Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation of CBCT-based HU correction for synthetic CT generation in head-and-neck radiotherapy: FOV-stratified and peripheral-region analysis

Khashayar Heshmati Jannat Magham
Ferdowsi University of Mashhad
Laleh Rafat-Motavalli
rafat@um.ac.ir

Ferdowsi University of Mashhad
Hashem Miri-Hakimabad
Ferdowsi University of Mashhad
Mahdieh Dayyani
Reza Radiotherapy and Oncology Center
Mohammad Mohammadi
Royal Adelaide Hospital

Research Article

Keywords: Residual U-Net, synthetic CT, cone-beam CT, HU correction, FOV truncation, adaptive radiotherapy

Posted Date: September 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10791575/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Cone-beam computed tomography (CBCT) is essential for image-guided radiotherapy for head and neck (H&N) cancer but suffers from inaccurate Hounsfield units (HUs) due to scatter, artifacts, and field of view (FOV) truncation, limiting its use for dose calculation. We aimed to compare linear correction, a residual U-Net model (ResUNet), and a paired cycle-consistent conditional GAN (PCC-cGAN) framework with residual U-Net generators for HU correction and synthetic CT (sCT) generation, the deep learning models trained in a paired setting, with the CycleGAN variant adapted from its original unpaired design to exploit cycle-consistency in addition to supervised training. Our evaluation emphasizes FOV stratification (complete vs. truncated) and spatial localization of residual errors. This study provides an FOV-stratified and peripheral-aware evaluation of CBCT-based HU correction, highlighting residual error patterns that may be overlooked by global MAE analysis alone.

Methods

A total of 100 CBCT/CT pairs from 67 H&N patients were divided into complete FOV (19 pairs) and truncated FOV (81 pairs) groups. Evaluation was performed within three spatial mask regions: intersection, eroded intersection, and peripheral. An independent-patient sensitivity analysis was conducted to assess patient-level overlap. Preliminary body-level dosimetric consistency was evaluated by comparing raw CBCT and ResUNet sCT against the reference planning CT (pCT) dose using gamma pass rates (2%/2mm and 3%/3mm) and dose Mean Absolute Error & Root Mean Squared Error (MAE/RMSE).

Results

ResUNet achieved the lowest HU-MAE across the evaluated regions in both FOV groups. In the independent-patient subset, this pattern was preserved for intersection and eroded mask regions, with ResUNet performing best in nearly all of the patient-region combinations. The peripheral MAE remained substantially higher than the intersection and eroded MAE for all methods (e.g., ResUNet: 171.51 vs. 76.93 HU in the complete FOV). Body-level gamma pass rates (GPR) and dose metrics were generally maintained or modestly improved by ResUNet, but HU improvements did not translate proportionally to dosimetric gains across all patients.

Conclusions

ResUNet provided the lowest HU-MAE among the evaluated methods under the conditions of this study for CBCT-based sCT generation in H&N radiotherapy, with persistent residual errors concentrated at the

peripheral evaluation region, where the highest residual MAE was observed. The FOV-stratified and peripheral-aware evaluation framework introduced here informs future DVH-based validation and dose-of-the-day adaptive workflows.

1. Background

Head and neck (H&N) cancer patients undergoing radiotherapy frequently experience anatomical changes, weight loss, tumor shrinkage, and soft tissue deformation, that can alter target and organ at risk (OAR) relationships relative to the planning CT (pCT), potentially compromising dose delivery (1, 2). Cone-beam computed tomography (CBCT) has become an integral component of image-guided radiation therapy (IGRT) for H&N cancer in advanced radiotherapy, enabling daily visualization of patient anatomy directly on a linear accelerator when the patient is set up for treatment through medical linear accelerator (3). However, despite its clinical utility for patient positioning, compared to pCT, raw CBCT suffers from scatter, beam hardening, artifacts, and elevated noise, resulting in unreliable Hounsfield unit (HU) that preclude accurate dose calculation (4, 5). Additionally, finite detector size in standard C-arm Linac systems (e.g., Varian) often causes field of view (FOV) truncation, clipping shoulders or peripheral anatomy and introducing further artifacts (6). These limitations motivate synthetic CT (sCT) generation to convert CBCT into CT-equivalent images with accurate HUs while preserving daily anatomy (7).

To address the limitations of CBCT for accurate dose calculation, sCT generation has emerged as a promising strategy in adaptive radiotherapy (8–10). The primary objective is to preserve patient-specific anatomy from CBCT while recovering CT-accurate HU values and electron density information (11). Early approaches relied on physics-based scatter correction techniques, image-domain corrections, or deformable image registration (DIR) between pCT and CBCT to estimate CT-equivalent images (12), though performance was limited by registration accuracy, residual artifacts, and anatomical changes (13). Deep learning has since advanced the field substantially: U-Net and residual U-Net improved HU accuracy and anatomical consistency (14), followed by GAN-based frameworks including conditional GANs (cGAN) and CycleGANs, capable of learning CBCT-to-CT mappings even without paired training data (15, 16). Most recently, diffusion and flow-based models have further enhanced image fidelity, structural preservation, and uncertainty modeling (17, 18) and benchmarks like SynthRAD2023 provide standardized comparisons across sites (19). These developments establish sCT generation as a central research direction for CBCT-guided adaptive radiotherapy (20).

Many existing studies on CBCT-based sCT generation report their results on pooled datasets without explicitly distinguishing between complete and truncated FOV acquisitions. However, in H&N radiotherapy, FOV truncation, which most commonly involves the shoulders being clipped outside the scanned volume, is a real and frequent clinical challenge, particularly in standard C-arm Linac systems with a limited detector size (6). Evaluating sCT methods exclusively on a complete FOV or carefully selected cases may overestimate performance and fail to address the conditions encountered in routine clinical practice.

Beyond the choice of correction method, a second gap in the literature concerns how HU accuracy is evaluated. Many CBCT to sCT studies report global image quality metrics such as mean absolute error (MAE), peak signal-to-noise ratio (PSNR), or structural similarity index (SSIM), which provide an overall assessment of performance. Although some studies have complemented these metrics with tissue-specific or region of interest (ROI)-based analyses, the spatial distribution of HU errors relative to valid reconstruction regions and FOV boundaries has received comparatively less attention (8, 21). This issue is particularly relevant for narrow- or truncated-FOV CBCT, where incomplete projection data and truncation artifacts can introduce spatially varying errors within the reconstructed volume. Previous studies have primarily focused on mitigating truncation effects or addressing their impact on downstream tasks such as image registration and adaptive radiotherapy. For example, Zhen et al. demonstrated that truncation artifacts and missing anatomy outside the CBCT FOV can impair DIR accuracy, particularly in regions close to the truncated boundaries, posing challenges for adaptive radiotherapy workflows (22). Similarly, Aootaphao et al. reported that truncated projection data introduce reconstruction artifacts and reduce image quality within the reconstructed FOV, motivating the development of truncation-correction techniques (23). However, these studies did not characterize residual HU errors according to spatial validity regions within the FOV. Recent work has addressed FOV truncation from a complementary perspective. Spinat et al. proposed a diffusion-based CBCT-to-CT translation framework with an extended FOV, in which missing peripheral anatomy outside the original CBCT FOV is inpainted using information from the deformably registered planning CT (24). In contrast, the present study evaluates HU correction only within the valid shared CT/CBCT region and does not reconstruct missing anatomy outside the original CBCT FOV.

The aim of this study was to evaluate CBCT-based HU correction for sCT generation in H&N radiotherapy using FOV-stratified and peripheral-aware analysis rather than relying only on global image-level accuracy metrics. CBCT is attractive for dose-related workflows in adaptive radiotherapy due to its availability on the treatment machine; however, HU inaccuracy and FOV truncation remain major barriers to its clinical use for dose calculation. Reporting only overall MAE or other global metrics may obscure spatially varying residual errors that are clinically relevant, particularly near peripheral regions like FOV boundaries and tissue interfaces. To address this gap, we systematically compared three HU correction methods (linear HU correction, a paired cycle-consistent cGAN-style residual U-Net, and a ResUNet) against a raw CBCT baseline. All evaluations were performed separately for the complete FOV and the truncated FOV subgroups to account for the clinical reality of FOV truncation in standard C-arm Linac systems. Rather than reporting a single global HU error, we stratified the analysis into three spatial mask regions (intersection, eroded intersection, and peripheral) to isolate whether residual errors persist at the peripheral evaluation region after correction. To address the concern of patient-level overlap between training and test sets (due to some patients contributing multiple CBCT/CT pairs from repeated CTs (reCTs)), we performed an independent-patient sensitivity analysis on a subset of test cases with no patient overlap. Finally, a preliminary body-level dose and gamma analysis was conducted to assess whether improvements in HU accuracy translate consistently into dosimetric agreement with the pCT. This study focuses exclusively on HU correction within the valid shared CBCT/CT anatomical region and

does not attempt to reconstruct anatomy outside the original CBCT FOV. Additionally, this work should be interpreted as a methodological imaging evaluation, not as a full clinical validation of adaptive radiotherapy or dose-of-the-day decision-making.

2. Methods

This section describes the patient cohort, imaging data, preprocessing pipeline, HU correction methods, evaluation framework, and preliminary dosimetric analysis used in this study. All codes used in this study were developed and executed with Python (version 3.10) and key libraries of TensorFlow (2.10.1), Keras (2.10.0), scikit-learn (1.7.1), SimpleITK (2.5.2), and SciPy (1.15.3), with CUDA support via cudatoolkit 11.3.1 and cuDNN 8.9.2.

2.1 Study cohort and FOV groups

This study was conducted on a retrospectively collected dataset of H&N cancer patients treated at the Reza Radiotherapy & Oncology Center (RROC), located at Mashhad, Iran, between 2023 and 2026. The clinical site investigated in this study was H&N cancer, a region particularly susceptible to anatomical changes during radiotherapy and challenging for CBCT-based imaging due to the proximity of the shoulders and the use of immobilization masks.

The study cohort consisted of 100 CBCT/CT image pairs derived from 67 unique H&N cancer patients. To explicitly investigate how FOV limitations affect sCT generation and the spatial distribution of residual errors, the cohort was divided into two distinct groups based on the CBCT FOV:

- Complete FOV group: 19 CBCT acquisitions where the relevant anatomical volume was fully contained within the scan field.
- Truncated FOV group: 81 CBCT acquisitions exhibiting FOV truncation, a common occurrence in standard C-arm Linac systems where the shoulders or peripheral anatomy are clipped.

The division allows us to systematically evaluate whether residual errors after HU correction, particularly at the peripheral evaluation region, differ between complete and truncated cases, and whether the relationship between HU accuracy and gamma pass rates (GPRs) is affected by FOV truncation.

All pCT images were acquired using a Siemens SOMATOM Definition AS scanner with a tube voltage of 120 kVp, in-plane pixel spacing of 0.9766 × 0.9766 mm^2, slice thickness of 3.0 mm, and a matrix size of 512 × 512 pixels.

CBCT images were acquired using the On-Board Imager (OBI) system integrated into a linear accelerator (Varian Medical Systems), with a tube current of 15 mA and a 512 × 512 acquisition matrix. Based on the FOV status, acquisitions were divided into two groups: complete FOV and truncated FOV. The complete FOV acquisitions were obtained at 125 kVp with a voxel spacing of 0.9080 × 0.9080 × 1.9885 mm^3. The

truncated FOV acquisitions were obtained at 80 and 100 kVp with a voxel spacing of 0.5112 × 0.5112 × 1.9897 mm^3; among these, three acquisitions were at 80 kVp and the remaining 78 at 100 kVp.

The dataset was randomly divided into training, validation, and test sets at the image-pair level. For the truncated FOV group (81 acquisitions), the split resulted in 58 training pairs, 11 validation pairs, and 12 test pairs. For the complete FOV group (19 acquisitions), the split resulted in 13 training pairs, 3 validation pairs, and 3 test pairs.

