Received: 13 May 2022 | Revised: 15 March 2023 | Accepted: 20 March 2023

DOI: 10.1002/mp.16388

RESEARCH ARTICLE

MEDICAL PHYSICS

Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy

Mark Gardner[1] | **Youssef Ben Bouchta**[1] | **Adam Mylonas**[1] | **Marco Mueller**[1] | **Chen Cheng**[1] | **Phillip Chlap**[2,3,4] | **Robert Finnegan**[3,5,6] | **Jonathan Sykes**[7] | **Paul J Keall**[1] | **Doan Trang Nguyen**[1,8]

[1]ACRF Image X Institute, The University of Sydney, Eveleigh, New South Wales, Australia

[2]South Western Sydney Clinical School, University of New South Wales, Liverpool, New South Wales, Australia

[3]Ingham Institute for Applied Medical Research, Liverpool, New South Wales, Australia

[4]Liverpool Cancer Therapy Centre, Liverpool Hospital, Liverpool, New South Wales, Australia

[5]Northern Sydney Cancer Centre, Royal North Shore Hospital, St Leonards, New South Wales, Australia

[6]Institute of Medical Physics, University of Sydney, Camperdown, New South Wales, Australia

[7]Blacktown Cancer & Haematology Centre, Blacktown Hospital, Sydney, New South Wales, Australia

[8]University of Technology, Sydney, New South Wales, Australia

Correspondence
Mark Gardner, ACRF Image X Institute, The University of Sydney, Eveleigh, New South Wales 2015, Australia.
Email: mark.gardner@sydney.edu.au

Funding information
Cancer Australia, Grant/Award Number: APP1180776

Abstract

Background: Using radiation therapy (RT) to treat head and neck (H&N) cancers requires precise targeting of the tumor to avoid damaging the surrounding healthy organs. Immobilisation masks and planning target volume margins are used to attempt to mitigate patient motion during treatment, however patient motion can still occur. Patient motion during RT can lead to decreased treatment effectiveness and a higher chance of treatment related side effects. Tracking tumor motion would enable motion compensation during RT, leading to more accurate dose delivery.

Purpose: The purpose of this paper is to develop a method to detect and segment the tumor in kV images acquired during RT. Unlike previous tumor segmentation methods for kV images, in this paper, a process for generating realistic and synthetic CT deformations was developed to augment the training data and make the segmentation method robust to patient motion. Detecting the tumor in 2D kV images is a necessary step toward 3D tracking of the tumor position during treatment.

Method: In this paper, a conditional generative adversarial network (cGAN) is presented that can detect and segment the gross tumor volume (GTV) in kV images acquired during H&N RT. Retrospective data from 15 H&N cancer patients obtained from the Cancer Imaging Archive were used to train and test patient-specific cGANs. The training data consisted of digitally reconstructed radiographs (DRRs) generated from each patient's planning CT and contoured GTV. Training data was augmented by using synthetically deformed CTs to generate additional DRRs (in total 39 600 DRRs per patient or 25 200 DRRs for nasopharyngeal patients) containing realistic patient motion. The method for deforming the CTs was a novel deformation method based on simulating head rotation and internal tumor motion. The testing dataset consisted of 1080 DRRs for each patient, obtained by deforming the planning CT and GTV at different magnitudes to the training data.

The accuracy of the generated segmentations was evaluated by measuring the segmentation centroid error, Dice similarity coefficient (DSC) and mean surface distance (MSD). This paper evaluated the hypothesis that when patient motion occurs, using a cGAN to segment the GTV would create a more accurate segmentation than no-tracking segmentations from the original contoured

GTV, the current standard-of-care. This hypothesis was tested using the 1-tailed Mann-Whitney U-test.

Results: The magnitude of our cGAN segmentation centroid error was (mean ± standard deviation) 1.1 ± 0.8 mm and the DSC and MSD values were 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively. Our cGAN segmentation method reduced the segmentation centroid error ($p < 0.001$), and MSD ($p = 0.031$) when compared to the no-tracking segmentation, but did not significantly increase the DSC ($p = 0.294$).

Conclusions: The accuracy of our cGAN segmentation method demonstrates the feasibility of this method for H&N cancer patients during RT. Accurate tumor segmentation of H&N tumors would allow for intrafraction monitoring methods to compensate for tumor motion during treatment, ensuring more accurate dose delivery and enabling better H&N cancer patient outcomes.

KEYWORDS
cGAN, radiotherapy, segmentation

1 | INTRODUCTION

Radiation therapy (RT) is indicated for 74% of head and neck (H&N) cancer patients.[1] H&N RT has a higher risk of adverse side effects than treatments to other sites since there are many important organs located near the planning target volume (PTV). Recent advances in RT, including intensity-modulated RT, which confirms the high dose to the complex shapes of the target volume and minimizes dose to organs at risk (OAR), has led to improved survival rates and reduced toxicities.[2–5] However, treatment-related toxicities still occur and can become a serious health risk if the dose received by organs close to the PTV exceeds certain thresholds.[4–6]

To minimize the dose delivered to healthy organs, H&N cancer patients are required to wear a skin-tight immobilization mask that minimizes patient motion.[7,8] However, despite the restrictive nature of immobilization masks, motion in the order of several millimeters of the tumor and surrounding tissue can still occur during and between treatment fractions.[9,10] This motion can be caused by a change in the mask's fit due to patient weight loss between fractions, imperfections in the mask manufacturing and fitting process, tumor shrinkage, or treatment-related oedema.[7] The current standard of care is to use PTV margins of 2–5 mm rather than motion tracking to account for motion and changes in the target volume.[11] This margin, combined with intrafraction motion, leads to increased dose to surrounding healthy tissue, as well as decreased dose to the target.[12]

Tracking the tumor position during treatment would enable the use of more accurate radiotherapy techniques such as gating or real-time beam adaptation technology and could lead to being able to reduce the PTV margins. One method used in other treatment sites such as prostate, lung and liver to track the tumor position during treatment involves surgically implanting fiducial markers into the tumor which can then be tracked using kV projection images acquired during treatment.[13–15] However, surgically implanting fiducial markers into H&N tumors has a high risk of complication and the large and complex shape of H&N tumors makes implanting markers difficult. If intrafraction tumor monitoring is to be implemented for H&N tumors, a markerless approach to detecting the tumor position in kV images needs to be implemented.

Markerless tumor detection methods have been previously proposed for the lungs,[16–22] liver,[22] prostate,[23,24] and pancreas.[25] As detecting the Gross Tumor Volume (GTV) in kV images is difficult due to the low contrast between the GTV and surrounding tissue, most markerless tracking methods are based on deep learning.[22,25] Template-matching or feature based registration has also been used for markerless tumor tracking, however these methods use high-gradient or high intensity features such as bone as a surrogate for tumor motion.[26,27]

4DCT is commonly used to train deep learning based markerless segmentations methods because the multiple volumes provide large amounts of training data showing how the images will change with motion.[22,23,28] However, 4DCTs are not routinely acquired for H&N RT, so in this paper we present a way of training a deep-learning network to segment H&N tumors from only the planning CT.

