



OPEN

Synthetic CT generation based on CBCT using improved vision transformer CycleGAN

Yuxin Hu¹, Han Zhou^{2,4}, Ning Cao¹, Can Li^{3✉} & Can Hu^{1✉}

Cone-beam computed tomography (CBCT) is a crucial component of adaptive radiation therapy; however, it frequently encounters challenges such as artifacts and noise, significantly constraining its clinical utility. While CycleGAN is a widely employed method for CT image synthesis, it has notable limitations regarding the inadequate capture of global features. To tackle these challenges, we introduce a refined unsupervised learning model called improved vision transformer CycleGAN (IVI-T-CycleGAN). Firstly, we integrate a U-net framework that builds upon ViT. Next, we augment the feed-forward neural network by incorporating deep convolutional networks. Lastly, we enhance the stability of the model training process by introducing gradient penalty and integrating an additional loss term into the generator loss. The experiment demonstrates from multiple perspectives that our model-generated synthesizing CT(sCT) has significant advantages compared to other unsupervised learning models, thereby validating the clinical applicability and robustness of our model. In future clinical practice, our model has the potential to assist clinical practitioners in formulating precise radiotherapy plans.

Image-guided radiotherapy(IGRT) technology serves as the cornerstone of accurate radiation therapy, facilitating the real-time monitoring of imaging changes in tumors and normal organs before and during treatment^{1,2}. CBCT, as the primary imaging modality in IGRT, offers advantages such as lightweight and open architecture, along with reduced radiation exposure³. Consequently, CBCT finds extensive application in image-guided patient positioning and image registration within the clinical oncology radiation therapy workflows^{4,5}. However, CBCT often suffers from noise and artifacts, making the process of segmentation more challenging^{6,7}. Furthermore, CBCT images are less reliable compared to computed tomography (CT) when it comes to providing accurate CT/ED information⁸. These limitations hinder the clinical application of CBCT in adaptive radiotherapy. To overcome these shortcomings, we aim to enhance the quality of CBCT images and improve the accuracy of tumor treatment while ensuring better protection of organs at risk. This will be accomplished by generating synthetic CT (sCT) using the density from CT and the detailed anatomical information obtained from CBCT.

CT synthesis methods can be divided primarily into supervised and unsupervised approaches. Supervised learning relies on paired data and utilizes reconstruction loss to guide the network in generating sCT. The most representative techniques include fully convolutional networks (FCNs)^{9,10} and variations of the U-Net^{11–15}. Dong et al.¹⁰ employed FCNs for the end-to-end nonlinear mapping from MRI to CT. Kida et al.¹⁴ applied a 2D U-Net deep convolutional neural network to the pelvis to synthesize CT from CBCT. Similarly, Li et al.¹⁵ utilized an enhanced U-Net structure with residual blocks to generate CT from CBCT in the head and neck region. Xie et al.¹⁶ presented a groundbreaking scatter artifact mitigation technique employing convolutional neural networks, complemented with a context-aware loss function, specifically designed to address artifact reduction in CBCT images. Liu et al.¹⁷ employed a U-Net architecture, trained on registered images, to generate intermediate CT representations from CBCT scans, with the objective of tackling artifact challenges and paving the way for subsequent unsupervised processing. Kihwan et al.¹⁸ introduced a self-supervised model, leveraging a blind input network to map partially obscured projections back to their original form, ultimately aiming to enhance the quality of CBCT denoised reconstructions. However, obtaining fully paired data in the clinical setting is challenging due to factors such as setup errors and organ movement. Consequently, supervised methods

¹School of Computer and Software, Hohai University, Nanjing 211100, China. ²School of Electronic Science and Engineering, Nanjing University, NanJing 210046, China. ³Engineering Research Center of TCM Intelligence Health Service, School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing 210023, China. ⁴Department of Radiation Oncology, The Fourth Affiliated Hospital of Nanjing Medical University, Nanjing 210013, China. ✉email: lican@njucm.edu.cn; hucan@hhu.edu.cn

necessitate pre-processing to establish pairing prior to network training, as well as a high demand for precise image registration algorithms.

While unsupervised CycleGAN and its derivatives continue to hold relevance, alternative methods contribute significantly to image synthesis. Contrastive learning¹⁹ has emerged as a powerful technique, harnessing existing data to supervise model training in image generation tasks. Wang et al.²⁰ developed a contrast learning-based approach for CBCT scatter correction, which employs a low-pass filter to remove unnecessary image details, isolates scatter projections, and then subtracts scattering artifacts from the original projections to produce improved images, with the aim of enhancing CBCT image quality. Their DCLGAN²¹ model further refines this concept by filtering strip artifacts, using inverse coordinate transformations to address ring artifacts, and generating corrected images by subtracting these artifacts from the original CT scans, striking a balance between artifact removal and detail preservation. Liu et al.²² introduced a weakly-supervised strategy that models the degradation between low-dose and normal-dose images in the latent space, focusing on generating high-quality low-dose images. Diffusion models, too, have garnered attention, with Li et al.²³ frequency-guided diffusion model (FGDM) being a notable example. This model employs frequency-domain filters to guide the diffusion process, enabling more accurate cross-modal image transformations. Liu et al.²⁴ presented a cascaded unconditional diffusion model that integrates low-dose CT images into the diffusion process, iteratively solving multiple maximum a posteriori problems to tackle image denoising effectively. Muzaffar et al.²⁵ introduced SynDiff, an adversarial diffusion model that facilitates a gradual mapping of noise and source images to the target during conditional diffusion, demonstrating superior performance in MRI-CT conversion tasks. Lastly, Peng et al.²⁶ employed a conditional denoising diffusion probabilistic model (DDPM) with a U-net architecture incorporating residual and attention blocks. This model converts Gaussian white noise samples into target CT images conditioned on CBCT, successfully achieving synthetic CT synthesis. This paper focuses on the classical unsupervised CycleGAN-based model.

Unsupervised learning differs from supervised learning in that it does not rely on reconstruction loss but instead leverages constraints rooted in its own network structure or other structures²⁷. In the context of CT synthesis, the most common unsupervised models are cycle-consistent generative adversarial networks (CycleGAN)²⁸ and its variations. Liu et al.²⁹ employed a 3D CycleGAN with attention gates to generate CT images from non-paired pelvis data in CBCT synthesis. Harms et al.⁴ introduced residual blocks to the generator to achieve end-to-end mapping from CBCT to CT. Likewise, dense blocks^{30–33} are frequently integrated into the U-net architecture to efficiently extract and generate more accurate images. Moreover, meticulously designed loss functions are widely used to guide the process of image generation. Zeng et al.³⁴ proposed a hybrid GAN that incorporates a weakly supervised mechanism for synthesizing CT from MRI when paired data is limited.

CycleGAN and its variant methods have been extensively explored for CT synthesis. However, there are still several challenges that need to be addressed. Previous studies have shown that the generator of CycleGAN and its variants are limited to preserving and transmitting local feature information, leading to low image quality and authenticity.