It should be noted that some patients contributed more than one CBCT/CT pair to the dataset. This occurred when a patient underwent a reCT during their treatment course, followed by additional CBCT acquisitions registered to the new reCT. Consequently, two pairs originated from some patients. Specifically, among the 12 test pairs in the truncated FOV group, 6 cases were derived from patients who also had a pair present in the training or validation set. Similarly, 1 of the 3 test pairs in the complete FOV group originated from such a patient. Thus, while the initial split was performed at the image-pair level, to account for the effect of patient overlap, an independent-patient sensitivity analysis was conducted. This analysis presents the key results only on a subset of the test data consisting exclusively of patients whose pairs did not appear in the training or validation sets (i.e., no patient overlaps between training/validation and testing). The results from this analysis were compared against the primary results to verify the generalizability of our findings.

It is important to clarify that for truncated FOV cases, the sCT images were not generated as a complete reconstruction of the missing anatomy (e.g., shoulders or peripheral regions). Instead, sCT generation and all subsequent evaluations were performed only within the common valid volume where both the original CBCT and the pCT provided reliable information. Consequently, assessing HU correction in truncated FOV cases did not imply anatomical completion, but rather reflected performance strictly within the overlapping, non-truncated region of the CT/CBCT pair. Given the limited number of test cases, particularly in the complete FOV subgroup, the results were interpreted descriptively and no formal hypothesis testing was performed.

2.2 Image preprocessing and spatial harmonization

All pCT and CBCT volumes underwent a standardized preprocessing pipeline. DICOM data were de-identified and converted to compressed NIfTI format. To ensure uniform spatial resolution, CT images were resampled to a consistent voxel spacing of 1×1×3 mm^3 using linear interpolation. Each CBCT image was then rigidly registered to its corresponding CT image using Elastix (via SimpleITK) to correct for patient setup differences while preserving anatomy (25).

Treatment beds and couch structures were removed using a predominantly automated geometry-driven pipeline involving thresholding, morphological operations, and connected component analysis, which requires only a single patient-specific cropping parameter. Finally, all the images were harmonized to a common FOV by cropping to the maximum shared spatial extent and padding the exterior with air (–

1000 HU). All the CBCT and CT images underwent another series of preprocessing steps to ensure compatibility for deep learning model training.

2.2.1 Resizing

All the volumes in the dataset originally had a spatial dimension of 64 × 512 × 512 (number of slices × height × width). However, due to computational memory constraints during model training, each volume was downsampled to a uniform size of 64 × 256 × 256. Linear interpolation (via SimpleITK's ResampleImageFilter, sitkLinear) was used for image resizing, as no mask existed at this stage (26). This resizing step reduced the memory footprint while preserving the essential anatomical structures necessary for HU correction.

2.2.2 HU clipping and normalization

To reduce the impact of extreme HU values and to stabilize network training, all pCT and CBCT images were first clipped to a fixed HU window. Based on the typical HU range of human tissues in radiotherapy, the clipping range was set from − 1000 HU (air) to 3000 HU (dense bone). Following clipping, images were normalized to the [0, 1] range using min-max normalization:

$$I_{norm} = \frac{I_{clipped} - (-1000)}{3000 - (-1000)} = \frac{I_{clipped} + 1000}{4000}$$

where $I_{clipped}$ represents the image after HU clipping. This normalization scheme maps air to approximately 0 and dense bone to 1, with soft tissues falling in the intermediate range.

2.2.3 Body mask generation

To focus the loss computation on relevant anatomical regions and ignore background air (which contains no clinical information), a body mask was generated for each slice. The mask was derived from the normalized CT and CBCT images using a simple thresholding and hole-filling procedure. Specifically, for each axial slice of the normalized CT image (values in [0,1]), pixels with intensities greater than 0.01 were initially identified as candidate body regions. Following thresholding, morphological hole filling was applied to fill any cavities within the body contour (e.g., air pockets in the trachea or sinuses), ensuring that internal soft tissue structures were included in the mask.

For the logical OR mask configuration used in this step, the final body mask was computed as the element-wise logical OR of the CT-derived mask and the CBCT-derived mask (union mask). This approach ensured that the loss was computed only on pixels that contained anatomical information in either modality, effectively ignoring background air while including soft tissue and bone regions.

2.2.4 2D slice extraction

After preprocessing, each volume was converted into a set of 2D axial slices. Each slice became an independent training sample with shape (1, 256, 256). The corresponding CT slice served as the ground

truth, and the corresponding mask slice defined the region of interest for loss computation. This slice-wise approach enables 2D convolutional neural network training with a manageable memory footprint.

2.3 HU correction and sCT generation methods

To establish a hierarchy of correction complexity and isolate the contributions of different modeling choices, we compared three distinct approaches for converting CBCT to sCT alongside a baseline using raw uncorrected CBCT.

2.3.1 Raw CBCT baseline

Raw CBCT intensities were used directly as a baseline without any correction. This allowed us to quantify the extent of HU inaccuracy present in the original acquisitions and to serve as a lower performance bound for all correction methods.

2.3.2 Linear correction

As a simple yet interpretable baseline, a voxel-wise linear calibration model was fitted to map raw CBCT intensities to CT HUs. The model assumed the following form:

$$CT = a \times CBCT + b$$

Importantly, separate coefficients (a, b) were estimated independently for the complete FOV and truncated FOV groups, as the relationship between CBCT and CT intensities may differ systematically. To fit these coefficients, voxel pairs from all training CT/CBCT volumes within each FOV group were aggregated separately, filtering out background voxels where CBCT intensity was below − 900 HU (i.e., air). All available voxel pairs from the training and validation dataset were used to fit the linear calibration model. Ordinary least squares linear regression was used to estimate a and b for each group (from the sklearn library) (27).

The linear calibration coefficients after fitting on the training data were a = 0.7416 and b = 15.82 HU for the complete FOV and a = 0.7066 and b = − 1.93 HU for the truncated FOV.

2.3.3 ResUNet residual correction

We implemented a residual U-Net architecture (ResUNet) that directly predicts a voxel-wise correction map (28, 29). The network takes a raw CBCT image as input and outputs a residual map δ, which is then added to the input to produce the corrected sCT:

$$sCT = CBCT + \delta$$

2.3.3.1 Architecture

The model follows an encoder-decoder structure with residual blocks in the bottleneck. Specifically, the encoder consists of four convolutional blocks with instance normalization and ReLU activation,

progressively increasing the filter count from 64 to 512 while downsampling via strided convolutions (30, 31). The bottleneck contains two residual blocks (each with two convolutional layers and a skip connection) followed by a dropout layer (rate 0.3) for regularization. The decoder employs transposed convolutions for upsampling, with skip connections concatenating corresponding encoder feature maps to preserve fine spatial details. A final 7×7 convolutional layer with tanh activation produces the correction map δ, which is scaled by a fixed factor of 0.1 before being added to the input. This scaling factor stabilizes training by limiting the magnitude of each correction step (Fig. 1).

2.3.3.2 Loss function and training details

The network was trained using only MAE loss between the final sCT and the ground truth CT. Training was performed for a maximum of 200 epochs with a batch size of 32 and Adam optimizer with a learning rate of 1×10^{-4}, using early stopping with a patience of 25 epochs based on validation loss to prevent overfitting. For the complete FOV group, which contained substantially fewer training samples, the model pretrained on the truncated FOV data was subsequently fine-tuned on the complete FOV data using a batch size of 8 and a learning rate of 5×10^{-5}.

This design choice allows us to evaluate whether a deterministic, spatially aware regression model can achieve accurate HU correction without the complexity of a GAN, focusing purely on intensity accuracy rather than perceptual realism, while also addressing the data imbalance between FOV groups through transfer learning.

2.3.4 Paired cycle-consistent conditional GAN-style residual U-Net

It is important to note that our implementation is not a classical unpaired CycleGAN (32). Instead, we designed a paired cycle-consistent conditional GAN (PCC-cGAN) that operates on registered CT/CBCT image pairs and employs a residual learning strategy similar to the ResUNet described above. The only architectural difference between the generators and ResUNet is the inclusion of attention blocks in two of the U-Net skip connections.

2.3.4.1 Generator architecture

Both generators (G_A: CBCT → sCT and G_B: CT → CBCT) share an identical attention-augmented residual U-Net structure. The encoder is mostly similar to the ResUNet encoder. The decoder employs transposed convolutions for upsampling, with attention gates that selectively emphasize relevant features from the encoder skip connections before concatenation. A final 7×7 convolutional layer with tanh activation produces a correction map δ, which is scaled by a fixed factor of 0.1 and added to the input.

2.3.4.2 Discriminator architecture

Two conditional PatchGAN discriminators (70×70 receptive field) are used. D_A distinguishes real CBCT from fake CBCT (generated from CT), while D_B distinguishes real CT from sCT (generated from CBCT).

Each discriminator takes the concatenation of the input (CBCT or CT) and the target image as a conditional input and outputs a patch-wise classification map (Fig. 2) (33).

2.3.4.3 Loss functions

The total generator loss combines four components: reconstruction loss, cycle loss, identity loss, and adversarial loss. MAE and SSIM loss were used in reconstruction loss, and gradient difference loss (GDL) was added to them for cycle loss (34, 35).

generator loss = (9 × reconstruction loss) + (3 × cycle loss) + (1 × identity loss) + ([0→0.6 in 30 epochs] × adversarial loss)

where:

- adversarial loss = Binary Cross Entropy
- reconstruction loss = (5 × MAE) + (1 × SSIM loss)
- cycle loss = (1 × MAE) + (0.25 × SSIM loss) + (1 × GDL)
- identity loss = MAE

The weight of the adversarial loss was gradually increased during training. starting from 0 at the beginning of training, it was linearly increased to 0.6 by epoch 30 and then held constant at 0.6 for the remainder of the training.

The discriminator loss is standard binary cross-entropy with label smoothing (real labels set to 0.95) and Gaussian noise ($\sigma = 0.02$) added to inputs during discrimination to improve stability.

2.3.4.4 Training details

Training continues for up to 200 epochs with a batch size of 6, and Adam optimizers with learning rates of 1×10^{-4} are used for both generators and discriminators. due to the high computational demand of the PCC-cGAN architecture and memory constraints, the 69 training/validation pairs were processed in two sequential stages. The first stage used 35 pairs, including 29 training and 6 validation pairs. The best checkpoint from this stage was then used to initialize the second stage, which used the remaining 34 pairs, including 29 training and 5 validation pairs. For the second stage, just the generators' learning rates was reduced from 1×10^{-4} to 5×10^{-5}. Validation loss was monitored independently in each stage, and the checkpoint with the lowest validation reconstruction loss was saved for evaluation. In both stages, training was stopped when the validation loss did not improve for 15 consecutive epochs. The final checkpoint from the second stage was used for testing.

For the complete FOV group, the final model pretrained on the truncated FOV data was subsequently fine-tuned on the complete FOV data using the same batch size of 6 and a learning rate of 5×10^{-5} for generators and 1×10^{-4} for discriminators.

Validation was performed using the same reconstruction loss (5×MAE + 1×SSIM loss) which was also used for early stopping and model checkpoint selection. Specifically, training was stopped if the validation reconstruction loss did not improve for 15 consecutive epochs, and the model parameters that achieved the lowest validation loss were restored as the final model for testing.

2.4 Evaluation masks and HU metric

To enable a clinically meaningful evaluation of HU correction accuracy, we defined several anatomical regions of interest using binary masks derived from the patient's CT and CBCT volumes. Rather than computing metrics globally over the entire image, we focused our analysis on specific regions where HU accuracy directly impacts dose calculation.

