Physics and Imaging in Radiation Oncology 34 (2025) 100744

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

Physics-based data augmentation for improved training of cone-beam computed tomography auto-segmentation of the female pelvis

Yvonne J.M. de Hond [a,*], Paul M.A. van Haaren [a], An-Sofie E. Verrijssen [a], Rob H.N. Tijssen [a], Coen W. Hurkmans [b]

[a] Department of Radiation Oncology, Catharina Hospital Eindhoven, Eindhoven, the Netherlands
[b] Department of Electrical Engineering and Department of Applied Physics and Science Education, Technical University Eindhoven, Eindhoven, the Netherlands

ARTICLE INFO

Keywords:
Auto-segmentation
Cone-beam computed tomography
Cervix
Deep-Learning
nn-UNet
Female Pelvis

ABSTRACT

Background and Purpose: Labeling cone-beam computed tomography (CBCT) images is challenging due to poor image quality. Training auto-segmentation models without labelled data often involves deep-learning to generate synthetic CBCTs (sCBCT) from planning CTs (pCT), which can result in anatomical mismatches and inaccurate labels. To prevent this issue, this study assesses an auto-segmentation model for female pelvic CBCT scans exclusively trained on delineated pCTs, which were transformed into sCBCT using a physics-driven approach.
Materials and Methods: To replicate CBCT noise and artefacts, a physics-driven sCBCT (Ph-sCBCT) was synthesized from pCT images using water-phantom CBCT scans. A 3D nn-UNet model was trained for auto-segmentation of cervical cancer CBCTs using Ph-sCBCT images with pCT contours. This study included female pelvic patients: 63 for training, 16 for validation and 20 each for testing on Ph-sCBCTs and clinical CBCTs. Auto-segmentations of bladder, rectum and clinical target volume (CTV) were evaluated using Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95). Initial evaluation occurred on Ph-sCBCTs before testing generalizability on clinical CBCTs.
Results: The model auto-segmentation performed well on Ph-sCBCT images and generalized well to clinical CBCTs, yielding median DSC's of 0.96 and 0.94 for the bladder, 0.88 and 0.81 for the rectum, and 0.89 and 0.82 for the CTV on Ph-sCBCT and clinical CBCT, respectively. Median HD95′s for the CTV were 5 mm on Ph-sCBCT and 7 mm on clinical CBCT.
Conclusions: This study demonstrates the successful training of auto-segmentation model for female pelvic CBCT images, without necessarily delineating CBCTs manually.

1. Introduction

Inter-fraction motion in the pelvic area can be substantial during radiation therapy [1–3]. To manage inter-fraction motion, online adaptive radiotherapy (ART) is emerging, particularly with the introduction of magnetic resonance linear accelerators (MR-Linac) and ring-based cone-beam computed tomography (CBCT) guided Linacs, such as the Ethos,Varian. Online ART is expanding to C-arm CBCT Linacs as well, despite the generally poor image quality of the CBCT compared to the CT. Within online ART, treatment plans are dynamically adjusted based on the daily anatomy on the CBCT[4,5]. This type of online ART involves the necessity of fast and accurate contours on the CBCT scan, which is challenging with manual contouring.

One approach to obtain contours on CBCT is via deformable image registration (DIR), by transferring the contours from the planning CT (pCT) to the CBCT. The accuracy of DIR, however, is limited by the substantial anatomical changes in the pelvic region and the suboptimal image quality of the CBCT [6]. More recent approaches involve deep-learning (DL) techniques for auto-segmentation[7–10]. For the training of these models, labeled CBCT data is often necessary. Unlike pCT scans, the delineation of CBCT scans is not typically integrated into the clinical workflow and is more tedious and error-prone due to poor image quality.

To avoid manually contouring vast amount of CBCT images for training, several methods have been explored. For example, DL-models are trained with labeled pCT and unlabeled CBCT data[7]. However, defining loss functions for both labeled and unlabeled data is complicated, which can result in overfitting with small labeled datasets or error

* Corresponding author at: Department of Radiation Oncology, Catharina Hospital Eindhoven, Michelangelolaan 2 5623 EJ, Eindhoven, the Netherlands.
E-mail address: yvonne.d.hond@catharinaziekenhuis.nl (Y.J.M. de Hond).

https://doi.org/10.1016/j.phro.2025.100744
Received 25 November 2024; Received in revised form 17 February 2025; Accepted 27 February 2025
Available online 7 March 2025

accumulation of the predicted contours during training. Another approach to train DL-models without labeling CBCT data involves DL-models for image synthesis, generating synthetic CBCT (sCBCT) data from pCT[9]. Subsequently, the contours of the pCT can be transferred to the sCBCT. These DL-sCBCT images are used as input for the auto-segmentation model. However, image synthesis can introduce anatomical mismatches between the pCT and DL-sCBCT[11]. This anatomical mismatch can cause incorrect labeled DL-sCBCTs, which decreases the performance of the auto-segmentation model. A DL-free method to generate sCBCT images could solve the incorrect labeled sCBCTs. Therefore, the aim of this study was to evaluate the performance of an auto-segmentation model for female pelvic CBCT scans exclusively trained on delineated pCTs, which were transformed into sCBCT using a physics-driven approach.

