DIVE-ART: a tool to guide clinicians towards Dosimetrically Informed Volume Editions of automatically segmented volumes in Adaptive Radiation Therapy

Benjamin Roberfroid[a,*], John A. Lee[a], Xavier Geets[a,b], Edmond Sterpin[a,c,d] and Ana M. Barragán-Montero[a]

[a]Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium

[b]Cliniques universitaires Saint-Luc, Department of Radiation Oncology, Brussels, Belgium

[c]KU Leuven – Department of Oncology, Laboratory of Experimental Radiotherapy, Leuven, Belgium

[d]Particle Therapy Interuniversity Center Leuven – PARTICLE, Leuven, Belgium

*Corresponding author at: Université catholique de Louvain (UCLouvain), Molecular Imaging, Radiotherapy and Oncology (MIRO), Avenue Hippocrate 54, Bte

B1.54.07, 1200 Brussels, Belgium.

E-mail address: benjamin.roberfroid@uclouvain.be

Purpose:

Automatic segmentation algorithms are key for online adaptive radiotherapy. However, the lack of quality assurance tools requires manual review of auto-segmented volumes. Our goal is to reduce these manual operations thanks to a Dosimetrically Informed Volume Edition (DIVE) map.

Methods:

The DIVE map for a given auto-segmented volume is generated as follows. First, the volume is deformed so that N deformation scenarios (Sd,d=1,..N) are generated. Second, the 3D dose distribution for the initial (S0) and deformed (Sd) scenarios is predicted with a deep convolutional neural network (DCNN). Third, an "impact of correction" (IOC) scalar value is computed for each Sd as the weighted sum of relevant dose-volume metrics differences between S0 and Sd. These IOC values are then spatially aggregated to form the DIVE-map.

Our DCNN was trained with 180 adaptive treatment fractions. DIVE performance for auto-segmented bladder and rectum was evaluated on 30 fractions. The evaluation assessed 1) the dose-volume metrics from automatically generated plans for either conventionally- or DIVE-corrected contours; 2) the decrease in the amount of corrections when using the DIVE-map. Metrics were evaluated on the conventionally corrected contours.

Results:

Using the DIVE-map, the number of corrected voxels was reduced by at least 50% in more than half of the 30 fractions (16 for rectum, 18 for bladder), while keeping the same dosimetric quality as the plan with conventionally corrected organs (differences in V60Gy below +2%).

Conclusion:

The proposed method successfully reduced the need for corrections in auto-segmented organs without impairing plan dosimetric quality.

1. Introduction

Automatic segmentation in radiation therapy has raised considerable interest for many years, since it allows the clinicians to automate the tedious task of manually contouring anatomical structures and tumor volumes. The accuracy of these algorithms is crucial, particularly in the context of online adaptive radiotherapy, where the time for review and correction should be as short as possible.

In recent years, the accuracy of auto-segmentation algorithms has reached a tipping point that enables a broad adoption in clinical environments, mainly thanks to the use of machine learning methods. In particular, deep convolutional neural networks (DCNNs) have demonstrated their ability to delineate various anatomical volumes quite accurately across multiple locations [1-8]. However, in their current state, DCNNs can still not be used clinically without human supervision, since they cannot guarantee consistently good accuracy for every single volume. Few studies have analyzed the dosimetric quality of treatment plans using uncorrected auto-segmented contours [9, 10]. While these studies have shown overall good dosimetric quality, some cases might remain, for which treating without correction can lead to severe consequences for the patient. In order to spot these cases, clinicians typically perform slice-by-slice visual inspection and manual correction of every volume to ensure patient's safety [11-15]. This review process can take from few minutes up to half an hour for complex cases, such as for head and neck cancer [2, 16], which considerably slows down or even hamper online adaptive workflows [7, 15, 17, 18].

Although cumbersome and inefficient, this review process is currently the only solution to an existing problem: the lack of quality assurance (QA) tools or metrics for auto-segmentation algorithms. Recently, some groups have started to approach this problem from different perspectives. For instance, predicting geometric indicators like the Dice similarity index [19-21]; using a secondary and independent algorithm (i.e., deformable registration or yet another DCNN) to detect segmentation errors [22, 23]; or estimating the uncertainty of the DCNN prediction [24, 25], among other approaches. However, all these remain purely focused on global segmentation and geometric quality, irrespective of subsequent steps and objectives in the process of treatment planning. Indeed, when planning a radiotherapy treatment, organ and target volume segmentation is not a goal per se and the actual endpoint remains the dosimetric quality of the plan. QA tools relying on geometric indicators are thus agnostic proxies that, in the end, may not correlate well with clinical quality [1, 26, 27]. From that perspective, it is expected that segmentation accuracy matters more in some regions of the anatomy and less in others. Within a restricted time budget for correction, like in online adaptive radiotherapy, it then makes sense to focus the contour editions on those regions that have higher dosimetric impact, whereas regions in the dose map that are hardly sensitive to segmentation quality become secondary.

In this work, we propose a QA tool for auto-segmentation algorithms designed from a dosimetric perspective instead of a geometric one: the DIVE (Dosimetrically Informed Volume Edition) map, which combines spatial deformations of the auto-segmented volumes and DCNN to predict the dosimetric relevance of different regions [28-32]. Our solution takes the form of a graphical indicator, in the spirit of a heatmap, which can be overlaid on the planning image (e.g., CT or CBCT). The DIVE-map could guide clinicians visually to disregard irrelevant corrections and focus on those with high dosimetric impact, significantly speeding up the contour review process. From a broad perspective, we believe that such a visual tool could mitigate the last remaining shortcomings of automatic segmentation, by further reducing the subsequent contour edition time in clinical conditions.

2. Materials and methods

This section describes the process to generate a DIVE-map and the data used for it. As a proof-of-concept , we analyze the added value of the DIVE-map for online adaptive treatments of prostate cancer patients using a commercially available adaptive solution: ETHOS, from Varian (Palo Alto, USA) [33]. The generation process walks through four main steps that are summarized in Figure 1. We focus on two main organs at risk (OAR): bladder and rectum, which were generated for each treatment fraction by the DCNN-based ETHOS auto-segmenter. Consequently, the reduction of manual correction was also evaluated per fraction.

2.1 Data

Our database consisted of 12 prostate cancer patients, without pelvic nodal irradiation. The dose prescription was 60 Gy to the prostate and 48 Gy to the seminal vesicles, delivered in 20 fractions. The CTV-to-PTV expansion consisted in an isotropic margin of 7 mm. CT images were acquired on a Toshiba Acquilion CT scanner with a slice thickness of 2mm. All plans were optimized with the ETHOS "intelligent optimization engine" (IOE), using the clinical goals template at Table 1 and a 9-fields IMRT beam configuration. For each treatment fraction, we have at our disposal both the raw auto-segmented volumes on the daily CBCT and the corresponding manually corrected contours that were clinically used. These corrections were done by the physician in charge, following the conventional process of slice-by-slice visual inspection and review. The corrected contours are referred hereafter to as "conventionally corrected".