The aim of this paper was to develop a realistic data augmentation approach for H&N images which would allow a deep learning method to detect and segment the primary GTV for H&N cancer patients in kV images using only the planning CT as training data. The deep learning network used was a conditional generative adversarial network (cGAN). The effectiveness of our cGAN segmentation method was evaluated by testing the hypothesis that our cGAN segmentation method improves GTV segmentation accuracy when compared to the current standard of care in which no GTV tracking

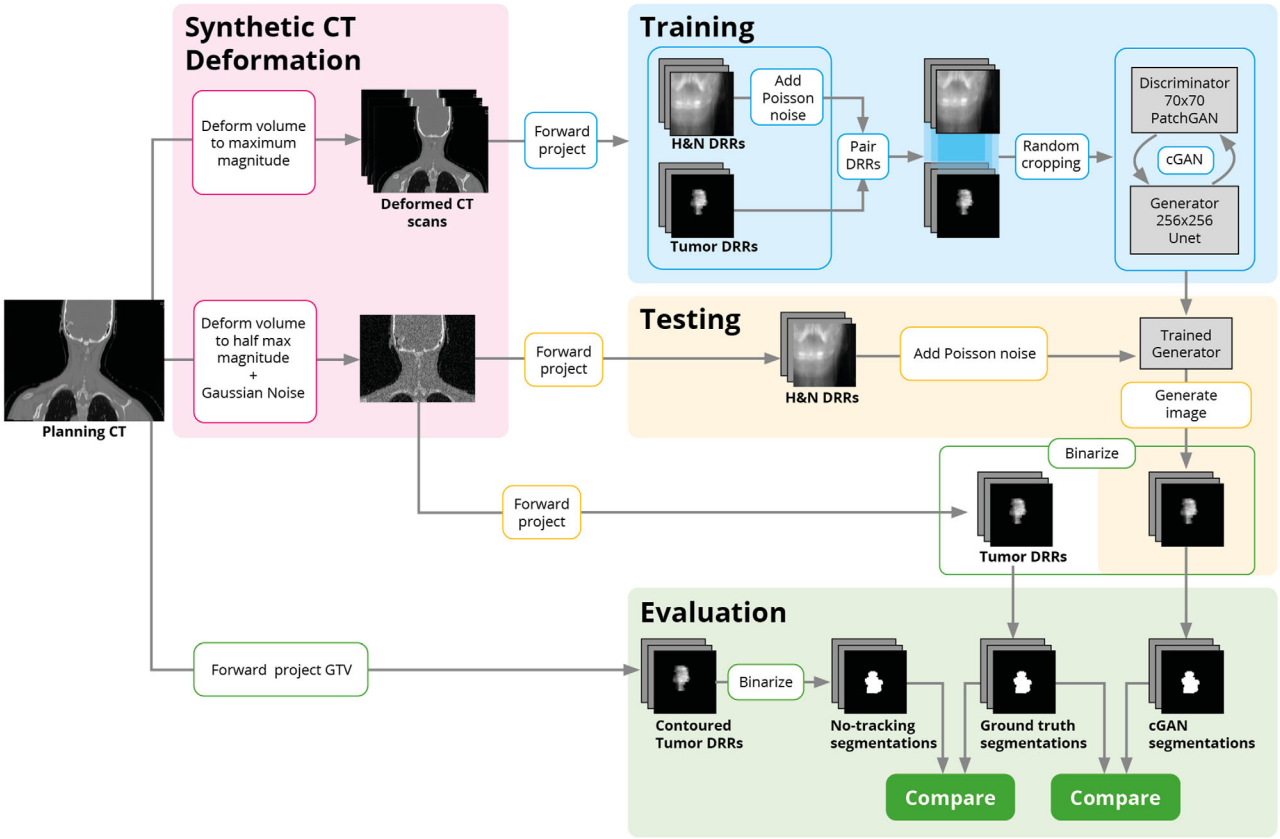


FIGURE 1 A flowchart showing the method used to evaluate the patient-specific deep learning method for one patient. First, the planning CT and contoured GTV is deformed multiple times and forward projected to create the training data. The same planning CT is deformed at half the magnitude of the training data to create the testing data. The original contoured GTV is also forward projected to create the no-tracking segmentations which assume no motion occurred in the testing data. DRR, digitally reconstructed radiograph; GTV, gross tumor volume.

is used. The data augmentation simulated realistic patient movement, which was achieved using a novel synthetic deformation method. To the authors knowledge, this paper describes the first implementation of markerless tumor detection of H&N tumors in kV images. This feasibility study is an important step in demonstrating the feasibility of markerless tracking of H&N tumors, and is an important step toward reducing the need for immobilization equipment during RT.

2 | METHODS

A flowchart detailing the training, testing and evaluation of our GTV tracking method is shown in Figure 1. The code used in this paper to generate these results is available on a public repository (https://github.com/ACRF-Image-X-Institute/MarkerlessHNGTVTracking). Most deep neural network-based markerless tumor tracking methods uses the large amount of data acquired in 4DCTs to train their neural network.[22,23] As 4DCTs are not routinely used for H&N treatments, our training dataset was generated from the planning CT by using a novel synthetic CT deformation method to deform each patient's planning CT to generate multiple CT volumes. From these multiple CT volumes, synthetic images in the form of digitally reconstructed radiographs (DRRs) were created and used to train a patient-specific cGAN to segment the GTV in the DRRs. To create the testing dataset, the planning CT volumes were again deformed by creating an additional realistic synthetic deformation. This additional deformation had different magnitudes to the deformations used to create the training data. The resultant deformed CT was then used to create a set of testing DRRs.

The cGANs used in this paper were trained using DRRs, which are simulated 2D fluoroscopy x-ray images created from a 3D CT volume.[2] Using a known projection geometry, DRRs can be created at different projection angles to simulate kV images acquired during RT. There are known differences between the noise properties and the image quality of kV images and DRRs,[29-31] however using DRRs to train the patient-specific cGANs allows for the networks to be trained without needing any additional images to be acquired. The use of DRRs for testing allows for the exact location of the ground truth GTV segmentations to be known in each projection and is a useful first step in evaluating the feasibility of our cGAN segmentation method described in this paper.

2.1 | Database and patient selection

The data for this paper involved 15 patients with head and neck squamous cell carcinoma (HNSCC) and were acquired from the HNSCC database[32-34] on The Cancer Imaging Archives (TCIA).[35] For each patient, the original data consisted of a planning CT and corresponding structure file, which contained the contoured primary GTV. The 15 patients were sequentially selected based on tumor location to ensure a range of primary tumor locations and to investigate the feasibility of the method described in this paper. The locations of the tumor were the oropharynx ($n = 5$), the larynx ($n = 5$), and the nasopharynx ($n = 5$). Patients with different tumor locations in the head and neck were selected to test the robustness of our patient-specific segmentation method.

2.2 | Synthetic CT deformations

Previous implementations of deep learning networks for markerless tumor detection and segmentation in kV images trained the network using data from 4DCT scans.[16,17,22,25] The advantage of using the 4DCT for training is that the network is trained on images showing how the tumor and surrounding tissue move and deform. H&N cancer treatment planning is typically done on a regular CT scan, which presents a challenge for training a patient-specific segmentation model because regular CT scans contain a single volume whereas 4DCTs contain multiple volumes. This reduces the training dataset and results in the network being less effective at detecting and segmenting the tumor when motion occurs.

To compensate for the lack of motion data, we developed a CT deformation-based data augmentation method that can be used to generate synthetic images showing realistic head motion. This data augmentation method allowed each patient's cGAN to be trained on a patient-specific dataset containing images of H&N motion without requiring additional CT scans. The assumption was made that two types of movements would be the primary sources of tumor motion during RT treatment: head rotation and internal tumor motion. A brief summary of the method used to create these synthetic deformations is included below, with a more detailed discussion provided in the Supplementary Materials.

2.2.1 | Head rotation

Six different training volumes representing the different types of head rotation were created: axial rotation (left-right rotation around the superior-inferior (SI) axis rotation), lateral bending (left-right rotation around the anterior-posterior (AP) axis), and flexion/extension (back-front rotation around the left-right (LR) axis). The head rotation deformation vector fields (DVFs) were created by rotating the volume around anatomical landmarks to replicate real head motion. A summary of the anatomical landmarks used as centres of rotation and realistic upper limits of the magnitude of patient head rotations during treatment is described in the Supplementary Materials. A head mask was created so the deformation would only be applied to the head and upper neck, with the inferior boundary on the neck being defined manually, as approximately halfway between the superior boundary of the shoulders, and the junction between the occipital bone in the neck and the top vertebrae.

2.2.2 | Tumor motion

To simulate tumor motion, the contoured GTV volume was rigidly shifted while the surrounding bones were kept stationary. The DVF used to shift the GTV had a Gaussian smoothing filter applied with a filter width $\sigma = 2\ mm$. The use of the smoothing filter created a more realistic deformation by ensuring that there was a gradual and continuous change in the DVF magnitude from the outside the boundary of the GTV to the centre of the GTV. Although σ is a heuristic value, the value chosen for this paper was based on similar instances of smoothing organ boundaries in CT scans.[36-38] The magnitudes of the applied tumor shifts were based on realistically large magnitudes of motion of the GTV during treatment and depended on the tumor location, as tumor motion during treatment is different for oropharyngeal, laryngeal, and nasopharyngeal tumors. A discussion on realistic magnitudes of tumor shifts during treatment is provided in the Supplementary Materials. Since minimal tumor motion has been observed in the LR direction the magnitude of these shifts was set to 0.[11,39] Similarly, since nasopharyngeal tumors have minimal tumor motion independent of the surrounding bony anatomy, the magnitude of the GTV shifts in all directions for nasopharyngeal tumors was set to 0. In total four volumes were created to simulate the tumor motion: an anterior and posterior GTV shift, and a superior and an inferior GTV shift.