2. Materials and methods

2.1. Patient data

This study included data of 109 consecutive female pelvic cancer patients, all treated at Catharina hospital, Eindhoven. pCT images were acquired between January 2020 and January 2023 on a Philips big bore CT (Philips, Eindhoven, Netherlands) with 120 kVP, 512 × 512 voxels, pixel size 1.17 mm and 3 mm slice thickness. The pCT without contrast agent in the bladder was used for this study. For testing the auto-segmentation model on clinical CBCT images, CBCT images were acquired on multiple Xray Volumetric Imaging (XVI) v5.0.6 and v5.0.7 systems (Elekta AB, Stockholm, Sweden) between January 2021 and September 2023. CBCT imaging settings were: medium Field-of-View, bowtie filter, 64 mA and 40 ms per projection, 5.5 frames/sec in a

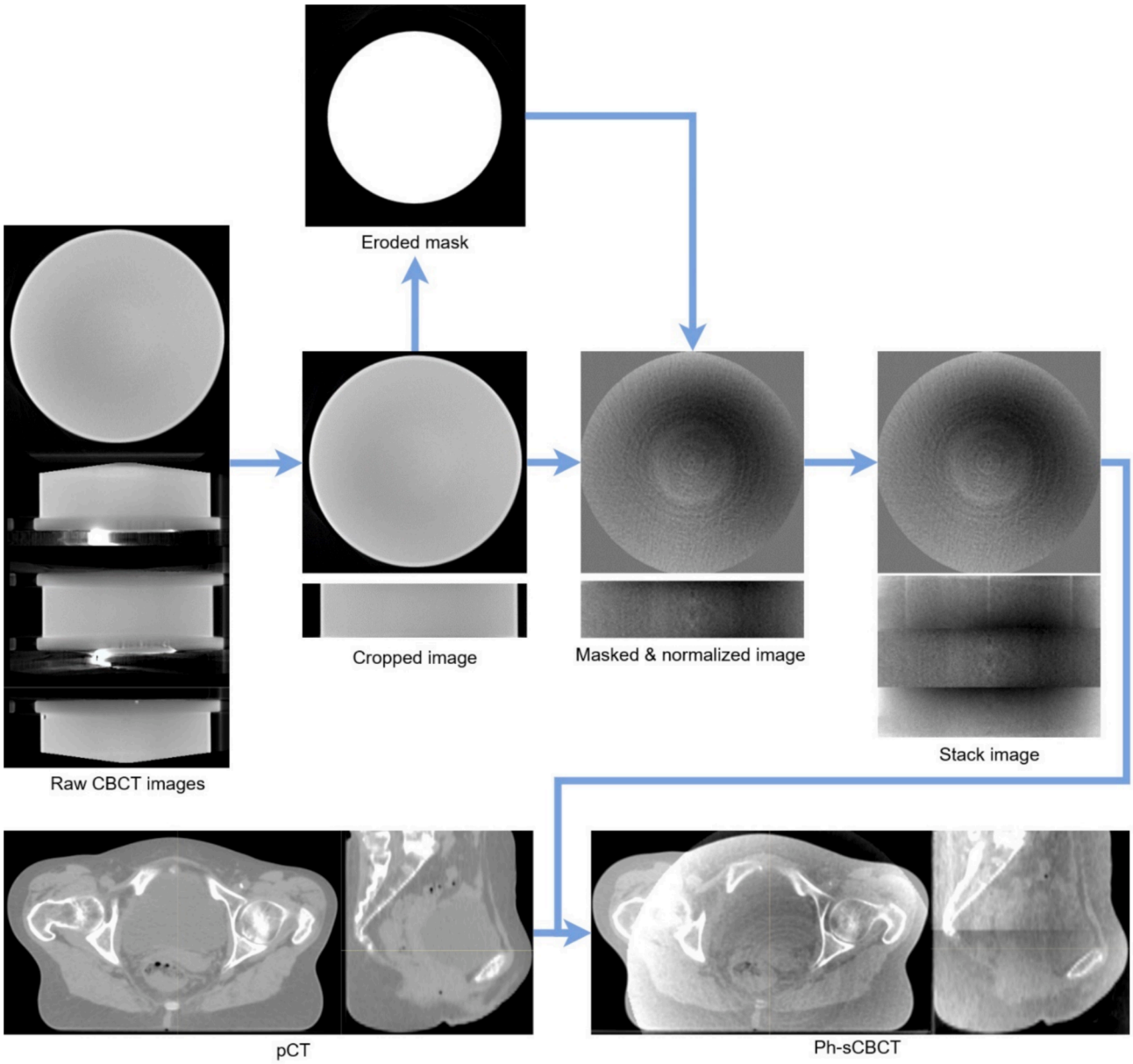

Fig. 1. Illustration of the pipeline to generate Ph-sCBCT images from pCT images. First, three CBCT scans of a water-phanthom were made. Second, the CBCT water images were cropped. Third, to only include water voxels, the image was masked. Fourth, the average value of the water voxels was set to zero. Fifth, the three water images were stacked to increase the image in craniocaudal direction. The transitions of the stacked scans were inhomogeneous, resulting in a noticeable inconsistency in the Ph-sCBCT image. Finally, the stacked water image was elementwise added to the pCT image in order to generated Ph-sCBCT images, which added CBCT artefacts and scatter. The entire pipeline is implemented using Python, and the corresponding code is available for academic and research purposes. It can be accessed at the following repository: https://github.com/ydehondcze/creating_Ph-sCBCT.

