A multi-channel cycleGAN for CBCT to sCT generation

Chelsea A. H. Sargeant[0000−0002−8737−4835], Edward G. A. Henderson[0000−0003−3752−4054], Dónal M. McSweeney[0000−0002−7099−8270], Aaron G. Rankin[0000−0002−0386−1119], and Denis Page[0009−0005−9015−180X]⋆

Division of Cancer Sciences, The University of Manchester
chelsea.sargeant@postgrad.manchester.ac.uk

Abstract. Image synthesis is used to generate synthetic CTs (sCTs) from on-treatment cone-beam CTs (CBCTs) with a view to improving image quality and enabling accurate dose computation to facilitate a CBCT-based adaptive radiotherapy workflow. As this area of research gains momentum, developments in sCT generation methods are difficult to compare due to the lack of large public datasets and sizeable variation in training procedure. To compare and assess the latest advancements in sCT generation, the *SynthRAD2023* challenge provides a public dataset and evaluation framework for both MR and CBCT to sCT synthesis. Our contribution focuses on the second task, CBCT-to-sCT synthesis. By leveraging a multi-channel input to emphasize specific image features, our approach effectively addresses some of the challenges inherent in CBCT imaging, whilst restoring contrast necessary for accurate visualisation of patients anatomy. Additionally, we introduce an auxiliary fusion network to further enhance the fidelity of generated sCT images.

Keywords: synthetic CT · multi-channel cycleGAN · CBCT quality

1 Introduction

In photon and proton radiotherapy, on-treatment cone-beam computed tomography (CBCT) imaging plays a crucial role in detecting anatomical changes and ensuring accurate patient positioning for precise treatment setup [7]. However, CBCT image quality often suffers from artefacts such as scatter, noise and beam hardening, resulting in inferior image quality compared to planning CT (pCT) scans. These artefacts lead to inaccuracies in Hounsfield unit (HU) values, rendering CBCT images unsuitable for accurate dose calculation.

To address these challenges, deep learning-based image synthesis techniques have been successfully employed to enhance CBCT image quality to a level comparable to CT scans. This involves using neural networks to discover a complex mapping between source (CBCT) and target (CT) image domains, enabling the generation of synthetic CT (sCT) images directly from CBCT data [9]. The primary objective is to produce sCT images that retain the anatomical structure

---

⋆ Authors listed in order of contribution.

of the CBCT scans while closely approximating the HU of the pCT images to facilitate accurate dose calculation using on-treatment imaging.

Cycle-consistent generative adversarial networks (cycleGANs) have emerged as the dominant methodology for the conversion of CBCT to CT images for various anatomical sites [1,3,4]. CycleGANs use two generator-discriminator network pairs, trained concurrently, to allow the model to learn the translation function from unpaired images [12]. The generative networks are responsible for converting between the CBCT and CT image domains. One generator produces CT-like output (a sCT) from CBCT input, while the other performs the reverse conversion. The discriminators learn to differentiate between real and synthetic images. Once trained, a single generator is employed for the desired translation, in this context, the CBCT-to-CT generator.

Our approach is an extension of the 2D cycleGAN architecture originally proposed by Zhu *et al.*. The generator architecture has been modified to incorporate long-range skip connections and self-attention layers to improve model performance. Crucially, we separate the input CBCT and CT images into multiple channels, each emphasizing specific image features, and introduce an auxiliary fusion network to recombine the multi-channel predictions. The incorporation of multi-channel information and subsequent fusion of these channels accounts for the diverse variety of input images and contributes to the enhanced fidelity of generated sCT images. Our method is robust to the challenges of CBCT imaging. It effectively mitigates the impact of significant scatter artefacts often encountered in CBCT scans, whilst successfully restoring contrast necessary for accurate visualisation of the soft tissues within the patients. This approach enables the generation of high-quality synthetic CT images from CBCT scans. With these extensions, we aim to produce sCT images of high enough quality to facilitate accurate dose calculations on data acquired from on-treatment CBCT imaging.

2   Method

2.1   Dataset

Imaging data of patients who underwent brain or pelvis radiotherapy were collected from three Dutch institutions: Radboud University Medical Center, University Medical Center Utrecht and University Medical Center Groningen for the *SynthRAD2023* challenge. Here, we focused on the second of two tasks, which involved generating synthetic CTs from CBCTs. The training, validation and testing cohorts were determined prior to the challenge by the organisers, more details can be found in Table 1 in Thummerer *et al.* [11]. A total of 360 CBCT-CT pairs were available, split evenly across the brain and pelvis.

All pre and post-processing was applied to 3D images, however, individual axial slices were extracted from all image volumes (CT and CBCT) for training. A random 80/20 split was used to create train-time training and validation sets. In our case, two separate site-specific models were trained: one for brain

scans and another for pelvis scans. Training and validation splits for each site categorised by center name (A, B, C; as in [11]) are shown in Table 1. For image acquisition parameters, see [11].

Table 1. Train-time training and validation splits. Values represent number of axial slices from each center.