To address these issues, we propose a new unsupervised method called improved vision transformer CycleGAN (IViT-CycleGAN). (1) In the generator, we introduce a U-net framework based on vision transformer (ViT) to extract and preserve essential features and detailed information. The use of skip connections helps mitigate information loss. (2) For ViT, we further incorporate a deep convolutional network in the feed-forward neural network and combine it with the self-attention mechanism of the transformer. This integration aims to automatically attend to information from different positions in the image during generation, better understanding the global structure of the image, and focusing on regions with more details, resulting in clearer and more realistic generated images. (3) To improve the training stability of the model, we introduce gradient penalization, ensuring that the network weights do not undergo significant changes when there are minor variations in the model inputs. (4) To enhance the consistency between the generated images and the source images, we add an additional loss term in the generator loss. This loss captures the differences between the source and generated images.

Methods

CycleGAN models

CycleGAN is shown in Fig. 1, where generator G aims to synthesize the input image x into a synthesized image $G(x)$ that is highly similar in quality to the image of target domain y . The discriminator D_y is designed to distinguish the real image in the target domain y from the image $G(x)$ synthesized by the generator G ³⁵.

The discriminator performs loss updating by backpropagation:

$$\mathcal{L}_{\text{disc,A}} = \mathbb{E}_{x \sim B} \ell_{\text{GAN}}(\mathcal{D}_A(\mathcal{G}_{B \rightarrow A}(x)), 0) + \mathbb{E}_{x \sim A} \ell_{\text{GAN}}(\mathcal{D}_A(x), 1), \quad (1)$$

$$\mathcal{L}_{\text{disc,B}} = \mathbb{E}_{x \sim A} \ell_{\text{GAN}}(\mathcal{D}_B(\mathcal{G}_{A \rightarrow B}(x)), 0) + \mathbb{E}_{x \sim B} \ell_{\text{GAN}}(\mathcal{D}_B(x), 1). \quad (2)$$

where ℓ_{GAN} is the classification loss function. 0 and 1 are the class labels of the generated and real images respectively³⁶.

The generators are updated by backpropagating loss from three sources: GAN loss, cycle-consistency loss, and identity-consistency loss³⁶. Using $\mathcal{G}_{A \rightarrow B}$ as an example:

$$\mathcal{L}_{\text{GAN,A}} = \mathbb{E}_{x \sim A} \ell_{\text{GAN}}(\mathcal{D}_A(\mathcal{G}_{A \rightarrow B}(x)), 1), \quad (3)$$

$$\mathcal{L}_{\text{cyc,A}} = \mathbb{E}_{x \sim A} \ell_{\text{reg}}(\mathcal{G}_{B \rightarrow A}(\mathcal{G}_{A \rightarrow B}(x)), x), \quad (4)$$

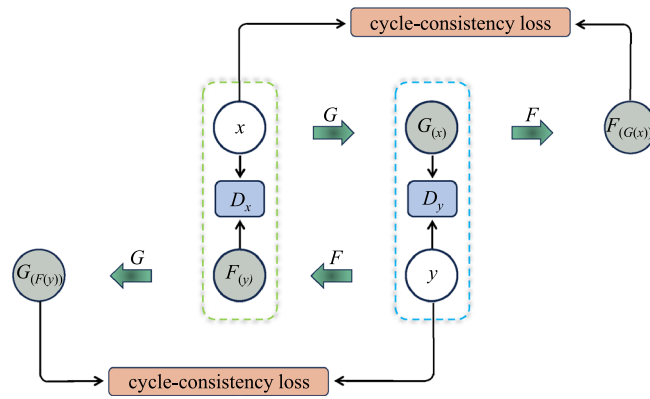


Figure 1. Schematic of the original CycleGAN model.

$$\mathcal{L}_{\text{idt},A} = \mathbb{E}_{x \sim A} \ell_{\text{reg}}(\mathcal{G}_{B \rightarrow A}(x), x). \tag{5}$$

And,

$$\mathcal{L}_{\text{gen},A \rightarrow B} = \mathcal{L}_{\text{GAN},A} + \lambda_{\text{cyc}} \mathcal{L}_{\text{cyc},A} + \lambda_{\text{idt}} \mathcal{L}_{\text{idt},A}, \tag{6}$$

$$\mathcal{L}_{\text{gen},B \rightarrow A} = \mathcal{L}_{\text{GAN},B} + \lambda_{\text{cyc}} \mathcal{L}_{\text{cyc},B} + \lambda_{\text{idt}} \mathcal{L}_{\text{idt},B}. \tag{7}$$

Here, ℓ_{reg} can be any regression loss function, to harmonize the comparison experiments, we refer to CycleGAN³⁷ parameter settings: $\lambda_{\text{cyc}}=10$ and $\lambda_{\text{idt}}=0.5$ ³⁶.

ViT-CycleGAN architecture

The original generator of CycleGAN can only retain and convey local feature information, lacking the ability to capture global features, thereby resulting in subpar image quality and authenticity. To address this limitation, this research incorporates a ViT-based U-net framework into the generator, as depicted in Fig. 2.

Firstly, the U-net architecture is employed to extract and retain crucial organizational features and detailed information, effectively resolving the issue of information loss through the utilization of skip connections. Subsequently, the self-attention mechanism of the transformer is employed to automatically prioritize information from various positions within the image during image generation, enhancing the comprehension of the global structure within organizational images. Lastly, a deep convolutional network is introduced into the feedforward neural network to concentrate on regions with more intricate details, resulting in clearer and more realistic

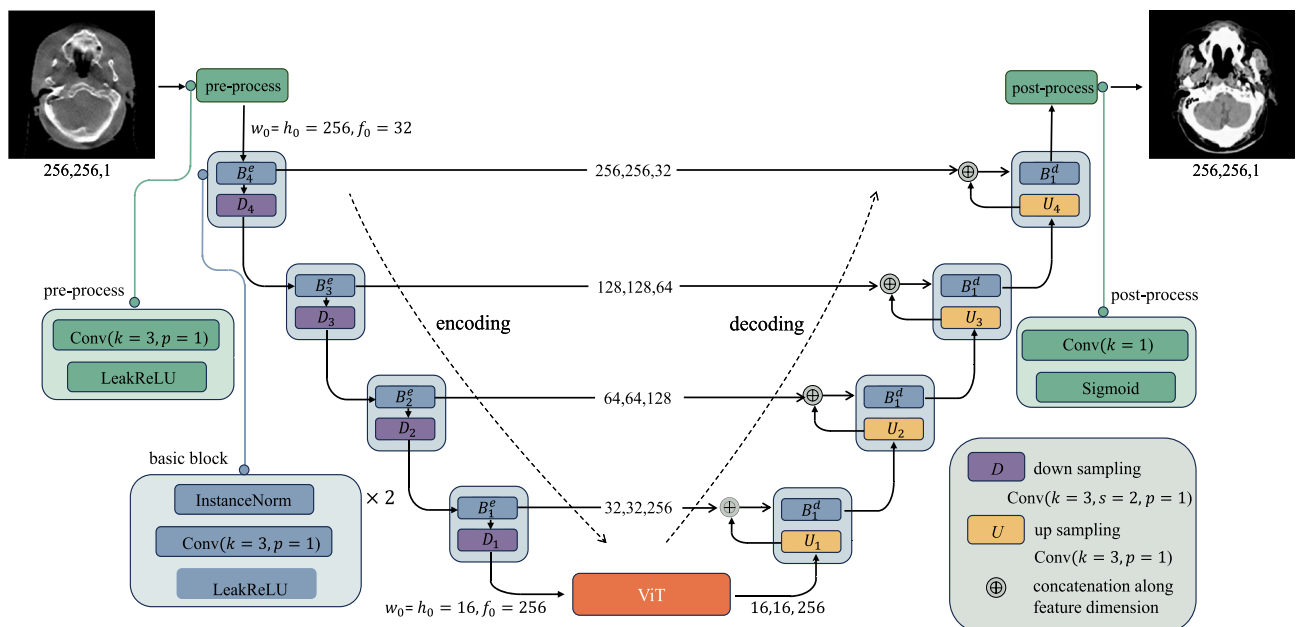


Figure 2. IViT-CycleGAN model generator section.