360° arc. First fraction CBCT images were used in this study. Patients with metal orthopedic implants and patients who had undergone hysterectomy were excluded. The included patients had a median age of 64 years [range: 29–91], and a median Body Mass Index of 24 kg/m^2 [range: 18 kg/m^2 – 34 kg/m^2]. This research used anonymized patient data and received approval under Dutch law for a medical research waiver (non-WMO legislation).

2.2. Training and validation data preparation

pCT images were delineated according to EMBRACE II guidelines for clinical purposes or retrospectively by radiotherapy technologists[12]. The organs-at-risk (OAR) and target structures were rectum, bladder, uterus, cervix, vagina and parametria. The clinical as well as retrospective delineations were verified by a radiation oncologist. The delineations of the uterus, cervix, vagina and parametria were combined to a target area, the clinical target volume (CTV). The bowel contrast was removed in pCT images, in which oral contrast was used during scanning. The contrast was removed by adjusting each voxel above 90 HU within the bowel structure to a random value between 0 HU and 60 HU to replicate bowel tissue and CT scatter.

The pCT images have reduced noise and artefacts compared to the CBCT scans. To incorporate the inherent noise and artefacts characteristics of CBCT scans into to the training of the auto-segmentation algorithm, the pCT image was transformed in a physics driven sCBCT (Ph-sCBCT) image. In Fig. 1, the pipeline to generate Ph-sCBCT images is illustrated. First, CBCT scans of a homogeneous medium were made to exclusively capture noise and artefacts generated by the x-ray beam and detector. As homogeneous medium, a Harmonized Phantom Body 40 mm Section (Philips Healthcare Co.) filled with demi water was used. Since the dimensions of this water-phantom were smaller than the CBCT field of view (FOV), three CBCT scans were performed per XVI system. The water-phantom was displaced by 80 mm per scan to encompass the complete longitudinal range of the detector. The CBCT acquisition protocol was identical to the patient CBCT acquisition protocol.

Second, the water-phantom scans were cropped and masked to exclude the treatment-table and phantom boundary artefacts. This way only the voxels within the water-phantom were used resulting in a water image. Third, the average pixel intensity value of the water image was normalized to zero. To amplify the noise and ring artefacts, the normalized water-image was multiplied by four. Fourth, the processed water images were stacked in the order of the scanning position to form a larger longitudinal range. The transitions of the stacked scans were inhomogeneous, resulting in a noticeable inconsistency in the Ph-sCBCT image.

Last step was adding the processed water images element wise to the pCT to create Ph-sCBCT images. The center of the water image coincided with the center of the CTV structure on the pCT. To improve the auto-segmentation model robustness for patient positioning, variability in CBCT artefact positions was introduced by randomly translating the water scan between −20 mm to 0 mm in the craniocaudal direction, and between −15 mm and 15 mm in the other two directions. In addition to this translation, the water scan was randomly rotated between −90° and 90°. Since the native space of the Ph-sCBCT images is equal to that of the pCT images, the delineated structures of the pCT were used. The entire pipeline is implemented using Python, and the is available for academic and research purposes on https://github.com/ydehondcze/creating_Ph-sCBCT.

Ring artefacts in CBCT images are typically caused by inconsistencies in the response of individual detector elements, leading to variations in signal intensity across the detector array. These intensity differences are detector panel- and XVI system-specific. To increase the auto-segmentation model's robustness for multiple XVI systems, the Ph-sCBCT generation pipeline was performed six times per patient, using water-phantom scans acquired on six different XVI systems.

2.3. Auto-segmentation model

The nn-UNet was selected as DL-algorithm for this study based on its described good performance and suitability for a large range of image types[13]. The automated configuration of nn-UNet operates autonomously, requiring no manual intervention, when applied to a novel dataset. For training, the 3D U-Net architecture was used. The DL-model was trained on full-resolution image data. Five folds cross-validation was used and an ensemble of the folds was used to generate the final prediction. In each fold, 63 patients were used for auto-segmentation training and 16 patients for validation. Since multiple Ph-sCBCT images were generated per patient, each fold of the Ph-sCBCT model consisted of train/validation images of 378/96 3D images, respectively. Training was performed on a RTX5000 graphic card and lasted 34 h per fold. Loss-functions were Dice and cross-entropy[13]. Loss-curves of the five folds training and validation are given in supplemental materials Fig. A1.