	Brain				*Pelvis*			
	A	B	C	**Total**	A	B	C	**Total**
Training	10,955	8,262	10,434	29,651	4,772	4,111	2,369	11,252
Validation	2,711	2,012	2,690	7,413	1,194	1,056	563	2,813
Total	13,666	10,274	13,124	37,064	5,916	5,167	2932	14,065

2.2   Pre-Processing

Preprocessing was performed by the challenge organisers resulting in 3D CT and CBCT images with corresponding binary masks of the patient. All images of an anatomical region were resampled to the same voxel spacing: $1 \times 1 \times 1\text{mm}^3$ and $1 \times 1 \times 2.5\text{mm}^3$ for the brain and pelvis, respectively. A full description of the preprocessing can be found in [11].

Following the initial preprocessing, images were processed as follows:

- Correction of masks and CBCT image range;
- Overflow correction for pelvis patients;
- Mask application;
- Multi-channel range selection and normalisation;
- Padding/Cropping.

These preprocessing steps were also applied on inference (with the exception of the mask correction). The challenge dataset was curated to represent the variation in a realistic multi-centre setting; this included variation in image protocol, CBCT quality, field-of-view (FOV) and severity of artefacts. To capture as much information as possible, we opted to include all available patients in the training set and handle the variability of the dataset with a multi-channel approach. Below, we present a more detailed description of this pipeline.

Mask and image range correction All provided images and masks were visually inspected to ensure consistency across the training cohort. For a large number of pelvis cases, the masks included large abnormalities resulting in missing regions within patient body and were deemed unsuitable for training. As a result, the masks were regenerated using thresholding and binary dilation and morphological closing functions from *SimpleITK*. The CBCTs from centre A had to be shifted by -1024 HU to ensure the CBCTs from all centres were contained within a consistent range of $[-1024, 3000]$ HU.

Overflow correction Upon visual inspection it was also noticed that CBCTs from 55/60 pelvis patients from Centre C suffered from high intensity artefacts on the surface of the patient that were not seen in the corresponding CT. We believed this to be due to a numerical overflow error occurring during the CBCT image reconstruction. This was corrected prior to training by overwriting the artefact voxels with air (-1024 HU). In detail, a distance transform was used to create a $\sim$ 40mm thick hull around the external contour of the patient to capture high-intensity values ($>$1000 HU) resulting from the overflow error.

Multi-channel input The CT and CBCT scans were normalised in three separate channels using windowing to enhance the contrast of structures within the patient. In the first channel, we use the full width of the image range $[-1024, 3000]$ HU. In channel two, we use a contrast setting used to view soft tissue structures within the region of interest; for the CT this was $[-150, 150]$ HU and $[-100, 100]$ HU for the pelvis and brain, respectively, chosen to be centred on the water peak. Due to a lack of calibration of the CBCT images, both within and among centres, an automated peak finder was implemented. For this we used the $find\_peaks$ function from the $SciPy$ package. The window width remained the same; either $\pm$150 HU and $\pm$100 HU depending on the image site, however, the level varied on a patient basis depending on the peak found within the image intensity histogram as shown in Figure 1. In the final channel, both the CT and CBCT images were clipped to $[600, 3000]$ HU to capture information about the high-density structures within the patient: high-density bone and any metal present due to artificial hips, pins, and dental work, for example. Each channel was then independently scaled to the range $[0, 1]$ using min-max normalisation to improve training stability. Figure 2 depicts the three-channel input for both the CBCT and CT images following range selection and min-max normalisation.

Cropping The three-channel images were padded or cropped to height and width of 448×448 voxels for the pelvis and 304×304 voxels for the brain.

Model input The model input consisted of three-channel CT/CBCT axial slices, paired only by rigid registration.

