Physics and Imaging in Radiation Oncology 32 (2024) 100658

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

Proton dose calculation on cone-beam computed tomography using unsupervised 3D deep learning networks

Casper Dueholm Vestergaard [a,*], Ulrik Vindelev Elstrøm [a], Ludvig Paul Muren [a], Jintao Ren [a], Ole Nørrevang [a], Kenneth Jensen [a], Vicki Trier Taasti [a,b]

[a] Danish Centre for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark
[b] Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, the Netherlands

ARTICLE INFO

Keywords:
Adaptive proton therapy
Deep-learning
Synthetic CT
Cone-beam CT
Proton dose calculation
Image quality improvement

ABSTRACT

Background and Purpose: Poor image quality of cone-beam computed tomography (CBCT) images can hinder proton dose calculation to assess the influence of anatomy changes. The aim of this study was to evaluate image quality and proton dose calculation accuracy of synthetic CTs generated from CBCT using unsupervised 3D deep-learning networks.

Materials and methods: A total of 102 head-and-neck cancer patients were used to train (N=82) and test (N=20) i) a cycle-consistent generative adversarial network, ii) a contrastive unpaired translation, and iii) a fusion of the two (CycleCUT). For patients in the test set, a repeat CT was deformably registered to a same-day CBCT to create a ground-truth CT for comparison. The proton plan was re-calculated on the ground-truth CT and synthetic CTs. The image quality of the synthetic CTs was evaluated using peak signal-to-noise ratio, structural similarity index measure, mean error, and mean absolute error (MAE). Proton dose calculation accuracy was assessed through 3D gamma analysis and dose-volume-histogram parameters.

Results: All synthetic CTs accurately preserved the CBCT anatomy (verified by visual inspection) while improving the image quality. The CycleCUT network had slightly improved image quality compared to the other networks (MAE in body: 53 Hounsfield units (HU) vs. 54/55 HU). All networks had similar proton dose calculation accuracy with gamma passing rate above 97%.

Conclusions: All three evaluated networks generated synthetic CT images with dose distributions comparable to those of conventional fan-beam CT. The synthetic CT generation was fast, making all networks feasible for adaptive proton therapy.

1. Introduction

The sharp distal fall-off of the proton dose deposition makes proton therapy sensitive to changes in patient positioning, anatomy and tissue densities [1,2]. A cone-beam computed tomography (CBCT) image is often acquired of the patient in the treatment position prior to the daily treatment delivery to verify the positioning of the patient and tumor, and to monitor changes in patient anatomy and/or the tumor (due to, e.g., weight loss or tumor shrinkage) [3,4]. Unfortunately, CBCT scans are typically subject to high levels of scattering and imaging artifacts compared to conventional fan-beam computed tomography (CT) scans, and they are therefore often inadequate for accurate proton dose calculations [5,6]. Patients are often referred to a rescan on a conventional CT scanner (a so-called repeat CT; rCT) if large anatomical changes are observed on the CBCT images. An improved workflow could be obtained if dose re-calculations could be performed directly on the CBCT [4,7].

Strategies based on deep-learning (DL) have been proposed as an alternative to the traditional CBCT correction methods for adaptive radiotherapy [8]. Many studies have found promising results for artifact reduction and improvements in spatial uniformity and image quality [9], and also for accurate CBCT-based photon [10] and proton [11] dose calculation by exploiting U-nets [12]. The U-net, however, relies on paired image data for training, which is challenging to obtain due to positional and anatomical changes [13]. Many have therefore investigated the cycle-consistent generative adversarial network (CycleGAN) [14] for synthetic CT (sCT) generation from CBCT, as the CycleGAN does not require paired data for training, allowing it to handle anatomical changes between CT and CBCT acquisitions effectively [13,15,16].

* Corresponding author.
E-mail address: cvo@clin.au.dk (C. Dueholm Vestergaard).

https://doi.org/10.1016/j.phro.2024.100658
Received 1 July 2024; Received in revised form 4 October 2024; Accepted 15 October 2024
Available online 26 October 2024