2.4. Test data preparation and evaluation of auto-segmentation model

To test the auto-segmentation model on Ph-sCBCT data, images of twenty patients were preprocessed as described in section 2.2. To evaluate the performance of the model on clinical CBCT data, additional XVI CBCT images of 20 patients were included in the test set. To evaluate the auto-segmentations, manual delineations on the clinical CBCT images were used as reference segmentations. The bladder, rectum and CTV were delineated. These delineations were verified by a radiation oncologist.

Auto-segmentations of the DL-model were compared to the reference segmentations using the volumetric Dice similarity coefficient (DSC), 95th percentile hausdorff distance (HD95) and average surface distance (ASD). These metrics were calculated in 3D using the MedPy library. To standardize the delineations, slices below the caudal part of the bladder and slices above the cranial part of CTV were excluded from the reference contour. Evaluation was only performed in areas where the reference contour was present. An example in which these caudocranial restrictions of the metric calculation was used is given in the sagittal view of the clinical CBCT segmentation (Fig. 2a). To statistically test if there was a significant difference in auto-segmentation performance on Ph-sCBCT images and clinical CBCT images, a Mann-Whitney *U* test was used [14]. A p-value below 0.05 was considered significant.

3. Results

3.1. Visual inspection of auto-segmentation performance

The examples in Fig. 2 were selected based on the combined DSC and HD95 scores of all three segmented structures. In the good case typical example of the Ph-sCBCT and clinical CBCT image (Fig. 2a), the manual and auto-segmented contours were in general comparable, except for the caudal part of the CTV and bladder on the clinical CBCT image as shown in the sagittal view. In the rectum auto-segmentation of the clinical CBCT image, more transversal slices were segmented compared to the manual segmentation, due to limited manual delineation as described in section 2.4. Although the manual contour and auto-segmented contour differ in the anterior part of the CTV edges in the average case (sagittal view in Fig. 2b), the auto-segmented contour resembled the anatomy on the CBCT better in the edge between CTV and bladder. Furthermore, the cranial part of the rectum was further delineated in the manual compared to the auto-segmentation. Except for these discrepancies, the auto-segmented contours generated by the DL-model were in general comparable to the manual contours. In the poor-case example of the Ph-sCBCT image (Fig. 2c), the auto-segmented rectum contour differed from the manual segmentation on the cranial part of the rectum as shown in the sagittal image. In addition, the auto-segmented bladder contour differed from the manual contour on the