2.2.3 | DVF creation

A flowchart showing how the final volumes were created is shown in Figure 2. For each deformed training volume, the head rotation and internal tumor motion DVFs were calculated separately and then combined to create the initial DVF estimation. For combined head rotation

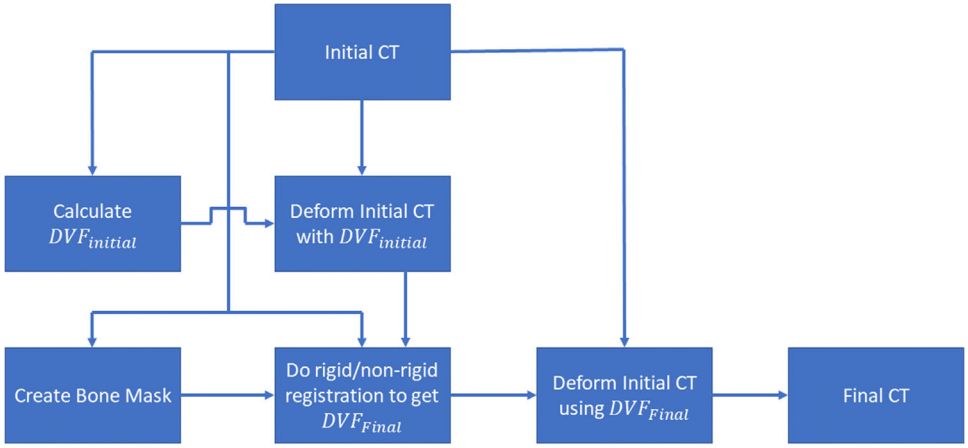


FIGURE 2 Flowchart for deforming the initial CT volume to simulate realistic motion.

motion (e.g., combination head flexion and internal tumor motion) the individual head rotation DVFs were similarly calculated for each type of motion separately (e.g., head flexion deformation calculated separately to internal tumor motion) and then combined to create the initial head motion DVF. The initial DVFs of the head rotation and the internal tumor motion were created using the platipy library for python.[40]

To create realistic DVFs that reduce boundary artefacts, a combined rigid/non-rigid registration method was used. The rigid/non-rigid registration method is a non-rigid registration method developed by Staring et al. which attempts to penalise non-rigid motion in user defined regions and is described in more detail in the Supplementary Material.[41] To implement this registration algorithm, we first applied the initial DVF generated by platipy to the planning CT to obtain an approximately deformed CT. We then created a bone mask from the planning CT and used it to define a region in which non-rigid motion was penalised. Finally, we obtained the final DVF by registering the original planning CT to the approximately deformed CT using the combined rigid/non-rigid registration method. The result was a more realistic transformation in which bones move rigidly and soft tissue non-rigidly. The registration process was completed using the elastix toolbox.[42] This synthetic deformation method was converted into a generalized software package for deforming the head and neck in CT scans and is available in a public repository (https://github.com/ACRF-Image-X-Institute/CTHeadDeformation).

2.2.4 | Synthetic deformations for creating training data

A table summarizing the different training volumes created (and the deformation magnitudes used to create those volumes) for each patient is shown in Table 1. For patients with oropharyngeal and laryngeal tumors, the training data for each patient was derived from all 11 volumes (as annotated in Table 1). However, for patients with nasopharyngeal tumors since no GTV shifts were added, only volumes 1–7 (as annotated in Table 1) were generated for each patient and used to produce the training data. To robustly demonstrate the effectiveness of the GTV segmentation method described in this paper, the same deformations cannot be used to create the training volume as well as the testing volume, but it is also unrealistic to have the motion of the testing volume larger than the training data. Hence, the head rotation and GTV shift magnitudes used for training (shown in Table 1) were twice as large as realistic upper limits of head and GTV motion (as described in the Supplementary Material), so the testing data could be created using realistic head and GTV deformations.

2.3 | cGAN segmentation method

To train a cGAN, two convolutional neural networks are trained simultaneously. A generator network *G* takes an input image *x* and creates a segmentation image *G*(*x*) based on the training data. The discriminator network *D* classifies whether the paired image *xy* came from the training set or the generator network, as shown in Figure 3.[43]

The cGAN was chosen for segmentation as the use of the discriminator network in the cGAN allows for the generation of unique and adaptive loss metrics.[44,45] The DRR images (and in future patient kV images) vary considerably for different patients depending on the tumor location, and even between patients depending on the projection angle. The ability of the cGAN to "learn" the loss function will ideally allow for the generation of patient-specific networks that can perform

TABLE 1 A summary of the training volumes created for each patient. The magnitude of the tumor motion depends on the tumor location and is shown in shown in more detail in the supplementary materials. No additional motion means that any tumor displacement resulting from the applied deformation is only as a result of the head moving, and not from any additional internal organ motion.

Volume number	Head rotation	Tumor motion	Motion magnitude (deg/mm)
1	None	No additional motion	0
2	Head rotation left	No additional motion	5.0°
3	Head rotation right	No additional motion	5.0°
4	Head tilt left	No additional motion	6.4°
5	Head tilt right	No additional motion	6.4°
6	Head nod up	No additional motion	6.2°
7	Head nod down	No additional motion	6.2°
8	None	Superior GTV shift	Oropharynx: 4 mm Larynx: 7.6 mm Nasopharynx: 0 mm
9	None	Inferior GTV shift	Oropharynx: 4 mm Larynx: 7.6 mm Nasopharynx: 0 mm
10	None	Anterior GTV shift	Oropharynx: 3.4 mm Larynx: 4.4 mm Nasopharynx: 0 mm
11	None	Posterior GTV shift	Oropharynx: 3.4 mm Larynx: 4.4 mm Nasopharynx: 0 mm

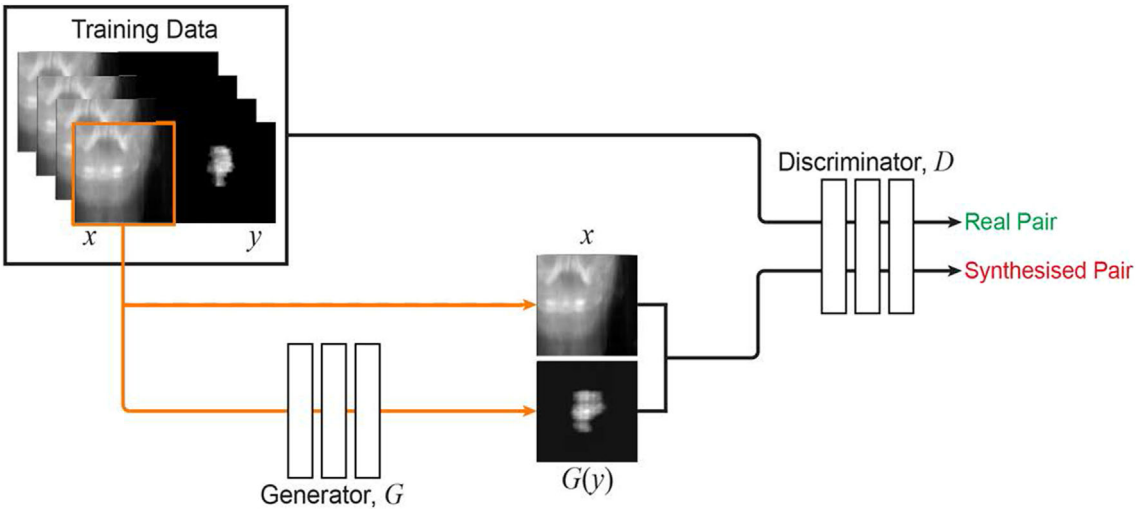


FIGURE 3 During training, the generator G creates image $G(x)$ from the image x that was passed to the discriminator D which aims to reject synthetic images and accept real segmentations y. This method efficiently trains the generator to produce accurate tumor segmentations from H&N DRRs.

accurate tumor segmentation of the unique and variable DRR images. In this paper, the aim of the cGAN was to allow for segmentation of the GTV in DRRs. However, to make the network optimization process more efficient, the cGAN was trained to generate an image showing a tumor DRR from a regular H&N DRR, as shown in Figure 3.[24] The output image generated by the cGAN was then binarized to create the GTV segmentation.