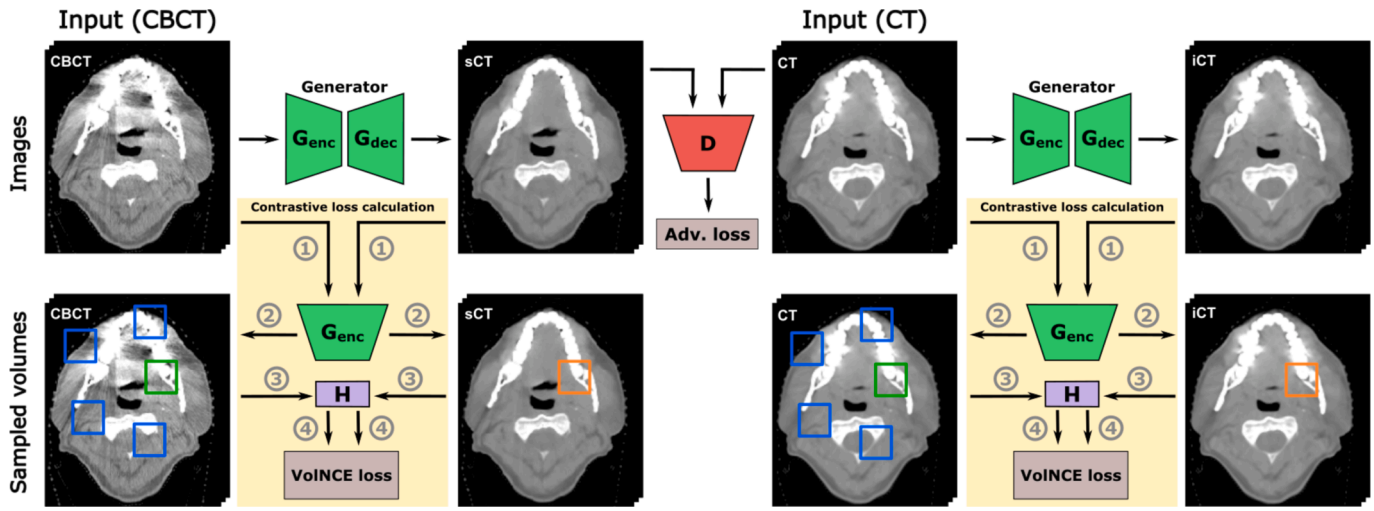

Fig. 1. Working principle of the training of the CUT network. CBCT and CT images (input) are translated into synthetic CT (sCT) and identity CT (iCT) images, respectively, using the generator (shown in green), which consists of an encoder (G_{enc}) and decoder (G_{dec}). The input CT and generated sCT are used as input to the discriminator (D, shown in red) to calculate the adversarial loss (Adv. Loss; Eq. (4)). The schematic for the contrastive loss calculation is shown in the yellow box where the numbers (1–4) correspond to the following steps: (1) Feature extraction. (2) Negative volumes (blue squares) and one positive volume (green square) are sampled from each feature map of the CBCT and CT images and compared to a sampled volume-of-interest (orange square) on the sCT and iCT images, respectively. The positive volume and volume-of-interest are placed at corresponding locations, while the negative volumes are randomly distributed over the rest of the image stack. (3) The sampled volumes are projected to a shared latent space, using the multilayer perceptron, H (purple). (4) The output of H is used to calculate the volume-based noise contrastive estimation (VolNCE) loss.

An alternative approach for unpaired image translation, the contrastive unpaired translation (CUT) network, was recently proposed [17]. This unsupervised network uses contrastive loss to maximize mutual information between corresponding patches in input and output images. In contrast to CycleGAN, the CUT network employs a one-directional mapping, resulting in simpler, faster, and more stable training with lower GPU memory requirements. The simplified architecture avoids the strict cycle-consistency constraint of the CycleGAN which sometimes can cause artifacts or limit the variety in the generated images [17]. A two-dimensional (2D) CUT network has been explored for sCT generation from CBCT [18] and magnetic resonance imaging (MRI) [19]. The latter group also successfully integrated the contrastive loss into the CycleGAN framework to produce the so-called CycleCUT network, which leverages the strengths of both the CycleGAN and CUT network [19].

Acknowledging the progress made with the 2D CUT network in sCT generation, the complexity and spatial dependencies of anatomical structures in images necessitate the use of 3D models to capture context and variations in all directions more accurately. Thus, 3D approaches can potentially enhance the precision of sCTs for proton therapy clinics. To our knowledge, no study has previously investigated the CUT and CycleCUT networks in 3D with applications in proton therapy.

In this study, we therefore evaluated the quality of sCTs generated from CBCT by three 3D DL models, the CycleGAN, CUT, and CycleCUT network, in terms of CT number and proton dose calculation accuracy, with the purpose of enabling CBCT-based daily proton dose calculations for head-and-neck cancer patients.

2. Methods and materials

2.1. Imaging protocol and pre-processing

CT and CBCT data from a total of 103 head-and-neck cancer patients treated with proton therapy, were available for this study. All scans were visually reviewed for severe artifacts, and one patient was excluded due to abnormal anatomy from osteosynthetic implants, resulting in 102 patients for training and evaluation of the DL networks. Permission to use the patient data was issued by The Central Denmark Region Committees on Health Research Ethics (case number: 2119125). All patients were treated with intensity modulated proton therapy (IMPT) and according to the Danish Head and Neck Cancer Group (DAHANCA) guidelines [20].

The CT scans were acquired with a SOMATOM Definition Edge CT scanner (Siemens Healthineers, Forchheim, Germany) in dual-energy CT (DECT) twin-beam acquisition mode. The images were reconstructed using the Qr40 kernel with beam hardening correction for bone and iterative reconstruction (ADMIRE, strength level 3). The DECT images were used to create 90 keV virtual monoenergetic images applying the Monoenergetic Plus algorithm in Siemens *syngo.via* software.

The CBCT scans were acquired in treatment position on a ProBeam gantry (Varian Medical Systems, Palo Alto, CA, USA). A standard ProBeam head-and-neck protocol, with full fan bowtie and half scan acquisition, was used (tube voltage: 100 kV; exposure: 889 mAs). The images were reconstructed using filtered-back projection with the FDK (Feldkamp-Davis-Kress) algorithm [21]. The CBCT field-of-view was 21x25 cm (crania-caudal x transversal).

The patient cohort was randomly split for training/validation/testing with 77/5/20 patients, respectively. The validation set was used to monitor the training process, while the test set was used to evaluate the trained networks. Each patient had several CT scans (range: 5–8; including planning CT (pCT) and weekly rCTs) and CBCT scans (range: 33–50). A total of 467/32 CT scans and 2781/176 CBCT scans were used for training/validation, respectively. Details on the image preprocessing of the CBCT/CT scans can be found in supplementary material S1. For the test patients, fourteen patients were bilateral, and six patients were unilateral (two patients were re-irradiated with additional constraints on the beam angles). The prescription doses for the high-risk targets were between 60 and 68 Gy (see Table S1 in the supplementary material for additional information). One CBCT/rCT pair acquired on the same day was selected for each test patient. The rCT and CBCT images were selected based on visual slice-wise inspection to ensure anatomical correspondence between the two images.

