \pm 0.07	0.03 \pm 0.05	0.03 \pm 0.05
Geographical Miss Index 3D	0.07 \pm 0.15	0.16 \pm 0.12	0.08 \pm 0.04	0.02 \pm 0.03	0.03 \pm 0.03
Center of volume comp. [mm]	4.1 \pm 5.0	8.9 \pm 6.8	4.7 \pm 2.6	2.0 \pm 1.3	2.1 \pm 1.7
Mean surface dist. (MSD) [mm]	1.8 \pm 2.2	2.5 \pm 1.6	1.6 \pm 0.7	0.8 \pm 0.8	1.0 \pm 0.8
Residual mean surface dist. (RMSD) [mm]	3.1 \pm 3.7	4.6 \pm 3.1	2.2 \pm 0.93	1.8 \pm 1.7	1.8 \pm 1.3
Hausdorff Distance (HD ₉₅) [mm]	6.7 \pm 8.8	10.8 \pm 8.4	4.4 \pm 2.0	2.7 \pm 2.7	2.9 \pm 2.7
Right Boundary [mm]	2.9 \pm 4.7	− 0.9 \pm 4.1	0.0 \pm 1.4	− 4.9 \pm 19.2	0.8 \pm 1.2
Left Boundary [mm]	0.1 \pm 3.7	2.5 \pm 16.1	− 0.3 \pm 1.8	− 0.26 \pm 1.0	2.5 \pm 6.0
Anterior Boundary [mm]	2.0 \pm 5.1	5.2 \pm 10.5	− 0.4 \pm 2.1	0.6 \pm 1.5	0.7 \pm 1.6
Posterior Boundary [mm]	− 1.9 \pm 11.7	1.7 \pm 5.8	0.0 \pm 2.0	− 0.1 \pm 2.2	− 0.2 \pm 1.9
Inferior Boundary [mm]	− 0.3 \pm 4.8	7.0 \pm 13.2	− 8.5 \pm 4.8	− 1.8 \pm 4.1	− 1.5 \pm 3.3
Superior Boundary [mm]	− 1.8 \pm 10.6	− 8.7 \pm 16.9	− 4.9 \pm 8.7	− 0.7 \pm 2.0	− 0.5 \pm 1.6

Overlap and distance measurements between manual and automatic contours as mean values \pm standard deviation (SD)

manual corrections before the automatic contours could be used for radiotherapy planning, however even in these cases the time needed by a human expert for editing the autogenerated organ contours would likely be less than generating the complete contour manually from scratch.

Delpon et al. [23] compared five commercial atlas-based segmentation software solutions for radiation planning to manual segmentation for 10 patients for bladder and rectum: ABAS (Elekta Oncology Systems, Crawley, UK), WorkFlow Box (Mirada Medical Ltd., Oxford, UK), MIM Maestro (MIM Software Inc., USA), SPICE (Philips N.V., Netherlands) and RayStation (RaySearch Laboratories AB, Sweden). For the rectum, the software systems achieved a volume ratio of 0.9–1.3 (this study: 0.9) and a mean DSC of 0.49–0.75 (this study: 0.79). For the bladder, the atlas-based software systems achieved a volume ratio of 1.01–1.62 (this study: 0.9) and a mean DSC of 0.62–0.81 (this study: 0.88). In another study using the atlas-based system ABAS, Kim et al. reported DSC values for the bladder below 0.6 [24]. These findings indicate that the algorithm presented in this work is able to outperform atlas-based algorithms for organ segmentation.

In recent years, a significant focus was put on the development and evaluation of machine learning based algorithms. Feng et al. [25] developed a deep convolutional neural network for autosegmentation of thoracic organs and reported DSC 0.972/0.979 and HD₉₅ 2.103/3.958 mm for left/right lung. Regarding these metrics, our algorithm achieved an almost identical performance (DSC 0.97/0.97, HD₉₅ of 2.7/2.9 mm). Cardenas et al. reported

in their review about deep learning autosegmentation architecture types [26] mean DSC values between 0.89 and 0.93 for the heart, 0.93–0.98 for the lungs and 0.7–0.84 for bladder. In comparison, the algorithm presented here achieved equivalent results for the heart (DSC 0.92) and right/left lung (DSC 0.97), and superior results for the bladder (DSC 0.88). Considering that the evaluations were not performed on the same data sets, we conclude that the accuracy of the algorithm presented here is at least comparable to other modern machine and deep learning-based algorithms.