Our cGAN implementation used was based on the pix2pix method which provides a general solution to train a network to perform image-to-image translation (https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix).[44] The cGAN was initialized with random parameters and trained to optimize the loss function

$$G^* = arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (1)$$

where λ is a constant (set to 100 for this implementation) and for all expected values of x (E_x):

$$\mathcal{L}_{\text{cGAN}}(G, D) = E_{x,y}[\log D(x, y)] + E_x[\log(1 - D(x, G(x)))] \quad (2)$$

$$\mathcal{L}_{L1} = E_{x,y}\|y - G(x)\|_1 \quad (3)$$

A custom data loader was used to allow 16-bit grayscale images to be used as input. A 70 × 70 PatchGAN[44,46] was used for the discriminator architecture and a 256 × 256 UNet[44,47] for the generator architecture (architecture shown in Supplementary Material). The training images of 550 × 550 pixels in size with a pixel width of 0.388 mm were randomly cropped to a size of 512 × 512 pixels during network training, and the testing images had a size of 512 × 512 pixels. Both training and testing images had random Poisson noise ($100 \leq \lambda \leq 200$) added to mimic the noise patterns of a kV image.[31]

Each patient-specific model was trained with a batch size of 8 and a learning rate of 0.0002 using the Adam optimiser. Each network was trained until the loss functions reached an equilibrium.[48,49] A plot of the loss functions for one patient is shown in the Supplementary Material. The models were trained on a desktop computer with two Intel Xeon Gold 6130 processors (2.1 GHz) with 64 GB RAM and a NIVIDIA Quadro P6000 Graphics Processing Unit (GPU). A patient-specific network was trained for each patient as training an effective patient-specific network requires less data than a similar generalized network and will be applicable to patients imaged using different imaging systems.

2.3.1 | Training data

The training data consisted of sets of paired images: a training H&N DRR and a corresponding training tumor DRR obtained from the 3D contour of the primary GTV. The training H&N and tumor DRRs were created by forward x-ray projecting each CT and GTV volume respectively using the RTK library.[50] For each CT volume, 3600 pairs of DRRs were generated evenly in a 360° arc around the volume. For each oropharyngeal and laryngeal patient, the training data were derived from 11 CT volumes (the original planning CT and 10 deformed CTs) and for nasopharyngeal patients seven CT volumes (the original planning CT and six deformed CTs). In total, the training data consisted of 39 600 paired DRRs for patients with oropharyngeal and laryngeal tumors and 25 200 paired DRRs for patients with nasopharyngeal tumors.

2.3.2 | Testing data

To generate the testing data for each patient, the planning CT and contoured GTV were deformed again three separate times, using the synthetic deformation method used to create the training data, but with different magnitudes of head rotations and GTV shifts. The magnitudes for the head rotations and GTV shifts used to create the three testing CT volumes are shown in Table 2. Additionally, the testing CT volumes had Gaussian noise ($\sigma = 10^{-5}\text{mm}^{-1}$) added to further differentiate the testing volume from the training CT volumes.[51] Each testing CT volume and GTV volume were used to create the testing dataset consisting of 360 paired DRRs evenly spaced around a 360° range leading to a total of 1080 testing DRRs for each patient.

For each testing H&N DRR created, the trained generator network produced an image $G(x)$ which was an estimation of the testing tumor DRR. To obtain segmentations from the predicted image $G(x)$ and the ground truth testing tumor DRR, both sets of images were normalized and then binarized using a threshold of 0.1. These binarized segmentations were then compared to evaluate the accuracy of our cGAN segmentation method.

2.3.3 | No-tracking segmentations

The original GTV contour from the planning CT was forward projected to create an additional set of testing segmentations (referred to as no-tracking segmentations), which assumed no motion occurs and hence does not track the GTV after the testing DVF is applied (as shown in Figure 1). The size of the no-tracking dataset was the same as the testing dataset. The no-tracking segmentations were used to test the hypothesis that the cGAN segmentation method improves the GTV segmentation accuracy when compared with the current standard of care (no GTV tracking). Hence the hypothesis is true if the accuracy of the cGAN segmentations is significantly greater than the accuracy of the no-tracking segmentations.

2.4 | Accuracy evaluation

To compare the accuracy of each segmentation, three metrics were used: the centroid error, Dice Similarity Coefficient (DSC), and mean surface distance (MSD). In this paper, we hypothesised that our cGAN segmentation method would improve the accuracy of GTV segmentations when compared with the no-tracking segmentations. The mean of each metric was calculated for every patient for both our cGAN segmentation method and the no-tracking segmentations. The

TABLE 2 The magnitude and direction of the applied head rotations and tumor shifts used to deform the planning CT and contoured GTV to create the testing data.

Testing volume number	Head rotation	Rotation magnitude	Tumor shift direction	Tumor shift magnitude
1 (25% of training magnitude)	Axial Rotation—left	1.25°	Superior	Oropharynx—1.0 mm Larynx—1.9 mm Nasopharynx—0 mm
	Lateral Bending—left	1.6°	Anterior	Oropharynx—0.85 mm Larynx—1.1 mm Nasopharynx—0 mm
	Head extension	1.55°	L/R	0 mm
2 (50% of training magnitude)	Axial Rotation—left	2.5°	Superior	Oropharynx—2.0 mm Larynx—3.8 mm Nasopharynx—0 mm
	Lateral Bending—left	3.2°	Anterior	Oropharynx—1.7 mm Larynx—2.2 mm Nasopharynx—0 mm
	Head extension	3.1°	L/R	0 mm
3 (75% of training magnitude)	Axial Rotation—left	3.75°	Superior	Oropharynx—3.0 mm Larynx—5.7 mm Nasopharynx—0 mm
	Lateral Bending—left	4.8°	Anterior	Oropharynx—2.55 mm Larynx—3.3 mm Nasopharynx—0 mm
	Head extension	4.65°	L/R	0 mm

averaged metric values were used to test the hypotheses using a paired 1-tail Mann-Whitney U test with a significance level of $\alpha = 0.05$.

The segmentation centroid error was used to measure the localization accuracy of our cGAN segmentation when compared to the ground truth segmentation. The centroid error was calculated in the kV imaging frame of reference where u is in the lateral direction and v is in the SI direction. The combined centroid error in the u and v directions was also calculated and is reported as the absolute centroid error magnitude.

The DSC was used to estimate how similar the shape of our cGAN and no-tracking segmentations were to the ground truth segmentations.[52] Similarly the MSD was also calculated, which measured the average distance (in mm) between the outline of the predicted and ground truth segmentations.[53] The MSD decreases as the accuracy of the segmentation increases.

3 | RESULTS

The centroid error, DSC and MSD of our cGAN segmentations for all testing data for all patients are shown in Table 3. Analysis of the change in segmentation accuracy due to the different magnitudes of head rotations and GTV shifts is shown in the Supplementary Material. For all patients the mean ± standard deviation cGAN segmentation centroid error in the u (lateral) and the v (vertical) directions was −0.0 ± 1.0 mm and 0.1 ± 0.9 mm, respectively. Additionally, for all patients, the 95th percentile centroid error in the u and v directions were [−1.4, 1.3] mm and [−1.4, 1.6] mm, respectively. The absolute centroid error of our cGAN segmentation 1.1 ± 0.8 mm was for all patients, with the 95th percentile error of [0.3, 2.3] mm. This is shown in comparison to the centroid error of the no-tracking segmentations in Figure 4. Using a paired 1-tail Mann-Whitney U test, our cGAN segmentation method significantly reduced the absolute segmentation centroid error when compared to the no-tracking segmentations ($p < 0.001$). Each patient-specific cGAN network took an average of 3 h to train.