Local ethics committee approval by the CEHF (Comité d'Ethique hospitalo-facultaire) was obtained for the use of all patient data under the agreement "Learning from the past: the MIRO treatment planning database".

2.2 DIVE-map generation

2.2.1. General workflow

For a given auto-segmented volume, the DIVE-map is generated in four steps (Figure 1). First, in order to simulate possible contouring errors of the given volume, N spatial deformations are applied to its contour (the planning image remains untouched), following different amplitudes and directions (section 2.2.2). The deformed volume is added to the rest of the planning structures, namely, the non-deformed OARs and targets, resulting in a total of N deformation scenarios ($S_{d,d=1,..N,}$). Second, we used a DCNN to predict a fast and precise approximation of the 3D dose distribution for the initial (S_0) and deformed (S_d) scenarios (section 2.2.3). This aims at evaluating the dosimetric impact of the simulated contouring errors in the different S_d scenarios. The third step consists in computing a surrogate of the global dosimetric impact of generating a plan with scenario S_0 rather than scenario S_d (assuming S_d would be a correction of S_0). This is referred hereafter to as "impact of correction" (IOC). The IOC associated with each S_d is a scalar value that is computed as the weighted sum of dose-volume metrics differences between S_0 and S_d (section 2.2.4) and it is spread uniformly back over the deformed volume of S_d. In the fourth step, the various IOCs that are spread spatially in the differently deformed volumes are averaged locally in each voxel to get the complete DIVE-map (section 2.2.4), highlighting the buildup of strong IOCs, or alternatively the lack thereof.

DIVE-maps for bladder and rectum are independently generated, i.e., there is no combination of deformation between them. In total, 60 DIVE-maps (30x2 OAR) were generated for the 10 first fractions of 3 randomly chosen patients in our database. An example of generated DIVE maps is depicted in Figure 2.

2.2.2 Deformation scenarios

Our volume deformation process uses the RBF mesh deformation of the PyGeM Python toolbox [34]. By varying the direction and the amplitude of deformation, we generated a total of *N* = 162 deformation scenarios S_d per organ and per fraction. See Supplementary Materials (Appendix A.) for further details.

2.2.3 DCNN dose prediction model

Our DCNN was a hierarchically densely connected U-Net (HD-UNET) [35-37], with 9 input channels containing the CBCT scan, the PTV volumes, and seven OARs (the anal canal, bladder, rectum, left femoral head, right femoral head, bowel, and penile bulb). Notice that when predicting the dose for the initial S_0 and deformation scenarios S_d of bladder or rectum, the corresponding input channel contained the initially auto-segmented or deformed volume of the organ, respectively. Notice also that PTV of prostate and seminal vesicles are rigidly registered from the initial CT to the CBCT since, in the Ethos workflow, target volumes for daily CBCT are not yet available when correcting Ethos auto-segmented OARs.

The model was trained with data from 180 adaptive fractions of 9 patients (9x20) randomly chosen among the 12 initial ones. See Supplementary Materials (Appendix B.) for more details.

2.2.4 Dosimetric Impact Of Correction value (IOC) computation and aggregation

Each of the *N* deformation scenarios of a given volume is characterized by an IOC scalar value. The IOC formula consists in a weighted sum of *M* dose-volume metrics differences between S_0 and S_d (M=11, see Table 1). Notice that metrics for deformed scenarios are extracted from the predicted dose with our DCNN (section 2.2.3). The weights for each metric difference, α_i, are the inverse of the mean standard deviation of the evaluated metrics on the 20 fractions *Fx* of each 9 training patients *P*: $IOC = \sum_{i=1}^{M} \alpha_i |Metric_{i,S0} - Metric_{i,Sd}|$; $\alpha_i = \frac{1}{\bar{\sigma}_i}$ with ; $\bar{\sigma}_i = \frac{\sum_{j=1}^{P} \sigma_{j,i}}{P}$; $\sigma_{j,i} = \sqrt{\frac{1}{Fx} \sum_{k=1}^{Fx} \left(Metric_{i,j,k} - \overline{Metric}_{i,j}\right)^2}$

The idea behind taking the inverse of the standard deviation of each metric is to normalize each metric difference according to their sensitivity to volume changes. Consequently, weighted metric differences might better reflect their clinical importance, rather than the sensitivity to volume changes.

The computed IOC value for each S_d is then assigned uniformly to all voxels inside the mask of the deformed volume of interest to form a 3D map. The final DIVE-map results from averaging out these 3D IOC maps over all S_d starting from the smallest deformation amplitude *A*, then locking the concerned voxels, and continuing iteratively to the next deformation amplitudes, like a tree grows bigger wood rings every year. For more information, see Supplementary Materials (Appendix C. and D.)

2.3 Contours corrections guided by the DIVE-map and Ethos plan generation

The capability of our DIVE-map to indicate dosimetrically relevant regions for correction was assessed with an automatic slice-by-slice correction strategy based on the values of the DIVE-map and the conventionally corrected contours (considered as our reference). We define a "misclassified" voxel when the auto-segmenter did not assign the correct label to this specific voxel: either auto-segmentation considered this voxel as belonging to the nearby OAR while it should not or, conversely, auto-segmentation did not include this voxel into the OAR while it actually belongs to it. In our DIVE-correction strategy, if all misclassified voxels of a given slice had corresponding DIVE-

map voxels values (DMi) below a threshold (*t*), no correction was performed on the slice. In contrast, when there was at least one misclassified voxel with DMi above *t*, the entire slice was corrected in the same way as in the conventionally corrected contour. Three thresholds *t* were evaluated to trigger the corrections guided by the DIVE-map, i.e., *t*=1, *t*=1.5, and *t*=2.0.

To evaluate the dosimetric impact of using DIVE-corrected contours, we generated treatment plans for each of the 10 first fractions of the remaining 3 patients not used to train our DCNN dose prediction model. Plans were automatically optimized by the Ethos IOE with the sclinical goals in Table 1, the synthetic CT (I.e., the planning CT registered on the daily CBCT) and the custom structure set composed of the DIVE-corrected bladder and rectum, as well as the other OAR and conventionally corrected PTVs. Notice that, in our institution, other volumes than bladder, rectum, and targets are hardly corrected, consequently, those were kept uncorrected.

2.4 Results analysis

To evaluate the added-value of our DIVE-map we present two types of results: dosimetric values and relative number of corrected voxels. On the one hand, we analyze the clinical goals (i.e., dose-volume metrics, Table 1) for the treatment plans generated with DIVE-corrected and conventionally corrected volumes. We specifically focus on high dose metrics, i.e., V60Gy for OAR and D95% to PTV_prostate, since they are of the highest relevance for the plan clinical quality. The goal here is to demonstrate that the dosimetric quality of the plan for DIVE-corrected volumes is the same as the one for conventionally corrected volumes. On the other hand, we report the relative number of corrected voxels for auto-segmented rectum and bladder according to the value of *t*. In this context, a relative correction of 100% indicates that all voxels that were corrected to obtained the conventionally corrected contour were deemed necessary when looking at the DIVE-map. In contrast, a value of 0% means that no corrected voxels are actually necessary, consequently, the contour remains fully uncorrected.