2.2. Deep-learning architectures

The three DL networks (CycleGAN, CUT, and CycleCUT) were

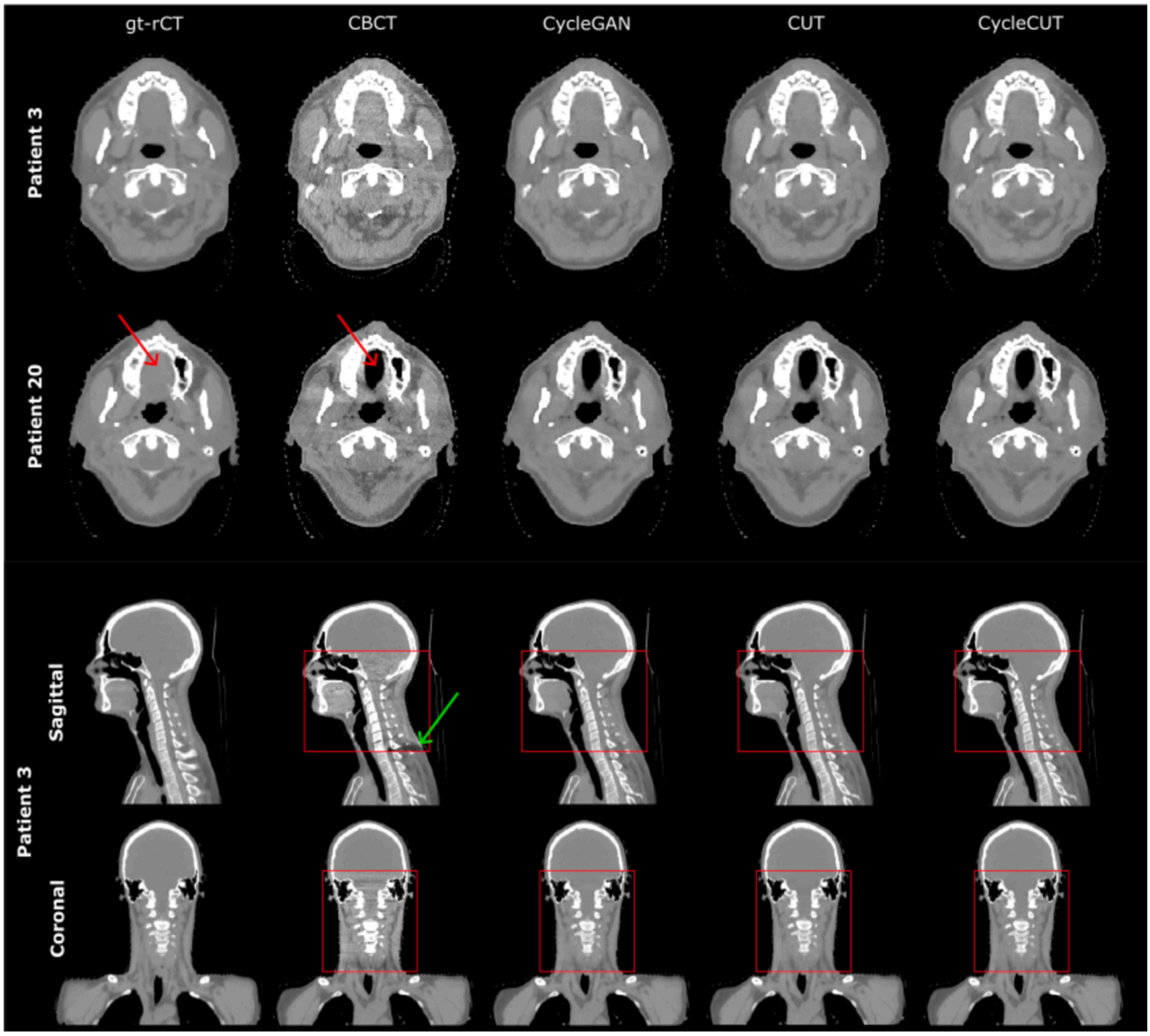

Fig. 2. (Top) Axial slices of gt-rCT, CBCT, and sCTs for two patients (Patient 3 and 20, see Section S2 in supplementary material), representing a good (Patient 3) and less optimal (Patient 20) anatomical correspondence between the gt-rCT and CBCT (red arrow on gt-rCT and CBCT for Patient 20). (Bottom) Sagittal and coronal views of Patient 3. The green arrow on the sagittal view of the CBCT shows an artifact that has been corrected on all sCTs. The CBCT field-of-view is indicated by a red box. All images are displayed in the same window level/width setting (0 HU/800 HU).

implemented in PyTorch (v 1.13.0) in Python. The working principles of the networks are briefly outlined in the following. Further details on the networks and the training details can be found in supplementary material S3.

The CycleGAN consisted of two generators, G_{CT} and G_{CBCT}, which generated sCTs and synthetic CBCTs (sCBCTs), respectively, and two discriminators, D_{CT} and D_{CBCT}, which tried to distinguish real images (CT/CBCT) from synthetic images (sCT/sCBCT). The loss function of the generators, $\mathcal{L}_G$, consisted of three parts: 1) cycle consistency loss, $\mathcal{L}_{cycle}$, which was used to preserve anatomy of the input image, 2) adversarial loss, $\mathcal{L}_{adv}$, which was used to enforce correspondence between voxel intensities in the synthetic images and real images, and 3) identity loss, $\mathcal{L}_{identity}$, which enforced that no changes were made to the input if it resembled an image from the target domain. The full loss function used to train the two generators is given by:

$$\mathcal{L}_G = \mathcal{L}_{adv,G_{CT}} + \mathcal{L}_{adv,G_{CBCT}} + \lambda_c \cdot \left(\mathcal{L}_{cycle,G_{CT}} + \mathcal{L}_{cycle,G_{CBCT}}\right) + \lambda_i \cdot \left(\mathcal{L}_{identity,G_{CT}} + \mathcal{L}_{identity,G_{CBCT}}\right) \quad (1)$$

where λ_c and λ_i, are parameters used to scale the cyclic and identity loss with respect to the adversarial loss. Following the original CycleGAN paper [14], we set $\lambda_c = 10$ and $\lambda_i = 5$.

The loss function for the discriminators, $\mathcal{L}_D$, was based purely on adversarial loss, and given by:

$$\mathcal{L}_D = \frac{1}{2} \cdot \left(\mathcal{L}_{adv,D_{CT}} + \mathcal{L}_{adv,D_{CBCT}}\right) \quad (2)$$

Similarly to the CycleGAN, the CUT network [17] contained an adversarial loss ($\mathcal{L}_{adv,G}/\mathcal{L}_{adv,D}$), which strived to correct the CT numbers of the CBCT. However, in the CUT network, the cycle consistency and identity losses were replaced with a volume-based contrastive loss function, $\mathcal{L}_{VolNCE}$ (NCE: noise contrastive estimation). This loss function aimed to train an encoder (first half of the generator) to associate spatially corresponding sub-volumes of the input (CBCT) and output (sCT) images, while disassociating them from other randomly sampled sub-volumes in the CBCT image. Similarly, real CT images were compared to identity CT images (iCT; output of generator when inputting a CT image; see Fig. 1) to prevent the generator from imposing

changes to CT images, like the identity loss of CycleGAN. The loss functions used to train the generator, $\mathscr{L}_G$, and discriminator, $\mathscr{L}_D$, of the CUT network is given by:

$$\mathscr{L}_G = \mathscr{L}_{adv,G} + \frac{1}{2}\cdot(\mathscr{L}_{VolNCE}(CBCT, sCT) + \mathscr{L}_{VolNCE}(CT, iCT)) \quad (3)$$

$$\mathscr{L}_D = \mathscr{L}_{adv,D} \quad (4)$$

Additional details on the calculation of the volume-based noise contrastive estimation (VolNCE) loss, $\mathscr{L}_{VolNCE}$, can be found in the supplementary material S3.2.