It is important to note that all the network predictions were initially generated at a training resolution of 256 × 256 pixels. For evaluation, the predicted sCT volumes were resized back to their original resolution of 512 × 512 pixels using linear interpolation. To quantify the potential error introduced by the resolution conversion used during evaluation, an additional resampling analysis was performed. The original CT images at 512 × 512 pixels were resized to 256 × 256 pixels and subsequently restored to 512 × 512 pixels using the same linear interpolation procedure. The MAE between the original and twice-resampled CT images was calculated to estimate the intensity difference attributable solely to the resizing process (Complete results are provided in Additional file 1, Table S1).

2.4.1 Mask generation

For each patient volume, a body mask was generated independently from the CT and CBCT volumes, using an identical threshold-based procedure. For each volume, a threshold was defined as the minimum voxel value in that volume (-1000 HU) plus 100 HU, which separates the patient's body from the background air. Binary holes within the resulting mask were then filled to include internal soft tissue structures. To exclude disconnected artifacts, only the largest connected 3D component was retained. Then a valid FOV mask was defined as the intersection of the CT body mask and the CBCT body mask, ensuring that only voxels containing valid anatomy in both modalities were considered.

An additional exploratory sensitivity analysis using sCT-specific masks was performed to assess the effect of mask definition on the observed results. The sensitivity analysis showed that the overall ranking of the methods, including the superior performance of ResUNet, was preserved. further details are provided in Additional file 1 (Tables S2, S3, and S4).

2.4.2 Derived mask regions

Using the CT mask, CBCT mask, and valid FOV mask three distinct evaluation regions were defined:

- Intersection mask: Representing voxels consistently identified as bodies in both modalities.

Intersection mask = Valid FOV mask = CT mask ∩ CBCT mask

- Eroded intersection mask: The intersection mask was eroded by 5 iterations using the default 2D structuring element of "binary_erosion" from scipy library, applied independently to each axial (x-y) slice, with no erosion along the z-axis (slice direction) (36). Erosion was performed at the evaluation resolution of 512×512 pixels, with in-plane pixel spacing of 1 × 1 mm². The physical extent of erosion was determined by the implemented structuring element and voxel spacing; therefore, the number of erosion iterations was not interpreted as a direct millimeter measurement. This isolates the interior region away from tissue-air interfaces and FOV boundaries.
- Peripheral mask: This mask isolates the peripheral region of the body contour, including tissue-air interfaces.

Peripheral mask = Intersection mask – Eroded mask

The same patient-specific evaluation masks were applied identically to linear correction, PCC-cGAN, and ResUNet outputs. These masks allowed us to test the hypothesis that residual HU errors after correction are concentrated at the peripheral region rather than being uniformly distributed throughout the anatomy.

2.4.3 HU accuracy metric

HU correction performance was quantified using MAE between the CBCT/sCT image and the ground truth pCT image. The MAE was computed volume-wise at the patient level to provide a clinically meaningful per-patient accuracy measure:

$$MAE = \frac{1}{N_{mask}} \sum_{i=1}^{N} I_{mask}(i) \cdot |y_i - \hat{y}_i|$$

where y_i and $\widehat{y}_i$ are the CT and CBCT/sCT HU values at voxel i, respectively. $I_{mask}(i)$ is an indicator function for the mask region, and N_{mask} is the total number of voxels within that mask.

All the metrics for sCT images were computed after 0 to 1 clipping and denormalizing the network outputs back to the original HU range of [-1000, 3000] HU, followed by resizing the images back to their original dimensions of 512 × 512 pixels per slice with linear interpolation (the size before computational downsampling for training). This ensures that the reported MAE values are in clinically meaningful HUs and are consistent with the original image resolution.

2.5 Independent-patient sensitivity analysis

Some patients contributed more than one CBCT/CT pair. Consequently, a subset of test cases shared the same patient identity as cases in the training set. This overlap may introduce a risk of optimistic bias in the evaluation metric, as the model could have learned patient-specific anatomical features during training.

To assess the potential impact of this patient-level overlap on our results, we performed an independent-patient sensitivity analysis. A dedicated subset of the test data was constructed by excluding all test cases where the patient also appeared in the training or validation sets.

MAE was calculated in this independent-patient subset for each of the three mask regions. The results from this analysis were compared against the primary results to verify whether the observed improvements and error patterns remained consistent when patient overlap was eliminated. A substantial difference between the primary and independent-patient results would indicate that patient overlap influenced the evaluation; conversely, consistent results would support the stability of the observed performance pattern in the independent-patient subset.

2.6 Preliminary dose and gamma analysis

To assess the dosimetric consistency of the ResUNet sCT images, a preliminary dose analysis was performed. We emphasize that this analysis serves as an initial dosimetric consistency check and is not a substitute for a full dose-volume-histogram-based (DVH-based) clinical validation, which would require a larger patient cohort and more comprehensive treatment planning evaluation (37).

2.6.1 Dose calculation

Dose calculations were performed on three image sets for each test patient: the reference pCT, the CBCT, and the ResUNet sCT. All the dose calculations were conducted using the same treatment plan, beam configuration, and dose calculation algorithm to ensure comparability. The dose distributions were calculated on the original pCT geometry as the ground truth reference.

Dose calculations were performed using the Eclipse™ treatment planning system (version 16.1.5) (Varian Medical Systems, Palo Alto, CA). The anisotropic analytical algorithm (AAA, version 16.12) was used for dose computation. The resulting RT Dose distributions had a native voxel spacing of 2.5 × 2.5 × 3.0 mm^3 (x, y, z). All plans employed 6 MV photon beams delivered using intensity-modulated radiation therapy (IMRT). The same CT-to-density calibration table was applied consistently for pCT, CBCT, and sCT.

As mentioned before, after model prediction, the sCT outputs were upsampled back to 512 × 512 pixels using linear interpolation to restore the original image resolution for evaluation and dose calculation. For dose calculation, the CBCT and sCT volumes were rigidly registered to the pCT space and resampled using nearest-neighbor interpolation. Nearest-neighbor interpolation was selected for registration during dose calculation to avoid HU averaging and slice-boundary artifacts observed with linear interpolation. We also evaluated an all-nearest-neighbor resampling strategy for the sCT upsampling step (following prediction). This approach was not used for the primary analysis; it was examined only as a sensitivity check, and it resulted in degraded GPRs compared with those of the chosen pipeline. Therefore, the selected pipeline (linear interpolation for sCT upsampling combined with nearest-neighbor for registration/resampling) was adopted for the primary analysis. Nevertheless, the impact of different interpolation strategies on dosimetric metrics warrants further investigation. Notably, a similar nearest-neighbor resampling approach has also been employed in recent sCT literature for returning sCT images to the original CT coordinate space (18).

For dose comparison, the RT Dose distributions were resampled in-plane to the pCT image grid using linear interpolation, while the 3 mm slice spacing was retained. Thus, the dose arrays used for gamma analysis had a physical voxel spacing of approximately 0.9766 × 0.9766 × 3.0 mm^3. The physical voxel spacing in all three dimensions was explicitly incorporated into the gamma distance-to-agreement calculation.

2.6.2 Dosimetric accuracy metrics and gamma analysis

Dosimetric accuracy was evaluated using voxel-wise MAE and Root Mean Square Error (RMSE), in gray (Gy), computed within the body region. To assess dose distribution agreement, three-dimensional gamma analysis was performed using 2%/2 mm and 3%/3 mm criteria with global normalization. The dose-difference criterion was defined relative to the global maximum dose of the reference pCT distribution, and only reference voxels receiving ≥ 10% of this global maximum dose were included in the GPR calculation. The physical voxel spacing of the dose distributions was incorporated into the three-dimensional distance-to-agreement calculation. GPRs were computed using an in-house Python implementation. Analysis was performed for ResUNet sCT and raw CBCT as a baseline. For exploratory localization of gamma failures, failures were further classified according to the spatial dose gradient in the reference dose distribution, with 0.5 Gy/mm used as the threshold to distinguish high- and low-gradient regions. The 0.5 Gy/mm value was used as an exploratory stratification threshold rather than a validated clinical cutoff. In addition, the mean dose gradient was calculated separately at gamma-failure and gamma-passing locations for each patient.

2.6.3 Limitations

This preliminary dosimetric analysis has several limitations. The analysis does not include full treatment plan re-optimization on the sCT, which would be required for online adaptive radiotherapy. DVH-based metrics (e.g., D95%, D98%, V20) were not evaluated. Therefore, the results presented here should be interpreted as an initial consistency check rather than a comprehensive clinical validation. A more extensive DVH-based evaluation of a larger independent cohort is warranted for future work.

3. Results

3.1 HU correction performance in the complete and truncated FOV groups

To account for the impact of FOV limitations on HU correction, all evaluations were performed separately for the truncated FOV and complete FOV subgroups. Table 1 shows the MAE for each correction method across the intersection, Eroded, and peripheral mask regions.

A resize-only control was also performed by resizing the original 512 × 512 pCT images to 256 × 256 and then back to 512 × 512 using the same linear interpolation procedure. This control produced peripheral, eroded, and intersection MAE values of approximately 29–46 HU, 9–15 HU, and 12–18 HU, respectively.

These findings indicate that part of calculated MAE specially the peripheral MAE measured after sCT generation arise from the resizing/interpolation process itself rather than from the HU correction model alone (Additional file 1, Table S1).

ResUNet achieved the lowest MAE among the evaluated methods in both the complete and truncated FOV groups for all the mask regions. However, for all methods including ResUNet, the peripheral MAE remained substantially higher than the intersection and eroded MAE. For example, in the complete FOV group, ResUNet's peripheral MAE (171.51 HU) was more than double its intersection MAE (76.93 HU). This elevation in error at the boundary region was observed in both the complete and truncated FOV groups, indicating that the peripheral evaluation region remains a persistent source of measured residual error after HU correction, although the resize-only control indicates that part of this error may also arise from the required resampling procedure.

MAE values were consistently lower in the complete FOV group across all methods in intersection and eroded regions but for peripheral region it was lower just for the linear correction method (Table 1). In terms of HU correction performance, the methods ranked from highest MAE (worst) to lowest MAE (best) as follows: raw CBCT, linear correction, PCC-cGAN, and ResUNet.

Table 1
Mean ± standard deviation of MAE-HU comparison in the complete and truncated FOV groups.

Method	Complete Intersection	Complete Eroded	Complete Peripheral	Truncated Intersection	Truncated Eroded	Truncated Peripheral
CBCT	107.29 ± 6.77	95.56 ± 8.44	201.61 ± 13.27	130.85 ± 21.74	122.04 ± 20.12	196.86 ± 39.78
Linear Correction	93.81 ± 5.71	81.74 ± 6.60	190.71 ± 17.32	110.79 ± 19.28	99.75 ± 20.07	193.08 ± 22.17
PCC-cGAN	80.89 ± 2.75	67.62 ± 2.75	188.02 ± 4.13	91.66 ± 12.05	80.60 ± 10.94	174.07 ± 25.63
ResUNet	76.93 ± 2.48	65.21 ± 1.92	171.51 ± 6.29	88.22 ± 11.81	77.18 ± 10.58	170.51 ± 26.42

3.2 Independent-patient sensitivity analysis

As noted in the Materials and Methods section, some patients contributed two CBCT/CT pairs to the dataset due to reCT acquisitions during their treatment course. Consequently, among the 12 test pairs in the truncated FOV group, 6 cases were derived from patients who also had a pair present in the training or validation set. Similarly, 1 of the 3 test pairs in the complete FOV group originated from such a patient. To assess whether this patient-level overlap influenced our findings, we performed an independent-patient sensitivity analysis on a subset of the test data consisting exclusively of patients whose pairs did not appear in the training or validation sets.

A separate sensitivity analysis was also performed to compare the use of a fixed mask versus an sCT-specific mask; this analysis confirmed that the overall ranking of the methods remained unchanged. further details are provided in Additional file 1 (Tables S2, S3, and S4).