3. Results

Dosimetric differences of plans generated with conventionally corrected versus DIVE-corrected and uncorrected bladder and rectum are illustrated in Figure 3, for OAR V60Gy and PTV_prostate D95%. Using uncorrected contours led to an OAR V60Gy difference above +2% for 13 fractions, while this was reduced to 6 fractions for the DIVE-corrected contours with threshold *t*=2.0. The best trade-off between dosimetric quality and reduced number of corrections was found with *t*=1.5, since no fraction showed a V60Gy above 2% difference for the considered OAR. The lowest difference for D95% on PTV_prostate was -0.5%, which still ensures an acceptable target coverage of 97.7% of the prescription.

Table 2 reports the OARs mean metric differences between the plans using uncorrected/DIVE-corrected and conventionally corrected contours. Note that, except for the V60Gy in uncorrected cases, no mean metric difference exceeds +1%. Concerning the bowel D0.1cm^3, although it remains below the 60Gy clinical goal, it presents a large difference range. Such dose metric is naturally unstable due to the tiny volume evaluated by the D0.1cm^3 over the large bowel volume.

Figure 4 reports the relative number of corrected voxels according to the DIVE-map threshold value. As expected, the higher the threshold, the smaller the number of corrections. We can notice global trends. For the rectum, fractions from patient_1 seem to require the highest amount of corrections, followed by patient_2 and patient_3. Concerning the bladder, fractions from patient_3 are those requiring more corrections, followed by patient_1 and patient_2. Notice that fractions from patient_2 do not require any bladder correction for t=1.5 and 2.0. Overall, the amount of corrections

could be reduced by more than 50% in 16 and 18 fractions for rectum and bladder, respectively, when using the DIVE-map with t=1.5. However, for some fractions, plans generated with uncorrected contours achieved similar dosimetric quality as the conventionally corrected contours. This was the case for 6 fractions in patient_1, all fractions in patient_2 and one fraction in patient_3. For these fractions, the DIVE-corrected rectum with *t*=1.5 required more than 50% of corrections for 5 out of 6 fractions of patient_1, 4 out of 10 fractions of patient_2 and no fraction in patient_3. For the bladder, it resulted in one out of 6 fractions for patient_1, and no fractions for patient_2 and patient_3.

4. Discussion

This work presents a "heatmap" tool to assist clinicians in correcting auto-segmented contours, by indicating dosimetrically relevant regions when it is overlaid on the (re-)planning image. We coined it the DIVE-map, standing for dosimetrically informed volume edition. We evaluated it on auto-segmented bladder and rectum from 3 prostate cancer patients (10 fractions each) receiving adaptive radiotherapy with ETHOS (Varian).

It was first shown that the use of uncorrected rectum and bladder volumes auto-segmented by ETHOS would result in 13 plans out of 30 where OAR V60Gy values would strongly exceed their homologous values achieved with conventionally-corrected contours. This highlighted the dosimetric importance of correcting auto-segmented volumes. Then, using the DIVE-map with a reasonable correction threshold *t* of 1.5, we demonstrated the preservation of the dosimetric quality while reducing the amount of corrections. In particular, there were 6 and 11 fractions where, following the DIVE-map indications, no correction had to be applied to the rectum and bladder, respectively. Moreover, the amount of corrections was reduced by more than 50% in 16 and 18 fractions for rectum and bladder, respectively. However, for 9 rectums and 1 bladder, the DIVE-map marked as relevant more than 50% of conventional corrections, while uncorrected contours presented the same dosimetric quality. This highlighted a good sensitivity of the necessary corrections when using *t*=1.5, while the specificity was more limited, especially for the rectum. This could hint that some of the used rectum deformations are not fully relevant. Usage of more realistic rectum deformation to enhance the correction specificity could be a next investigation step.

Concerning the DIVE-threshold *t*, our results suggested to use a value between 1.5 and 2. This value might not generalize well to other locations and specific treatments, since it is influenced by the quality of the predicted doses, the quality of the auto-segmenter, or the dose-volume metrics used to compute IOC factors. Therefore, anyone wishing to use this tool in their own clinical conditions will have to go through a calibration step to find the optimal *t* value, which can take some time and requires representative data.

A third hitch concerns the DIVE-map generation time. It took around one hour to generate a single DIVE-map, using a purely sequential process, where each deformation scenario (0.5 seconds) and predicted dose (10-20 seconds) is computed one after the other. Such generation timing is incompatible with online adaptive treatments. Therefore, parallelization of these processes and reduction of the number of deformation scenarios (e.g., by keeping only the most relevant ones) could be a next step towards the clinical use of such a tool. Alternatively, it could be feasible to build a database of DIVE-maps and then train a DCNN model to predict them, with a workflow and timing similar to the dose prediction process.

Another potential weakness of our approach is that the DIVE-maps of bladder and rectum are generated independently, whereas both volumes are likely to be correlated in terms of dose

optimization. Ideally, this would require making deformation combinations with both volumes, which would, however, add a lot of scenarios and further slowdown the generation process. In our case, with only two volumes and globally good segmentation quality, DIVE-maps gave sufficient information to perform reliable corrections. However, for other cases with more volumes to correct or lower auto-contouring quality, performance remains to be evaluated.

This study did not evaluate the time that a clinician could spare by using the DIVE-map, but the tool is designed to be used for that purpose and that is the reason why its performance was evaluated in the same way as humans would probably use it, namely, one slice after the other. Future work could investigate the joint use of an uncertainty map [24], indicating the most probable regions of contouring error for DCNN auto-segmentation, with the DIVE-map. This would allow us to highlight the regions where contouring errors are more probable instead of the entire space where the deformations were evaluated.

On a side note, the time for manual corrections might not be the only concern when using automatic segmentation in adaptive conditions. Some studies have demonstrated the presence of automation bias, which results in noticeable differences between auto-segmented contours followed by manual corrections and fully manually segmented contours. While it may have a beneficial impact by decreasing the inter-observer contouring variability [2, 38], it might also result in over-reliance of clinicians on automated contouring [1, 38], jeopardizing the contouring and dosimetric quality. Such an excessive reliance on automation has already been observed in other medical fields [39-41]. Consequently, DIVE-maps might also help to prevent possible omission of dosimetrically important corrections and again ensuring plan clinical quality.

In conclusion, with the limited information from uncorrected contours and approximative position of target, the DIVE-map was able to discard most of irrelevant corrections. Even if some enhancements of the approach in terms of correction specificity and generation time seem possible, such tool could have great future potential in guiding physicians during manual contour corrections.