2.3   Proposed Method

Our method is based on a traditional cycleGAN architecture [12] with some specific modifications for this challenge. CycleGANs consist of two sets of competing generator and discriminator networks trained simultaneously to enable unpaired training. The generators learn to produce CT-like output from CBCT input (a sCT) and vice versa (synthetic CBCT (sCBCT) output from CT input). The discriminators learn to classify CT/sCT and CBCT/sCBCT input as real and fake. Additionally, cycle consistency was enforced by inputting the synthetic image into the reverse generator to reverse the translation back to the original image modality.
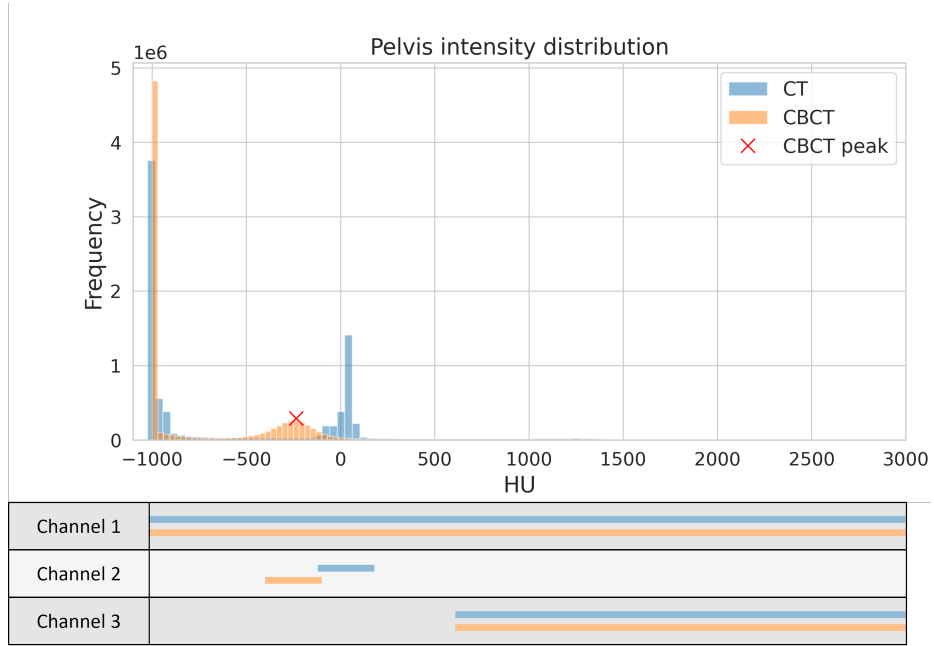
Fig. 1. CBCT and CT image histogram for a pelvis scan. Below the histogram shows the ranges of each of the three input channels. Channel one accounts for the entire image. Channel two captures soft tissue information by centering the channel on either the water peak in the case of the CT image or on a peak found on a per-patient basis on the CBCT. In either case a window of ± 150 was used. High-density structures, such as certain bones or artificial metal implants, were accounted for in the third channel.

Disciminator network The discriminator used in our modified cycleGAN was standard and directly adapted from the PatchGAN [2]. The total loss for the discriminators is

$$\mathcal{L}_D = \mathcal{L}_{D_{CT}} + \mathcal{L}_{D_{CBCT}} \tag{1}$$

where D_{CT} is the discriminator aiming to classify sCT from pCT and D_{CBCT} is the discriminator aiming to classify sCBCT from CBCT. The loss function used was binary cross-entropy.

Generator network The generator architecture used for our cycleGAN was based on the convolutional neural network detailed by Zhu *et al.* [12]. This has three convolutional layers with pooling, followed by nine residual convolutional blocks and then three resize convolutional layers, described by Odena *et al.*, to re-establish the original spatial dimensions [6]. We added UNet-like long-range skip connections between corresponding down- and up-sampling convolutional levels to preserve contextual information [8]. Finally, we added attention gates to the skip connections to emphasize salient features propagated forward from earlier in the network [10].
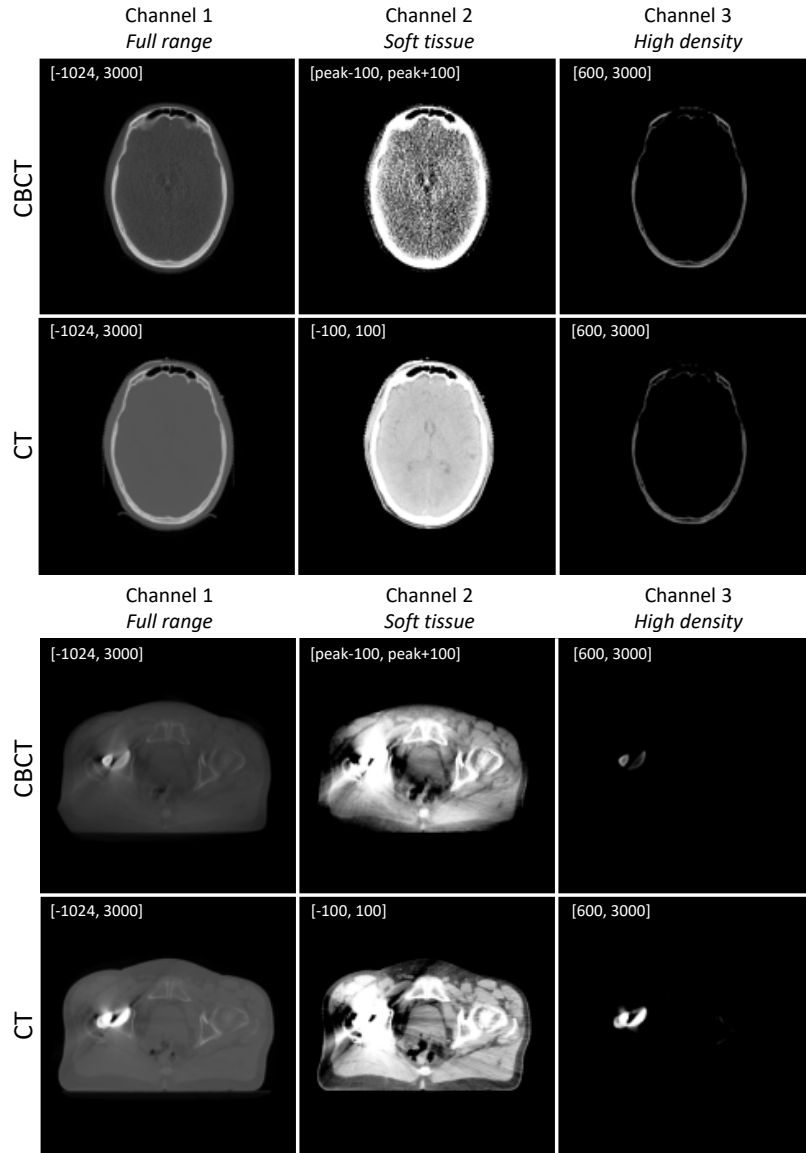
Fig. 2. Three channel input for the brain and pelvis model. Channel one captures the full image window, channel two is the soft tissue window and channel three captures high-density structures such as high-density bone in the brain patient and metal implants in the pelvis patient displayed. The range is reported in HU, and each channel is independently normalised to [0, 1].