Table 2 reports the MAE for each method on this independent-patient subset across three mask regions: intersection, eroded, and peripheral. The analysis included 24 patient-region combinations, corresponding to 8 independent patients and 3 mask regions. In the independent-patient subset, ResUNet showed the lowest MAE across most patient-region comparisons, with the lowest value observed in 22 of 24 patient-region comparisons and the PCC-cGAN model was best in only two combinations (one patient in eroded region and one in peripheral region). In terms of mean MAE values, ResUNet consistently achieved the lowest errors across all three mask regions.

The overall pattern of results remained consistent with the primary analysis presented in Table 1. ResUNet continued to achieve the lowest MAE across most patient-region comparisons. The mean MAE values also showed a consistent advantage for ResUNet across the evaluated regions. An exception for the non-learning method was observed in the peripheral region, where raw CBCT showed a lower mean MAE than linear correction. These findings are descriptive, as the small number of independent patients limits the strength of conclusions regarding generalizability.

Table 2
Independent-patient subset analysis

Method	Intersection MAE mean ± SD	Eroded MAE mean ± SD	Peripheral MAE mean ± SD
CBCT	115.15 ± 13.66	106.67 ± 14.90	179.67 ± 22.18
Linear Correction	97.14 ± 10.28	85.48 ± 11.09	184.76 ± 17.57
PCC-cGAN	83.49 ± 7.49	72.74 ± 7.53	164.71 ± 17.13
ResUNet	80.48 ± 6.75	70.40 ± 6.80	156.72 ± 12.15

In summary, ResUNet consistently reduced the HU-MAE compared with raw CBCT, linear correction, and PCC-cGAN. However, peripheral-region errors remained substantially greater than intersection-region errors across all methods.

3.3 Visual comparison and boundary behavior

To visually demonstrate the spatial distribution of HU correction errors, representative axial slices from a complete FOV case and a truncated FOV case are shown in Fig. 3 and Fig. 4, respectively. Each figure presents the reference pCT, CBCT, ResUNet-generated sCT, and the corresponding absolute error map. The peripheral mask region is overlaid on the error map to highlight the boundary area, while the intersection mask is used to define the region of interest for error computation, and the eroded mask is shown to indicate the interior region away from tissue-air interfaces.

3.4 Preliminary body-level dosimetric and gamma analysis

To assess the dosimetric consistency of the ResUNet-generated sCTs, we performed a preliminary gamma analysis at the body level for the raw CBCT and ResUNet sCT. Table 3 reports the GPRs and dose MAE/RMSE for each test patient, separately for the complete FOV and truncated FOV cases. Dose MAE showed substantial inter-patient variability. Across the reported test cases, ResUNet sCT dose MAE ranged from 0.99 to 2.90 Gy, whereas raw CBCT ranged from 1.23 to 3.38 Gy. Thus, although ResUNet generally reduced dose error, the magnitude of improvement varied considerably between patients.

In the complete FOV group, ResUNet showed slight improvements in GPRs. For example, in Patient05, the 2%/2mm pass rate improved from 92.98% (CBCT) to 94.83% (sCT), and the dose MAE decreased from 1.25 Gy to 0.99 Gy. However, in Patient08, the differences between CBCT and ResUNet were minimal and dose MAE showed little change.

In the truncated FOV group, while ResUNet improved GPRs in several patients. For example, Patient01 and Patient11 are among the most pronounced improvement cases (Patient01: 2%/2mm improved from 67.58% to 70.70%; Patient11: 2%/2mm improved from 60.92% to 65.06%). In contrast, Patient05 showed a small deterioration, with dose MAE increasing from 1.23 to 1.41 Gy and the 2%/2 mm GPR decreasing from 92.69% to 91.28%. Patient08 and Patient10 showed minimal changes, as well.

The mean dose gradient at gamma-failure locations was numerically higher than at passing locations in all 15 sCT evaluations. The same pattern was observed in 14 of 15 CBCT evaluations, with Patient12 being the only exception. The additional threshold-based analysis using 0.5 Gy/mm showed patient-to-patient variation in the distribution of failures between high- and low-gradient regions. These findings indicate an association between gamma failures and the local dose-gradient environment, while also showing that dose-gradient category alone does not explain the observed patient-level variability. Detailed results are provided in Additional file 1 (Tables S5 and S6 and Figure S1).

These observations show that substantial improvements in HU accuracy do not necessarily translate proportionally to improvements in GPRs. Despite ResUNet achieving the lowest HU errors as shown in Table 1 and Table 2, the corresponding dosimetric improvements were not uniformly observed across all patients or all gamma criteria.

It must be emphasized that this analysis is an exploratory body-level dosimetric assessment and is not a substitute for a full DVH-based clinical validation. A comprehensive evaluation of target coverage and OAR sparing would require plan re-optimization on sCT and patient-specific dose-volume histogram analysis, which is beyond the scope of this study. An exploratory organ-level gamma summary was also prepared and is provided in Additional file 1 (Table S7). The organ-level results showed heterogeneous behavior across OARs, supporting the need for future DVH-based validation.

It should be noted that this dosimetric analysis compared only the ResUNet sCT against the raw CBCT baseline. No dosimetric conclusions regarding the linear correction or PCC-cGAN methods can be drawn

from these results.

Table 3
Preliminary body-level dose and GPR.

FOV	Patient	Dose MAE/RMSE [Gy]		3D Global 3%/3 mm GPR		3D Global 2%/2 mm GPR	
		sCT	**CBCT**	**sCT**	**CBCT**	**sCT**	CBCT
Complete	Patient02	1.57 / 3.72	1.65 / 4.11	91.93	91.34	85.02	85.02
	Patient05	0.99 / 2.61	1.25 / 3.27	97.47	96.37	94.83	92.98
	Patient08	2.03 / 4.65	2.14 / 5.01	89.08	88.37	81.64	81.97
Truncated	Patient01	2.90 / 5.85	3.38 / 7.03	80.94	77.80	70.70	67.58
	Patient02	1.59 / 4.01	1.83 / 4.27	92.88	91.70	85.44	83.15
	Patient03	1.57 / 3.56	1.77 / 4.21	92.73	92.19	86.25	85.67
	Patient04	2.26 / 5.35	2.39 / 5.85	90.00	88.30	82.07	80.25
	Patient05	1.41 / 3.38	1.23 / 3.06	96.74	97.03	91.28	92.69
	Patient06	2.47 / 4.75	2.57 / 4.91	85.70	85.30	70.60	69.24
	Patient07	1.36 / 2.80	1.35 / 3.07	95.53	95.24	89.55	90.20
	Patient08	1.99 / 4.40	1.93 / 4.36	90.78	90.86	83.99	84.20
	Patient09	2.22 / 4.39	2.25 / 4.35	89.73	87.12	73.51	69.56
	Patient10	2.35 / 5.44	2.23 / 5.39	90.88	91.36	82.13	83.90
	Patient11	2.65 / 4.85	3.14 / 5.68	81.17	75.69	65.06	60.92
	Patient12	2.37 / 3.97	2.67 / 4.45	80.89	76.00	61.34	57.29

4. Discussion

4.1 Principal findings

In this study, we systematically compared three methods for HU correction and sCT generation from CBCT in H&N cancer patients: linear HU correction, PCC-cGAN, and ResUNet. ResUNet achieved the lowest HU-MAE across the evaluated regions in both complete and truncated FOV groups. This pattern was also observed in the independent-patient sensitivity analysis, suggesting that the observed performance advantage was not an artifact of patient-level data leakage. Despite the substantial HU improvements achieved by ResUNet, residual errors remained elevated at the peripheral region compared with the interior anatomy. Furthermore, the improvements in HU accuracy did not translate proportionally to improvements in GPRs at the body level, with variable dosimetric outcomes across patients. These findings describe the performance obtained under the specific training and evaluation protocols used in this study and should not be interpreted as evidence that the ResUNet architecture alone caused the observed advantage.

4.2 Positioning relative to prior sCT studies

A growing body of literature has investigated deep learning-based sCT generation from CBCT for H&N radiotherapy. GAN-based frameworks, including cGANs and CycleGANs, have been widely adopted due to their ability to learn complex CBCT to CT mappings, sometimes even without paired training data (15, 38). More recently, emerging generative paradigms such as diffusion models have shown promising results in improving image fidelity, structural preservation, and uncertainty modeling (24, 39). Large-scale benchmarking efforts such as the SynthRAD2025 challenge have also comprehensively evaluated sCT methods on multi-center datasets (40).

Our study does not aim to directly compete with these advanced generative models in terms of raw HU error minimization. Rather, the primary distinction of our work lies in the evaluation framework: we explicitly stratify performance by FOV status (complete vs. truncated) and by spatial mask region (intersection, eroded intersection, and peripheral). Although prior studies have incorporated various regional and tissue-specific evaluation strategies, relatively few have examined HU errors according to the FOV status (complete versus truncated) and spatially defined mask regions such as intersection, eroded intersection, and peripheral. Our FOV-stratified and peripheral-aware evaluation framework addresses this gap, providing a more clinically realistic assessment of sCT performance and identifying the peripheral region as the principal remaining source of error after HU correction.

4.3 Factors potentially underlying the observed ResUNet performance

The superior performance of ResUNet over linear HU correction is primarily explained by the inherent limitations of a global intensity mapping approach. Linear correction applies a single slope-intercept transformation to all voxels, assuming a spatially uniform relationship between CBCT and CT intensities. However, CBCT artifacts, including scatter-induced cupping and shading, beam hardening, dental metal artifacts, and FOV boundary truncation, are highly spatially heterogeneous. Global linear mapping cannot

account for these localized intensity distortions, leading to systematic under-correction in some regions and over-correction in others. Consequently, while linear correction reduces the overall HU bias compared with raw CBCT, it fails to resolve the spatially varying errors that are critical for accurate dose calculation.

The comparison between ResUNet and PCC-cGAN should be interpreted in the context of the training protocols used in this study. Under the respective training and optimization protocols, ResUNet achieved lower HU-MAE than PCC-cGAN across the evaluated regions. The two models were not trained under identical conditions: ResUNet was trained using a direct masked MAE objective, whereas PCC-cGAN combined reconstruction, cycle-consistency, identity, and adversarial losses and was trained using a two-stage procedure because of memory constraints. The direct MAE objective may be more closely aligned with the voxel-wise HU correction metric used for evaluation, potentially contributing to the observed performance difference. However, because the training protocols and loss formulations differed and no ablation study was performed, the present results do not establish that the ResUNet architecture itself was responsible for the lower MAE.

To put it briefly, it can be said that compared with linear correction and raw CBCT, the PCC-cGAN achieved substantially lower HU-MAE. However, within the conditions and training protocols evaluated in this study, the supervised ResUNet provided the lowest HU-MAE among the compared correction methods.

4.4 Effects of FOV truncation and boundary regions

Within the common CT/CBCT anatomical region, the truncated FOV group showed higher MAE than the complete FOV group for all evaluated methods in intersection and eroded mask regions. For ResUNet, the intersection MAE increased from 76.93 HU in the complete FOV group to 88.22 HU in the truncated FOV group, while the eroded MAE increased from 65.21 HU to 77.18 HU (Table 1). Similar increases were observed for CBCT, linear correction, and PCC-cGAN, indicating that the higher error associated with truncated acquisitions was also present within the shared anatomical region.

The peripheral region showed a less consistent FOV-dependent pattern. Peripheral MAE was slightly higher in the truncated FOV group only for linear correction (193.08 HU vs. 190.71 HU), whereas it was lower for CBCT (196.08 HU vs. 201.61 HU), PCC-cGAN (174.07 HU vs. 188.02 HU), and ResUNet (170.51 HU vs. 171.51 HU). Thus, the peripheral response was not uniformly determined by FOV status and varied across correction methods. This variability may reflect both patient- and acquisition-dependent differences in the peripheral region.