Sultana et al. [27] used a two-step hierarchical convolutional neural network segmentation strategy for automatic contouring of multiple organs of the pelvis, combining an UNet architecture with a generative adversarial network. They reported excellent mean DSC values for bladder of 0.95 (this study: 0.88) and rectum of 0.90 (this study: 0.79), however based on a single center and relatively small cohort of 290 training and 15 test cases.

Lustberg et al. [28] compared a prototype version of the “Mirada DLC Expert” (Mirada Medical Ltd., Oxford, UK), which utilizes convolutional neural networks, to the atlas-based autosegmentation “WorkFlow Box” of the same company and to manual segmentation, and found that the DLC expert showed promising results for automatically generating high quality contours, providing a greater time saving compared to existing solutions.

We consider the comparably large unseen evaluation cohort size of 237 patients for the thorax region and 102 patients for the pelvic region to be a strength of

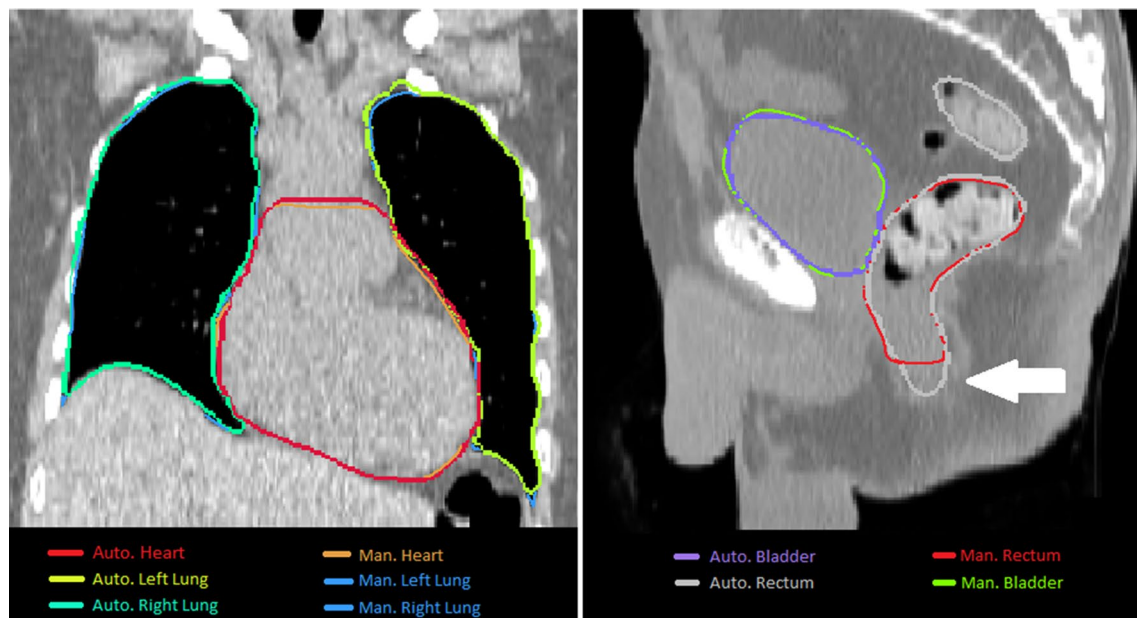


Fig. 5 Exemplary cases. Left panel: Exemplary case with manual and automated contours of left and right lung and heart. Right panel: Exemplary case with manual and automated contours of rectum and bladder. The arrow indicates the difference at the inferior boundary of the rectum

this study. We also aimed at a comprehensive geometric evaluation to facilitate the comparison with other studies. A limitation is that for the thorax region only female patients were used. We assume that the findings are transferable to male patients, however this needs to be confirmed by further studies and by thorough inspection in clinical practice. Another limitation is that only one human reader was used. Multiple human readers would allow the assessment of inter-observer variations, which can be quite substantial [14, 29, 30]. For another algorithm it has already been shown that the accuracy of deep-learning based autosegmentation is comparable to inter-observer variability [31]. It can be speculated that automatic algorithms might even become able to contour organs at risk with a higher reproducibility and accuracy than humans, especially when less experienced readers are included [1, 2, 32, 33]. This could, in addition to the time savings, also increase the quality of the radiation treatment planning, like it has been discussed for head and neck patients in [34].

The algorithm in the software prototype used in this study corresponds to the algorithm implemented in two products by Siemens Healthineers (Erlangen, Germany), the server-based “syngo.via RT Image Suite” (version VB50) and the cloud-based “AI-Rad Companion Organs RT” (version VA20), for lungs, rectum and bladder. The heart model evaluated in this study is slightly improved with regard to the latest released product versions at the time of submission of this publication. Both products are

commercially available and are certified for clinical use, making this study relevant for clinical practice.