The generator receives three-channel input and produces three-channel output. This is performed for ease of implementation of the cycle-consistency aspect, where the output of each generator is also sent through the reverse generator.

Fusion network An additional fusion network was added onto the cycleGAN to fuse the output three channels from the CBCT-sCT generator into a single-channel greyscale image. This final network had an identical architecture to the generators but contained only a single residual block, and short-range residual connections across all convolutional layers as in Milletari *et al.* to speed up convergence [5].

A mean square error loss term was added to train the fusion network to maximise the similarity between the output single-channel image and the CT.

The total loss for the generators is

$$
\begin{aligned}
\mathcal{L}_G = {} & \mathcal{L}_{G_{\text{GAN-sCT}}} + \mathcal{L}_{G_{\text{GAN-sCBCT}}} \\
& + \alpha(\mathcal{L}_{G_{\text{cycle-CT}}} + \mathcal{L}_{G_{\text{cycle-CBCT}}}) \\
& + \beta(\mathcal{L}_{G_{\text{identity-CT}}} + \mathcal{L}_{G_{\text{identity-CBCT}}}) \\
& + \mathcal{L}_{G_{\text{fusion-sCT}}}
\end{aligned}
\tag{2}
$$

where $\alpha = 10$ and $\beta = 5$. Each generator loss was calculated using the mean squared error. The $\mathcal{L}_{G_{\text{GAN}}}$ loss terms reward the generator for deceiving the discriminator. The $\mathcal{L}_{G_{\text{cycle}}}$ loss terms, or consistency loss, are performed between CT and sCT (or CBCT and sCBCT) pairs where the synthetic image has been passed through both generators in a cycle. $\mathcal{L}_{G_{\text{identity}}}$ are the identity losses which is calculated between the real images (CT or CBCT) and the same real image after it has been passed through the opposite generator (CBCT-to-sCT or CT-to-sCBCT respectively) to enforce an identity constraint.

3 Implementation Details

The model was implemented in PyTorch Lightning version 1.5.10 and two networks were trained independently for each anatomical site. Each model was trained using a single NVidia GeForce RTX 3090 GPU with 24GB VRAM. The training protocols used to train our brain and pelvis cycleGANs are shown in Table 2.

3.1 Post-Processing / Channel Fusion

Upon inference, the trained three-channel CBCT-to-sCT generator network was applied to preprocessed CBCTs (as described in 2.2) to generate a three-channel sCT. After a forward pass of the CBCT-to-sCT generator, the three-channel sCT was input into the fusion network to generate a single-channel fused sCT.