The observed differences should also be interpreted in the context of the acquisition protocols. Complete-FOV CBCT examinations were acquired at 125 kVp with a voxel spacing of 0.9080 × 0.9080 × 1.9885 mm^3, whereas truncated-FOV examinations used 80 and 100 kVp with a voxel spacing of 0.5112 × 0.5112 × 1.9897 mm^3. Therefore, the comparison between complete and truncated groups reflects

differences associated with both FOV status and the corresponding acquisition protocols, and cannot be attributed exclusively to FOV truncation.

Despite this variability, the peripheral region remained a substantially higher-error region than the interior for all methods and FOV groups. For example, ResUNet showed a peripheral MAE of 171.51 HU compared with 76.93 HU in the complete-FOV intersection, and 170.51 HU compared with 88.22 HU in the truncated-FOV intersection. This persistent elevation indicates that residual HU errors are concentrated near peripheral and tissue-air interface regions even after correction.

Several factors likely contribute to the persistently high peripheral MAE. Our models were trained on 2D axial slices, which lack full volumetric context. At the peripheral evaluation region, the relationship between CBCT and CT intensities may depend on structures above or below the current slice. A 2D model cannot access this information, potentially leading to greater errors at these boundaries. The peripheral region encompasses areas where intensity gradients are steep and small misalignments or intensity fluctuations translate into large HU errors. These regions are inherently more difficult to map than the homogeneous interior. The observed patient- and acquisition-dependent variation may reflect the heterogeneous nature of peripheral regions, where scatter, truncation artifacts, tissue-air interfaces, and local intensity gradients vary substantially between patients and acquisitions.

Scattering, beam hardening, and truncation artifacts do not follow a simple or spatially uniform pattern near the FOV peripheral. Their magnitude and distribution depend on patient size, positioning, and specific scan geometry. Learning stable mapping in this region requires substantially more data and perhaps dedicated architectural modifications. Our masked MAE loss treats all voxels within the mask equally. It does not assign additional weights to the peripheral region, even though this region is more difficult to deal with. A loss that emphasizes boundary accuracy might reduce the peripheral MAE, possibly at the expense of interior performance.

Taken together, these findings suggest that while ResUNet effectively restores HU accuracy in the interior anatomy, the peripheral region remains a persistent source of residual error. Future work should consider 3D architectures, peripheral-weighted loss functions, or dedicated post-processing to address this limitation.

4.5 Robustness in the independent-patient subset

As mentioned before, some patients contributed multiple CBCT/CT pairs to the dataset due to reCT acquisitions during their treatment course. This introduced a potential risk of patient-level overlap between the training/validation and test sets. To assess whether this overlap biased our results, we performed an independent-patient sensitivity analysis on a subset of test cases with no patient overlap.

In this independent-patient subset, ResUNet remained the best-performing method in most patient-region combinations (8 independent patients × 3 mask regions). The overall pattern of HU accuracy was preserved, with ResUNet achieving the lowest mean MAE across the evaluated regions and persistently higher MAE in the peripheral region (Table 2). CBCT showed a lower mean peripheral MAE than linear

correction (non-learning approach) in this independent-patient subset. This difference highlights some patient- and region-dependent variability in the relative performance of conventional HU correction approaches. Given the small number of independent patients, this finding should be interpreted descriptively rather than as evidence of a systematic advantage of raw CBCT over linear correction. Overall, the consistency of the main ranking supports the robustness of the primary findings despite potential patient-level overlap.

Nevertheless, we acknowledge that a fully patient-independent split (where no patient appears in more than one set) is the gold standard for model evaluation. The independent-patient subset showed a similar descriptive performance pattern, although the small sample size (8 patients) limits firm conclusions regarding generalizability. Future work should employ a fully patient-independent split on a larger, multi-institutional cohort to further validate these findings.

4.6 HU accuracy versus dosimetric accuracy

The relationship between HU accuracy and dosimetric accuracy is not straightforward. Table 3 shows that ResUNet sCT generally maintained or modestly improved body-level GPRs and dose MAE/RMSE compared with raw CBCT. However, this improvement was neither uniform across patients nor directly proportional to the magnitude of HU error reduction achieved by ResUNet.

In several patients, ResUNet clearly improved dosimetric agreement with the reference CT. For example, in Patient11, the 2%/2mm GPR increased from 60.92% (CBCT) to 65.06% (ResUNet), and the dose MAE decreased from 3.14 Gy to 2.65 Gy is one of the most pronounced improvement cases. Similarly, in Patient01 and Patient12, the differences between CBCT and ResUNet were in favor of sCT. In some other cases, however, raw CBCT performed comparably or even slightly better. For Patient05, the 2%/2mm GPR decreased from 92.69% (CBCT) to 91.28% (ResUNet), and the dose MAE increased from 1.23 Gy to 1.41 Gy. Similarly, in Patient08 and Patient10, the differences between CBCT and ResUNet were minimal or marginally in favor of CBCT.

Several factors explain why substantial improvements in the HU-MAE do not translate proportionally to improvements in GPR. As mentioned above, the residual HU errors after ResUNet correction are concentrated in the peripheral region. The dosimetric impact of an HU error depends not only on its magnitude but also on its spatial location relative to the dose distribution and beam arrangement. The observed gamma results further indicate that the relationship between HU accuracy and dosimetric agreement is spatially dependent. The exploratory failure-localization analysis showed that gamma failures were associated with the local dose-gradient environment; for example, 88.4%, 80.4%, 77.4%, and 73.0% of sCT failures occurred in high-gradient regions in Patient05, 08, 10, and 11, respectively. Although the proportion of failures occurring in high-gradient regions varied among patients, gamma-failure locations generally exhibited higher mean dose gradients than passing locations. However, this pattern was not universal, with a comparatively large contribution from low-gradient regions in Patient09 and few other patients. Consistent with this spatial distribution, the mean dose gradient at failure locations was higher than at passing locations in all patients in sCT images and CBCT images have

same pattern for 14 of 15 calculations. Global HU-MAE alone was insufficient to predict dosimetric agreement. Gamma failures occurred, on average, in regions with higher local dose gradients, suggesting that the local dose-distribution context contributes to the relationship between image-intensity errors and gamma agreement. However, the patient-dependent distribution of failures indicates that dose gradient alone does not explain the observed variability. Therefore, improvements in overall HU-MAE do not necessarily translate into proportional improvements in gamma passing rate. The gamma analysis used global normalization, with only reference voxels receiving ≥ 10% of the global maximum dose included in the calculation.

The specific beam arrangement (gantry angles, field sizes, and modulations) determines which anatomical regions are irradiated. An HU error in a region that is not traversed by any beam will have no effect on the calculated dose, regardless of its magnitude. Conversely, an error in a region intersected by multiple beams may be amplified. It is possible for an HU error in one region to partially compensate for an error elsewhere, leading to a dose distribution that accidentally agrees well with the reference despite substantial local HU inaccuracies.

It is essential to emphasize that this analysis is an exploratory body-level dosimetric assessment and is not a substitute for a DVH-based clinical validation. A comprehensive evaluation would require treatment plan re-optimization on the sCT, patient-specific DVH metrics, and assessment of target coverage and OAR sparing. Such an analysis is beyond the scope of this study but represents a critical direction for future work before the clinical implementation of sCT-based adaptive radiotherapy.

4.7 Limitations and future work

Several limitations of this study should be acknowledged, along with directions for future research. All the data were collected from a single institution. The patient cohort, while representative of routine clinical practice at our center, may not capture the full variability in anatomy, CBCT acquisition protocols, or image quality encountered at other institutions. Future work should validate our findings on multi-center, multi-vendor datasets to assess generalizability.

The complete FOV group contained only 19 acquisitions, with 3 test cases. This limited sample size constrains the statistical power of comparisons within this subgroup. Larger cohorts of complete FOV H&N CBCTs are needed to confirm the observed performance differences. Although the initial split was performed at the image-pair level, some patients contributed multiple pairs due to reCT scans. We addressed this concern through an independent-patient sensitivity analysis, which supported that the main findings were not driven by patient overlap. Nevertheless, a fully patient-independent split with no patient appearing in more than one set remains the gold standard. Future studies should prospectively design such splits with larger cohorts.

The resize-only control also showed that linear interpolation during the 256 × 256 to 512 × 512 reconstruction can contribute approximately 29–46 HU of peripheral MAE; therefore, the peripheral MAE

(intersection and eroded as well) reported in this study should be interpreted as reflecting both model-related HU errors and a potential contribution from the resampling process.

As discussed in Section 4.4, the use of 2D axial slices limits the models' volumetric context, particularly at FOV boundaries and tissue-air interfaces. A 3D architecture therefore represents a promising direction for future work.

Our dose and gamma analyses were limited to body-level metrics and served only as exploratory body-level dosimetric assessments. A comprehensive dosimetric evaluation was beyond the scope of this study. We did not perform plan re-optimization on sCT, nor did we evaluate DVH-based clinical endpoints such as D95%, D98%, target coverage, or OAR dose metrics. These metrics are essential for determining whether sCT is clinically acceptable for adaptive replanning. Future work should include full DVH-based validation with plan re-optimization and, ideally, phantom or clinical endpoint studies.

In a complete adaptive radiotherapy workflow, sCT generation is only the first step. To estimate the actual delivered dose, sCT must be combined with accurate contour propagation or auto-segmentation of target volumes and OARs. The current study focused exclusively on HU correction and sCT generation. Future work should integrate ResUNet sCT with an auto-contouring tool to enable a fully automated pipeline for daily dose accumulation and adaptive decision-making. Without this integration, the clinical utility of sCT for dose tracking remains incomplete.

5. Conclusions

In this study, multiple HU correction strategies, including linear correction, PCC-cGAN, and ResUNet were compared against a raw CBCT baseline for CBCT-based sCT generation in H&N radiotherapy using FOV-stratified and mask-defined evaluation. ResUNet provided the lowest HU-MAE in both the complete and the truncated FOV groups, and this advantage was preserved in an independent-patient sensitivity analysis. The residual error after HU correction was concentrated at the peripheral evaluation region rather than within the interior body region, with the peripheral region showing the highest residual MAE. Preliminary body-level dosimetric analysis indicated that ResUNet sCT generally maintained or modestly improved dose agreement with pCT, although improvements in HU accuracy did not translate proportionally to improvements in GPR. Under the evaluated conditions, ResUNet-based correction reduced residual HU errors compared with the evaluated baselines; however, remaining peripheral discrepancies and the need for further clinical validation should be considered. These findings support the need for FOV-stratified and peripheral-aware evaluation before CBCT-based sCT workflows are used for dose-related applications.

Abbreviations

H&N
head and neck
OAR

organ at risk
pCT
planning CT
CBCT
cone-beam computed tomography
IGRT
image-guided radiation therapy
HU
Hounsfield unit
FOV
field of view
sCT
synthetic CT
DIR
deformable image registration
GAN
generative adversarial network
cGAN
conditional generative adversarial network
MAE
mean absolute error
PSNR
peak signal-to-noise ratio
SSIM
structural similarity index
ResUNet
residual U-Net
ROI
region of interest
reCT
repeat CT
RROC
Reza Radiotherapy & Oncology Center
GPR
gamma pass rate
GDL
gradient difference loss
DVH
dose-volume-histogram
AAA

anisotropic analytical algorithm
IMRT
intensity-modulated radiation therapy
RMSE
root mean square error
Gy
Gray

Declarations

Ethics approval and consent to participate. This retrospective study was approved by the Research Ethics Committee of Ferdowsi University of Mashhad, Mashhad, Iran (approval no. IR.UM.REC.1405.194). The study was conducted in accordance with the Declaration of Helsinki and relevant institutional guidelines. Written informed consent was obtained from all patients before treatment, including consent for the use of their clinical data and imaging records for educational and research purposes. All CT and CBCT data used in this study were de-identified before analysis.