For all patients the mean ± standard deviation DSC and MSD values for our cGAN segmentation were 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively, with the 95th percentile error for the DSC and MSD being [0.85, 0.94] and [0.9, 2.5] mm, respectively. The distribution of both the DSC and MSD values for our cGAN segmentation method are shown in comparison to the no-tracking segmentations in Figure 4. Using a paired 1-tail Mann-Whitney U test, our cGAN segmentation significantly reduced the MSD ($p = 0.031$) when compared to the no-tracking segmentations. The cGAN segmentation method did not significantly improve the DSC when compared to the no-tracking method ($p = 0.294221$) and only significantly improved the DSC for the Oropharynx tumors ($p < 0.0001$) but not for the Larynx ($p = 0.203$) and reduced the DSC for the Nasopharyx ($p < 0.0001$).

TABLE 3 The centroid error, DSC and MSD values for the predicted cGAN segmentations. All values are mean ± standard deviation. The best result for each tumor location group for each metric is highlighted by*.

Patient number	Centroid error—*v* (mm)	Centroid error—*u* (mm)	Absolute centroid error magnitude (mm)	DSC	MSD (mm)
Oropharynx					
1	0.7 ± 0.7	0.1 ± 0.9	1.2 ± 0.5	0.90 ± 0.01	1.5 ± 0.2
2	−0.3 ± 0.6*	0.2 ± 0.6	0.9 ± 0.4*	0.92 ± 0.01	1.1 ± 0.2*
3	−0.8 ± 0.8	−0.5 ± 0.7	1.2 ± 0.5	0.92 ± 0.01	1.6 ± 0.2
4	0.7 ± 0.5	−0.2 ± 0.7	1.0 ± 0.5	0.89 ± 0.03	1.6 ± 0.4
5	0.7 ± 0.7	−0.1 ± 0.7*	1.0 ± 0.5	0.92 ± 0.01	1.9 ± 0.3
All oropharynx	0.2 ± 0.9	−0.1 ± 0.7	1.1 ± 0.5	0.91 ± 0.02	1.5 ± 0.4
Larynx					
6	−0.6 ± 0.5	0.1 ± 0.8	1.0 ± 0.5	0.90 ± 0.02	1.3 ± 0.3*
7	0.7 ± 0.6	0.2 ± 0.6	1.0 ± 0.5	0.87 ± 0.02	1.6 ± 0.3
8	−0.2 ± 0.9*	0.0 ± 0.6*	1.0 ± 0.6	0.92 ± 0.02*	1.3 ± 0.3*
9	−0.7 ± 1.0	0.0 ± 0.6*	1.2 ± 0.6	0.86 ± 0.03	1.5 ± 0.3
10	−0.3 ± 0.4	0.1 ± 0.6	0.7 ± 0.4*	0.91 ± 0.02	1.3 ± 0.3*
All larynx	−0.2 ± 0.9	0.1 ± 0.7	1.0 ± 0.5	0.89 ± 0.03	1.4 ± 0.3
Nasopharynx					
11	0.9 ± 0.8	0.2 ± 1.5	1.7 ± 0.9	0.91 ± 0.02	2.3 ± 0.5
12	−0.3 ± 0.5*	0.0 ± 0.9	0.9 ± 0.5	0.90 ± 0.02	2.1 ± 0.4
13	0.9 ± 1.0	−0.2 ± 1.1	1.4 ± 1.0	0.92 ± 0.02	2.0 ± 0.6
14	−0.4 ± 0.5	0.0 ± 0.5*	0.7 ± 0.4*	0.94 ± 0.01*	1.4 ± 0.3*
15	0.1 ± 0.6	−0.3 ± 2.3	1.3 ± 2.0	0.89 ± 0.06	1.4 ± 0.3
All nasopharynx	0.3 ± 0.9	−0.1 ± 1.4	1.2 ± 1.2	0.91 ± 0.04	1.8 ± 0.7
Total					
All H&N patients	0.1 ± 0.9	−0.0 ± 1.0	1.1 ± 0.8	0.90 ± 0.03	1.6 ± 0.5

An example of our cGAN segmentation for several DRRs at different projection angles is shown in Figure 5 and video showing our cGAN segmentations for all projection angles is shown in Video S1. Figure 5 demonstrates an example of the typical accuracy of the cGAN segmentations achieved by this method for one H&N cancer patient.

4 | DISCUSSION

In this paper, we investigated the feasibility of a realistic data augmentation methods to assist in markerless tumor segmentation method for detecting and segmenting the primary GTV in DRRs for H&N cancer patients. We have demonstrated that we can use a synthetic CT deformation method to augment the training data by realistically deforming the planning CT. The results demonstrated that our cGAN segmentation method increased the accuracy of the GTV segmentation compared to the current standard of care. Our cGAN segmentation method detected the centroid of the GTV in DRRs with an accuracy of 1.1 ± 0.8 mm and segmented tumors in DRRs with DSC and MSD value of 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively.

Table 3 shows the magnitude of the mean centroid error for all patients was approximately 1 mm. This suggests that our cGAN segmentation method can accurately localize the GTV in DRR images. Accurate detection of the GTV is a crucial step in the implementation of image-guided RT methods.[54] Similarly, the mean DSC and MSD values of 0.90 and 1.6 mm, respectively suggest that our cGAN segmentation method can accurately segment the GTV in DRR images. The ability to adapt to the changing shape of the GTV would enable real-time adaptive RT that can not only modify the treatment beam location but also adapt to changes in the shape of the GTV volume. Future work will look to evaluate whether our cGAN segmentation method can be used to implement markerless image-guided RT or adaptive RT for H&N patients.

Most markerless tracking results described in other literature present the 3D localization error instead of the error in each 2D image, making direct accuracy

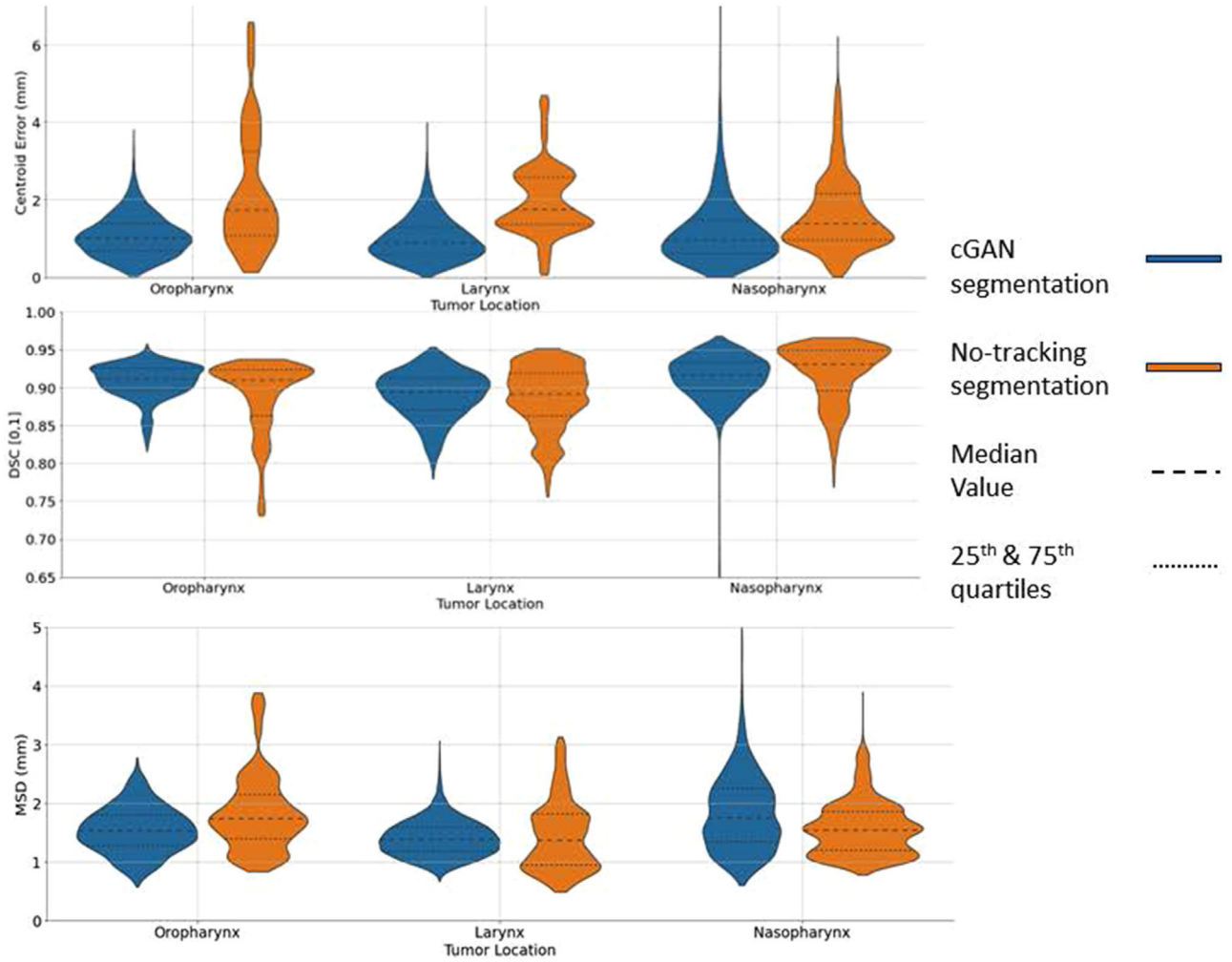
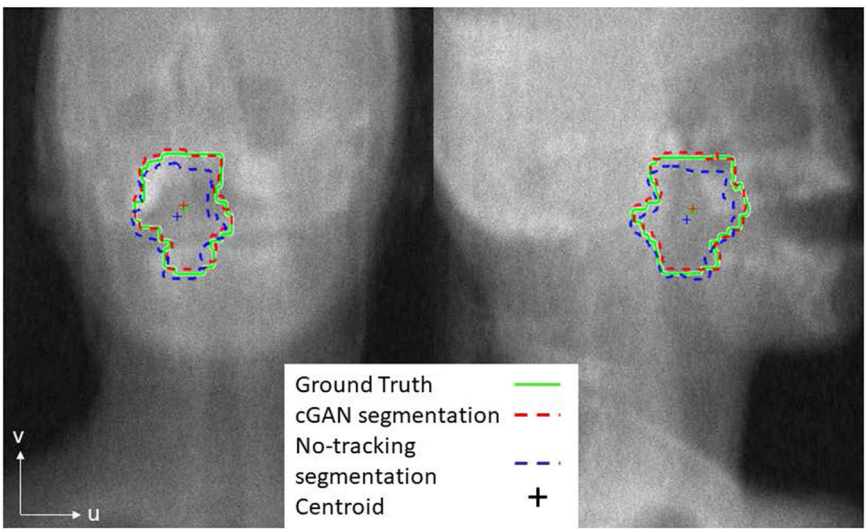