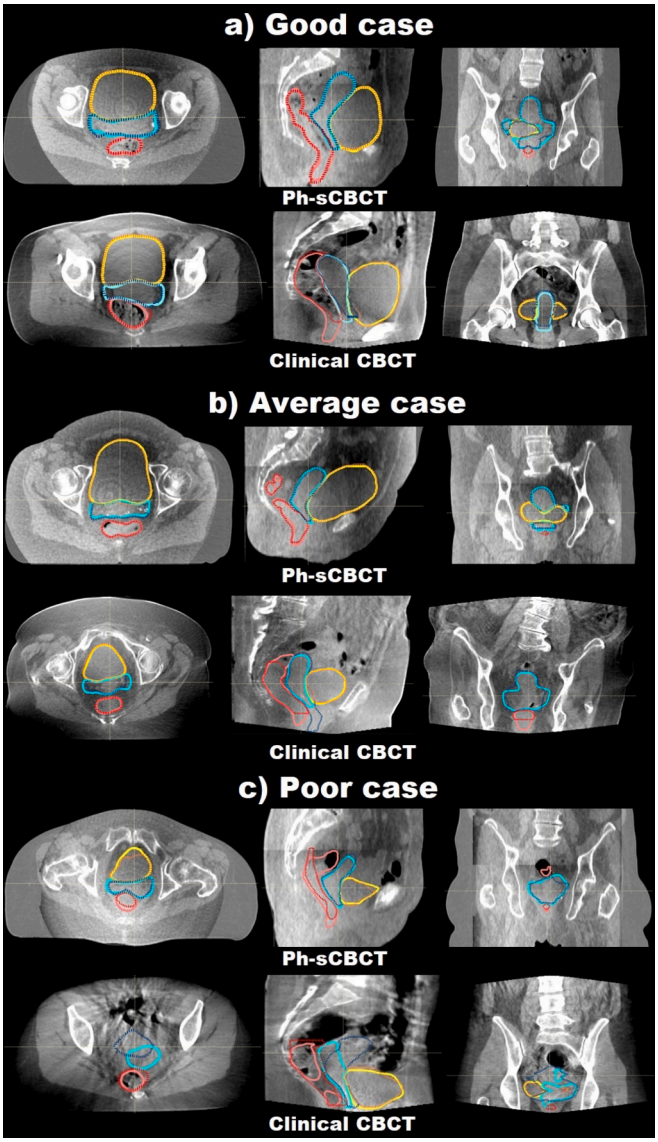

Fig. 2. Typical examples of a good (a), average (b) and poor (c) case performance of auto-segmentation of the bladder (yellow), CTV (blue) and rectum (red) on Ph-sCBCT and clinical CBCT images. From left to right, transversal image, sagittal image and coronal image. The auto-segmented contours generated by the nn-UNet trained on Ph-sCBCT images are illustrated with solid lines and darker color. Manual contours are illustrated with a dotted line and lighter color. For Ph-sCBCT, the DSCs in good, average, and poor cases were 0.98/0.97/0.91 for the bladder, 0.93/0.87/0.71 for the rectum, and 0.92/0.87/0.83 for the CTV. The HD95 in these cases were 2 mm/2 mm/3 mm for the bladder, 3 mm/4 mm/19 mm for the rectum, 4 mm/ 6 mm/ 6 mm for the CTV. For clinical CBCT, the DSCs in good, average, and poor cases were 0.96/0.92/0.93 for the bladder, 0.88/0.89/0.61 for the rectum, and 0.87/0.80/0.59 for the CTV. The HD95 in these cases were 3 mm/4 mm/4 mm for the bladder, 8 mm/7 mm/ 16 mm for the rectum, 5 mm/7 mm/24 mm for the CTV. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

anterior part of the bladder (transversal view). In the poor-case example of the clinical CBCT a discrepancy was observed in the rectum and uterus part of the CTV as shown in sagittal and transversal image. The auto-segmented CTV did not include the anterior part of the uterus.

3.2. Quantitative analysis of auto-segmentation performance

Overall, the performance of the auto-segmentations on the clinical CBCT was significantly less compared to the auto-segmentations on native Ph-sCBCT images, albeit still achieving high scores (Fig. 3). The median DSC values for clinical CBCT reached 0.94 for the bladder, 0.81 for the rectum, and 0.82 for CTV, with HD95 and ASD scores also indicating reasonable alignment with manual contours (Table 1). Although the performance of the auto-segmentation model was slightly lower when applied on clinical CBCT images compared to Ph-sCBCT images, the manual and auto-segmented contours on the clinical CBCT images were in general comparable.

4. Discussion

This study demonstrates the feasibility of training auto-segmentation models for female pelvis CBCT images exclusively trained on Ph-sCBCT images with pCT contours by using a physics-driven approach. In contrast to auto-segmentation training on manually delineated CBCT images, this approach exclusively requires clinical contoured CBCT patient data for testing the auto-segmentation model, an advantage that facilitates agile updating of the segmentation model to new imaging settings.

The DSC of auto-segmentation of the CTV on native data, Ph-sCBCT, were above 0.80. This DSC was within the reported manual delineation variability of cervical cancer CTV delineation on pCT images, ranging between 0.72 and 0.90, in a quantification study among radiotherapy centers in the United Kingdom[15]. On non-native clinical CBCT data, the segmentation model performed less compared to testing on native Ph-sCBCT data. Although the auto-segmentation model was exclusively trained on native Ph-sCBCT data and clinical CBCT data was new for the segmentation model, the DSC of the CTV on clinical CBCT was, in 90 % of the cases, within this delineation variability on pCT images. Therefore, the generalization of the auto-segmentation model to non-native clinical CBCT data was sufficient despite the significant lower contrast-to-noise ratio (CNR) in the clinical CBCT compared to the Ph-sCBCT images (supplementary materials A2). A higher proportion of segmentations on clinical CBCT images obtained through auto-segmentation (77 % of cases) met the AAPM TG-132 guidelines, which recommend a DSC tolerance of 0.8–0.9, compared to only 20 % of cases where DIR was used to deform pCT segmentations onto clinical CBCT images (see supplementary materials A3)[16]. Therefore, the evaluated auto-segmentation model is more suitable for dose calculations compared to acquiring contours through DIR.

The DSC and HD95 metrics for the CTV auto-segmentation were comparable to the uterus segmentation metrics on Ethos CBCT images [4]. These metrics for the CTV auto-segmentation were acceptable given the number of factors that challenges CTV contouring. On the one hand, the image quality of the XVI clinical CBCT images is poor compared to Ethos CBCT images. On the other hand, CTV delineating is more challenging compared to a single organ, such as the uterus, as the CTV encompasses multiple structures, including the cervix and parametria. This complexity contributes to a higher degree of interobserver variability in defining the cervix-uterus CTV compared to the uterus alone, as a result of variations in the inclusion of the parametria and vagina[17]. In several poor auto-segmented CTV cases, discrepancies in the parametria region were observed between manually and auto-segmented CTV (supplementary materials A4). This difference in segmentation was most likely caused by variability in delineation of the parametria as the definition of the CTV is partially motivated by the clinical experience [17].

The median DSC (0.94) and HD95 (4 mm) of the bladder segmentations were comparable to metrics reported in a study where an auto-segmentation model was trained and tested on Ethos CBCT images (0.94 and 4 mm)[4]. Although the rectum DSC and HD95 was lower compared to auto-segmentation on Ethos CBCT, our rectum DSC (median: 0.81) and HD95 (median: 9 mm) was comparable or better to auto-segmentation model trained on manual delineated CBCT images (median: 0.70–0.75 and 14 mm)[4,8,18]. The median DSC of auto-