Consent for publication. Not applicable.

Availability of data and materials. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests. The authors declare that they have no competing interests.

Funding. This work was supported by Ferdowsi University of Mashhad under Grant No. 3/63186.

Authors' contributions. KHJM developed the Python codes, performed the computational analyses, generated the figures, and drafted the manuscript under the joint supervision of LRM and HMH. LRM and HMH jointly conceived and supervised the study, contributed to the study design and methodology, guided the research and analytical workflow, interpreted the results, managed the scientific development of the project, and critically reviewed and revised the manuscript. MD contributed as a radiation oncology advisor, provided the clinical imaging data, and offered clinical guidance for the interpretation of the findings. MM contributed as a medical physics advisor and provided guidance on the physical and dosimetric interpretation of the results. All authors read and approved the final manuscript.

Acknowledgements. The authors would like to thank the Reza Radiotherapy and Oncology Center (RROC) for providing clinical imaging data for this study.

Additional files. Additional file 1.docx

Title of data: Supplementary analyses of image resampling, evaluation-mask sensitivity, gamma-failure localization, and organ-level gamma passing rates.

Description of data: Table S1 presents the MAE introduced by resizing the original CT images from 512 × 512 to 256 × 256 pixels and subsequently restoring them to 512 × 512 pixels using linear interpolation, providing an estimate of the intensity differences attributable to the resampling procedure. Tables S2, S3, and S4 present sensitivity analyses comparing the constant evaluation mask with sCT-derived body masks for Linear Correction, PCC-cGAN, and ResUNet, respectively. Table S5 presents the proportions of gamma failures occurring in high- and low-dose-gradient regions for the 3%/3 mm criterion. Table S6 presents the mean dose gradient for failure and passing gamma for the 3%/3 mm criterion. Figure S1 illustrates the spatial localization of gamma failures for representative complete- and truncated-FOV cases. Table S7 presents organ-level gamma passing rate comparisons between sCT and CBCT using 3%/3 mm and 2%/2 mm criteria.

References

1. Barker JL Jr, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. Int J Radiat Oncol Biol Phys. 2004;59(4):960–70.
2. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? Radiother Oncol. 2015;115(3):285–94.
3. Boda-Heggemann J, Lohr F, Wenz F, Flentje M, Guckenberger M. kV cone-beam CT-based IGRT: a clinical review. Strahlenther Onkol. 2011;187(5):284–91.
4. Schulze R, Heil U, Groβ D, Bruellmann DD, Dranischnikow E, Schwanecke U, et al. Artefacts in CBCT: a review. Dentomaxillofac Radiol. 2011;40(5):265–73.
5. Giacometti V, Hounsell AR, McGarry CK. A review of dose calculation approaches with cone beam CT in photon and proton therapy. Phys Med. 2020;76:243–76.
6. Xia D, Chang Y-B, Manak J, Siddiqui AH, Zhang Z, Chen B, et al. Reduction of angularly-varying-data truncation in c-arm cbct imaging. Sens Imaging. 2018;19(1):14.
7. Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. Med Phys. 2021;48(11):6537–66.
8. O'Hara CJ, Bird D, Al-Qaisieh B, Speight R. Assessment of CBCT–based synthetic CT generation accuracy for adaptive radiotherapy planning. J Appl Clin Med Phys. 2022;23(11):e13737.
9. Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJ, Langendijk JA, Both S, Spadea MF, Knopf AC. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol. 2020;65(9):095002.
10. Xue X, Ding Y, Shi J, Hao X, Li X, Li D, Wu Y, An H, Jiang M, Wei W, Wang X. 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.

11. Maspero M, Houweling AC, Savenije MH, van Heijst TC, Verhoeff JJ, Kotte AN, 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.
12. Kurz C, Kamp F, Park YK, Zöllner C, Rit S, Hansen D, et al. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. Med Phys. 2016;43(10):5635–46.
13. Veiga C, Lourenço AM, Mouinuddin S, Van Herk M, Modat M, Ourselin S, et al. Toward adaptive radiotherapy for head and neck patients: uncertainties in dose warping due to the choice of deformable registration algorithm. Med Phys. 2015;42(2):760–9.
14. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys. 2017;44(4):1408–19.
15. Zhang Y, Ding S-g, Gong X-c, Yuan X-x, Lin J-f, Chen Q, et al. Generating synthesized computed tomography from CBCT using a conditional generative adversarial network for head and neck cancer patients. Technol Cancer Res Treat. 2022;21:15330338221085358.
16. Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. Phys Med Biol. 2019;64(12):125002.
17. Peng J, Qiu RL, Wynne JF, Chang CW, Pan S, Wang T, et al. CBCT-Based synthetic CT image generation using conditional denoising diffusion probabilistic model. Med Phys. 2024;51(3):1847–59.
18. Hadzic A, Joham SJ, Urschler M. Flow Matching for Conditional MRI-CT and CBCT-CT Image Synthesis. arXiv preprint arXiv:2510.04823. 2025.
19. Thummerer A, Van der Bijl E, Galapon A Jr, Verhoeff JJ, Langendijk JA, Both S, et al. SynthRAD2023 Grand Challenge dataset: Generating synthetic CT for radiotherapy. Med Phys. 2023;50(7):4664–74.
20. Rusanov B, Hassan GM, Reynolds M, Sabet M, Kendrick J, Rowshanfarzad P, et al. Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: a systematic review. Med Phys. 2022;49(9):6019–54.
21. Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—application in brain proton therapy. Int J Radiat Oncol Biol Phys. 2019;105(3):495–503.
22. Zhen X, Yan H, Zhou L, Jia X, Jiang SB. Deformable image registration of CT and truncated cone-beam CT for adaptive radiation therapy. Phys Med Biol. 2013;58(22):7979–93.
23. Aootaphao S, Thongvigitmanee SS, Puttawibul P, Thajchayapong P. Truncation effect reduction for fast iterative reconstruction in cone-beam CT. BMC Med Imaging. 2022;22(1):160.
24. Spinat Q, Duran A, Teboul O, Paragios N, Komodakis N. Diffusing Boundaries: CBCT-to-CT Translation with Extended Field of View. Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. LNCS 15972. Springer Nature Switzerland; 2025. pp. 149–59.

25. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging. 2009;29(1):196–205.
26. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of SimpleITK. Front neuroinformatics. 2013;7:45.
27. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc IEEE Conf Comput Vis Pattern Recognit (CVPR). 2016. pp. 770–8.
29. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. MICCAI 2015. Springer; 2015. pp. 234–41. LNCS 9351.
30. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022. 2016.
31. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML); 2010. pp. 807–14.
32. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc IEEE Int Conf Comput Vis (ICCV); 2017. pp. 2223–32.
33. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proc IEEE Conf Comput Vis Pattern Recognit (CVPR); 2017. pp. 1125–34.
34. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12.
35. Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. arXiv preprint arXiv:1511.05440. 2015.
36. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, Van Der Walt SJ. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72.
37. Drzymala R, Mohan R, Brewster L, Chu J, Goitein M, Harms W, et al. Dose-volume histograms. Int J Radiat Oncol Biol Phys. 1991;21(1):71–8.
38. Ryu S, Kim JH, Choi YJ, Lee JS. Generating synthetic CT images from unpaired head and neck CBCT images and validating the importance of detailed nasal cavity acquisition through simulations. Comput Biol Med. 2025;185:109568.
39. Altalib A, Li C, Perelli A. EqDiff-CT: Equivariant Conditional Diffusion model for CT Image Synthesis from CBCT. arXiv preprint arXiv:2509.21913. 2025.
40. Thummerer A, van der Bijl E, Galapon AJ, Kamp F, Savenije M, Muijs C, et al. SynthRAD2025 Grand Challenge dataset: Generating synthetic CTs for radiotherapy from head to abdomen. Med Phys. 2025;52(7):e17981.

Figures

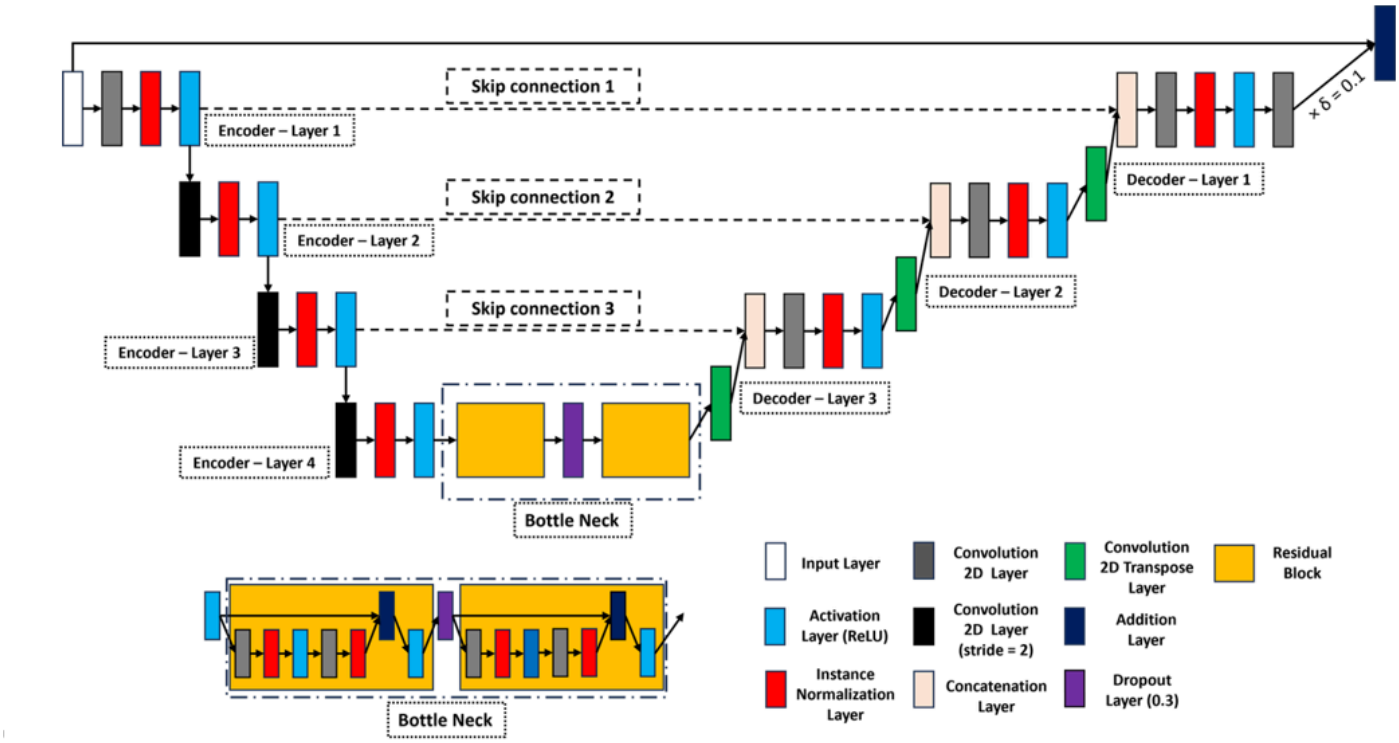


Figure 1

A schematic view of the ResUNet network. The only architectural difference between ResUNet and the generators of the PCC-cGAN is that the generator model used in the PCC-cGAN employs attention blocks in place of the standard skip connections at layers 2 and 3.