The CycleCUT network introduced contrastive learning into the framework of the CycleGAN. Hence, in the CycleCUT network, the structural consistency and anatomical preservation were enforced by having both a cycle consistency loss (as in CycleGAN) and a volume-wise contrastive loss (as in CUT). As for the CycleGAN, CycleCUT had two generators (G_{CT}/G_{CBCT}). The full generator loss function had four contributions from the VolNCE loss (see Fig. S3 in supplementary material S3.3), whereas the CUT only had two (see Eq. (3)). The full generator loss then is given by:

$$\begin{aligned}\mathscr{L}_G = {} & \mathscr{L}_{adv,G_{CT}} + \mathscr{L}_{adv,G_{CBCT}} + \\ & \lambda_c\cdot(\mathscr{L}_{cycle,G_{CT}} + \mathscr{L}_{cycle,G_{CBCT}}) + \lambda_i\cdot(\mathscr{L}_{identity,G_{CT}} + \mathscr{L}_{identity,G_{CBCT}}) + \\ & \lambda_{NCE}\cdot(\mathscr{L}_{VolNCE}(CBCT, sCT) + \mathscr{L}_{VolNCE}(CT, sCBCT) + \\ & \mathscr{L}_{VolNCE}(sCBCT, CyCT) + \mathscr{L}_{VolNCE}(sCT, CyCBCT))\end{aligned} \quad (5)$$

where $\lambda_c = 10$, $\lambda_i = 5$, and $\lambda_{NCE} = 0.1$. The loss of the discriminator was the same as for the CycleGAN network (Eq. (2)).

2.3. Evaluation

The trained generators were applied to the CBCTs from the test set to create the sCTs used in the evaluation. The CBCTs were passed through the trained generators in a sliding-window manner, and the overlapping volumes were averaged to generate the final 3D sCT. The clinical rigid registration between the CBCT and pCT was used to stitch the sCT onto the pCT to add image information outside the CBCT field-of-view. The rCT was first rigidly registered to the pCT and afterwards deformably registered to the CBCT to further reduce anatomical and setup differences between the rCT and the CBCT. The deformable image registration was performed in Velocity® (Varian Medical Systems, Palo Alto, CA, USA). The deformed rCT was seen as the ground truth (gt-rCT) in the evaluations. The structure set from the pCT was deformably propagated to the sCT and gt-rCT, and the clinical proton plan (created on the pCT in Eclipse (v.16.1), Varian – a Siemens Healthineers company, Palo Alto, CA, USA) was re-calculated on the sCT and gt-rCT.

The sCTs were compared to the gt-rCTs to evaluate the CT number and proton dose calculation accuracy on the sCTs. To quantitatively evaluate the CT number accuracy and image quality of the sCTs, the peak signal-to-noise ratio (PSNR), structural similarity measure (SSIM), mean absolute error (MAE), and mean error (ME) between the gt-rCTs and CBCTs/sCTs were calculated (the equations can be found in supplementary material S4). The volumes-of-interest (VOIs) used to calculate these metrics were: the high-risk target (CTV1, homogenous), the oral cavity (subject to many artifacts, e.g., from dental implants), mandible (high density), and the body (inside the CBCT field-of-view). Additionally, CT number line profiles and histogram distributions were calculated for the sCTs and gt-rCT for two representative patients.

The dose calculation accuracy was evaluated by comparing the proton dose distributions calculated on the sCTs and gt-rCT (reference) through differences in dose-volume-histogram (DVH) parameters for the target and organs-at-risk, and gamma analysis [22]. In the DVH comparison, the following DVH parameters were evaluated: $D_{99\%}$ for the three targets (CTV1-3, with different prescribed doses); D_{max} (max in 0.027 cm^3 volume [20]) for the high-risk target (CTV1), spinal cord, body, and mandible; and D_{mean} for the oral cavity, parotids (left and right), pharyngeal constrictor muscles (PCM; low/middle/upper), and submandibular glands (left and right combined). For the gamma analysis, gamma pass rates were calculated for passing criteria of 1%/2mm, 2%/2mm, 3%/2mm (used clinically at our institute for plan quality assurance), and 3%/3mm, all with a lower dose cut-off of 10% of the maximum dose in gt-rCT dose distribution (global evaluation). The Wilcoxon signed-rank test was used to determine the significance of any observed differences ($p \leq 0.05$) [23].

Table 1
Median [minimum (min); maximum (max)] values of peak-signal to noise ratio (PSNR; in decibels (dB)), structural similarity index measure (SSIM), mean absolute error (MAE; in Hounsfield units (HU)), and mean error (ME; in HU) calculated between the gt-rCT and CBCT/sCTs within the high-risk target, oral cavity, mandible, and the entire body (inside CBCT field-of-view). The best value for each metric (i.e. the highest value for the median PSNR and SSIM, the lowest median MAE, or the median ME closest to zero) have been highlighted in bold if the value was significantly different from both of the two other networks according to the Wilcoxon signed rank test ($p \leq 0.05$).