generated images. Specifically, the coding path of U-net extracts features from the input through four layers of convolution and downsampling, and passes the extracted features from each layer to the corresponding layer of the decoding path through skip connections. In the encoding path of U-net, the preprocessing layer converts the image into a tensor with dimensions (w_0, h_0, f_0) , and the preprocessed tensor halves the width w_0 and the height h_0 in each downsampled block while the feature dimension f_0 is doubled³⁶.

For the ViT module, as shown in Fig. 3, ViT is composed primarily of a stack of transformer encoder blocks. To construct an input to the stack, the ViT first flattens an encoded image along the spatial dimensions to form a sequence of tokens. The token sequence has length $w \times h$, and each token in the sequence is a vector of length f . It then concatenates each token with its two-dimensional Fourier positional embedding of dimension f_p and linearly maps the result to have dimension f_v ³⁶.

For the feedforward neural network, as shown in Fig. 4, we use a deep convolutional network instead of the original fully connected layer. The input, i.e. a sequence of tokens is first reshaped to a feature map rearranged on a 2D lattice. Then two 1×1 convolutions along with a depth-wise convolution are applied to the feature map³⁸. After that, the feature map is reshaped to a sequence of tokens which are used as by the self-attention of the network transformer layer. To improve the Transformer convergence, we adopt the rezero regularization scheme and introduce a trainable scaling parameter α that modulates the magnitudes of the nontrivial branches of the residual blocks. The output from the transformer stack is linearly projected back to have dimension f and unflattened to have width w and h . In this study, we use 12 transform encoder blocks³⁶.

Discriminator loss with gradient Penalty (GP)

To improve the training stability, we introduce a generalized GP^{36,39} form with the following \mathcal{D}_A loss formula:

$$\mathcal{L}_{\text{disc,A}}^{\text{GP}} = \mathcal{L}_{\text{disc,A}} + \lambda_{\text{GP}} \mathbb{E} \left[\frac{(\|\nabla_x \mathcal{D}_A(x)\|_2 - \gamma)^2}{\gamma^2} \right] \tag{8}$$

where $\mathcal{L}_{\text{disc,A}}$ is defined as in Eq. 1, and $\mathcal{L}_{\text{disc,B}}$ follows the same form. In our experiments, this γ -centered GP regularization provides more stable training and less sensitive to the hyperparameter choices³⁶.

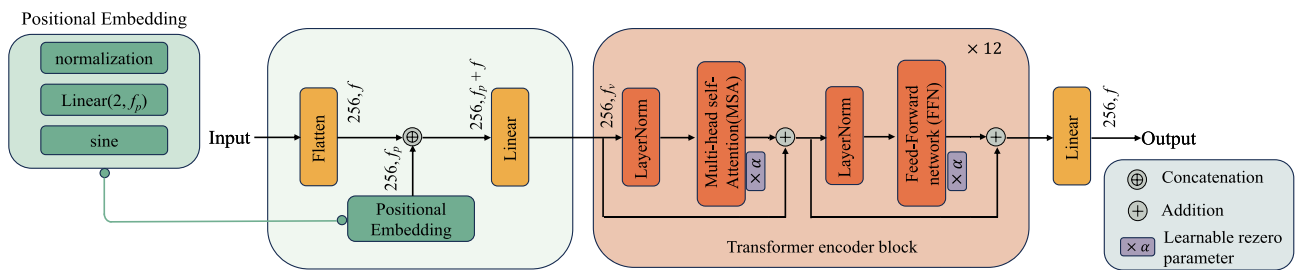


Figure 3. ViT module in IViT-CycleGAN.

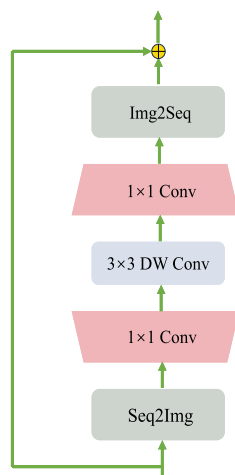


Figure 4. Depth-Wise Convolution Network. DW denotes depth-wise convolution. To cope with the convolution operation, the conversion between sequence and image feature map is added by Seq2Img and Img2Seq³⁸.

Pixel-wise consistency loss

To improve the consistency of the generated and source images, we experiment with the addition of an extra term $\mathcal{L}_{\text{consist}}$ ⁴⁰ to generator loss. This term captures the \mathcal{L}_1 difference between the downsized versions of the source and translated images. For example, for images of domain A :

$$\mathcal{L}_{\text{consist},A} = \mathbb{E}_A \ell_1(F(\mathcal{G}_{A \rightarrow B}(a)), F(a)) \quad (9)$$

where F is a resizing operator down to 32×32 pixels (low-pass filter). We add this term to the generator loss with a magnitude λ_{consist} for both domains⁴⁰.

Ethical statement

We confirmed that all methods were carried out in accordance with relevant guidelines and regulations, and informed consent for patients was waived by the Research Ethics Committee of the Nanjing Medical University. All experimental protocols and data in this study were approved by the Research Ethics Committee of the Nanjing Medical University. Approval number: NMUE2021301.

Experiments

Data acquisition

In this study, we test our proposed method in two datasets provided by a cooperative tertiary hospital.

H & N dataset

The CBCT and CT images were selected from 30 patients who were received volume modulated arc therapy (VMAT) in the head and neck (H & N) for nasopharyngeal and hypopharyngeal (NPC, HC) cancer from October 1, 2021 and September 1, 2023. The CT volumes were obtained with the dimensions 512×512 on the axial plane with a pixel size of $0.625 \times 0.625 \text{ mm}^2$ and a slice thickness of 2.75 mm using GE discovery positioning system. The CBCT volumes were obtained using Elekta XVI Systems. The applied images protocol were the following parameters: 200 degrees gantry rotation, 100 kVp, 10 mA, 10 ms, and F0S20 collimator. And images had a size of 384×384 on the axial plane. Every patient contain with 1 planning CT volume taken in the positioning before treatment and 3 CBCT volumes every week taken between treatment. We randomly divided the training set and test set according to 8:2.