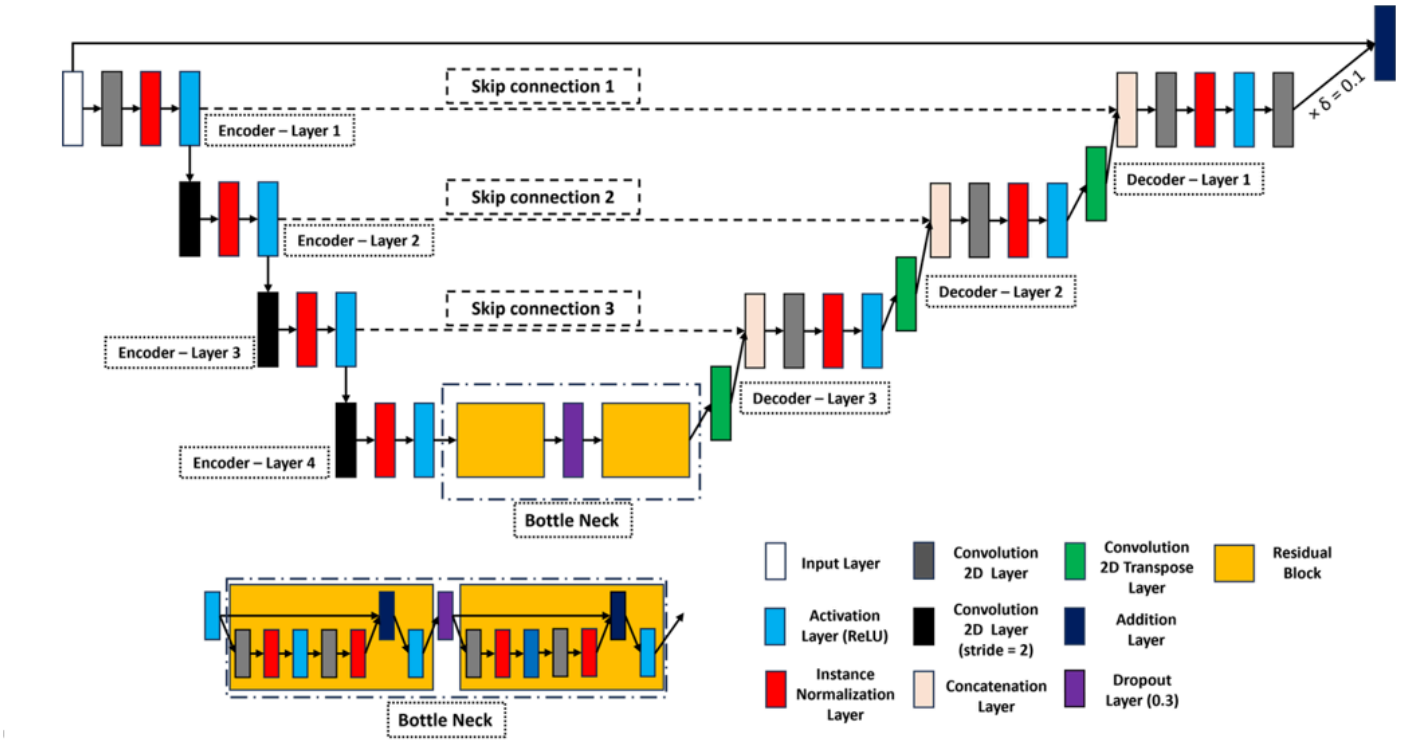


Figure 1

A schematic view of the ResUNet network. The only architectural difference between ResUNet and the generators of the PCC-cGAN is that the generator model used in the PCC-cGAN employs attention blocks in place of the standard skip connections at layers 2 and 3.

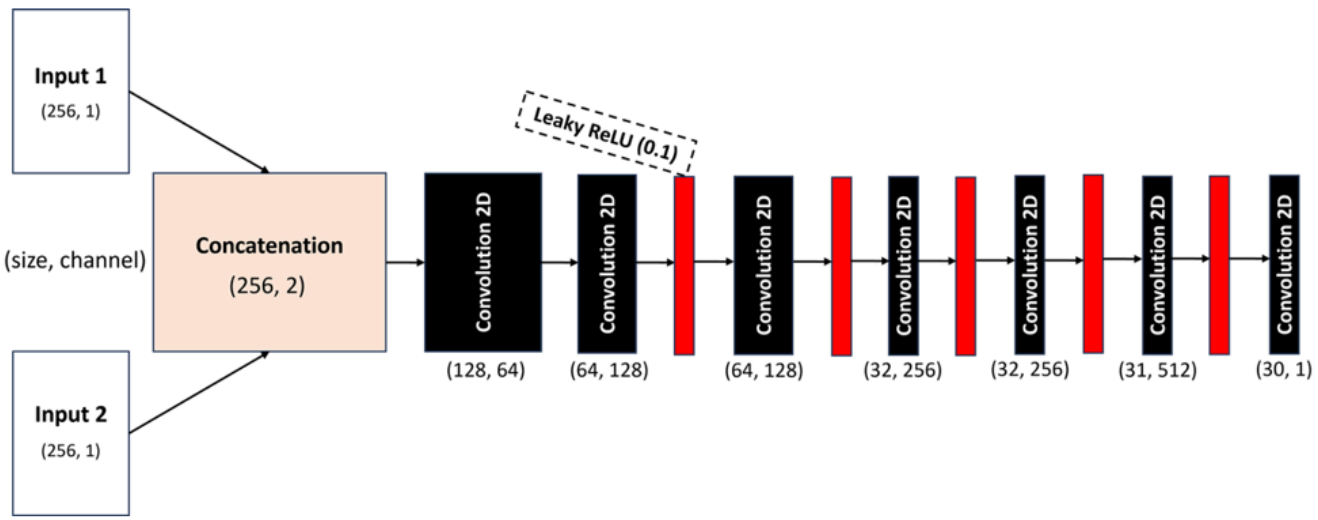


Figure 2

A schematic view of the conditional PatchGAN discriminator.

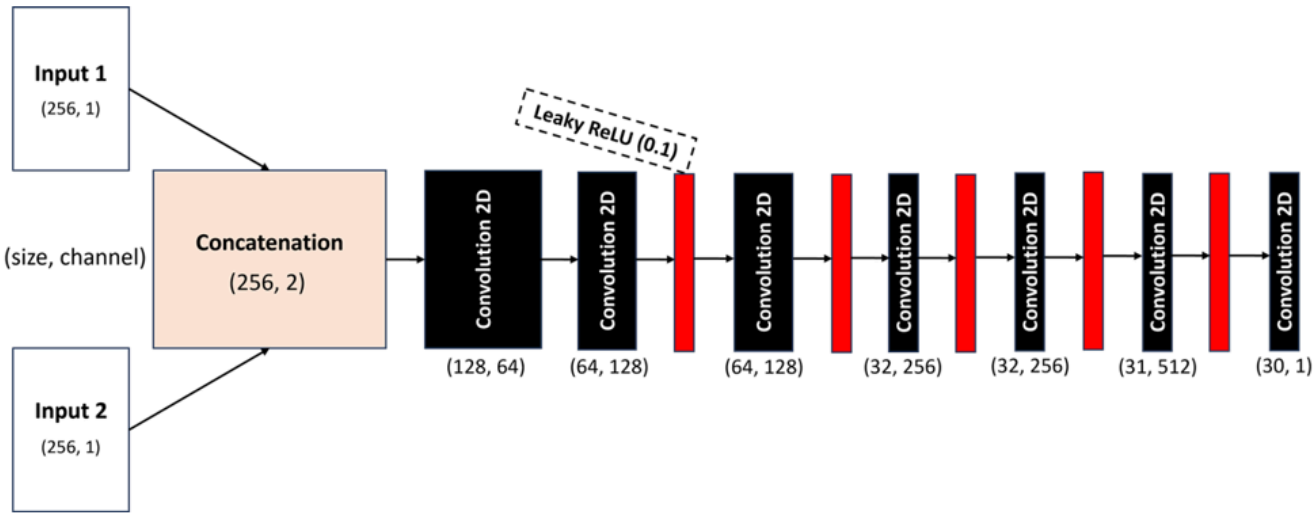


Figure 2

A schematic view of the conditional PatchGAN discriminator.

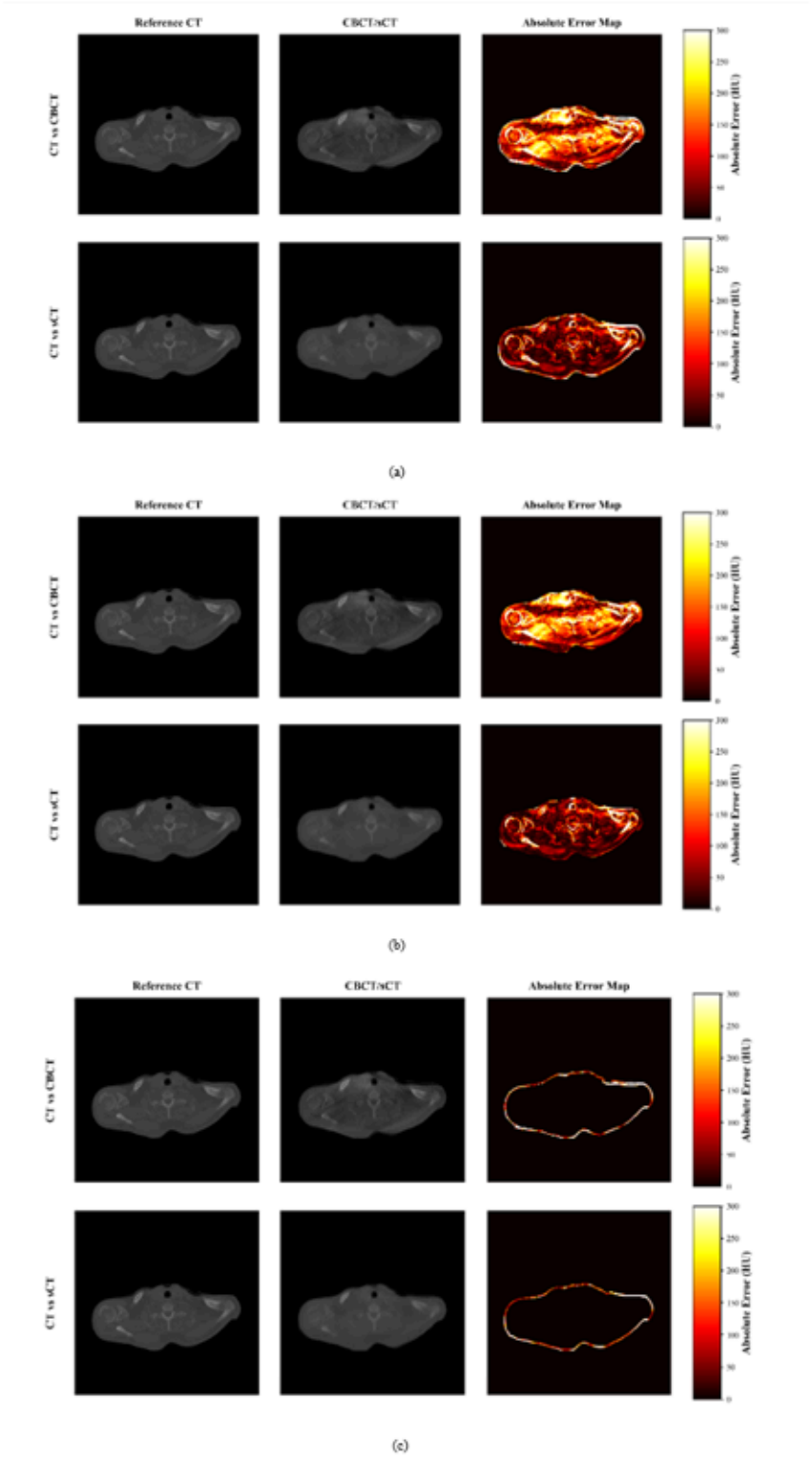


Figure 3

Reference pCT, CBCT, ResUNet sCT, and absolute error maps for a complete FOV case, overlaid with the (a) intersection, (b) eroded, and (c) peripheral masks.