	PSNR (dB) ↑ Median [min; max]	SSIM ↑ Median [min; max]	MAE (HU) ↓ Median [min; max]	ME (HU) Median [min; max]
	Target			
CBCT	36.5 [27.9; 40.8]	1.00 [0.96; 1.00]	41.2 [30.8; 129.7]	−20.4 [-1.7; −95.4]
CycleGAN	41.7 [33.2; 48.6]	1.00 [0.99; 1.00]	18.1 [11.6; 38.3]	−5.0 [0.1;19.1]
CUT	41.6 [33.2; 48.0]	1.00 [0.99; 1.00]	17.3 [11.2; 33.9]	3.3 [-1.0; 21.4]
CycleCUT	**42.4** [33.2; 49.0]	1.00 [0.99; 1.00]	**16.1** [10.1; 36.3]	−2.1 [0.0; −37.2]
	Oral Cavity			
CBCT	29.8 [18.3; 37.0]	0.98 [0.89; 1.00]	75.8 [31.8; 247.4]	−36.7 [0.0; −178.1]
CycleGAN	36.9 [26.5; 44.0]	1.00 [0.97; 1.00]	26.8 [12.9; 72.8]	**0.8** [-0.3; 45.8]
CUT	36.9 [26.3; 44.8]	1.00 [0.97; 1.00]	26.2 [12.8; 70.7]	5.3 [-0.3; 51.0]
CycleCUT	36.7 [26.3; 44.4]	1.00 [0.97; 1.00]	25.8 [11.7; 72.7]	2.6 [-0.8; 48.0]
	Mandible			
CBCT	18.6 [13.6; 22.2]	0.90 [0.78; 0.95]	373.6 [241.7; 694.5]	−315.1 [-197.3; −592.6]
CycleGAN	26.1 [19.0; 31.9]	0.97 [0.89; 0.99]	145.5 [77.5; 344.5]	−21.4 [-0.2; −61.7]
CUT	26.2 [19.2; 31.0]	0.97 [0.90; 0.99]	141.6 [83.7; 335.4]	22.7 [2.6; 66.1]
CycleCUT	**26.3** [19.1; 32.7]	0.97 [0.90; 1.00]	**137.2** [68.0; 336.6]	**9.5** [-2.3; 46.9]
	Body			
CBCT	25.6 [22.0; 35.5]	0.95 [0.92; 0.97]	137.5 [114.3; 194.3]	20.9 [-7.4; −130.7]
CycleGAN	31.8 [26.8; 36.1]	0.97 [0.94; 0.99]	54.2 [35.8; 88.5]	−6.9 [0.6; –23.5]
CUT	31.8 [27.0; 35.4]	0.97 [0.94; 0.99]	55.2 [35.6; 85.3]	5.3 [0.9; 22.0]
CycleCUT	31.8 [26.8; 36.1]	0.97 [0.94; 0.99]	**53.4** [34.1; 87.2]	**2.8** [0.1; 18.0]

3. Results

3.1. CT number evaluations

Fig. 2 presents sCTs generated by the three networks (CycleGAN, CUT and CycleCUT), along with the original CBCT image and the gt-rCT for two test patients, one with good and one with less optimal anatomical correspondence between the CBCT and gt-rCT. All three networks correctly replicated the anatomy of the CBCT, while obtaining an image quality comparable to the gt-rCT in all three views (axial/sagittal/coronal). CT number differences between the gt-rCT and the CBCTs/sCTs can be seen in the supplementary material S5.1.

All three sCT networks greatly improved the PSNR, ME, and MAE for all four structures, as well as the SSIM for the oral cavity, mandible, and

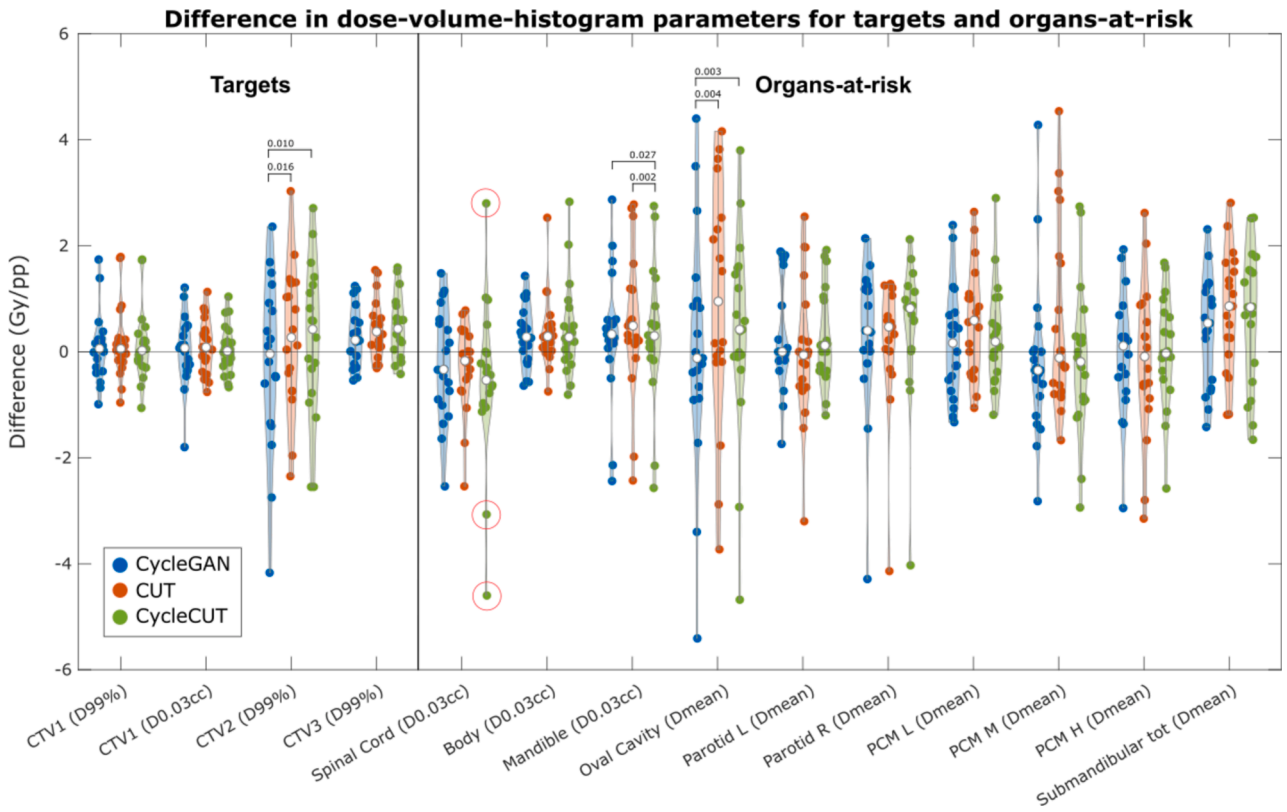

Fig. 3. Violin plots over the differences for dose-volume-histogram parameters (in Gray (Gy) or percent point (pp)) between the dose distribution calculated on the gt-rCT (reference) and sCTs, for the targets (left) and organs-at-risk (right). The three red circles for the spinal cord D_{max} show outlier values for the CycleCUT which were not seen for the CycleGAN and CUT sCT. The p-value is shown above a DVH parameter of one of the models if the value was significantly different from both other networks according to the Wilcoxon signed rank test ($p \leq 0.05$).

Table 2
Median [25th; 75th percentile] of the global gamma pass rate calculated between the dose distribution on the gt-rCT and sCTs. A lower dose cutoff level of 10% was used, and the analysis was restricted to the body structure of the gt-rCT. No significant differences were found between the sCTs for any of the gamma pass rate according to Wilcoxon signed rank test ($p \leq 0.05$).