The padding or cropping performed during preprocessing was reversed to restore the sCTs to the original CBCT image size. Next, the HU intensity range of each channel was restored using a linear re-scaling: the fused sCT and channel one of the three-channel sCT were rescaled to [-1024, 3000] HU; channel two was re-scaled to [-100, 100] or [-150, 150] HU for the brain and pelvis, respectively, and channel three was re-scaled to [600, 3000] HU.
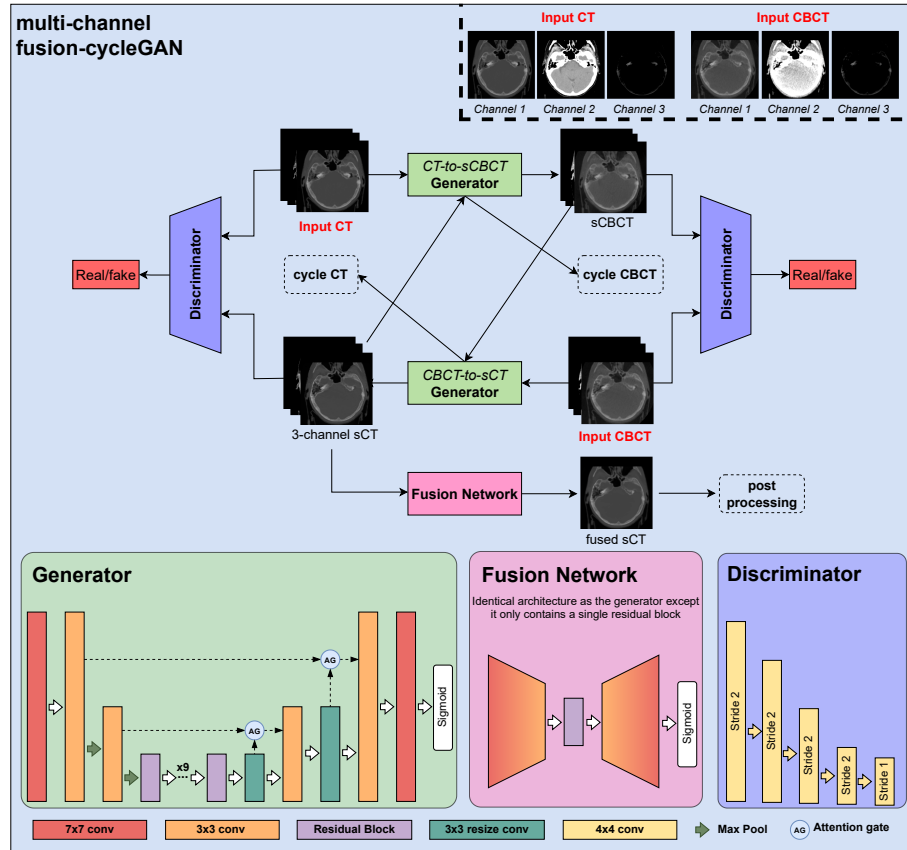
Fig. 3. The cycleGAN architecture is used to convert between CBCT and CT image domains, with an additional fusion network following CBCT-to-CT synthesis. Three-channel CBCT and CT images are input into the cycleGAN generators to produce three-channel synthetic CT (sCT) and CBCT output (sCBCT). The fusion network is only applied to the sCT images to recombine channels into a single greyscale sCT.

Given the three-channel output from our adapted cycleGAN and the single channel from the fusion bolt-on network, two methods of combining the multi-channel information were explored in preliminary testing. Initially, the fusion network output alone was evaluated. Secondly, a post-processing approach was explored to manually combine the three-channel output into a single-channel image. This performed by inserting the second and third (soft tissue and high-density) channels into a full window range output. We tested using both the fused sCT or the first channel sCT from the adapted cycleGAN as the reference full-width intensity range for this approach.

Table 2. Training protocols.

Data augmentation	none
Batch size	1
Maximum epochs	200
Optimizer	Adam
Initial learning rates	Generator: 0.0001 Discriminator: 0.0002
Learning rate decay schedule	After 5 epochs, decay to 80% of learning rate every 2 epochs
Stopping criteria, and optimal model selection criteria	Early stopping when the total generator validation loss does not improve for 20 epochs. Optimal model is chosen based on best image similarity metrics calculated on train-time validation data
Loss functions	Mean-squared error on the generators and binary cross-entropy on the discriminators.
Training time	Brain: $\sim$ 9 hours per epoch Pelvis: $\sim$ 3 hours per epoch

3.2 Checkpoint Selection

To select the best model, the top five epochs were saved during training according to the total generator validation loss (Eq. 2).

Independently, the validation data is loaded and image similarity metrics (listed in section 4) are calculated between all slices of the sCT and corresponding CT for each checkpoint. The metrics are averaged across the validation set and ranked according to performance. We select the top-performing model across image similarity metrics.

4   Evaluation

For the validation phase, the generated sCTs are assessed according to image similarity metrics:

- Mean absolute error (MAE)
- Structural similarity index (SSIM)
- Peak signal-to-noise ratio (PSNR)

5   Results

5.1   sCT fusion

In preliminary testing, we evaluated the direct sCT output from the fusion network. While the direct fused output produced sCTs in a reasonable HU intensity

range, the image quality was fairly poor; the soft tissue contrast was subpar and the network failed to reproduce regions of air as shown in Figure 4. As a result, the three channels output by the CBCT-to-CT generator were recombined manually. However, as the cycleGAN generator and fusion network were trained end-to-end, we still anticipate that the fusion network may have been a useful meta-learning task which could have aided the model to leverage all three channels by learning their true grey-level relations.

For the pelvis patients, we used channel one of the three-channel output of our model (channel one sCT) as the full range image and reinserted the second and third channels. For the brain, we used the output from the fusion network (fused sCT) as our base image.
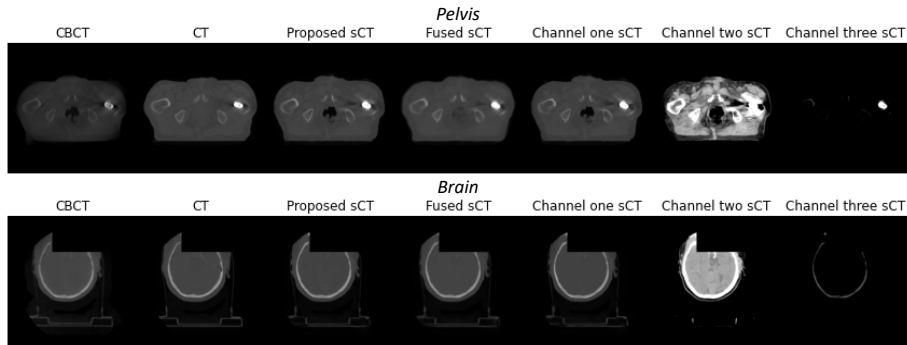

Fig. 4. All available synthetic images; the proposed sCT, fused sCT, and individual channel sCTs, with input CBCT and ground truth CT. These images were generated during train-time validation. The first five images are displayed in HU range [-1024, 3000], while channel three is displayed in the range [600, 3000] and the channel two images are displayed in HU ranges [-150, 150] and [-100, 100] for the pelvis and brain, respectively.