Figures and tables

Table 1: IOE clinical goals and "Impact of correction" (IOC) metrics. Each line refers to an optimization metric for a specific volume with a user-defined optimization priority, if used for IOE. Femur_L = left femoral head, Femur_R = right femoral head, PTV_prostate= PTV volume for the prostate, PTV_vs = PTV volume for the seminal vesicles. IOC, metrics will be used later for the DIVE-map generation (see section 2.2).

Volume	Metric + goal if used for IOE	Priority if used for IOE	Usage for IOC (Y/N)
PTV_prostate	D95% >= 98%	1: Most Important	Y
PTV_prostate	D2% <= 104%	1: Most Important	N
PTV_vs	D95% >= 98%	1: Most Important	Y
PTV_vs	D2% <= 104%	1: Most Important	N
Rectum	V60Gy < 1%	2: Very Important	N
Anal canal	V60Gy<1%	2: Very Important	N
Bladder	V60Gy < 3%	2: Very Important	N
Rectum	V52.8Gy < 15%	2: Very Important	Y
Rectum	V30Gy < 35%	2: Very Important	Y
Bladder	V50Gy < 20%	2: Very Important	Y
Bladder	V40.8Gy < 20%	2: Very Important	Y
Bowel	D0.1cm^3<60Gy	2: Very Important	N
PTV_prostate	D50% >= 99%	3: Important	N
PTV_prostate	D50% <= 101%	3: Important	N

PTV_vs	D50% >= 99%	3: Important	N
PTV_vs	D50% <= 101%	3: Important	N
Bowel	V55Gy <3%	3: Important	N
Bowel	V20Gy <300cm^3	3: Important	N
Anal canal	V35Gy<25%	3: Important	N
Anal canal	V20Gy<50%	3: Important	N
Left femoral head	V40Gy <20%	4: Less Important	N
Left femoral head	V35Gy < 5%	4: Less Important	N
Right femoral head	V40Gy <20%	4: Less Important	N
Right femoral head	V35Gy < 5%	4: Less Important	N
Penile bulb	V42.5Gy < 50%	4: Less Important	N
Penile bulb	V54.1Gy < 10%	4: Less Important	N
Skin	D1cm^3<50Gy	4: Less Important	N
Rectum	D1%	Not used	Y
Bladder	D1%	Not used	Y
Anal canal	D1%	Not used	Y
CTV_prostate	D98%	Not used	Y
CTV_vs	D98%	Not used	Y

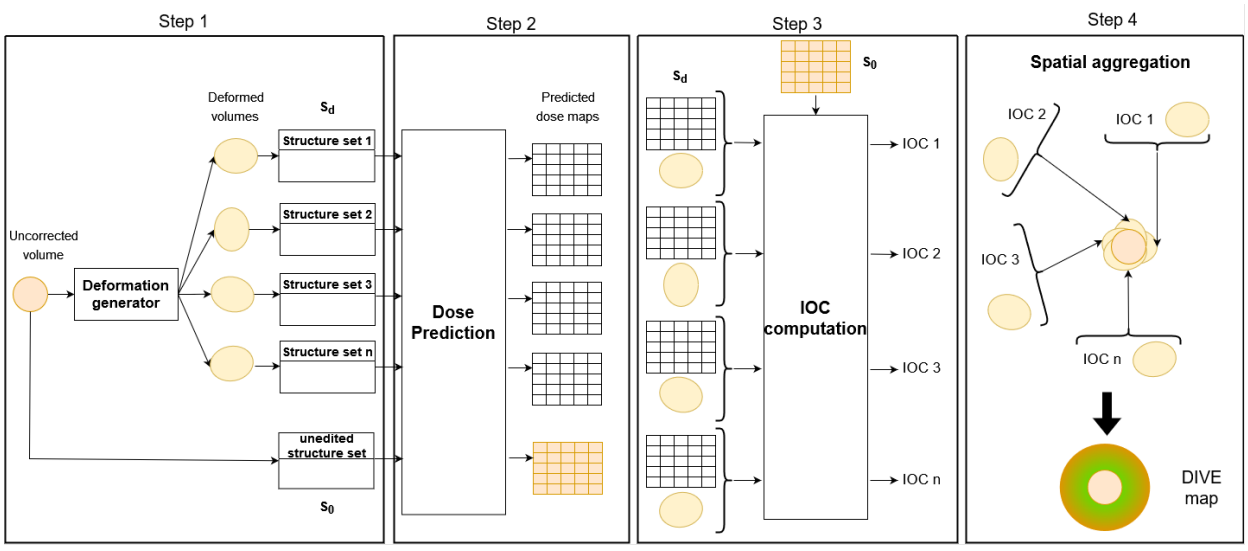

Figure 1: DIVE-map generation process for a given auto-segmented volume. Uncorrected = raw auto-segmented volume, without any modifications, S_0. IOC = "Impact of correction". Step 1: deformation of the considered uncorrected volume, incorporation of the deformed volume to the structure set (other volumes are not deformed), resulting in S_d scenarios. Step 2: Dose is predicted for each deformation scenario as well as the S_0 scenario. Step 3: "Impact of correction" IOC scalars are computed as a weighted sum of dose-volume metric differences between S_0 and each Sd. Step 4: IOC and deformed volumes are spatially aggregated around the uncorrected volume to generate the DIVE-map.

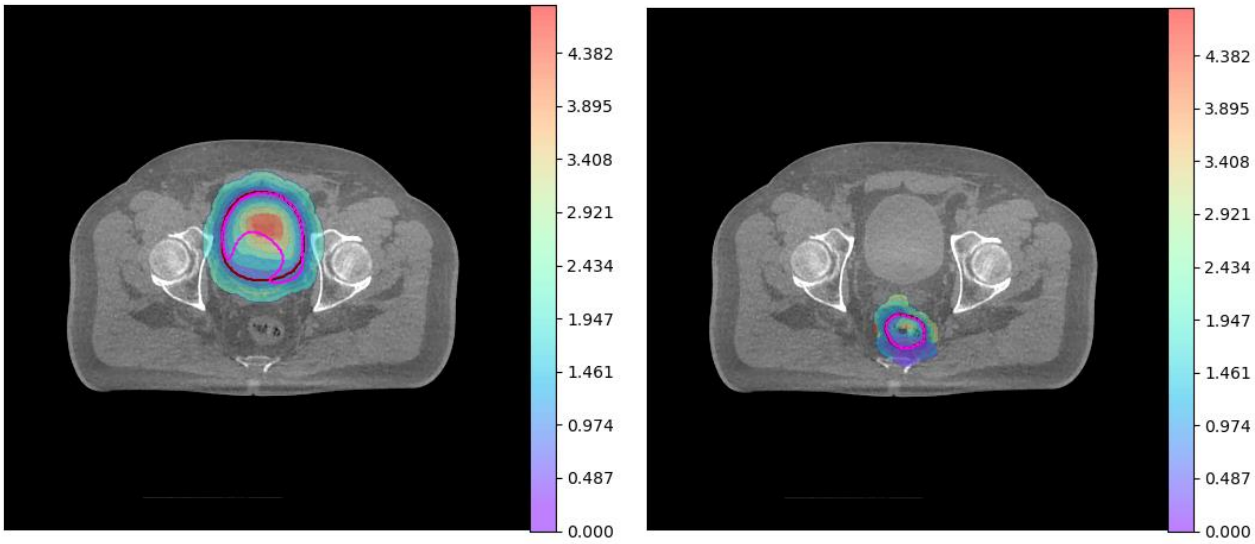

Figure 2: Example of generated DIVE-map for bladder(left) and rectum(right), for which the values are denoted by a colormap ranging from purple to red. maroon = uncorrected contour; fuchsia= conventionally corrected contour.