Chest dataset

The Chest dataset had the same acquisition time as the H & N dataset. The CT and CBCT parameters had few differences. It consisted of 30 patients, each with 1 planning CT volume taken before treatments and 3–5 CBCT volumes per week taken between treatments. The CT volumes were obtained with the dimensions 512×512 on the axial plane with a pixel size of $0.625 \times 0.625 \text{ mm}^2$ and a slice thickness of 5 mm using GE discovery positioning system. And the CBCT images were acquired with the following parameter: 360 degrees gantry rotation, 100 kVp, 10 mA, 10 ms, and F0M10 collimator. and the CBCT volumes were reconstructed at medium-resolution ($1 \times 1 \times 1 \text{ mm}^3$ voxels) on a $410 \times 410 \times 120$ matrix. We randomly divided the training set and test set according to 8:2.

Data processing

During the scanning process of CBCT and CT, non-human structures (such as treatment beds, fixation devices, and masks) are captured in the resulting images. These structures not only impede model training speed but also compromise the quality of synthesized images. To mitigate these issues, denoising is essential to eliminate the interference of irrelevant information prior to model training. In this study, the outlines of CT contours manually annotated by doctors serve as masks. These masks are subsequently multiplied with the corresponding CT images to generate clean CT images. Likewise, the masks are also multiplied with the CBCT images to produce clean CBCT images suitable for training purposes. Distinctive characteristics distinguish CBCT from CT, encompassing variations in imaging hardware, clinical protocols, and scanning configurations. Matching scans from the same patient often presents a challenge due to the inherent differences. Recognizing the stability of organ positions and reduced tissue mobility during data acquisition, we leveraged the open-source advanced normalization tools (ANTs) for affine registration, with the primary objective of ensuring alignment between each CBCT and CT pair for the purpose of model test.

Evaluation

To accurately compare the similarity between the sCT images generated by different models and the CT images, we introduced quantitative evaluation metrics such as mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). A lower MAE value indicates less difference between sCT and CT, resulting in more realistic image generation. Conversely, higher values of PSNR and SSIM indicate greater similarity between sCT and CT, leading to higher construction quality and more realistic images. These metrics are defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - G_{CBCT \rightarrow CT}(y_i)| \quad (10)$$

n denotes the number of testing slices, and x and y denote the CT and the CBCT, respectively. The CBCT generates the predicted CT after the CBCT generator $G_{CBCT \rightarrow CT}$, and then calculates the absolute value error with the CT.

$$PSNR = 20 \log_{10} \frac{MAX_{rCT}}{MSE} \quad (11)$$

MAX_{rCT} denotes the maximum pixel value in the sCT, and MSE is the mean square error. a larger PSNR indicates a higher similarity between the generated CT and the real CT, which means that the quality of the generated CT is better.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (12)$$

where x and y represent the CT and the fake-CT generated by the CBCT after the generator $G_{CBCT \rightarrow CT}$, respectively, and μ_x and μ_y represent the mean of x and y , σ_x^2 and σ_y^2 represent the variance of x and y , and σ_{xy} represents the variance of x and covariance of y , while c_1 and c_2 are the two constants used to maintain stability. The value of SSIM ranges from -1 to 1 , and the larger the value, the more similar the two images are.

Network training

In the experiments, all images are normalized to $[-1,1]$ and resized to 256×256 . we train the generator five times and then train the discriminator once. Its parameters are set to $epoch = 200$ and $batch\ size = 5$. The other comparison methods are implemented based on the code and details provided by the authors and have the same hyperparameter settings as ours. All algorithms in this study were implemented on a Linux system equipped with four NVIDIA Tesla V100s using Python 3.6 (<https://www.python.org/downloads/release/Python-362/>) and Tensorflow 1.14 (<https://tensorflow.google.co.uk/versions>) implementations. Figure 5 shows a plot of the discriminator loss function versus the number of iterations on the two datasets, H & N and Chest, which shows that the proposed method in this paper (Ours) has faster convergence and is better trained.

Result and discussion

Comparison of different methods

Tables 1 and 2 present a comparison of quantitative results between our method and several contrast algorithms. Our proposed method demonstrates superior performance to CycleGAN and its variations in all three evaluation metrics mentioned above on the H & N dataset. This is attributed to the introduction of an improved ViT-based U-net framework. This framework enables the extraction and preservation of essential features and detailed information, automatic focus on information from different positions in the images, better comprehension of the global structure of the images, and emphasis on regions with more details. Consequently, the generated images are clearer and more realistic. Furthermore, we apply our method to the Chest dataset without altering any parameters and report the experimental results in Table 2. Our method exhibits considerable advantages on the Chest dataset as well, outperforming CycleGAN and its variations in the evaluated metrics. The key factor behind this success is the incorporation of GP and additional pixel-wise consistency loss, which enhance the stability and robustness of the model. These experiments validate the applicability of our method not only to the head and neck region but also to other parts of the body.

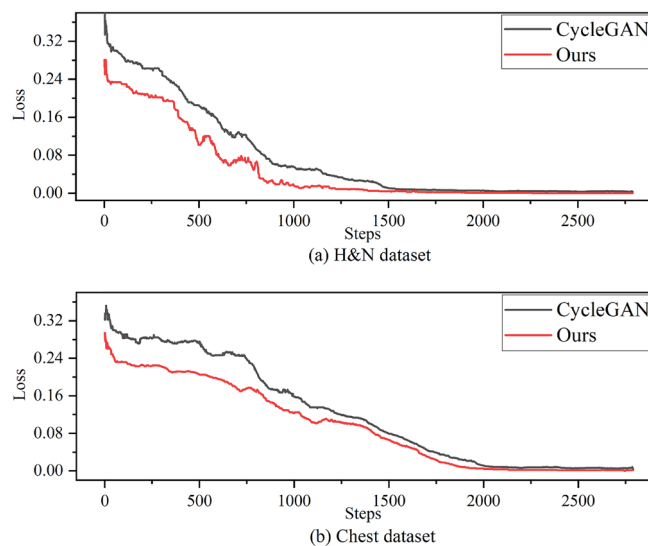


Figure 5. Plot of loss function vs. number of iterations on both datasets.

Method	MAE ↓	PSNR ↑	SSIM ↑
CycleGAN ⁴	15.13	29.1536	0.9712
DualGAN ⁴¹	18.59	28.5969	0.9624
AttentionGAN ²⁹	25.68	27.1325	0.9513
RegGAN ⁴²	14.74	29.8562	0.9742
ADCycleGAN ³⁵	13.56	30.1598	0.9762
Ours	10.94	31.8567	0.9834

Table 1. Comparison of metrics of different methods on the H &N dataset. Significant values are in bold.

Method	MAE ↓	PSNR ↑	SSIM ↑
CycleGAN ⁴	20.89	27.6615	0.9612
DualGAN ⁴¹	27.46	25.3512	0.9429
AttentionGAN ²⁹	23.29	26.1259	0.9526
RegGAN ⁴²	17.36	28.1523	0.9631
ADCycleGAN ³⁵	15.12	28.4563	0.9649
Ours	13.46	29.1289	0.9715

Table 2. Comparison of metrics of different methods on the Chest dataset. Significant values are in bold.