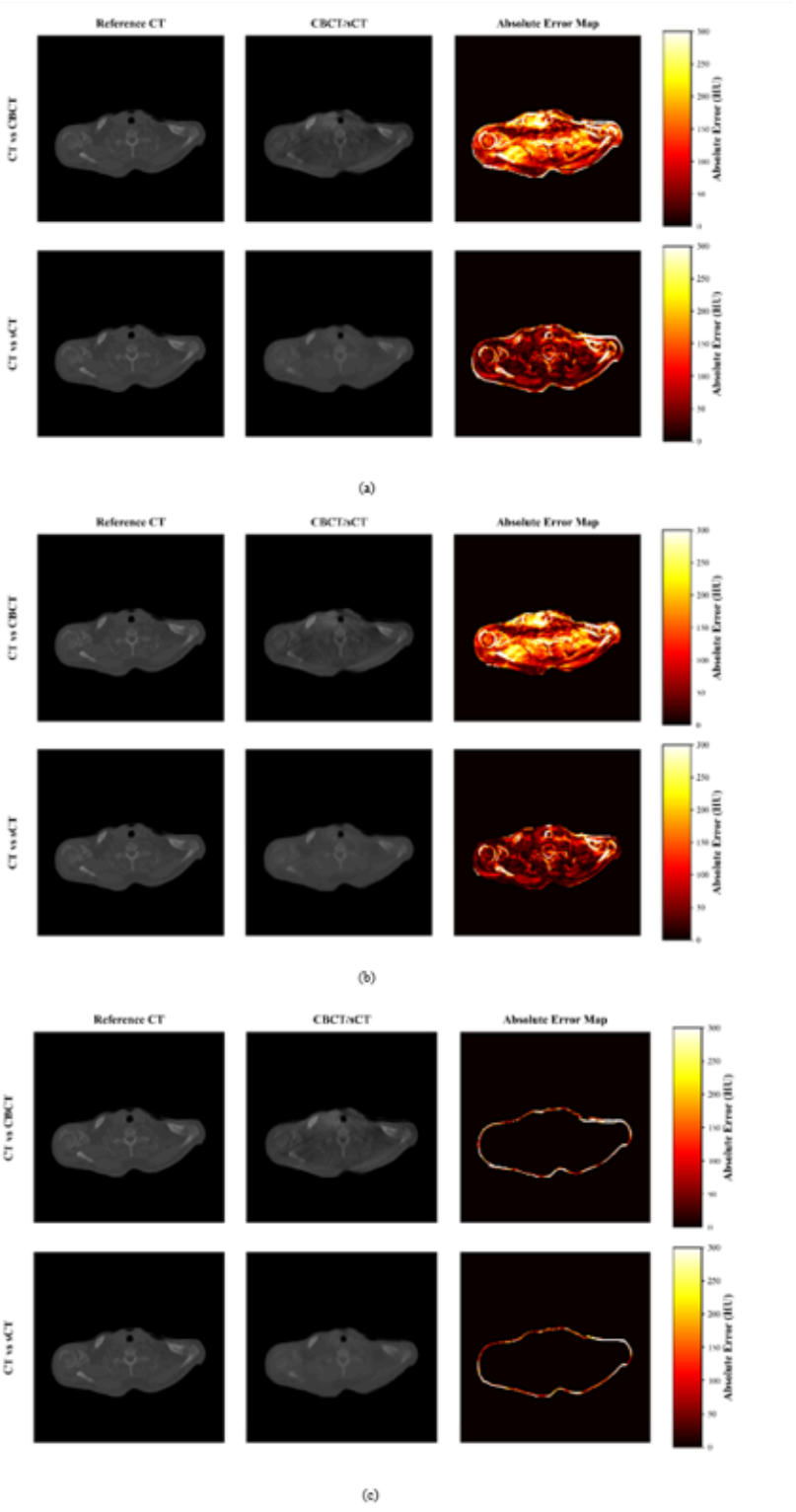


Figure 3

Reference pCT, CBCT, ResUNet sCT, and absolute error maps for a complete FOV case, overlaid with the (a) intersection, (b) eroded, and (c) peripheral masks.

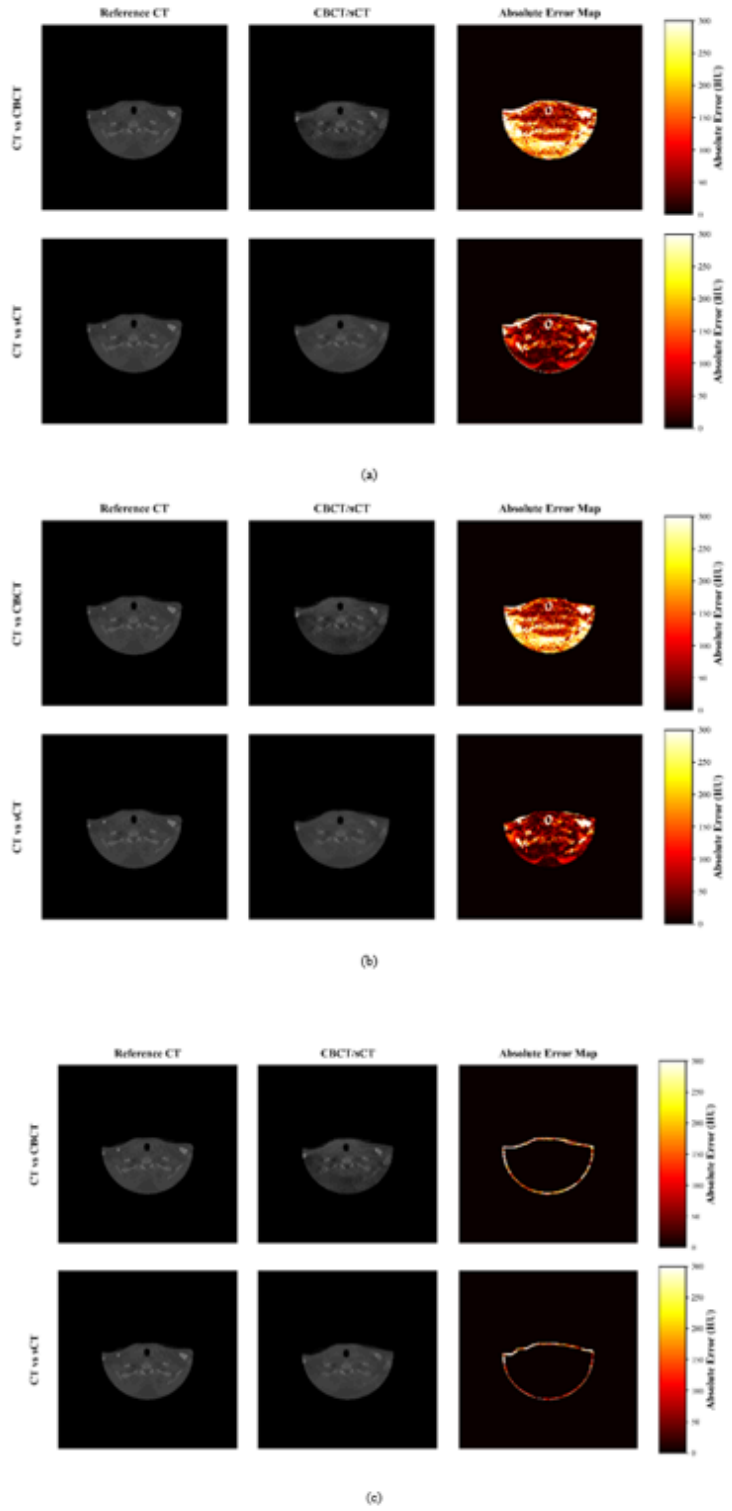


Figure 4

Reference pCT, CBCT, ResUNet sCT, and absolute error maps for a truncated FOV case, overlaid with the (a) intersection, (b) eroded, and (c) peripheral masks.

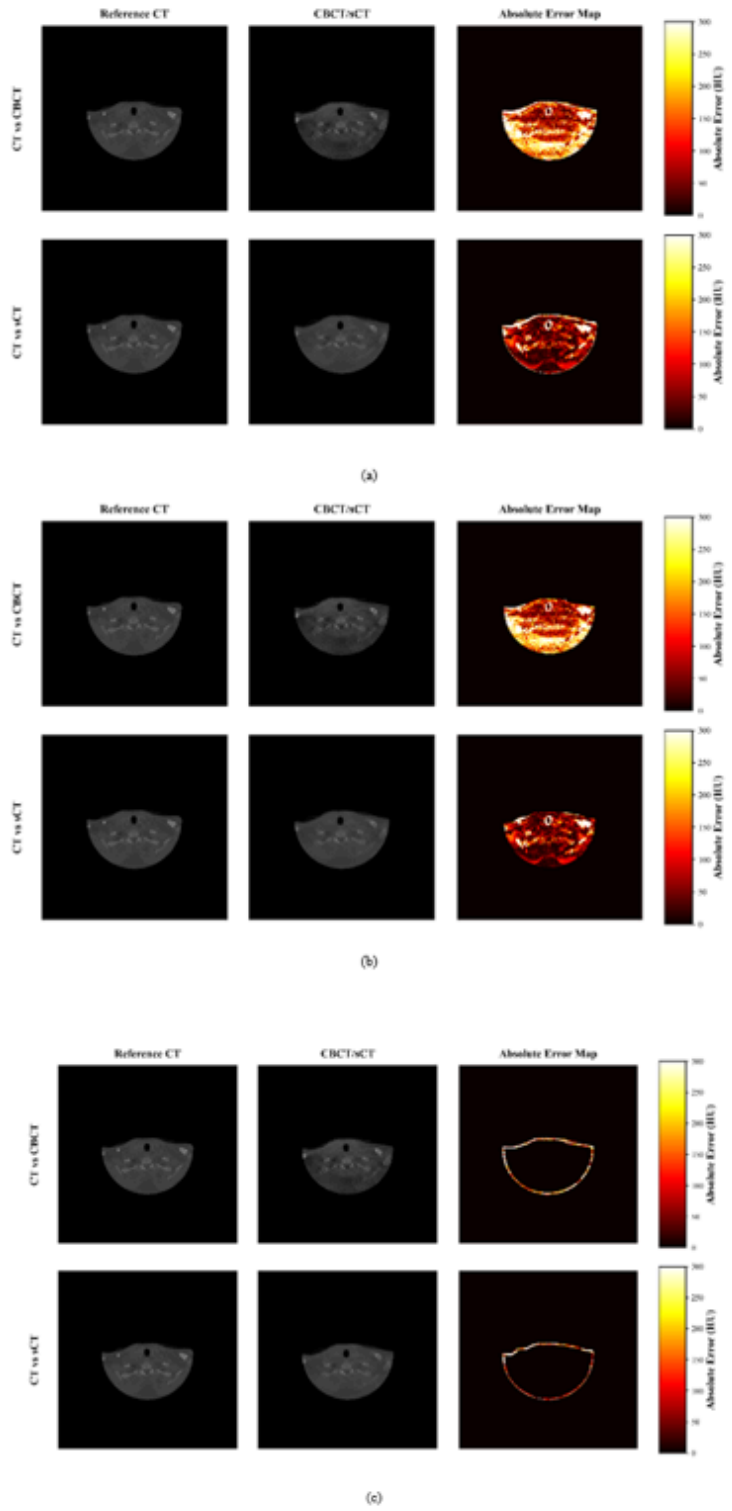


Figure 4

Reference pCT, CBCT, ResUNet sCT, and absolute error maps for a truncated FOV case, overlaid with the (a) intersection, (b) eroded, and (c) peripheral masks.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile1.docx