	Gamma pass rates			
	1%/2mm (%)	2%/2mm (%)	3%/2mm (%)	3%/3mm (%)
CycleGAN	97.5 [93.1; 98.9]	98.1 [95.3; 99.4]	98.8 [96.7; 99.6]	99.6 [99.3; 99.9]
CUT	97.4 [92.5; 98.8]	98.3 [94.7; 99.3]	98.9 [96.3; 99.5]	99.6 [99.1; 99.9]
CycleCUT	97.4 [92.3; 99.0]	98.1 [94.3; 99.5]	98.8 [96.0; 99.7]	99.6 [99.0; 99.9]

body compared to the CBCT (Table 1). The SSIM was similar for all three sCT networks, but the CycleCUT network had a significantly ($p \leq 0.05$) better PSNR and MAE for the target, PSNR, MAE and ME for the mandible, and MAE and ME for the body, while the CycelGAN had a had a significantly ($p \leq 0.05$) better ME for the oral cavity. Additional CT number comparisons (CT number line profiles and histogram distributions of the high-risk target) can be found in supplementary material S5.2 and S5.3.

3.2. Dose evaluations

For all three networks, the DVH differences between the gt-rCT and sCTs were low for the targets, especially for the high-risk target (CTV1; Fig. 3). For the OARs, the median deviations were all close to zero, however, for a few patients, larger deviations were seen. For most DVH parameters, the same patient was found to be an outlier for all three sCTs. However, the CycleCUT network had large deviations in the D_{max} for the spinal cord for three patients, which were not seen for the other networks (red circles in Fig. 3).

For all three networks and all evaluated gamma criteria, the median gamma passing rates over the twenty test patients were above 97%, and there were no significant differences ($p > 0.05$) between any of the networks in the gamma analysis (Table 2; Fig. 4).

4. Discussion

This study investigated image quality improvements of CBCT scans for head-and-neck cancer patients using three 3D DL networks: CycleGAN, CUT, and a fusion of the two, CycleCUT. The CycleCUT network had slightly improved CT number accuracy compared to the two other networks, especially in bony structures (Table 1). The anatomical preservation for the CUT network was found to be almost as good as for the CycleGAN and CycleCUT network (verified by visual inspection) despite the lack of a voxel-wise loss function for anatomical preservation. All networks performed equally well in terms of proton dose evaluations (Figs. 3-4 and Table 2).

For the test patients, the rCT and CBCT were acquired on the same day, and they were chosen based on being anatomically similar (verified by visual inspection). Moreover, the rCT was deformed to the CBCT (to create the gt-rCT) with the purpose of further minimizing the anatomical differences. Despite this effort, it was difficult to obtain an accurate anatomical correspondence for some patients, leading to an incorrect ground-truth (see Fig. 2). Additionally, we deformably transferred structures from the pCT to the gt-rCT and CBCT/sCT, which gave some discrepancies in the contours. Since most evaluated structures were small, contour differences sometimes affected the mean CT number and dose, especially for structures near dose gradients. This was the case for