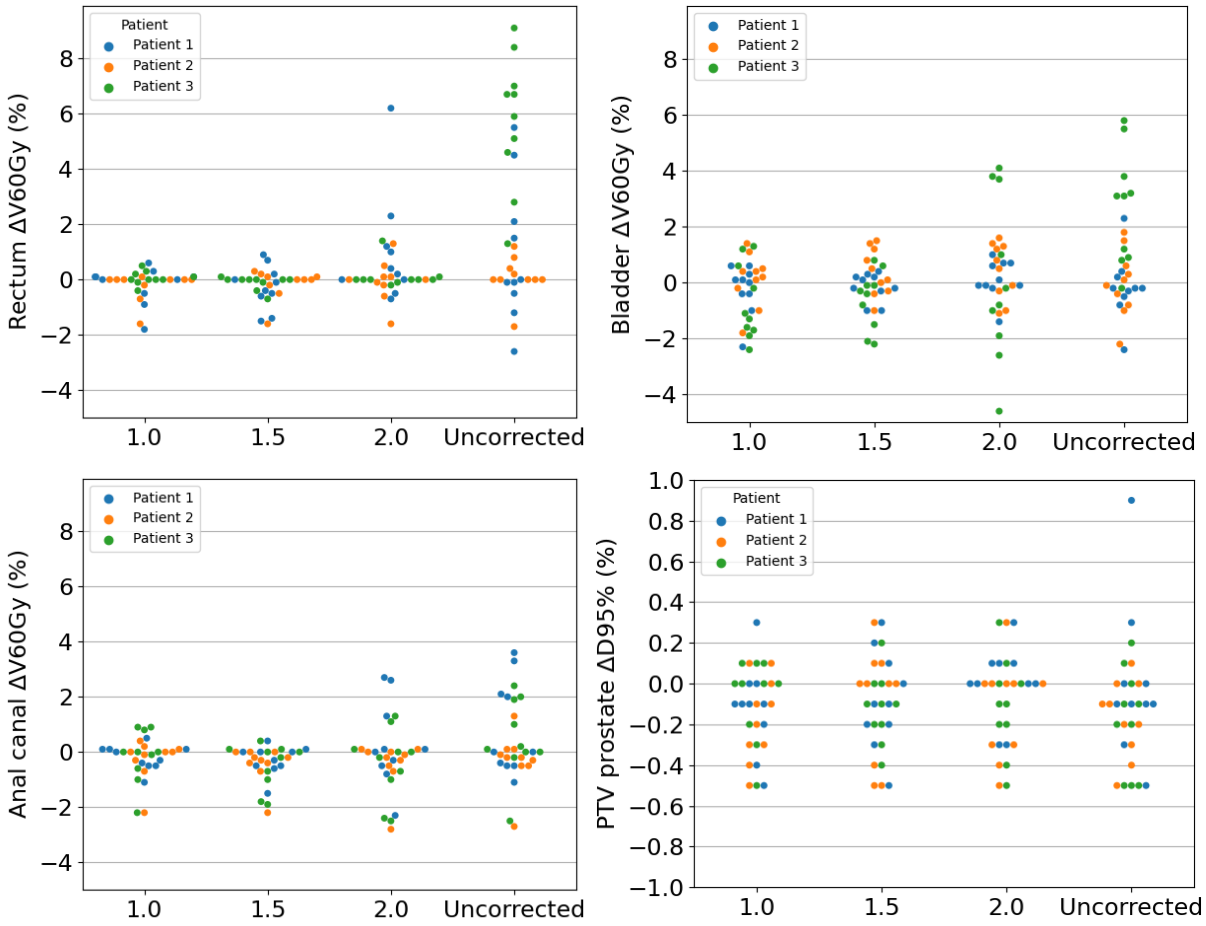

Figure 3: Dosimetric differences between plans generated with DIVE-corrected (t=1.0, 1.5, and 2.0)/uncorrected volumes and plans generated with conventionally-corrected volumes. Individual fraction values are denoted by a dot. Upper left: rectum V60Gy differences. Upper right: bladder V60Gy differences. Lower left: Anal canal V60Gy differences. Lower right: PTV prostate differences. "1.0", "1.5", and "2.0" denote the t threshold value of the DIVE-map below which no slice

correction is triggered. "Uncorrected" corresponds to the case where no corrections are operated on the bladder and rectum contours.

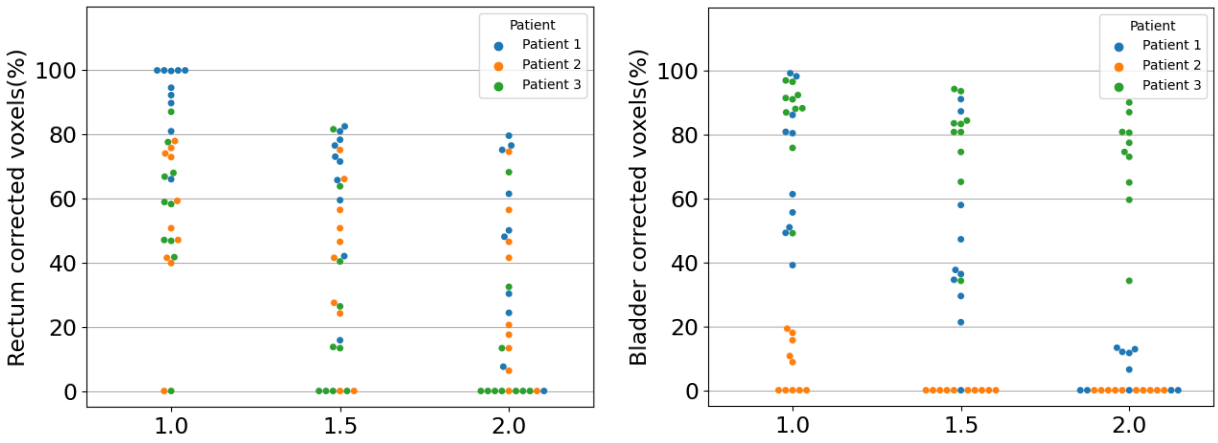

Figure 4: Relative number of corrected voxels for each fraction according to each DIVE-map threshold value. A value of 100% means that all the corrected voxels from conventionally corrected contour are deemed necessary, 0% means that no correction is done, and the corresponding volume remains fully uncorrected. Left: Rectum relative corrected voxels. Right: Bladder relative corrected voxels. Notice the increasing number of fractions requiring no correction at all as the correction threshold increases.