5.2   Results for challenge validation set

The sCTs generated by our method are displayed in Figure 6 for a subset of patients for each anatomical site alongside the original CBCT. Visually, the methods preserve the anatomical structure of the CBCT images and the contrast in soft tissue is enhanced for both pelvis and brain patients. For the brain, high-quality sCTs are generated despite the significant decrease in quality of CBCTs from center B.

Observed on the CBCT for pelvis patient 2PB069 in Figure 6 are significant artefacts on the axial CBCT slice that have been considerably suppressed by our method. Whilst we see a large reduction in streaking artefacts, improving visualisation of internal anatomy, the artefacts appear to have influenced the sCT body contour. This is also noticed for patient 2PA033.

Our method is also robust to scatter artefacts caused by high-density features such as metal hip implants. The CBCT and sCT for a pelvis patient with an artificial hip are shown in Figure 7, demonstrating accurate presentation of the metal and reduction in streaking in the nearby tissue.

Table 3 and Figure 5 summarise the quantitative results obtained by image similarity metrics; MAE, PSNR and SSIM. For both anatomical sites, our proposed methods achieve the highest performance on patients from center C, with reduced MAE and increased PSNR and SSIM. Whilst all centers suffer from outliers, our model performs least consistently for patients from center A.

Table 3. Quantitative comparison between the sCTs and ground truth, the pCT.

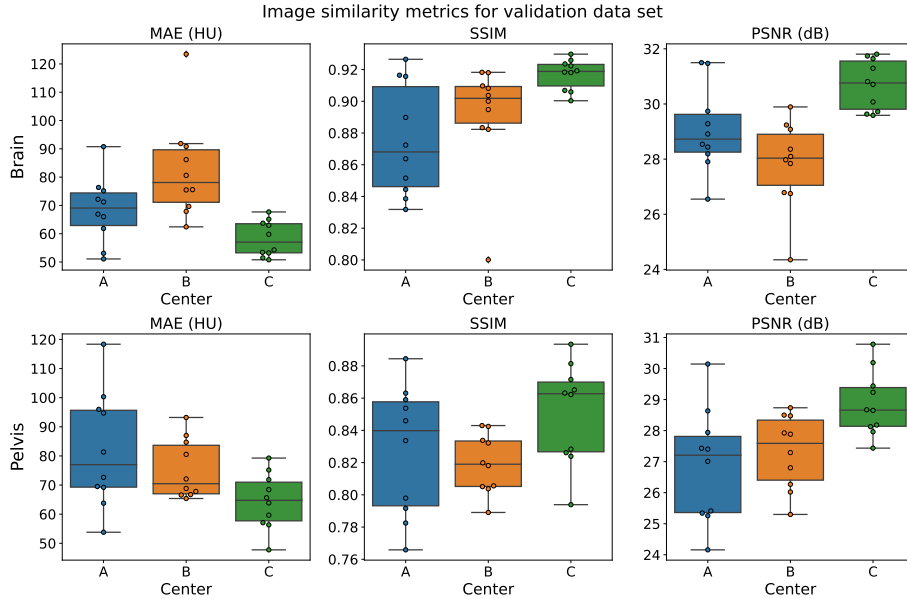| Anatomical Site	MAE (HU)	PSNR (dB)	SSIM
Brain	69.70 ± 15.54	29.20 ± 1.77	0.89 ± 0.03
Pelvis	73.91 ± 15.13	27.69 ± 1.57	0.83 ± 0.03
Full Dataset	71.83 ± 15.0	28.44 ± 1.85	0.86 ± 0.05