Ablation studies for IViT-CycleGAN

To thoroughly assess the efficacy of each module, we employed CycleGAN-ViT as our backbone and conducted a module stacking analysis to gauge their individual impact on the overall performance. The abbreviations used are as follows: Depth-Wise Convolution Network (DCN), Discriminator Loss with Gradient Penalty (GP), and Pixel-wise Consistency Loss (PL). The results from the H &N dataset are compiled in Table 3, while those from the Chest dataset are presented in Table 4.

Method	MAE ↓	PSNR ↑	SSIM ↑
CycleGAN-ViT	14.88	29.2561	0.9726
CycleGAN-ViT-DCN	13.12	29.8626	0.9768
CycleGAN-ViT-GP	13.48	29.3915	0.9752
CycleGAN-ViT-PL	14.64	29.3321	0.9744
CycleGAN-ViT-DCN-GP	11.78	31.4931	0.9825
CycleGAN-ViT-DCN-PL	12.62	30.3524	0.9788
CycleGAN-ViT-GP-PL	12.33	31.0162	0.9808
Ours	10.94	31.8567	0.9834

Table 3. Quantitative results for ablations based on CycleGAN-ViT in H &N dataset. Significant values are in bold.

Method	MAE ↓	PSNR ↑	SSIM ↑
CycleGAN-ViT	16.88	27.9823	0.9633
CycleGAN-ViT-DCN	15.02	28.5621	0.9684
CycleGAN-ViT-GP	15.35	28.3589	0.9667
CycleGAN-ViT-PL	16.12	28.0124	0.9641
CycleGAN-ViT-DCN-GP	13.79	28.9652	0.9708
CycleGAN-ViT-DCN-PL	14.33	28.6814	0.9699
CycleGAN-ViT-GP-PL	14.01	28.7714	0.9701
Ours	13.46	29.1289	0.9715

Table 4. Quantitative results for ablations based on CycleGAN-ViT in Chest dataset. Significant values are in bold.

The experimental findings substantiate the contributions of each module to the overall performance. DCN due to its inherent local properties, complements the self-attention mechanism in ViT, enabling it to engage in both global and regional information exchange. This local focus facilitates the extraction of finer details, resulting in more vivid and realistic generated images. GP, through regularization controlled by the parameter β , improves model stability during training. PL, by measuring the L1 difference between source and generated images, enhances consistency, thereby enhancing image generation quality. A comparative analysis of metrics, including MAE, PSNR and SSIM, reveals that DCN exhibits the most significant performance boost compared to GP and PL.

Visualization

In addition to quantifying the sCT using the aforementioned evaluation indicators, we incorporate visualization techniques to explore the results from various perspectives and validate the effectiveness of our proposed method by comparing the outputs of different models. In Fig. 6, we present a comparison of synthesis results obtained from six algorithms, namely CycleGAN, DualGAN, AttentionGAN, RegGAN, ADCycleGAN, and Ours, in H &N patients. These results showcase the generated images of the cervical Bone (marked by green arrows), Nasopharynx (marked by blue arrows), Pituitary (marked by yellow arrows), and Eyes (marked by orange arrows) in a sequential left-to-right and top-to-bottom manner. Upon closer examination of the magnified images, it is evident that the alternative algorithms yield images with increased noise levels and significant loss of lesion details. In contrast, our proposed method generates images characterized by minimal noise, enhanced details accuracy, and closer resemblance to Real CT results. This remarkable outcome can primarily be attributed to the incorporation of the ViT-based U-net framework, which excels in feature extraction while preserving crucial detailed information. The framework also demonstrates improved comprehension of the image's global structure, resulting in the production of images that are significantly clearer and more realistic.

We also conducted a comparative analysis of the synthesis results obtained from the six algorithms on the Chest dataset, as shown in Fig. 7. The first and second rows illustrate the Bronchi (highlighted by green arrows) and the Sternum (highlighted by blue arrows) as the representative anatomical structures, respectively, while the third and fourth rows depict the lung tumor regions. A thorough examination of the partially enlarged images reveals that our method applied to the Chest dataset produces consistent outcomes with those observed in the H &N dataset. Specifically, the generated images exhibit diminished noise, enhanced detail accuracy, and a greater resemblance to Real CT scans. Notably, our proposed algorithm demonstrates a more striking similarity to Real CT scans in the tumor regions, which proves instrumental in discerning changes within the tumor areas and offering valuable image references for adaptive radiotherapy. The Hounsfield Units (HU) CT values, reflecting tissue density and X-ray absorption, were analyzed in the test set slices ranging from -500 to 500 HU. Figure 8 exhibits a comparative histogram of HU values for our method and CBCT, revealing that the curve shapes and peak positions of our approach more closely resemble those of real CT scans, suggesting a certain level of accuracy in the generated sCT. Figures 9 and 10 present difference plots comparing our method, CBCT, and ground truth CT. Utilizing a rainbow color mapping, with blue indicating minimal difference and red indicating maximum, the plots demonstrate that the discrepancies between our method and CT are significantly smaller than those between CBCT and CT, indicating that the sCT generated in this study achieves a CT-like quality to a considerable extent.

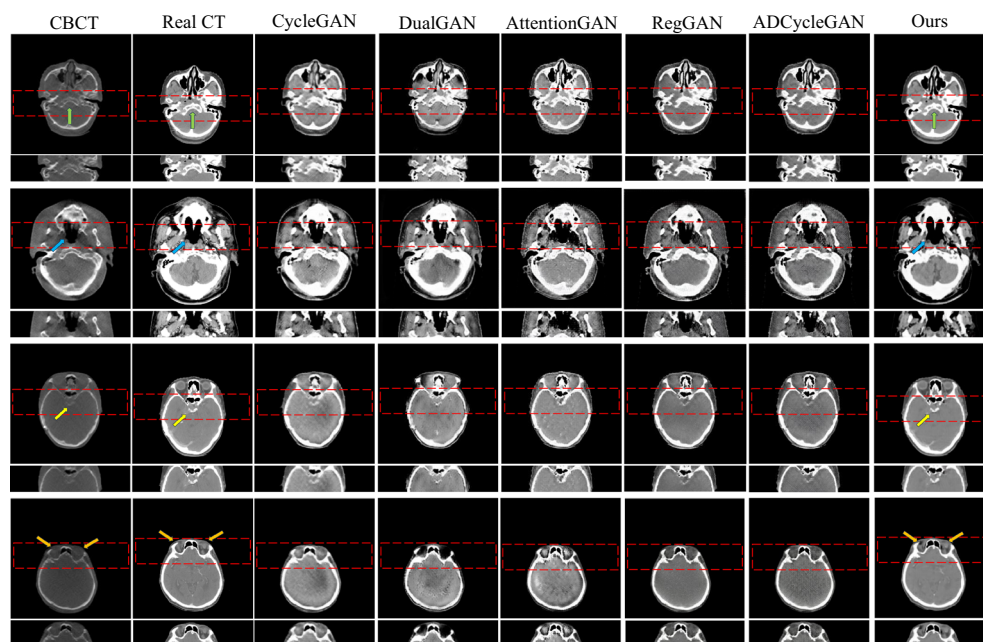


Figure 6. Visualization of sCT generation details on the H &N dataset test set.