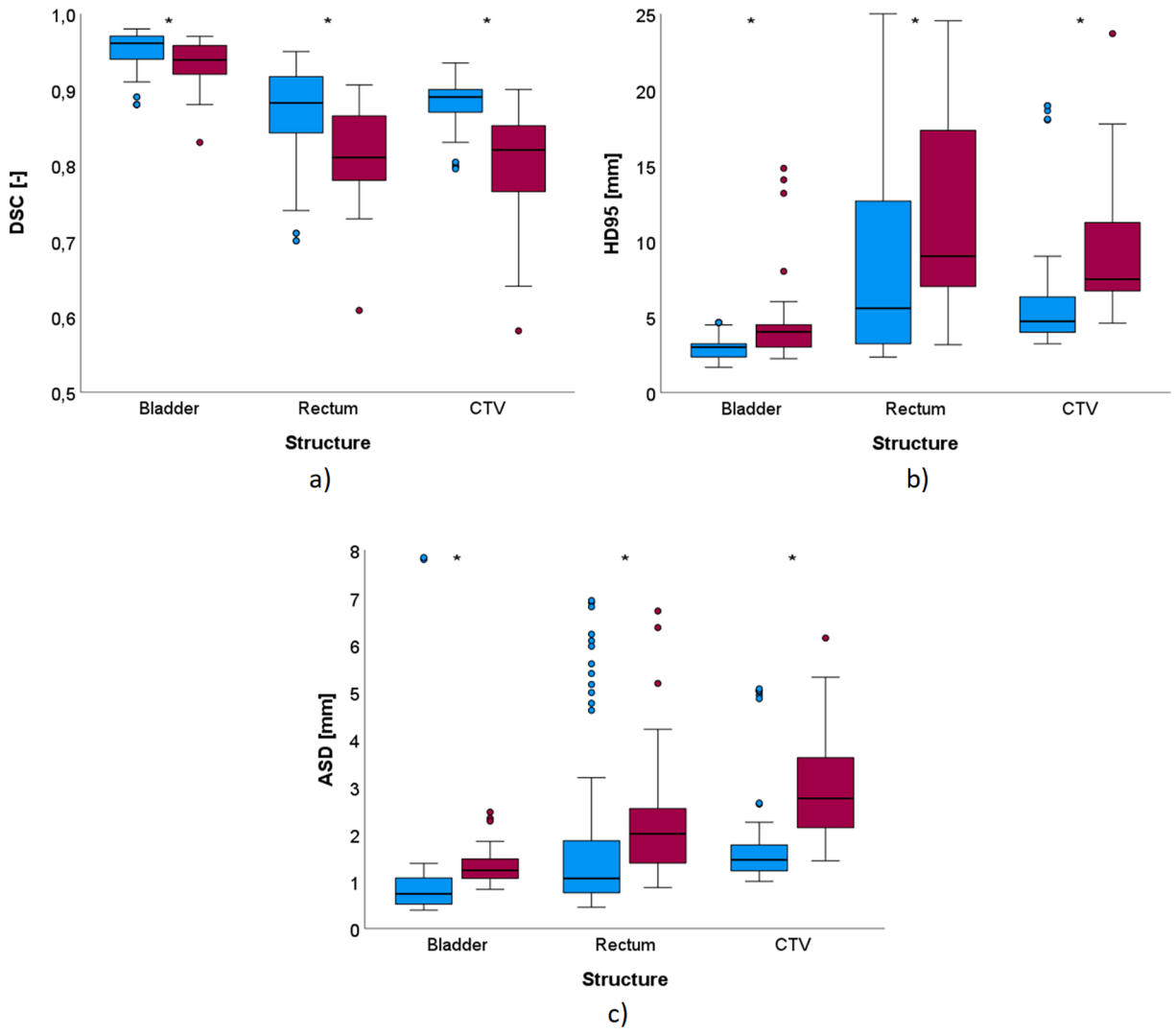

Fig. 3. Metrics of the bladder, rectum and CTV contours generated by the autosegmentation model on Ph-sCBCT (blue) and clinical CBCT images (red). a) DSC. b) HD95 in mm. c) ASD in mm. The upper and lower extents of the boxes represent the 25th and 75th percentile. Horizontal lines in boxes are medians, dots are outliers. Significant differences between Ph-sCBCT and clinical CBCT are indicated with a star. In Figure C, two outliers exceed the HD95 scale, with values of 43 mm and 44 mm for bladder segmentation on the Ph-sCBCT image. Examples of these outliers are provided in supplementary materials A4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of auto-segmentation performance on evaluation metrics. Metrics are reported as median [range]. P-values are reported of statistical test between performance on Ph-sCBCT and Clinical CBCT.

Structure	Test image type	DSC [-]	ASD [mm]	HD95 [mm]
Bladder	Ph-sCBCT	0.96 [0.88–0.98]	1 [0–8]	3 [2–44]
	Clinical CBCT	0.94 [0.88–0.97]	1 [1–2]	4 [2–15]
	p-value	<0.001	<0.001	<0.001
Rectum	Ph-sCBCT	0.88 [0.70 – 0.95]	1 [0–7]	6 [2–26]
	Clinical CBCT	0.81 [0.61–0.91]	2 [1–6]	9 [4–24]
	p-value	<0.001	0.001	0.006
CTV	Ph-sCBCT	0.89 [0.80 – 0.94]	1 [1–5]	5 [3–19]
	Clinical CBCT	0.82 [0.58–0.90]	3 [1–5]	7 [5–24]
	p-value	<0.001	<0.001	<0.001

segmentation of the bladder on clinical CBCT images (0.94) was within the range of reported average OAR delineation variability for volumes larger than 8 cc on pCT images (0.91)[19]. Although, the median DSC of the rectum auto-segmentation on clinical CBCTs (0.81) was lower than this reported variability, the median DSC of the rectum auto-segmentation on clinical CBCTs was within the range of reported values for bowel delineation on pCT images (0.77–0.88)[19].