FIGURE 4 Violin plots showing the distribution of the accuracy metrics for our cGAN segmentation (blue) compared with the no-tracking segmentations (orange) for the different tumor locations. The metrics shown are the magnitude of the absolute centroid error (top), the Dice Similarity Coefficient (middle) and mean surface distance (bottom). The width of the violin plot at each *y* value corresponds to the frequency of that value.

FIGURE 5 Example of our cGAN segmentation (red) and the no-tracking segmentation (blue) in comparison to the ground truth segmentation (green) for different projection angles.



comparisons difficult. Zhao et al. developed a deep learning method for identifying a rectangular bounding box containing the prostate location in DRRs and acquired kV projections at three specific angles (0°, 90°, 135°).[55] They were able to estimate the tumor position in the generated DRRs with a mean accuracy of 1.58-1.67 mm, however this accuracy was only demonstrated for those three projection angles. A similar method was applied for pancreatic tumors with a mean absolute distance between the predicted and actual tumor centroid of less than 2.6 mm.[25] A deep-learning method for generating GTV contours in DRRs was applied for lung tumors, which resulted in DSC values of 0.81–0.98. These results show that our cGAN segmentation method has similar levels of accuracy to other deep learning based markerless GTV detection methods.

The time taken for the trained network to generate an estimate of the GTV segmentation was approximately 0.01 s per image. This demonstrates that this method has a small latency, which is a feature of deep learning methods as has been noted previously.[23] This short latency, as well as a mean absolute accuracy of 1.1 mm suggest that real-time tracking of H&N tumors, which requires a system latency of less than 0.5 s, is feasible.[56] The training of the cGAN prioritized accuracy over training time to prove that cGANs can be used to accurately segment DRR images. Future work can optimize the training process to further reduce the training time, by experimenting with the number of training images, as well as the training parameters such as step length.

The images that were used to train and test our cGAN segmentation method in this paper were DRR images. This is a limitation of the HNSCC dataset that was used, as this dataset has only one CT per patient, with no available kV images. However, the use of DRRs instead of real kV images in this paper ensures an accurate ground truth location for each test image, enabling an accurate quantification of the GTV segmentation effectiveness. In a clinical environment, our cGAN segmentation method would detect and segment the GTV in kV images acquired during RT fractions. However, intrafraction motion during treatment would make it difficult to establish an accurate GTV location in any kV images acquired during treatment.[9,10] We view this study as a necessary but not sufficient study toward the long-term goal of clinical implementation. Future work will focus on clinically acquired data for which there is no direct ground truth.

DRRs have a greater quality than kV images acquired during treatment. Even though Poisson noise was added to make the DRRs used in this study more similar to kV images, Poisson noise does not model all additional sources of artefacts and noise.[31] The difference in image quality between DRR and kV images means that a cGAN network trained on DRRs may decrease in accuracy when applied to kV images. In the future, the methodology presented in this paper will be used to evaluate our cGAN segmentation method on kV images[23,24] or by adding more realistic noise to the DRRs to attempt to replicate kV images.[29,30]

Similarly, the deformation method used to create the testing data approximated patient motion based on available data on H&N motion. The rigid/non-rigid registration method used to generate DVFs attempts to keep the bones rigid while all other tissue can be non-rigidly deformed. Adil Al-Mayah et al. used a biomechanical based finite element analysis to simulate deformation of the upper vertebrae and H&N,[57] which could be investigated in the future to potentially refine the generated deformations and allow for realistic deformation of all tissue. Pukala et al. developed a method for realistically deforming a cone-beam CT image to match a second cone-beam CT image of the same patient from a different treatment fraction.[58,59] However, this method only allows for the transformation from one scan to another scan, and cannot be used to make additional, realistic deformations. Hence there is scope to further optimise the deformation algorithms to generate more realistic synthetic deformations.

During RT for H&N cancer, the PTV often includes nearby lymph nodes as targeting these nodes this can prevent cancer recurrence.[60,61] While this work focused on tracking the primary GTV in DRRs, our cGAN segmentation method could be easily adapted to segment nearby organs of interest which are also likely to be affected by patient motion. This is demonstrated in this paper by the accurate detection and segmentation results for patients with tumors of different shapes, sizes, and locations. Future work could investigate multi-target tracking methods to allow for the tracking of the primary GTV as well as surrounding organs for either targeting or avoidance.

Our cGAN segmentation method described in this paper was demonstrated using DRRs. As radiographs are a planar imaging modality, our method can only estimate motion in the same plane as the DRR. In marker-based tumor tracking applications, methods have been developed for using the marker locations in kV images to estimate the 3D position and rotation of the tumor.[62,63] It is thought that this 3D position estimation method can be adapted such that markerless 3D position and motion estimation is possible, either by using the location of the segmentation centroid or points on the segmentation outline instead of the marker locations.[64] An accurate estimation of the 3D position and motion of the tumor would allow for improved dose delivery through the use of beam adaptation methods[65] or gated treatments.[63]

Currently for H&N cancer RT treatments, kV images are mainly acquired prior to treatment to assist in accurately positioning the patient.[7] For real-time tumor tracking of H&N tumors, additional kV images would have to be acquired during treatment at regular intervals.[14] The acquisition of these images would result in extra imaging dose to the patient. Future development of this

method should quantify the increase in dose caused by the acquisition of these additional kV images and evaluate the risk associated with this additional dose.

The immobilization masks that are used to restrict motion of H&N patients during RT cause anxiety and distress in a significant number of patients.[66,67] If mask-free RT can be achieved without a reduction in the accuracy of the dose delivery, then this would lead to better H&N cancer patient experience. While our cGAN segmentation method described in this paper has been developed for patients wearing the mask during treatment, future work will look to extend the markerless tumor detection framework described in this paper to allow for H&N cancer radiation therapy treatments without the need for an immobilization mask.