As future work it is planned to include multiple human readers to assess inter-observer variations and an analysis of the dosimetric consequences of contour differences as a metric with more direct clinical impact than geometric parameters.

Conclusions

We described and evaluated a commercially available deep image-to-image network (DI2IN) algorithm for automatic contouring of organs at risk in radiation treatment planning. The automatic contours showed excellent agreements with manual contours drawn by an experienced radiation oncologist, with some deviations mostly at the lower part of the heart and both upper and lower parts of the rectum.

Appendix

The metrics used in this work were calculated as follows:

$$\text{Sensitivity} = \frac{AUTO \cap MAN}{MAN} \text{ (from [14])}$$

$$\text{Specificity} = \frac{AUTO \cap MAN}{AUTO} \text{ (from [14])}$$

$$AUTO = \text{Volume outside AUTO}, MAN = \text{Volume outside MAN}$$

$$\text{Jaccard Index (JCI)} = \frac{AUTO \cap MAN}{AUTO \cup MAN} \text{ (from [15, 16])}$$

$$\text{Discordance Index (DisI)} = 1 - \left(\frac{MAN \cap AUTO}{AUTO} \right) \text{ (from [18, 19])}$$

$$\text{Dice Similarity Coefficient} = \frac{2(AUTO \cap MAN)}{AUTO + MAN} \text{ (from [17])}$$

$$\text{SurfaceDistance}(DSM) = \frac{1}{n_s + n_{s'}} \left(\sum_{p=1}^{n_s} d(p, S') + \sum_{p'=1}^{n_{s'}} d(p', S) \right)$$

(from [14])

$$\text{ResidualMeanSurfaceDistance}(RMS) = \sqrt{\frac{1}{n_s + n_{s'}} \left(\sum_{p=1}^{n_s} d(p, S')^2 + \sum_{p'=1}^{n_{s'}} d(p', S)^2 \right)}$$

(from [20, 21])

$$\text{HausdorffDistance}(HD) = \max[d(S, S'), d(S', S)]$$

(from [22])

Acknowledgements

Not applicable.

This work was presented in the poster highlight session at the ESTRO Congress 2020 in Vienna, Austria.

Authors contributions

Conceptualization, SM, MD, CT, FV; methodology, SM, MD, AG, TH, CM; software, MD, ZX, SG, GC, BG, JR; validation, SM, AG, CT, TH; formal analysis, SM, MD, ZX, SG, GC, BG; investigation, SM, AG, CB, MN, CT; re-sources, CT, TH, FV, CB; data curation, SM, MD; writing—original draft preparation, SNM and CT; writing—review and editing, ZX, TH, FV, CM; visualization, SM, AG; supervision, TH, CM, CB, FV, CB, SC; project administration, CT, FV; funding acquisition, CT, CH. All authors have read and agreed to the published version of the manuscript.

Funding

Parts of this work were funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (project name ILAOS).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The CT data was anonymized for the scientific purpose of this work. This study complies with the declaration of Helsinki, Good Clinical Practice (GCP) and Good Epidemiological Practice (GEP). The data acquisition and analysis were in accordance with Bavarian hospital law (Art.27 Abs. 4 BayKrG).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany. ²Technology Excellence, Digital Technology & Innovation, Siemens Healthineers, Erlangen, Germany. ³Technology Excellence, Digital Technology & Innovation, Siemens Healthineers, Princeton, NJ, USA. ⁴Department of Radiation Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁵Cancer Therapy, Siemens Healthineers, Forchheim, Germany. ⁶Department of Radiation Oncology, LMU Klinikum, Marchioninistr. 15, 81377 München, Germany.

Received: 14 March 2022 Accepted: 28 June 2022

Published: 22 July 2022

References

- Nikolov S, Blackwell S, Mendes R, Fauw JD, Meyer C, Hughes C, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. 2018:1–31.