Zhang et al. published a comparable auto-segmentation study on cervical cancer CBCT images[8]. The used model (i.e., the nn-UNet described in[13]) was similar to the model used in this work. Further, the CBCT data was acquired with a CBCT system of the same vendor. In contrast to this study, however, the model in Zhang et al. was trained on a dataset of manually delineated CBCT images. The results presented by Zhang et al. showed slightly worse performance compared to the results presented in our work. To compare the influence of the different training methods, comparable datasets are required. However, 79 patients were used in our dataset compared to 23 patients.

To overcome the need to manually contour vast amount of CBCT images for training, Ph-sCBCT images were used in this study instead of the DL-sCBCT images as described by Schreier et al. or pCT images [9]. The performance with pCT images was significantly worse whereas Ph-sCBCT and DL-sCBCT images yielded comparable results (supplementary materials A5). However, the advantages of generating Ph-sCBCT images compared to DL-sCBCT images are 1) the smaller amount of CBCT data required and 2) the ensured anatomical accuracy. For the generation of DL-sCBCT images, a pCT to DL-sCBCT generator is necessary. Creating such a generator necessitates CBCT and pCT images

on top of the dataset used for training of the auto-segmentation model. This poses challenges in obtaining two large enough datasets of patients. Conversely, the Ph-sCBCT generator requires only a few water-phantom CBCT scans. These few scans facilitated the generation of multiple Ph-sCBCT images from one single patient, resulting in a larger augmented training dataset. Beyond differences in required patient data, the DL-sCBCT generator can induce anatomical differences, which can introduce bias in the auto-segmentation model[11].

The water-phantom was used to capture detector and scatter characteristics of the CBCT system. Ring artefacts in CBCT images are typically caused by inconsistencies in the response of individual detector elements, leading to variations in signal intensity across the detector array. These intensity differences are detector panel and XVI system specific. To enlarge the auto-segmentation model's robustness for multiple XVI systems, water-phantom scans of multiple XVI systems were made. Further research to adding multi-institutional data could enlarge model robustness to imaging settings and detector panels. This approach has the potential to enhance robustness for inter-machine variability, while simultaneously mitigating the need for patient data sharing, thereby simplifying data-sharing logistics and addressing privacy concerns.

To evaluate the clinical applicability of auto-segmentation on clinical CBCT images, further research on auto-segmentation in clinical processes is necessary. For example, qualitative analysis on auto-segmentation performance and the needed time to correct the auto-segmentations and applicability of the auto-segmentations for online adaptive radiotherapy on a C-arm CBCT based Linac is a subject that requires more analysis.

In conclusion, an auto-segmentation model was successfully trained for female pelvic CBCT images, without the need to manually delineate on CBCT images, achieving performance comparable to state-of-the-art models. The presented method avoids complex loss functions for labeled and unlabeled data and avoids anatomical disparities between CBCT and CT. Our model relies exclusively on pCT images and contours for training, generating Ph-sCBCT images with a physics-driven approach.

Credit author statement

All authors contributed to the conception of the work, analysis and interpretation of data. All authors contributed to and approved the final version of the manuscript. All authors agreed to be accountable for the work in ensuring that questions related to the accuracy or integrity of the manuscript are appropriately investigated and resolved.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used chat gpt 4.0 in order to check written sentences on language errors and to reduce the number of words in a section. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yvonne de Hond was financially supported by a research grant by Elekta (grant number SOW_20210426, Elekta Ltd., Crawley, UK).

Acknowledgments

We thank the radiation therapy technologists at Catharina Hospital for delineating the pCTs used in this study. Yvonne de Hond was financially supported by a research grant by Elekta (grant number SOW_20210426, Elekta Ltd., Crawley, UK).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2025.100744.

References

[1] Buchali A, Koswig S, Dinges S, Rosenthal P, Salk J, Lackner G, et al. Impact of the filling status of the bladder and rectum on their integral dose distribution and the movement of the uterus in the treatment planning of gynaecological cancer. Radiother Oncol 1999;52:29–34. https://doi.org/10.1016/s0167-8140(99)00068-7.

[2] Jadon R, Pembroke CA, Hanna CL, Palaniappan N, Evans M, Cleves AE, et al. A systematic review of organ motion and image-guided strategies in external beam radiotherapy for cervical cancer. Clin Oncol (R Coll Radiol) 2014;26:185–96. https://doi.org/10.1016/j.clon.2013.11.031.