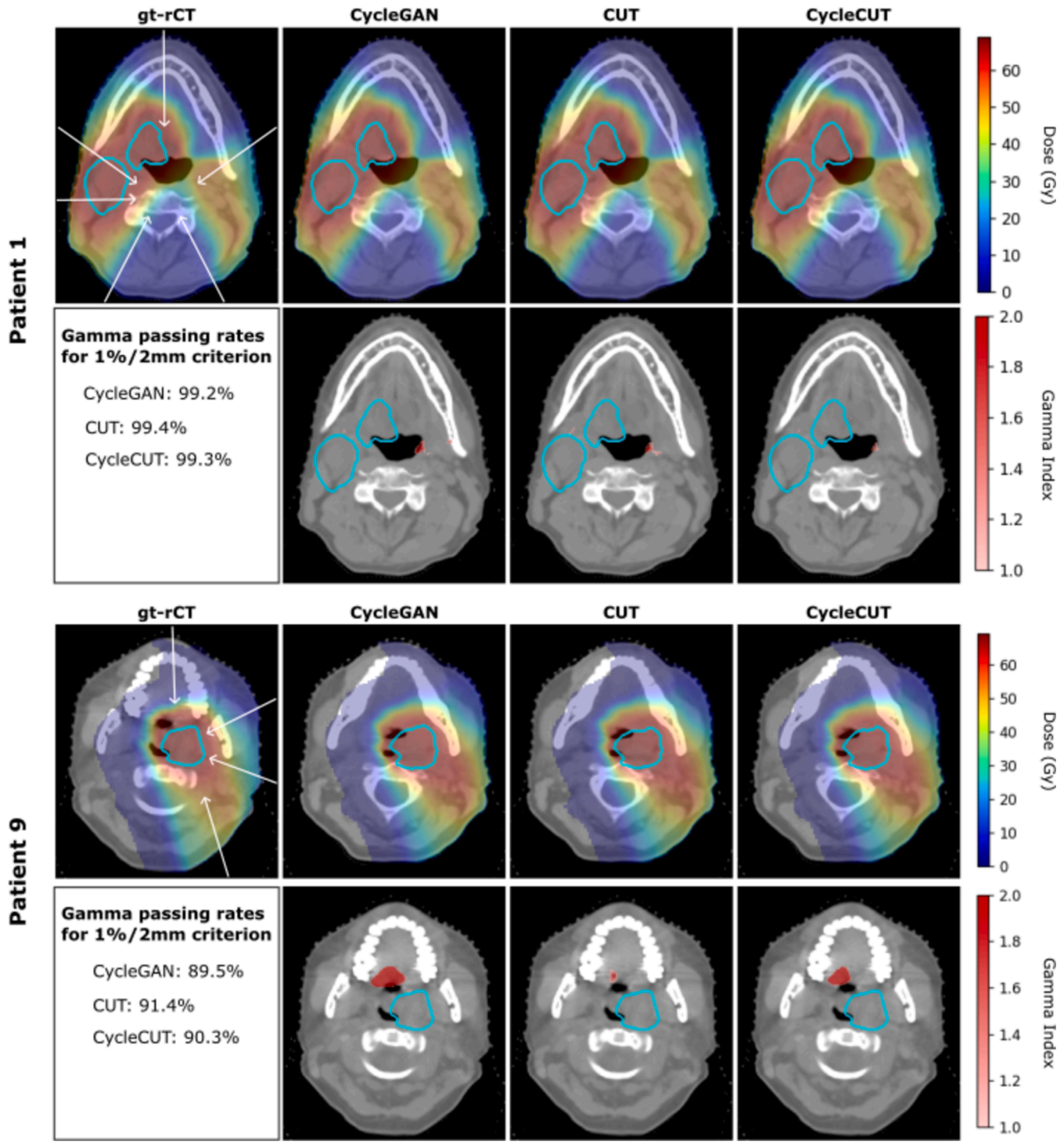

Fig. 4. Dose distributions calculated on the gt-rCT and sCTs and associated gamma maps (1%/2mm gamma passing criterion) for two patients: one with high gamma passing rates (Patient 1; see Section S2 in supplementary material), and one with low gamma passing rates (Patient 9; see Section S2 in supplementary material). In the gamma maps, only values above 1 are shown, indicating voxels where the gamma test fails. The high-risk target area is delineated in cyan. The beam directions are shown on the gt-rCT with white arrows (the anterior-posterior field contributes only below the mandible).

D_{max} of the spinal cord for the CycleCUT sCT in three patients (see encircled datapoints in Fig. 3). We found that rigidly transferring the spinal cord structure from the gt-rCT to the sCTs (such that the structure had the same size on the four images), instead of deformably transferring it from the pCT to each individual image, decreased the mean absolute deviation of the three encircled datapoints from 3.6 Gy to 0.6 Gy. This is a limitation, as some of the observed deviations originated from contour differences rather than the images themselves.

The CycleCUT network was originally developed for MRI-to-CT synthesis, where it significantly outperformed the CycleGAN and CUT in terms of image quality, but with similar dose calculation accuracies [19]. CBCT-to-CT synthesis using the CUT network has also been investigated, where the image quality was improved compared to the CycleGAN [18]. These results consistent with our findings. Both studies used the CUT network's NCE loss on a 2D patch-wise level, as originally developed. Our work demonstrates that it can also be applied effectively

at a 3D volume level. Consequently, our sCTs were found to have a good correspondence between consecutive slices (see Fig. 2), which sometimes can be difficult to achieve using 2D models [18,19,24]. Thus, the ability to capture the relationship between neighboring slices is an advantage of 3D models. However, 3D models require more GPU memory, a larger training dataset, and more training time compared to 2D models, which can limit the feasibility of 3D models for medical image analysis [25]. The lack of patient data can, however, be partially solved either by dividing the scans into smaller volumes (as done here), data augmentations [26], transfer learning [27], or by training a single network using data from multiple treatment sites [16]. Additionally, while the training time is longer for 3D models, we found that the average time needed for transforming a CBCT into a full field-of-view sCT (by stitching it onto the pCT) in dicom format, which can be directly imported into any treatment planning system for dose calculation, was just above 1 min (74 s) for all networks, making the sCT generation process suitable for adaptive proton therapy [1].

All three networks used in this study were based on the very successful GANs [28]. Several studies have previously showed that GAN-based DL can enhance CBCT image quality [29–33]. However, GANs are complex to train due to the simultaneous training of two competing networks, and they often suffer from mode collapse resulting in limited diversity of the generated images [34]. Alternative networks, such as diffusion models and transformers, have therefore recently been competing against the GANs in medical image analysis with promising results [33,35,36]. A recent study showed that a 2D conditional denoising diffusion probabilistic model had superior results in both visual quality and quantitative analysis compared to a 2D CycleGAN [36]. However, diffusion models rely on a time-consuming sampling procedure at inference (generation of sCT based on new data), which currently makes them impractical for an adaptive workflow.

In conclusion, we found that all three investigated sCT networks were able to generate high quality sCT images with proton dose distributions comparable to those of conventional fan-beam CT, allowing for proton dose calculations on the daily anatomy for dose verification. The sCT generation was fast (just above 1 min), making all investigated networks feasible for adaptive radiotherapy.

CRediT authorship contribution statement

Casper Dueholm Vestergaard: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Ulrik Vindelev Elstrøm:** Formal analysis, Investigation, Writing – review & editing. **Ludvig Paul Muren:** Conceptualization, Writing – review & editing, Supervision, Project administration. **Jintao Ren:** Software, Writing – review & editing. **Ole Nørrevang:** Writing – review & editing. **Kenneth Jensen:** Writing – review & editing. **Vicki Trier Taasti:** Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ludvig Muren author is an Editor-in-Chief for Physics and Imagin in Radiation Oncology and was not involved in the editorial review or the decision to publish this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2024.100658.

References

[1] Paganetti H, Botas P, Sharp GC, Winey B. Adaptive proton therapy. Phys Med Biol 2021;66:22TR01. https://doi.org/10.1088/1361-6560/ac344f.

[2] Bobić M, Lalonde A, Sharp GC, Grassberger C, Verburg JM, Winey BA, et al. Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy. Phys Med Biol 2021;66:055023. https://doi.org/10.1088/1361-6560/abe050.

[3] Baskar R, Lee KA, Yeo R, Yeoh K-W. Cancer and radiation therapy: current advances and future directions. Int J Med Sci 2012;9:193–9. https://doi.org/10.7150/ijms.3635.

[4] Hvid CA, Elstrøm UV, Jensen K, Grau C. Cone-beam computed tomography (CBCT) for adaptive image guided head and neck radiation therapy. Acta Oncol 2018;57:552–6. https://doi.org/10.1080/0284186X.2017.1398414.

[5] Harms J, Lei Y, Wang T, McDonald M, Ghavidel B, Stokes W, et al. Cone-beam CT-derived relative stopping power map generation via deep learning for proton radiotherapy. Med Phys 2020;47:4416–27. https://doi.org/10.1002/mp.14347.

[6] Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol 2020;65:095002. https://doi.org/10.1088/1361-6560/ab7d54.

[7] Elstrøm UV, Wysocka BA, Muren LP, Petersen JBB, Grau C. Daily kV cone-beam CT and deformable image registration as a method for studying dosimetric consequences of anatomic changes in adaptive IMRT of head and neck cancer. Acta Oncol 2010;49:1101–8. https://doi.org/10.3109/0284186X.2010.500304.