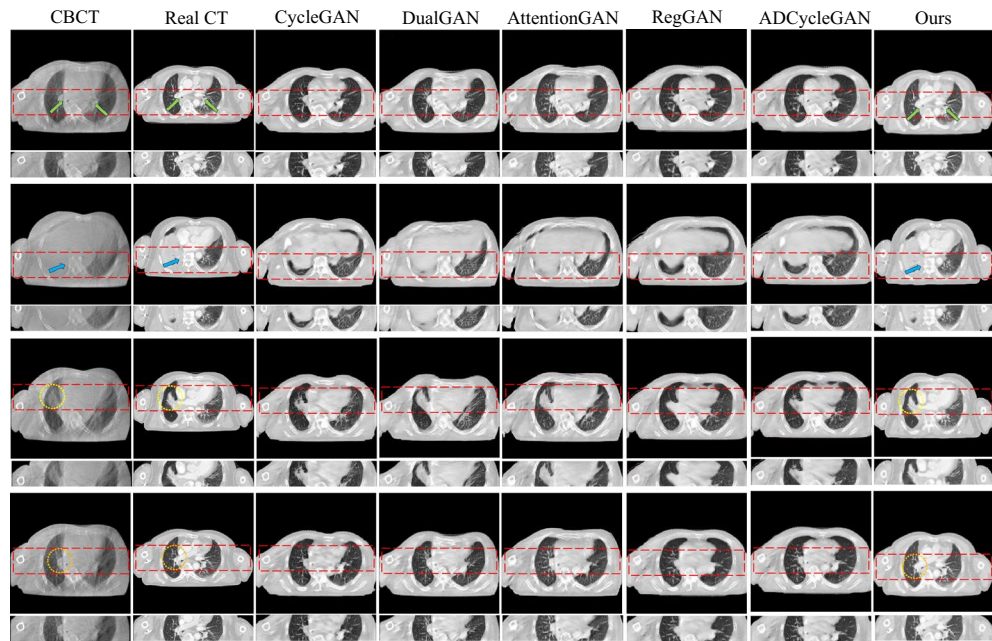


Figure 7. Visualization of sCT generation details on the Chest dataset test set.

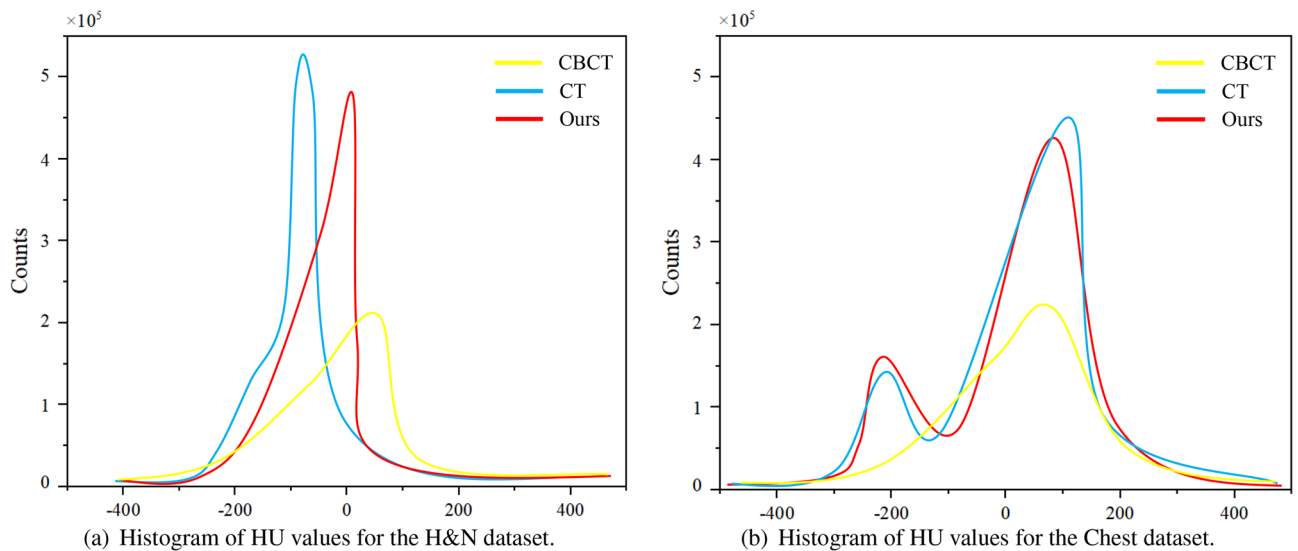


Figure 8. Histograms of HU values for the two types of datasets.

In addition to comparing the generated details, we also assess the performance of different methods using CT values. We use each pixel on the vertical and horizontal axes as a unit and calculate the average CT value at each pixel of all the data in the entire test set. Figures 11 and 12 showcase the distribution of average CT values across the vertical and horizontal axes. The x-axis denotes the pixel position, while the y-axis represents the average pixel value. The purple curve corresponds to the CT value distribution curve of Real CT, the red curve represents our proposed method, and the remaining colors indicate other methods. The outcomes highlight the congruence of CT value distribution trends between Real CT and the other five methods. Notably, our CT value distribution curve bears a stronger resemblance to Real CT when juxtaposed with the curves of the other five methods. Meanwhile, the CT values obtained from the other methods slightly surpass those of Real CT. In essence, our method generates sCT images that closely approximate Real CT, thus rendering them more authentically realistic.

Visualization of the synthetic CT 3D reconstruction

The CT’s imaging process, characterized by its distinct mechanism, yields two-dimensional data in the form of X-ray-derived slices. In clinical settings, three-dimensional reconstructions are crucial for multi-faceted analysis of patient lesions, facilitating accurate diagnosis and treatment. To evaluate the fidelity of our sCT, we conducted