[3] Li X, Wang L, Cui Z, Li Y, Liu P, Wang Y, et al. Online MR evaluation of inter- and intra-fraction uterus motions and bladder volume changes during cervical cancer external beam radiotherapy. Radiat Oncol 2021;16:179–01907. https://doi.org/10.1186/s13014-021-01907-1.

[4] Shelley CE, Bolt MA, Hollingdale R, Chadwick SJ, Barnard AP, Rashid M, et al. Implementing cone-beam computed tomography-guided online adaptive radiotherapy in cervical cancer. Clin Transl Radiat Oncol 2023;40:100596. https://doi.org/10.1016/j.ctro.2023.100596.

[5] Branco D, Mayadev J, Moore K, Ray X. Dosimetric and feasibility evaluation of a CBCT-based daily adaptive radiotherapy protocol for locally advanced cervical cancer. J Appl Clin Med Phys 2023;24:e13783. https://doi.org/10.1002/acm2.13783.

[6] Swamidas J, Kirisits C, De BM, Hellebust TP, Siebert FA, Tanderup K. Image registration, contour propagation and dose accumulation of external beam and brachytherapy in gynecological radiotherapy. Radiother Oncol 2020;143:1–11. https://doi.org/10.1016/j.radonc.2019.08.023.

[7] Brion E, Léger J, Barragán-Montero AM, Meert N, Lee JA, Macq B. Domain adversarial networks and intensity-based data augmentation for male pelvic organ segmentation in cone beam CT. Comput Biol Med 2021;131:104269. https://doi.org/10.1016/j.compbiomed.2021.104269.

[8] Zhang C, Lafond C, Barateau A, Leseur J, Rigaud B, Chan Sock Line DB, Yang G, Shu H, Dillenseger JL, de CR and Simon A. Automatic segmentation for plan-of-the-day selection in CBCT-guided adaptive radiation therapy of cervical cancer. Phys. Med. Biol 2022;67:245020. https://doi.org/10.1088/1361-6560/aca5e5.

[9] Schreier J, Genghi A, Laaksonen H, Morgas T, Haas B. Clinical evaluation of a full-image deep segmentation algorithm for the male pelvis on cone-beam CT and CT. Radiother Oncol 2020;145:1–6. https://doi.org/10.1016/j.radonc.2019.11.021.

[10] Beekman C, van BS, Stam J, Sonke JJ, Remeijer P. Improving predictive CTV segmentation on CT and CBCT for cervical cancer by diffeomorphic registration of a prior. Med Phys 2022;49:1701–11. https://doi.org/10.1002/mp.15421.

[11] De Hond YJM, Kerckhaert CEM, van Eijnatten MAJM, Van Haaren PMA, Hurkmans CW, Tijssen RHN. Anatomical evaluation of deep-learning synthetic computed tomography images generated from male pelvis cone-beam computed tomography. Phys Imaging Radiat Oncol 2023;25:100416. https://doi.org/10.1016/j.phro.2023.100416.

[12] Pötter R, Tanderup K, Kirisits C, de LA, Kirchheiner K, Nout R, Tan LT, Haie-Meder C, Mahantshetty U, Segedin B, Hoskin P, Bruheim K, Rai B, Huang F, Van LE, Schmid M, Nesvacil N, Sturdza A, Fokdal L, Jensen NBK, Georg D, Assenholt M, Seppenwoolde Y, Nomden C, Fortin I, Chopra S, van der HU, Rumpold T, Lindegaard JC and Jürgenliemk-Schulz I. The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies. Clin. Transl. Radiat. Oncol 2018;9:48-60. https://doi.org/10.1016/j.ctro.2018.01.001.

[13] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203–11. https://doi.org/10.1038/s41592-020-01008-z.

[14] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1947;18:50–60. https://doi.org/10.1214/aoms/1177730491.

[15] Eminowicz G, McCormack M. Variability of clinical target volume delineation for definitive radiotherapy in cervix cancer. Radiother Oncol 2015;117:542–7. https://doi.org/10.1016/j.radonc.2015.10.007.

[16] Brock KK, Mutic S, McNutt TR, Li H and Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. Med. Phys 2017;44:e43-e76. https://doi.org/10.1002/mp.12256.

[17] Lim K, Small Jr W, Portelance L, Creutzberg C, Jürgenliemk-Schulz IM, Mundt A, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. Int J

Radiat Oncol Biol Phys 2011;79(2):348–55. https://doi.org/10.1016/j.ijrobp.2009.10.075.

[18] Haensch A, Dicken V, Gass T, Morgas T, Klein J, Meine H and Hahn H. Deep learning based segmentation of organs of the female pelvis in CBCT scans for adaptive radiotherapy using CT and CBCT data. Proceedings of the 32nd International Congress and Exhibition of Computer Assisted Radiology and Surgery (CARS) 2018;179-80.

[19] Breunig J, Hernandez S, Lin J, Alsager S, Dumstorf C, Price J, et al. A system for continual quality improvement of normal tissue delineation for radiation therapy treatment planning. Int J Radiat Oncol Biol Phys 2012;83:e703–8. https://doi.org/10.1016/j.ijrobp.2012.02.003.