5 | CONCLUSION

In conclusion we have demonstrated the feasibility of our cGAN segmentation method in detecting and segmenting the primary GTV location in DRRs of H&N cancer patients. The centroid accuracy of our cGAN segmentation was -0.0 ± 1.0 mm in the u (Lateral) direction and 0.1 ± 0.9 mm and v (SI) directions. The DSC and MSD of our cGAN template segmentation was 0.90 ± 0.03 and 1.6 ± 0.5 mm, respectively when compared with the ground truth segmentation. When comparing the centroid error and MSD, the accuracy of our cGAN segmentation method was significantly greater than the no-tracking segmentations which assumed no motion. Accurate segmentation of the GTV can enable to tracking of the tumor during H&N RT treatment, leading to better dose delivery and better patient outcomes.

ACKNOWLEDGMENTS

The authors would like to thank Dr Helen Ball for her assistance with statistics, and Julia Johnson for her assistance with the figures. Finally, the authors would like to thank Dr Paul Liu and Dr Emily Hewson for helping to review the manuscript prior to submission.

CONFLICT OF INTEREST STATEMENT

Author Paul J Keall is an inventor on a patent application. A full list of disclosures is given at: https://imagex.sydney.edu.au/home/disclosures/.
Authors Adam Mylonas, Marco Mueller, and Doan Trang Nguyen are inventors on a patent application for deep learning-based markerless prostate segmentation.

REFERENCES

1. Barton MB, Jacob S, Shafiq J, et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiother Oncol*. 2014;112(1):140-144. doi:10.1016/j.radonc.2014.03.024
2. Lee AWM, Ng WT, Chan LLK, et al. Evolution of treatment for nasopharyngeal cancer - success and setback in the intensity-modulated radiotherapy era. *Radiother Oncol*. 2014;110(3):377-384. doi:10.1016/j.radonc.2014.02.003
3. Setton J, Caria N, Romanyshyn J, et al. Intensity-modulated radiotherapy in the treatment of oropharyngeal cancer: an update of the memorial sloan-kettering cancer center experience. *Int J Radiat Oncol Biol Phys*. 2012;82(1):291-298. doi:10.1016/j.ijrobp.2010.10.041
4. Nutting C, A'Hern R, Rogers MS, et al. First results of a phase III multicenter randomized controlled trial of intensity modulated (IMRT) versus conventional radiotherapy (RT) in head and neck cancer (PARSPORT: ISRCTN48243537; CRUK/03/005). *J Clin Oncol*. 2009;27(18 suppl):LBA6006-LBA6006. doi:10.1200/jco.2009.27.18_suppl.lba6006
5. Fang FM, Chien CY, Tsai WL, et al. Quality of life and survival outcome for patients with nasopharyngeal carcinoma receiving three-dimensional conformal radiotherapy vs. intensity-modulated radiotherapy-a longitudinal study. *Int J Radiat Oncol Biol Phys*. 2008;72(2):356-364. doi:10.1016/j.ijrobp.2007.12.054
6. Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2001;50(3):695-704. doi:10.1016/S0360-3016(01)01512-7
7. Leech M, Coffey M, Mast M, et al. ESTRO ACROP guidelines for positioning, immobilisation and position verification of head and neck patients for radiation therapists. *Tech Innov Patient Support Radiat Oncol*. 2017;1:1-7. doi:10.1016/j.tipsro.2016.12.001
8. Marsh R, Balter J, Evans VL, Eisbruch A. Design and analysis of an immobilization and repositioning system for treatment of neck malignancies. *Med Dosim*. 1997;22(4):293-297. doi:10.1016/S0958-3947(97)00102-7
9. Amelio D, Winter M, Habermehl D, Jäkel O, Debus J, Combs SE. Analysis of inter- and intrafraction accuracy of a commercial thermoplastic mask system used for image-guided particle radiation therapy. *J Radiat Res*. 2013;54(1):i69-i76. doi:10.1093/jrr/rrt038
10. Juan-Senabre XJ, López-Tarjuelo J, Conde-Moreno A, et al. Uncertainties and CTV to PTV margins quantitative assessment using cone-beam CT technique in clinical application for prostate, and head and neck irradiation tumours. *Clin Transl Oncol*. 2011;13(11):819-825. doi:10.1007/s12094-011-0740-8
11. Bruijnen T, Stemkens B, Terhaard CHJ, Lagendijk JJW, Raaijmakers CPJ, Tijssen RHN. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. *Radiother Oncol*. 2019;130:82-88. doi:10.1016/j.radonc.2018.09.015
12. Neubauer E, Dong L, Followill DS, et al. Assessment of shoulder position variation and its impact on IMRT and VMAT doses for head and neck cancer. *Radiat Oncol*. 2012;7(1):19. doi:10.1186/1748-717X-7-19
13. Keall PJ, Ng JA, Juneja P, et al. Real-time 3D image guidance using a standard LINAC: measured motion, accuracy, and precision of the first prospective clinical trial of kilovoltage intrafraction monitoring-guided gating for prostate cancer radiation therapy. *Int J Radiat Oncol Biol Phys*. 2016;94(5):1015-1021. doi:10.1016/j.ijrobp.2015.10.009
14. Ng JA, Booth JT, Poulsen PR, et al. Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: first clinical results. *Int J Radiat Oncol Biol Phys*. 2012;84(5):e655-e661. doi:10.1016/J.IJROBP.2012.07.2367
15. Keall PJ, Aun Ng J, O'Brien R, et al. The first clinical treatment with kilovoltage intrafraction monitoring (KIM): a real-time image guidance method. *Med Phys*. 2014;42(1):354-358. doi:10.1118/1.4904023
16. Mueller M, Zolfaghari R, Briggs A, et al. The first prospective implementation of markerless lung target tracking in an experimental quality assurance procedure on a standard linear accelerator. *Phys Med Biol*. 2020;65(2):025008. doi:10.1088/1361-6560/ab5d8b