Table 2: Differences (uncorrected/DIVE-corrected minus conventionally-corrected) on OAR metrics. Note that the bowel D0.1cm³, despite showing very large range of differences, was significantly below the 60Gy threshold for all plans

Volume	Dose-volume metric	Uncorrected minus conventionally-corrected		DIVE(t=1.5) minus conventionally-corrected	
		Mean	Range	Mean	Range
Rectum	V60Gy (%)	2.3	[-2.6; 9.1]	-0.2	[-1.6; 0.9]
Rectum	V52.8Gy (%)	-0.6	[-2.8; 0.8]	-0.5	[-1.9; 1.1]
Rectum	V30Gy (%)	0.4	[-8.8; 7.3]	0.9	[-7.1; 7.3]
Bladder	V60Gy (%)	0.8	[-2.4; 5.8]	-0.1	[-2.2; 1.5]
Bladder	V50Gy (%)	0.0	[-4.2; 1.8]	0.3	[-2.5; 2.2]
Bladder	V40.8Gy (%)	0.2	[-5.5; 3.1]	0.4	[-3.6; 2.8]
Bowel	D0.1cm³ (Gy)	1.0	[-4.5; 18.9]	-0.9	[-15.2; 16.2]
Bowel	V55Gy (%)	0.0	[0.0; 0.0]	0.0	[0.0; 0.0]
Bowel	V20Gy (cm³)	0.0	[-3.8; 7.3]	0.7	[-3.2; 6.7]
Anal canal	V60Gy (%)	0.3	[-2.7; 3.6]	-0.4	[-2.2; 0.4]
Anal canal	V35Gy (%)	-4.9	[-11.0; 0.2]	-3.5	[-8.5; 2.3]
Anal canal	V20Gy (%)	-7.1	[-17.9; 2.2]	-5.9	[-13.4; 3.0]
Left femoral head	V40Gy (%)	0.0	[0.0; 0.0]	0.0	[0.0; 0.0]
Left femoral head	V35Gy (%)	-0.1	[-0.8; 0.2]	-0.1	[-0.8; 0.0]
Right femoral head	V40Gy (%)	0.0	[0.0; 0.0]	0.0	[0.0; 0.0]
Right femoral head	V35Gy (%)	-0.2	[-1.8, 0.1]	-0.2	[-2.5; 0.1]
Penile bulb	V54.1Gy (%)	-1.1	[-5.5; 3.2]	-1.1	[-6.4; 3.6]
Penile bulb	V42.5Gy (%)	-0.7	[-5.2; 9.2]	-0.2	[-5.0; 9.4]

Acknowledgements

Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fond de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region”

References

1. Cha, E., et al., *Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy.* Radiotherapy and Oncology, 2021. **159**: p. 1-7.
2. Van der Veen, J., et al., *Benefits of deep learning for delineation of organs at risk in head and neck cancer.* Radiotherapy and Oncology, 2019. **138**: p. 68-74.
3. Fu, Y., et al., *A review of deep learning based methods for medical image multi-organ segmentation.* Physica Medica, 2021. **85**: p. 107-122.
4. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. 2015. Springer.
5. Sibolt, P., et al., *Clinical implementation of artificial intelligence-driven cone-beam computed tomography-guided online adaptive radiotherapy in the pelvic region.* Phys Imaging Radiat Oncol, 2021. **17**: p. 1-7.
6. Byrne, M., et al., *Varian ethos online adaptive radiotherapy for prostate cancer: Early results of contouring accuracy, treatment plan quality, and treatment time.* J Appl Clin Med Phys, 2022. **23**(1): p. e13479.
7. Zwart, L.G.M., et al., *Cone-beam computed tomography-guided online adaptive radiotherapy is feasible for prostate cancer patients.* Physics and Imaging in Radiation Oncology, 2022. **22**: p. 98-103.
8. Chapman, J.W., et al., *Robustness and reproducibility of an artificial intelligence-assisted online segmentation and adaptive planning process for online adaptive radiation therapy.* Journal of applied clinical medical physics, 2022. **23**(8): p. e13702.
9. Moazzezi, M., et al., *Prospects for daily online adaptive radiotherapy via ethos for prostate cancer patients without nodal involvement using unedited CBCT auto-segmentation.* J Appl Clin Med Phys, 2021. **22**(10): p. 82-93.
10. Rigaud, B., et al., *Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer.* International Journal of Radiation Oncology* Biology* Physics, 2021. **109**(4): p. 1096-1110.
11. Green, O.L., L.E. Henke, and G.D. Hugo. *Practical clinical workflows for online and offline adaptive radiation therapy*. in *Seminars in radiation oncology*. 2019. Elsevier.
12. Tetar, S.U., et al., *Clinical implementation of magnetic resonance imaging guided adaptive radiotherapy for localized prostate cancer.* Physics and imaging in radiation oncology, 2019. **9**: p. 69-76.
13. van Herk, M., et al., *Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis.* International journal of radiation oncology, biology, physics, 2018. **101**(5): p. 1057-1060.
14. Winkel, D., et al., *Adaptive radiotherapy: the Elekta Unity MR-linac concept.* Clinical and translational radiation oncology, 2019. **18**: p. 54-59.
15. Gungor, G., et al., *Time Analysis of Online Adaptive Magnetic Resonance-Guided Radiation Therapy Workflow According to Anatomical Sites.* Pract Radiat Oncol, 2021. **11**(1): p. e11-e21.
16. Van der Veen, J., et al., *Deep learning for elective neck delineation: More consistent and time efficient.* Radiotherapy and Oncology, 2020. **153**: p. 180-188.
17. Astrom, L.M., et al., *Online adaptive radiotherapy of urinary bladder cancer with full re-optimization to the anatomy of the day: Initial experience and dosimetric benefits.* Radiother Oncol, 2022. **171**: p. 37-42.