Fig. 5. Image similarity metrics; mean absolute error (MAE), structural similarity index measure (SSIM) and peak signal-noise ratio (PSNR) for our submitted methods.
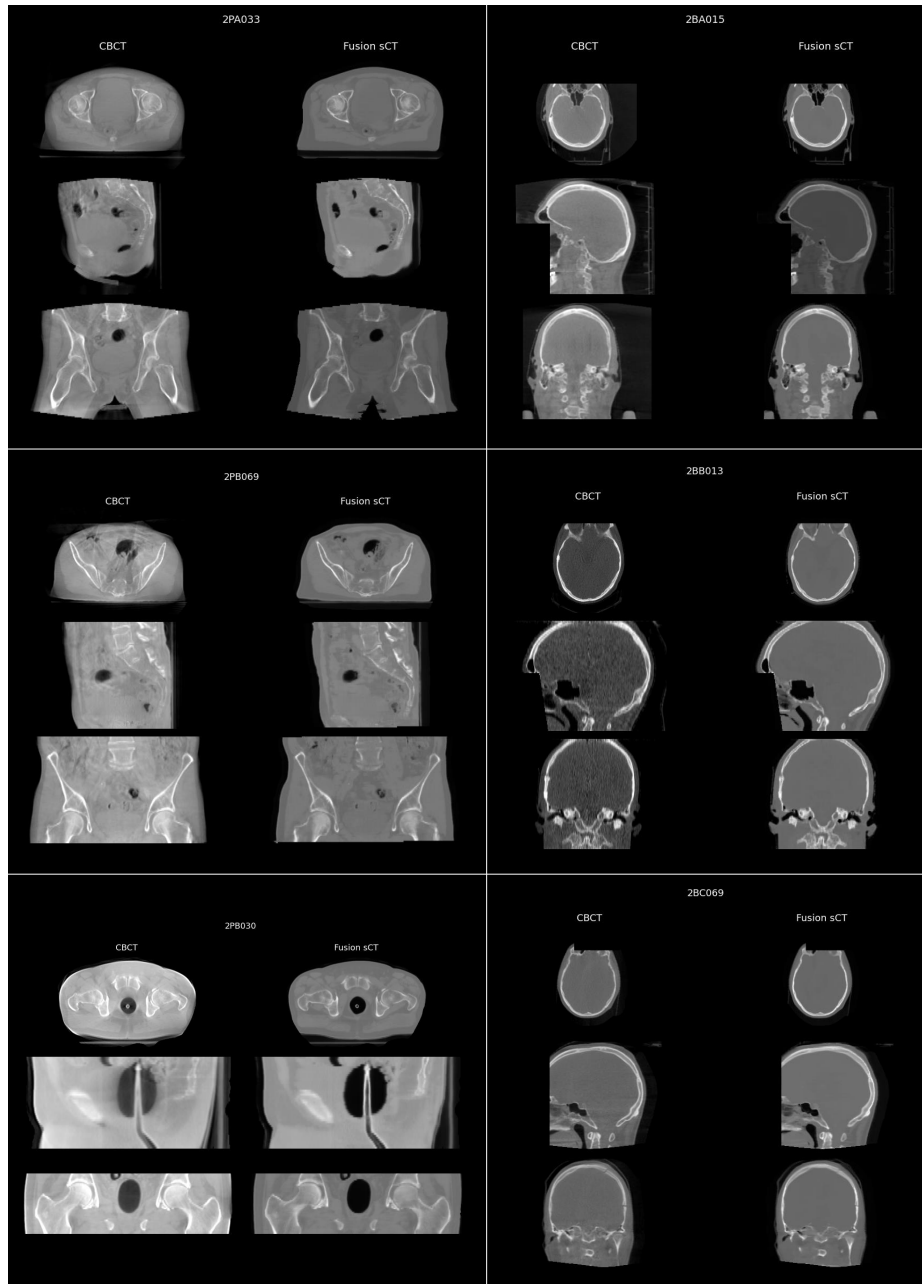
Fig. 6. CBCT and sCT generated by our methods. Pelvis and brain results are shown with HU scale of [-400,1200] and [250, 2500], respectively.
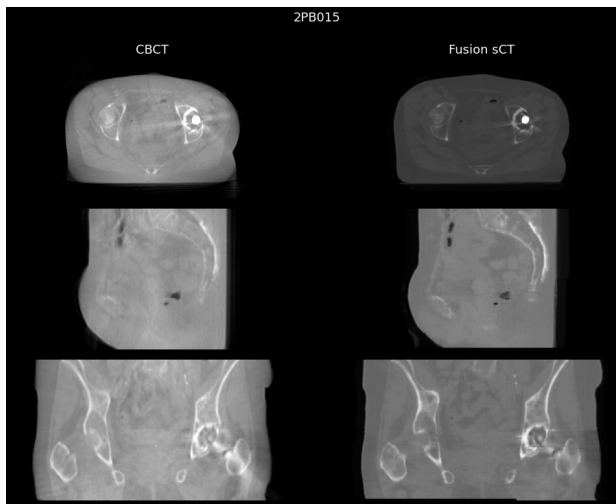
Fig. 7. CBCT and sCT generated by our methods for example pelvis patient with metal implant in the hip. Images are shown with HU scale of [-400,1200] and [250, 2500], respectively.

6   Discussion and Conclusion

Our approach demonstrates the successful generation of high-quality synthetic CT images from CBCT scans. The incorporation of image data at varying window/levels within a concatenated multi-channel input has allowed us to effectively address several challenges inherent in CBCT imaging. We observe notable improvements in the preservation of fine soft tissue details, suppression of artefacts such as streaking, and accurate representation of high-density features within the patient anatomy. Our reported image similarity metrics on the validation data set, an average mean absolute error of 71.83 ± 15.6 HU, PSNR 28.44 ± 1.85 dB and SSIM of 0.86 ± 0.05, indicate good similarity to the ground truth images, CTs registered by non-rigid image registration to the input CBCTs.