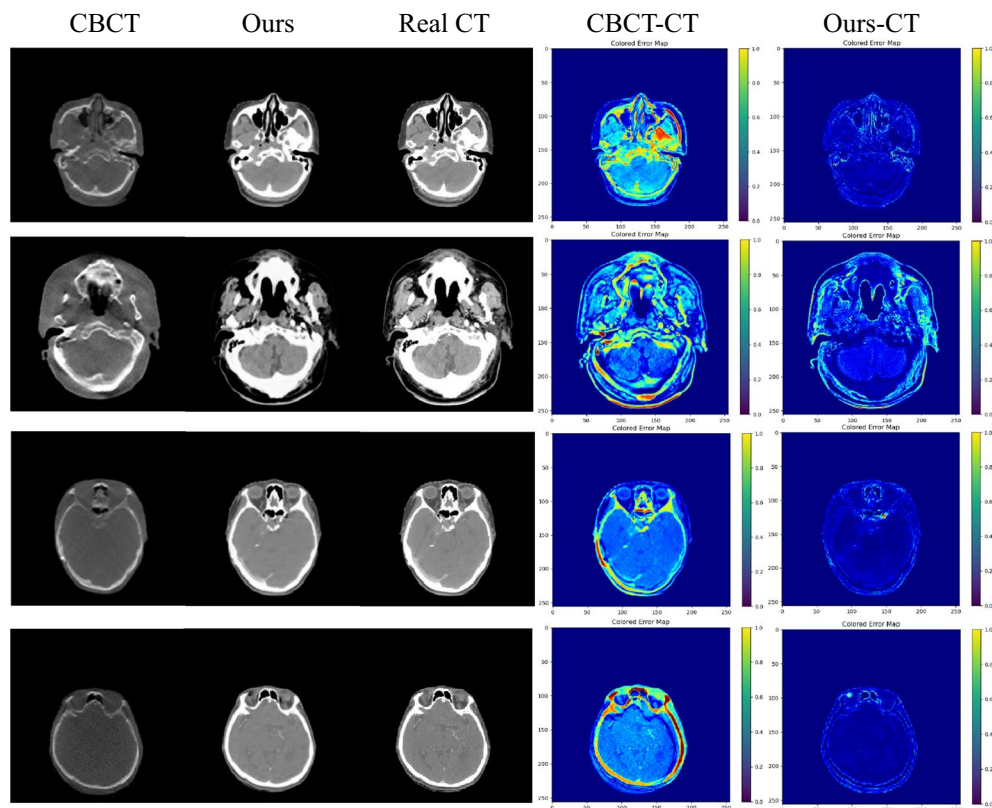


Figure 9. Error plots of CBCT vs. CT and the method of this paper vs. CT under the H & N dataset.

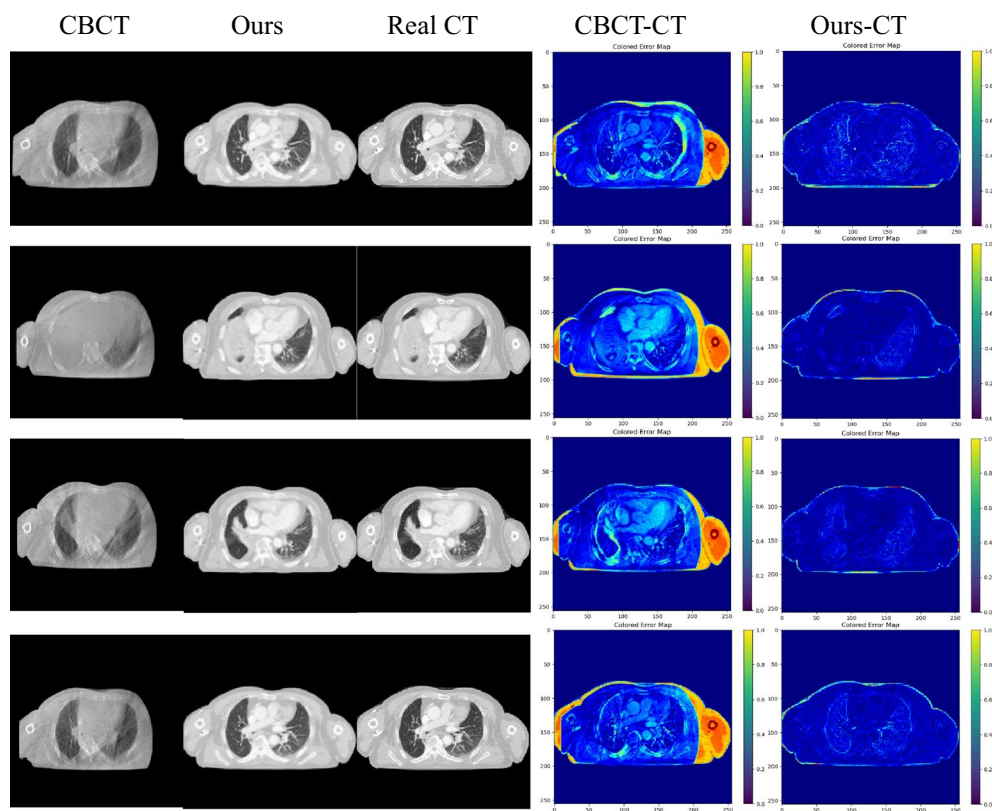


Figure 10. Error plots of CBCT vs. CT and the method of this paper vs. CT under the Chest dataset.

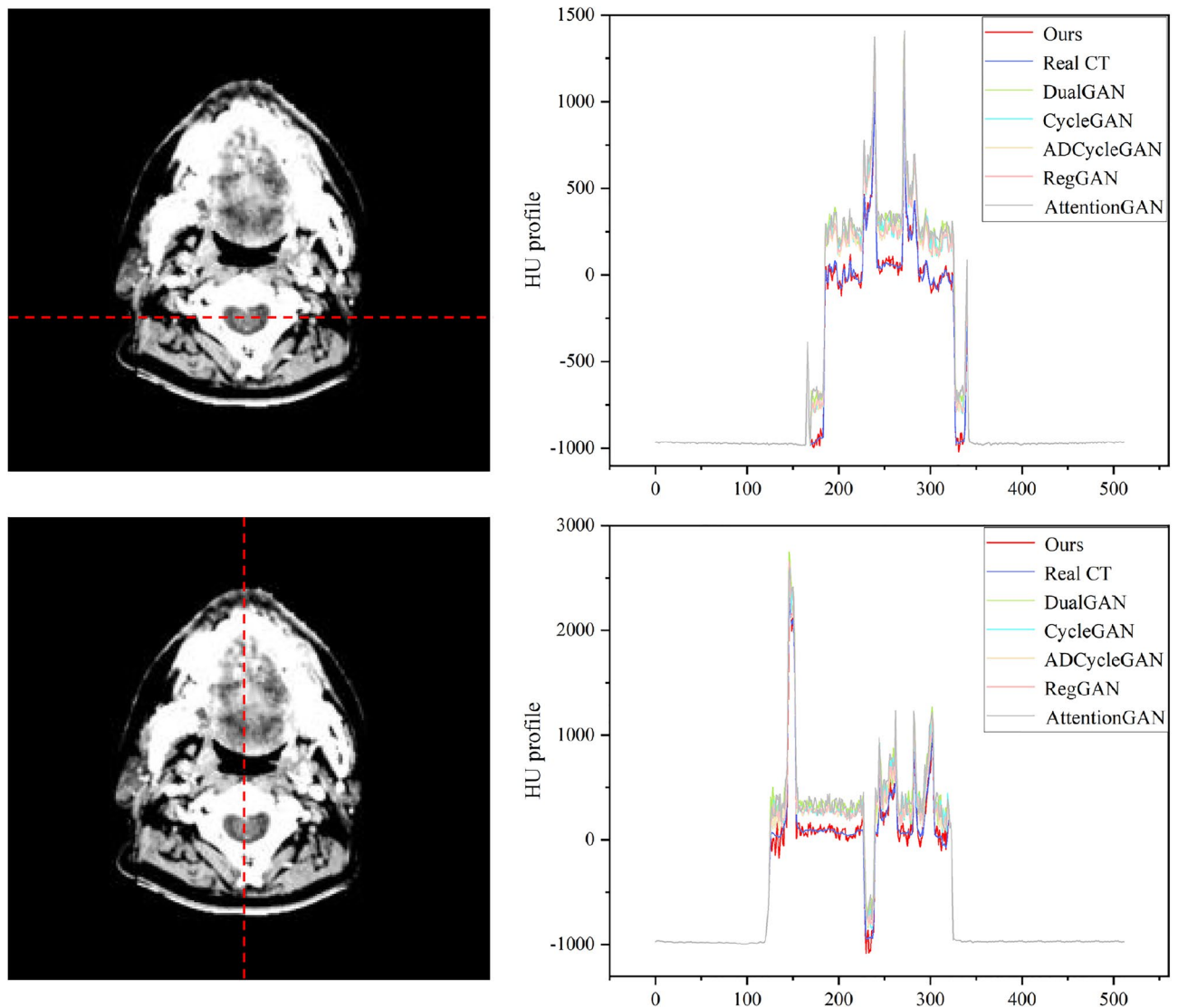


Figure 11. Visualization of CT value distribution on the H & N dataset test set.