[8] Taasti VT, Klages P, Parodi K, Muren LP. Developments in deep learning based corrections of cone beam computed tomography to enable dose calculations for adaptive radiotherapy. Phys Imaging Radiat Oncol 2020;15:77–9. https://doi.org/10.1016/j.phro.2020.07.012.

[9] Kida S, Nakamoto T, Nakano M, Nawa K, Haga A, Kotoku J, et al. Cone beam computed tomography image quality improvement using a deep convolutional neural network. Cureus 2018;10:e2548. https://doi.org/10.7759/cureus.2548.

[10] Yoganathan S, Aouadi S, Ahmed S, Paloor S, Torfeh T, Al-Hammadi Y, et al. Generating synthetic images from cone beam computed tomography using self-attention residual UNet for head and neck radiotherapy. Phys Imaging Radiat Oncol 2023;28:100512. https://doi.org/10.1016/j.phro.2023.100512.

[11] Thummerer A, de Jong BA, Zaffino P, Meijers A, Marmitt GG, Seco J, et al. Comparison of the suitability of CBCT- and MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients. Phys Med Biol 2020;65:235036. https://doi.org/10.1088/1361-6560/abb1d6.

[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015. https://doi.org/10.1007/978-3-319-24574-4_28.

[13] Kurz C, Maspero M, Savenije MHF, Landry G, Kamp F, Pinto M, et al. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. Phys Med Biol 2019;64:225004. https://doi.org/10.1088/1361-6560/ab4d8c.

[14] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. IEEE International Conference on Computer Vision (ICCV) 2017. https://doi.org/10.1109/ICCV.2017.244.

[15] Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med Phys 2019;46:3998–4009. https://doi.org/10.1002/mp.13656.

[16] Maspero M, Houweling AC, Savenije MHF, van Heijst TCF, Verhoeff JJC, Kotte ANTJ, et al. A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. Phys Imaging Radiat Oncol 2020;14:24–31. https://doi.org/10.1016/j.phro.2020.04.002.

[17] Park T, Efros AA, Zhang R, Zhu J-Y. Contrastive learning for unpaired image-to-image translation. Computer Vision – ECCV 2020. https://doi.org/10.1007/978-3-030-58545-7_19.

[18] Kang S-R, Shin W, Yang S, Kim J-E, Huh K-H, Lee S-S, et al. Structure-preserving quality improvement of cone beam CT images using contrastive learning. Comput Biol Med 2023;158:106803. https://doi.org/10.1016/j.compbiomed.2023.106803.

[19] Wang J, Yan B, Wu X, Jiang X, Zuo Y, Yang Y. Development of an unsupervised cycle contrastive unpaired translation network for MRI-to-CT synthesis. J Appl Clin Med Phys 2022;23:e13775. https://doi.org/10.1002/acm2.13775.

[20] Jensen K, Friborg J, Hansen CEE, Johansen J, Andersen M, et al. The Danish Head and Neck Cancer Group (DAHANCA) 2020 radiotherapy guidelines. Radiother Oncol 2020;151:149–51. https://doi.org/10.1016/j.radonc.2020.07.037.

[21] Rodet T, Noo F, Defrise M. The cone-beam algorithm of Feldkamp, Davis, and Kress preserves oblique line integrals. Med Phys 2004;31:1972–5. https://doi.org/10.1118/1.1759828.

[22] Li H, Dong L, Zhang L, Yang JN, Gillin MT, Zhu XR. Toward a better understanding of the gamma index: Investigation of parameters with a surface-based distance method. Med Phys 2011;38:6730–41. https://doi.org/10.1118/1.3659707.

[23] Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bull 1945;1:80–3. https://doi.org/10.2307/3001968.

[24] Liu Y, Lei Y, Wang T, Fu Y, Tang X, Curran WJ, et al. CBCT-based synthetic CT generation using generative adversarial networks with disentangled representation. Med Phys 2020;47:2472–83. https://doi.org/10.1002/mp.14121.

[25] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408–19. https://doi.org/10.1002/mp.12155.

[26] Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: A systematic literature review. Comput Biol Med 2023;152:106391. https://doi.org/10.1016/j.compbiomed.2022.106391.
[27] Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. BMC Med Imaging 2022;22:69. https://doi.org/10.1186/s12880-022-00793-7.
[28] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020;63:139–44. https://doi.org/10.1145/3422622.
[29] Xue X, Ding Y, Shi J, Hao X, Li X, Li D, et al. Cone beam CT (CBCT) based synthetic CT generation using deep learning methods for dose calculation of nasopharyngeal carcinoma radiotherapy. Technol Cancer Res Treat 2021;20:15330338211062415. https://doi.org/10.1177/15330338211062415.
[30] Jihong C, Kerun Q, Kaiqiang C, Xiuchun Z, Yimin Z, Penggang B. CBCT-based synthetic CT generated using CycleGAN with HU correction for adaptive radiotherapy of nasopharyngeal carcinoma. Sci Rep 2023;13:6624. https://doi.org/10.1038/s41598-023-33472-w.
[31] 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.
[32] Rusanov B, Hassan GM, Reynolds M, Sabet M, Rowshanfarzad P, Bucknell N, et al. Transformer CycleGAN with uncertainty estimation for CBCT based synthetic CT in adaptive radiotherapy. Phys Med Biol 2024;69:035014. https://doi.org/10.1088/1361-6560/ad1cfc.
[33] Huijben EMC, Terpstra ML, Galapon AJ, Pai S, Thummerer A, Koopmans P, et al. Generating synthetic computed tomography for radiotherapy. SynthRAD2023 Challenge Report 2024. https://doi.org/10.48550/arXiv.2403.08447.
[34] Li X, Jiang Y, Rodriguez-Andina JJ, Luo H, Yin S, Kaynak O. When medical images meet generative adversarial network: recent development and research opportunities. Discov Artif Intell 2021;1:5. https://doi.org/10.1007/s44163-021-00006-0.
[35] Pinaya WHL, Tudosiu P-D, Dafflon J, Da Costa PF, Fernandez V, Nachev P, et al. Brain Imaging Generation with Latent Diffusion Models. Deep Generative Models DGM4MICCAI workshop 2022. https://doi.org/10.1007/978-3-031-18576-2_12.
[36] Peng J, Qiu RLJ, Wynne JF, Chang C-W, Pan S, Wang T, et al. CBCT-Based synthetic CT image generation using conditional denoising diffusion probabilistic model. Med Phys 2024;51:1847–59. https://doi.org/10.1002/mp.16704.