18. Yoon, S.W., et al., *Initial Evaluation of a Novel Cone-Beam CT-Based Semi-Automated Online Adaptive Radiotherapy System for Head and Neck Cancer Treatment - A Timing and Automation Quality Study.* Cureus, 2020. **12**(8): p. e9660.
19. Chen, X., et al., *CNN-based quality assurance for automatic segmentation of breast cancer in radiotherapy.* Frontiers in Oncology, 2020. **10**: p. 524.
20. Luan, S., et al., *Machine Learning-Based Quality Assurance for Automatic Segmentation of Head-and-Neck Organs-at-Risk in Radiotherapy.* Technology in Cancer Research & Treatment, 2023. **22**: p. 15330338231157936.
21. Isaksson, L.J., et al., *Quality assurance for automatically generated contours with additional deep learning.* Insights into Imaging, 2022. **13**(1): p. 1-10.
22. Claessens, M., et al., *Machine learning-based detection of aberrant deep learning segmentations of target and organs at risk for prostate radiotherapy using a secondary segmentation algorithm.* Physics in Medicine & Biology, 2022. **67**(11): p. 115014.
23. Hann, E., et al., *Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping.* Medical image analysis, 2021. **71**: p. 102029.
24. Balagopal, A., et al., *A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy.* Medical image analysis, 2021. **72**: p. 102101.
25. Claessens, M., et al. *Quality assurance for AI-based applications in radiation therapy*. in *Seminars in radiation oncology*. 2022. Elsevier.
26. Kaderka, R., et al., *Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients.* Radiotherapy and Oncology, 2019. **131**: p. 215-220.
27. Fung, N.T.C., et al., *Automatic segmentation for adaptive planning in nasopharyngeal carcinoma IMRT: time, geometrical, and dosimetric analysis.* Medical Dosimetry, 2020. **45**(1): p. 60-65.
28. Nguyen, D., et al., *A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning.* Scientific reports, 2019. **9**(1): p. 1-10.
29. Kandalan, R.N., et al., *Dose prediction with deep learning for prostate cancer radiation therapy: model adaptation to different treatment planning practices.* Radiotherapy and Oncology, 2020. **153**: p. 228-235.
30. Gronberg, M.P., et al., *Dose prediction for head and neck radiotherapy using a three-dimensional dense dilated U-net architecture.* Medical physics, 2021. **48**(9): p. 5567-5573.
31. Zimmermann, L., et al., *Technical Note: Dose prediction for radiation therapy using feature-based losses and One Cycle Learning.* Med Phys, 2021. **48**(9): p. 5562-5566.
32. Liu, S., et al., *A cascade 3D U-Net for dose prediction in radiotherapy.* Medical Physics, 2021. **48**(9): p. 5574-5582.
33. Archambault, Y., et al., *Making on-line adaptive radiotherapy possible using artificial intelligence and machine learning for efficient daily re-planning.* Med Phys Intl J, 2020. **8**.
34. Tezzele, M., et al., *PyGeM: Python geometrical morphing.* Software Impacts, 2021. **7**: p. 100047.
35. Nguyen, D., et al., *Three-dimensional radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture.* arXiv preprint arXiv:1805.10397, 2018.
36. Barragan-Montero, A.M., et al., *Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations.* Med Phys, 2019. **46**(8): p. 3679-3691.
37. Draguet, C., et al., *Automated clinical decision support system with deep learning dose prediction and NTCP models to evaluate treatment complications in patients with esophageal cancer.* Radiotherapy and Oncology, 2022. **176**: p. 101-107.

38. Aselmaa, A., et al., *The influence of automation on tumor contouring.* Cognition, Technology & Work, 2017. **19**(4): p. 795-808.
39. Challen, R., et al., *Artificial intelligence, bias and clinical safety.* BMJ Quality & Safety, 2019. **28**(3): p. 231-237.
40. Bond, R.R., et al., *Automation bias in medicine: the influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms.* Journal of electrocardiology, 2018. **51**(6): p. S6-S11.
41. Philpotts, L.E., *Can computer-aided detection be detrimental to mammographic interpretation?* Radiology, 2009. **253**(1): p. 17-22.

Supplementary materials

Appendix A.

The deformation process consists in defining an initial mesh represented by six control points, two for each spatial direction, placed on a sphere that circumscribes the volume of interest. We define a deformation scenario S_d by the displacement of a single control point while conserving the initial position of the other control points. Displacing a single control point results thus in deforming the initial mesh along a specific direction, which in turn deform the volume of interest.

Figure S1 gives an example of 2D deformation scenarios generation for 3 directions and 3 amplitudes. In this paper, 18 different directions of deformation and 9 amplitudes were used resulting in 162 deformation scenarios.

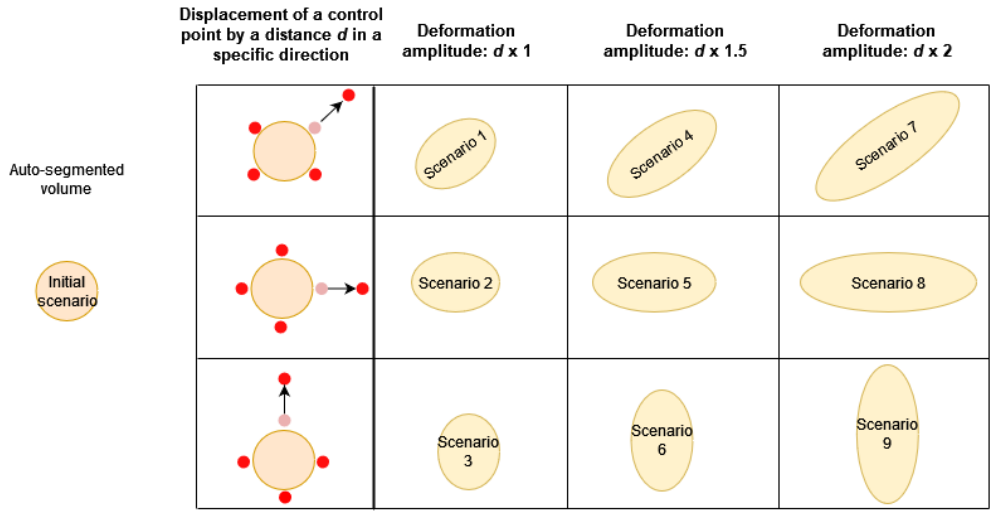

Figure S1: Illustration of deformation scenarios for a 2D case with 4 control points (red dots on the figure). Here, by making control point displacement in 3 directions and 3 amplitudes we obtained 9 deformation scenarios.

Appendix B.

Figure S2 illustrates the structure of the HD-UNET used for the dose prediction. There are 9 input channels: 1 for dose prescription mask, which consists of the prescription dose for the voxels inside the PTV prostate (60Gy), PTV seminal vesicles (48Gy) and 0 for voxels outside; and 7 for the OARs (the anal canal, bladder, rectum, left femoral head, right femoral head, bowel and penile bulb). Convolutional operations were performed with 16 filters and max-pooling was used to lower the resolution by a factor of two (2x2x2).

The dose distribution matrix, and OAR masks were interpolated to the CBCT grid. Our HD-UNET was trained using image patches with a size of [320 x 320 x 64]. Data were augmented by flipping training data in the left-right direction to double the dataset size. We trained it for 200 epochs and kept the model weights of the last computed epoch.