- van der Heyden B, Wohlfahrt P, Eekers DBP, Richter C, Terhaag K, Troost EGC, et al. Dual-energy CT for automatic organs-at-risk segmentation in brain-tumor patients using a multi-atlas and deep-learning approach. *Sci Rep*. 2019;9:4126.
- Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys*. 2019;46:576–89.
- Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. *Acta Oncol*. 2016;55:799–806.
- Feng M, Valdes G, Dixit N, Solberg TD. Machine learning in radiation oncology: opportunities, requirements, and needs. *Front Oncol*. 2018;8:110.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*;p. 234–41.
- Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017;44:547–57.
- Kearney V, Chan JW, Valdes G, Solberg TD, Yom SS. The application of artificial intelligence in the IMRT planning process for head and neck cancer. *Oral Oncol*. 2018;87:111–6.
- Ghesu FC, Georgescu B, Zheng Y, Grbic S, Maier A, Hornegger J, et al. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. *IEEE Trans Pattern Anal Mach Intell*. 2019;41:176–89.
- Yang D, DX, Zhou SK, Bogdan G, Mingqing C, Sasa G, et al. Automatic liver segmentation using an adversarial image-to-image network. *Springer, Cham*, 2017.
- Kingma DP, Ba J. Adam. A method for stochastic optimization. 2014;p. 1–15.
- Julia White AT, Douglas A, Thomas B, Shannon M, Lawrence M, Lori P, Abraham Recht RR, Alphonse T, Frank V, Wendy W, Allen Li X. Breast cancer atlas for radiation therapy planning: consensus definitions. *RTOG - Radiation Therapy Oncology Group*.
- Gay HA, Barthold HJ, O'Meara E, Bosch WR, El Naqa I, Al-Lozi R, et al. Pelvic normal tissue contouring guidelines for radiation therapy: a Radiation Therapy Oncology Group consensus panel atlas. *Int J Radiat Oncol Biol Phys*. 2012;83:e353–62.
- Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014;41:050902.
- van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys*. 2007;68:771–8.
- Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. *Physics in Medicine and Biology*. 2009;54:2863–73.
- Lorenzen EL, Taylor CW, Maraldo M, Nielsen MH, Offersen BV, Andersen MR, et al. Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: a multi-centre study from Denmark and the UK. *Radiother Oncol*. 2013;108:254–8.
- Kepka L, Bujko K, Garmol D, Palucki J, Zolciak-Siwinska A, Guzel-Szczepiorkowska Z, et al. Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy. *Radiotherapy and Oncology*. 2007;85:450–5.
- Holyoake DL, Robinson M, Grose D, McIntosh D, Sebag-Montefiore D, Radhakrishna G, et al. Conformity analysis to demonstrate reproducibility of target volumes for Margin-Intense Stereotactic Radiotherapy for borderline-resectable pancreatic cancer. *Radiother Oncol*. 2016;121:86–91.
- Heimann T, van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging*. 2009;28:1251–65.
- V.Y. Voiculescu I. An overview of current evaluation methods used in medical image segmentation. University of Oxford; 2015.
- Taha AA, Hanbury A. An efficient algorithm for calculating the exact Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 2015;37:2153–63.
- Delpont G, Escande A, Ruef T, Darreon J, Fontaine J, Noblet C, et al. Comparison of automated atlas-based segmentation software for postoperative prostate cancer radiotherapy. *Front Oncol*. 2016;6:178.

24. Kim N, Chang JS, Kim YB, Kim JS. Atlas-based auto-segmentation for postoperative radiotherapy planning in endometrial and cervical cancers. *Radiat Oncol.* 2020;15:106.
25. Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Med Phys.* 2019;46:2169–80.
26. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Semin Radiat Oncol.* 2019;29:185–97.
27. Sultana S, Robinson A, Song D, Lee J. Automatic multi-organ segmentation in computed tomography images using hierarchical convolutional neural network. *J Med Imaging (Bellingham).* 2020;7.
28. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol.* 2018;126:312–7.
29. Peulen H, Belderbos J, Guckenberger M, Hope A, Grills I, van Herk M, et al. Target delineation variability and corresponding margins of peripheral early stage NSCLC treated with stereotactic body radiotherapy. *Radiother Oncol.* 2015;114:361–6.
30. Joskowicz L, Cohen D, Caplan N, Sosna J. Inter-observer variability of manual contour delineation of structures in CT. *Eur Radiol.* 2019;29:1391–9.
31. Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiother Oncol.* 2020;144:152–8.
32. Hurkmans CW, Borger JH, Pieters BR, Russell NS, Jansen EPM, Mijnheer BJ. Variability in target volume delineation on CT scans of the breast. *Int J Radiat Oncol Biol Phys.* 2001;50:1366–72.
33. Anders LC, Stieler F, Siebenlist K, Schafer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. *Radiother Oncol.* 2012;102:68–73.
34. Kosmin M, Ledsam J, Romera-Paredes B, Mendes R, Moinuddin S, de Souza D, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer. *Radiother Oncol.* 2019;135:130–40.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