17. Shieh CC, Caillet V, Dunbar M, et al. A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. *Phys Med Biol*. 2017;62(8):3065-3080. doi:10.1088/1361-6560/aa6393
18. Mori S. Real-time image-processing algorithm for markerless tumour tracking using x-ray fluoroscopic imaging. *Br J Radiol*. 2014;87:20140001. doi:10.1259/bjr.20140001
19. Lewis JH, Li R, Watkins WT, et al. Markerless lung tumor tracking and trajectory reconstruction using rotational cone-beam projections: a feasibility study. *Phys Med Biol*. 2010;55(9):2505-2522. doi:10.1088/0031-9155/55/9/006
20. Hugo GD, Liang J, Yan D. Marker-free lung tumor trajectory estimation from a cone beam CT sinogram. *Phys Med Biol*. 2010;55(9):2637-2650. doi:10.1088/0031-9155/55/9/014
21. Fu D, Kahn R, Wang B, et al. Xsight lung tracking system: a fiducial-less method for respiratory motion tracking. *Treating Tumors that Move with Respiration*. Springer; 2007:265-282. doi:10.1007/978-3-540-69886-9_26
22. Hirai R, Sakata Y, Tanizawa A, Mori S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Phys Medica*. 2019;59:22-29. doi:10.1016/j.ejmp.2019.02.006
23. Zhao W, Han B, Yang Y, et al. Incorporating imaging information from deep neural network layers into image guided radiation therapy (IGRT). *Radiother Oncol*. 2019;140:167-174. doi:10.1016/j.radonc.2019.06.027
24. Mylonas A, Mueller M, Keall PJ, Booth JT, Nguyen DT. Towards real-time markerless prostate IGRT during VMAT treatment. In: American Association of Physicists in Medicine (AAPM). 2021.
25. Zhao W, Shen L, Han B, et al. Markerless pancreatic tumor target localization enabled by deep learning. *Int J Radiat Oncol Biol Phys*. 2019;105(2):432-439. doi:10.1016/j.ijrobp.2019.05.071
26. Xie Y, Xing L, Gu J, Liu W. Tissue feature-based intra-fractional motion tracking for stereoscopic x-ray image guided radiotherapy. *Phys Med Biol*. 2013;58(11):3615. doi:10.1088/0031-9155/58/11/3615
27. Lin T, Cervîo LI, Tang X, Vasconcelos N, Jiang SB. Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy. *Phys Med Biol*. 2009;54(4):981. doi:10.1088/0031-9155/54/4/011
28. Shieh CC, Keall PJ, Kuncic Z, Huang CY, Feain I. Markerless tumor tracking using short kilovoltage imaging arcs for lung image-guided radiotherapy. *Phys Med Biol*. 2015;60(24):9437. doi:10.1088/0031-9155/60/24/9437
29. Staub D, Murphy MJ. A digitally reconstructed radiograph algorithm calculated from first principles. *Med Phys*. 2013;40(1):011902. doi:10.1118/1.4769413
30. Dhont J, Verellen D, Mollaert I, Vanreusel V, Vandemeulebroucke J. RealDRR – rendering of realistic digitally reconstructed radiographs using locally trained image-to-image translation. *Radiother Oncol*. 2020;153:213-219. doi:10.1016/j.radonc.2020.10.004
31. Strid KG. Significance of quantum fluctuations in roentgen imaging. *Acta Oncol (Madr)*. 1980;19(2):129-138. doi:10.3109/02841868009130144
32. Grossberg AJ, Elhalawani H, Mohamed ASR, et al. HNSCC Dataset. Published online 2020. doi:10.7937/k9/tcia.2020.a8sh-7363
33. Grossberg AJ, Mohamed ASR, Elhalawani H, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. *Sci Data*. 2018;5(1):1-10. doi:10.1038/sdata.2018.173
34. Hesham E, Abdallah SRM, Aubrey LW, et al. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. *Sci Data*. 2017;4(1):1-14. doi:10.1038/sdata.2017.77
35. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045-1057. doi:10.1007/s10278-013-9622-7
36. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage*. 2009;45(1):S61-S72. doi:10.1016/J.NEUROIMAGE.2008.10.040
37. Eiben B, Bertholet J, Menten MJ, Nill S, Oelfke U, McClelland JR. Consistent and invertible deformation vector fields for a breathing anthropomorphic phantom: a post-processing framework for the XCAT phantom. *Phys Med Biol*. 2020;65(16):165005. doi:10.1088/1361-6560/ab8533
38. Papiez BW, Heinrich MP, Fehrenbach J, Risser L, Schnabel JA. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Med Image Anal*. 2014;18(8):1299-1311. doi:10.1016/J.MEDIA.2014.05.005
39. Gurney-Champion OJ, McQuaid D, Dunlop A, et al. MRI-based assessment of 3D intrafractional motion of head and neck cancer for radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;100(2):306-316. doi:10.1016/j.ijrobp.2017.10.016
40. Finnegan R, Chlap P. PlatiPy (Processing Library and Analysis Toolkit for Medical Imaging in Python). GitHub. Accessed June 21, 2021. https://github.com/pyplati/platipy
41. Staring M, Klein S, Pluim JPW. A rigidity penalty term for nonrigid registration. *Med Phys*. 2007;34(11):4098-4108. doi:10.1118/1.2776236
42. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205. doi:10.1109/TMI.2009.2035616
43. Mirza M, Osindero S. Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784 (2014) http://arxiv.org/abs/1411.1784
44. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. Proceedings of 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017. 2016; 5967-5976. Accessed April 19, 2021. http://arxiv.org/abs/1611.07004
45. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. 2019;58:101552. doi:10.1016/J.MEDIA.2019.101552
46. Li C, Wand M. Precomputed real-time texture synthesis with Markovian Generative Adversarial Networks. *Lecture Notes in Computer Science*. 2016:9907:702-716. doi:10.1007/978-3-319-46487-9_43
47. Ronneberger O, Fischer P, BroxT. U-Net: convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*. 2015;9351:234-241. Accessed August 18, 2021. https://arxiv.org/abs/1505.04597v1
48. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. 2014.
49. Kessler DA, MacKay JW, Crowe VA, et al. The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs. *Comput Med Imaging Graph*. 2020;86:101793. doi:10.1016/J.COMPMEDIMAG.2020.101793
50. Rit S, Vila Oliva M, Brousmiche S, Labarbe R, Sarrut D, Sharp GC. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). In: *Journal of Physics: Conference Series*. Vol 489. Institute of Physics Publishing; 2014:012079. doi:10.1088/1742-6596/489/1/012079
51. Singhrao K, Kirby N, Pouliot J. A three-dimensional head-and-neck phantom for validation of multimodality deformable image registration for adaptive radiotherapy. *Med Phys*. 2014;41(12):121709. doi:10.1118/1.4901523
52. Shamir RR, Duchin Y, Kim J, Sapiro G, Harel N. Continuous dice coefficient: a method for evaluating probabilistic segmentations. *arXiv preprint arXiv:1906*. 2019:11031.

53. Chalana V, Kim Y. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans Med Imaging*. 1997;16(5):642-652. doi:10.1109/42.640755
54. Santos JDL, Popple R, Agazaryan N, et al. Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. *Int J Radiat Oncol Biol Phys*. 2013;87(1):33-45. doi:10.1016/J.IJROBP.2013.02.021
55. Zhao W, Shen L, Han B, et al. Deep learning approach for markerless pancreatic tumor target localization. *Int J Radiat Oncol*. 2019;105(1):S202-S203. doi:10.1016/j.ijrobp.2019.06.268
56. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a). *Med Phys*. 2006;33(10):3874-3900. doi:10.1118/1.2349696
57. Al-Mayah A, Moseley J, Hunter S, et al. Biomechanical-based image registration for head and neck radiation treatment. *Phys Med Biol*. 2010;55(21):6491-6500. doi:10.1088/0031-9155/55/21/010
58. Pukala J, Meeks SL, Staton RJ, Bova FJ, Mañon RR, Langen KM. A virtual phantom library for the quantification of deformable image registration uncertainties in patients with cancers of the head and neck. *Med Phys*. 2013;40(11):111703. doi:10.1118/1.4823467
59. Varadhan R, Karangelis G, Krishnan K, Hui S. A framework for deformable image registration validation in radiotherapy clinical applications. *J Appl Clin Med Phys*. 2013;14(1):192-213. doi:10.1120/jacmp.v14i1.4066
60. Barkley HT, Gilbert Fletcher TH, Richard Jesse TH, Robert Lindberg TD. Management of cervical lymph node metastases in squamous ceil carcinoma of the tonsillar fossa, base of tongue, supraglottic larynx, and hypopharynx. *Am J Surg*. 1972;124(4):462-467.
61. Lee N, Harris J, Garden AS, et al. Intensity-modulated radiation therapy with or without chemotherapy for nasopharyngeal carcinoma: radiation therapy oncology group phase II trial 0225. *J Clin Oncol*. 2009;27(22):3684-3690. doi:10.1200/JCO.2008.19.9109
62. Poulsen PR, Cho B, Langen K, Kupelian P, Keall PJ. Three-dimensional prostate position estimation with a single x-ray imager utilizing the spatial probability density. *Phys Med Biol*. 2008;53(16):4331-4353. doi:10.1088/0031-9155/53/16/008
63. Nguyen DT, O'Brien R, Kim JH, et al. The first clinical implementation of a real-time six degree of freedom target tracking system during radiation therapy based on Kilovoltage Intrafraction Monitoring (KIM). *Radiother Oncol*. 2017;123(1):37-42. doi:10.1016/j.radonc.2017.02.013
64. Roggen T, Bobic M, Givehchi N, Scheib SG. Deep learning model for markerless tracking in spinal SBRT. *Phys Medica*. 2020;74:66-73. doi:10.1016/J.EJMP.2020.04.029
65. Zhang L, LoSasso T, Zhang P, Hunt M, Mageras G, Tang G. Couch and multileaf collimator tracking: a clinical feasibility study for pancreas and liver treatment. *Med Phys*. 2020;47(10):4743-4757. doi:10.1002/mp.14438
66. Clover K, Oultram S, Adams C, Cross L, Findlay N, Ponman L. Disruption to radiation therapy sessions due to anxiety among patients receiving radiation therapy to the head and neck area can be predicted using patient self-report measures. *Psychooncology*. 2011;20(12):1334-1341. doi:10.1002/pon.1854
67. Molassiotis A, Rogers M. Symptom experience and regaining normality in the first year following a diagnosis of head and neck cancer: a qualitative longitudinal study. *Palliat Support Care*. 2012;10(3):197-204. doi:10.1017/S147895151200020X

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gardner M, Bouchta YB, Mylonas A, et al. Realistic CT data augmentation for accurate deep-learning based segmentation of head and neck tumors in kV images acquired during radiation therapy. *Med Phys*. 2023;50:4206–4219. https://doi.org/10.1002/mp.16388