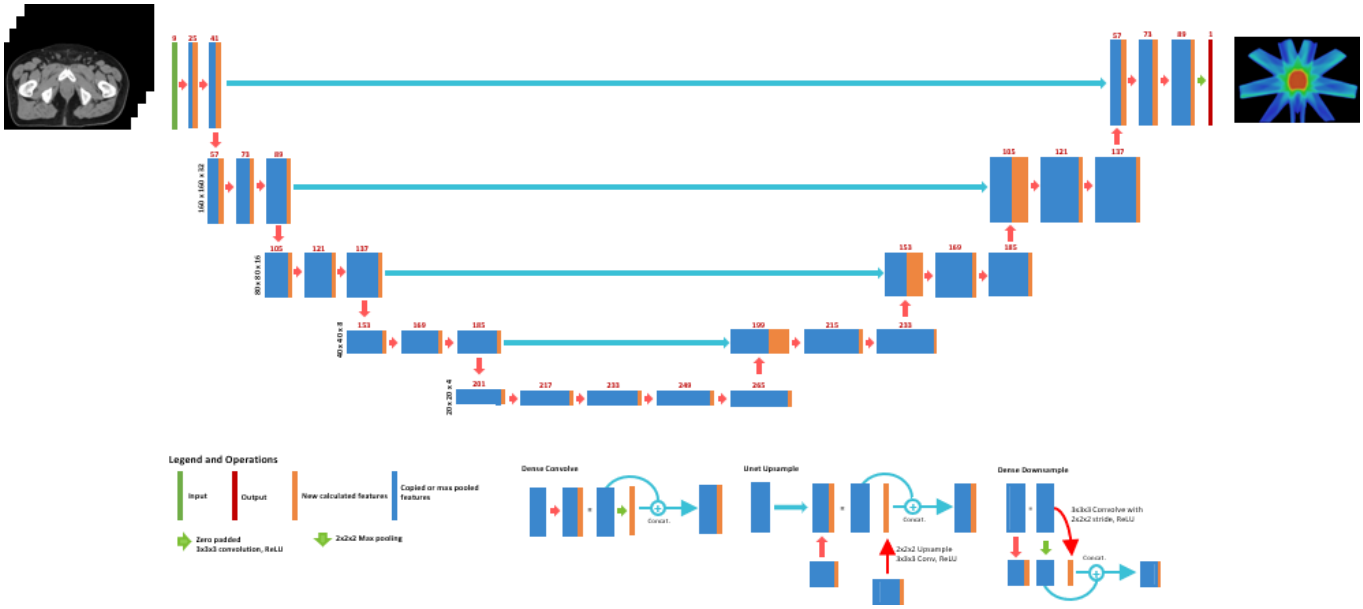

Figure S2: Structure of HD-UNET used for dose prediction.

Appendix C.

The α values used as weight for IOC computations are depicted on Table S3. Those were computed on the 180 fractions also used as training data for the DCNN dose prediction. When the dose objectives were well met, α_i were halved.

Table S3: α values for IOC computations for each of the considered metrics.

Volume	Metric	α	Metric condition to use α/2
CTV prostate	D98%	1.78	>58.8Gy
CTV seminal vesicles	D98%	1.90	>47.04Gy
PTV prostate	D95%	1.31	>57Gy
PTV seminal vesicles	D95%	1.69	>45.6Gy
Rectum	D1%	1.65	<59Gy
	V52Gy	0.56	<8Gy
	V30Gy	0.37	<25Gy
Anal canal	D1%	0.49	<59Gy
Bladder	D1%	1.76	<59Gy
	V50	0.28	<12Gy
	V40	0.20	<15Gy

Appendix D.

Figure S4 illustrates the aggregation process of the different IOC according to their deformation scenario. The aggregation process begins with the lowest deformation amplitude. The space is filled with the IOC value of each deformation scenario. When deformed masks overlap, the space is filled with averaged IOC. Once all the deformation scenario of the deformation amplitude has been used, the filled voxels are locked and can no longer be updated. The process continues with the deformation scenarios of the second lowest deformation amplitude and so on.

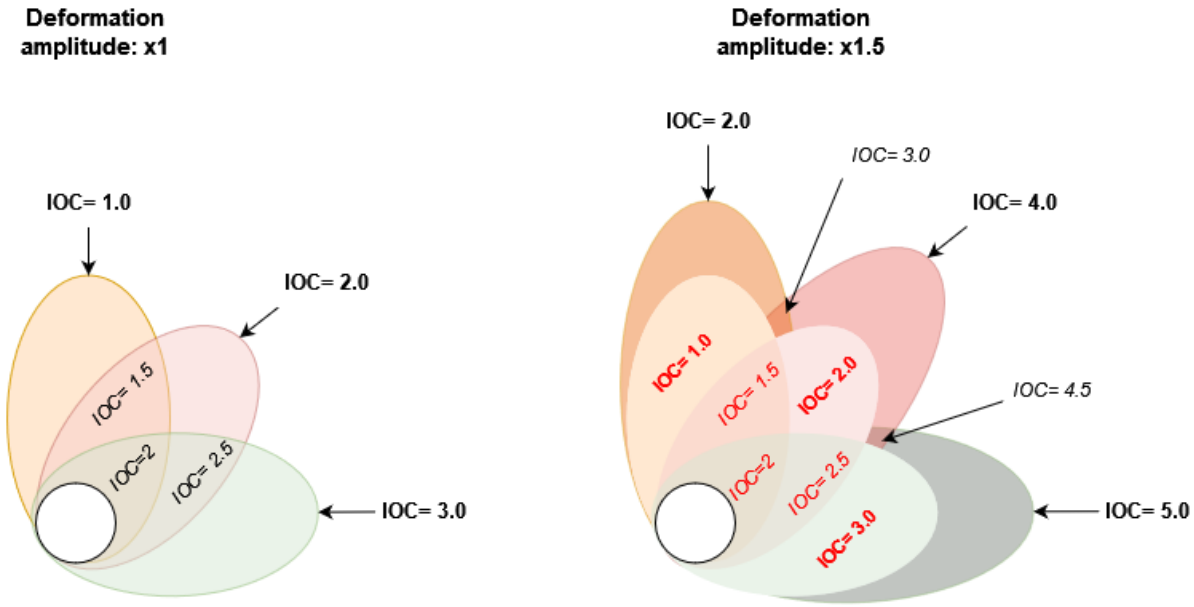

Figure S4: Illustration of heatmap aggregation for 3 directions of deformation and 2 amplitudes. Bolded IOC values are computed following the presented formula, italic ones are IOC average of overlapping scenarios. Once all the computations for an amplitude are done, all the updated voxels are locked. Those are represented by the red IOC values. Then, computations for the following deformation amplitude are performed and so on. Note that here in this example, contours color only helps to differentiate the scenarios and does not represent any dosimetric value.