Based on current validation metrics, we have observed that sCTs generated from center C outperform those from both center A and B for both brain and pelvis patients, despite center C having a reduced number of slices (approximately half) in the pelvis training set. Specifically, the MAE is 64.51 ± 9.11 HU for center C, compared to 82.21 ± 18.71 and 75.19 ± 9.62 for centers A and B, respectively. This outcome suggests that our model generalises effectively between centers, and we attribute the discrepancy partially to the fact that images from center C appear cropped, particularly in regions where higher errors are observed for other centers - the end slices. Since these slices lack clinical significance, the performance of center C is more indicative of our model effectiveness.

We chose to exploit a multi-channel input to capture specific information about both the CBCT and CT images. The first channel spans the full image

range for simplicity upon training, while the second channel focuses on narrow window widths to enhance soft tissue contrast. For high-intensity features and dense bone, the third channel is employed. Our investigations into training with different channel widths indicate a sensitivity of learned features to window width, necessitating further research into the optimal window/level for each channel as well as channel quantity.

Our fusion methods differ between brain and pelvis sites. In the case of brain images, the fused sCT can be used in the recombination of the three channels. However, it negatively affects the quality of the sCT images for pelvis patients. The fusion network architecture is modelled after the generator network in the cycleGAN. In this setup, the output is compared to a CT image registered only through rigid image registration. Since pelvis patients' anatomy can change unpredictably over even short time intervals, the semi-paired nature of our training could be inappropriate. On the other hand, brain patients experience fewer such changes, allowing the fusion network to perform well despite the absence of deformable image registration between the CBCT and CT images. While cycle-GANs are designed to operate in an unsupervised manner with unpaired data, there is an expectation that incorporating non-rigid image registration into the training process could lead to an overall performance improvement. Our future efforts will concentrate on refining the fusion process, including optimizing the design of the fusion network architecture

CBCT to sCT image synthesis holds the potential to enhance visualization and improve dose calculations for adaptive radiotherapy [9]. Our method succeeds in generating high-quality synthetic CT images from on-treatment CBCTs. We demonstrate that by incorporating multiple window/level inputs and effectively fusing them, we can enhance the quality of the resulting sCT images.

References

1. Harms, J., Lei, Y., Wang, T., Zhang, R., Zhou, J., Tang, X., Curran, W.J., Liu, T., Yang, X.: Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Medical Physics **46**(9), 3998–4009 (Jul 2019). `https://doi.org/10.1002/mp.13656`
2. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (Jul 2017). `https://doi.org/10.1109/cvpr.2017.632`

3. Kurz, C., Maspero, M., Savenije, M.H.F., Landry, G., Kamp, F., Pinto, M., Li, M., Parodi, K., Belka, C., van den Berg, C.A.T.: CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. Physics in Medicine & Biology **64**(22), 225004 (Nov 2019). `https://doi.org/10.1088/1361-6560/ab4d8c`

4. Maspero, M., Houweling, A.C., Savenije, M.H., van Heijst, T.C., Verhoeff, J.J., Kotte, A.N., van den Berg, C.A.: A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. Physics and Imaging in Radiation Oncology **14**, 24–31 (Apr 2020). `https://doi.org/10.1016/j.phro.2020.04.002`

5. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE (10 2016). `https://doi.org/10.1109/3dv.2016.79`

6. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. Distill **1**(10) (Oct 2016). `https://doi.org/10.23915/distill.00003`

7. Oldham, M., Létourneau, D., Watt, L., Hugo, G., Yan, D., Lockman, D., Kim, L.H., Chen, P.Y., Martinez, A., Wong, J.W.: Cone-beam-CT guided radiation therapy: A model for on-line application. Radiotherapy and Oncology **75**(3), 271.E1–271.E8 (Jun 2005). `https://doi.org/10.1016/j.radonc.2005.03.026`

8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Lecture Notes in Computer Science, pp. 234–241. Springer International Publishing (2015). `https://doi.org/10.1007/978-3-319-24574-4_28`

9. Rusanov, B., Hassan, G.M., Reynolds, M., Sabet, M., Kendrick, J., Rowshanfarzad, P., Ebert, M.: Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: A systematic review. Medical Physics **49**(9), 6019–6054 (Jul 2022). `https://doi.org/10.1002/mp.15840`

10. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis **53**, 197–207 (Apr 2019). `https://doi.org/10.1016/j.media.2019.01.012`

11. Thummerer, A., van der Bijl, E., Galapon, A., Verhoeff, J.J.C., Langendijk, J.A., Both, S., van den Berg, C.N.A.T., Maspero, M.: SynthRAD2023 grand challenge dataset: Generating synthetic CT for radiotherapy. Medical Physics **50**(7), 4664–4674 (Jun 2023). `https://doi.org/10.1002/mp.16529`

12. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE (Oct 2017). `https://doi.org/10.1109/iccv.2017.244`