3D reconstructions, focusing on the consistency across dimensions. Figures 13 and 14 illustrate the results for the H & N and Chest datasets, featuring axial, sagittal, and coronal perspectives. Our method generates the axial view, while the sagittal and coronal views are reconstructed from the sCT. The reconstructed slices from these datasets evidence that the generated sCT successfully retains the original anatomical integrity, ensuring a consistent representation of internal organ structures. For the Dose and Volume Histogram (DVH) on the right side of Figs. 13 and 14, where the solid line represents Real CT and the dashed line represents the Ours method, it can be seen that the close proximity of our method to the real clinical dose distribution validates the clinical applicability of our method.

Dose calculation

The primary purpose of sCT is to serve as a foundation for subsequent clinical tasks, particularly dose calculation. Hence, dose calculation offers the most precise approach to verifying the effectiveness of sCT generation and its clinical suitability. To this end, we conducted a comparison between the sCT generated by our method and the Real CT across different dose levels. The resulting discrepancies were then visualized in three-dimensional displays, as presented in Figs. 15 and 16. On the left side of Fig. 15, the differences between our method and the Real CT treatment plan for a nasopharyngeal cancer patient under different dose distributions are shown. On the right side of Fig. 15, the differences in DVH for the patient are displayed, with the solid line representing Real CT and the dashed line representing our method. It can be observed from the left side of Fig. 15 that our method closely approximates the actual clinical dose distribution under different dose distributions. Examination of the DVH on the right side of Fig. 15 reveals nearly no disparity in the preventive dose for the nasopharyngeal target area and lymphatic drainage area. The experiment successfully validates the clinical applicability of our method. Furthermore, for Chest patients, we observed the discrepancies between our method and Real CT in sCT under different dose distributions. Based on the 3D dose distribution and DVH in Fig. 16, it can be concluded that

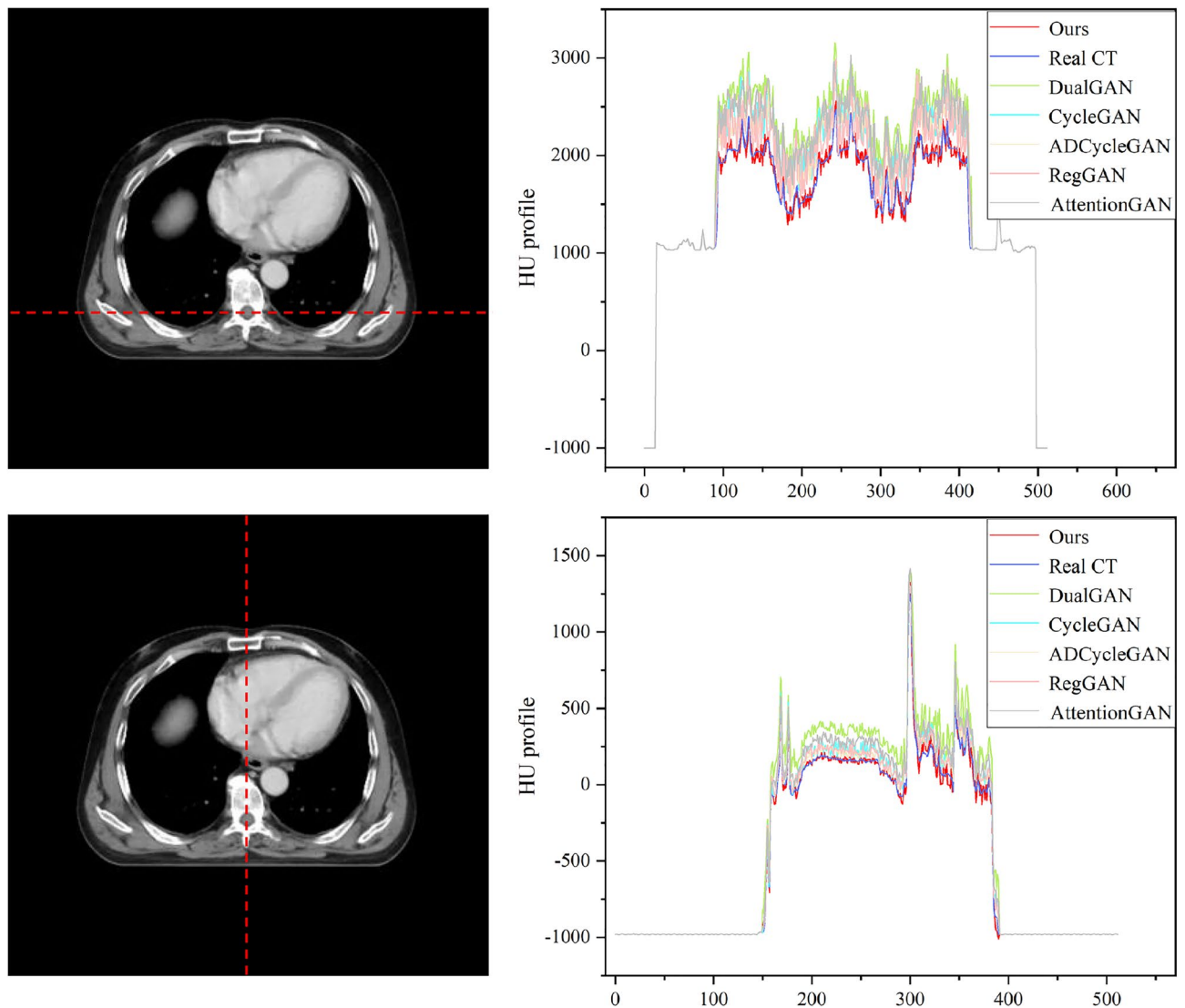


Figure 12. Visualization of CT value distribution on the Chest dataset test set.

there is also a slight disparity in the dose received by the target area and lung tissue, thereby further confirming the clinical applicability and robustness of our method.

Discussion

This research introduces IViT-CycleGAN, an unsupervised learning model designed to synthesize sCT from CBCT data. The selection of CycleGAN is driven by the practical challenge of obtaining paired CBCT and CT scans in clinical settings. Our approach enhances the original CycleGAN by incorporating a ViT-based U-Net generator, which effectively extracts and retains vital features and fine details. To further refine image generation, we integrate a deep convolutional network within the feedforward neural network, leveraging the Transformers' self-attention mechanism to enable automatic focus on diverse image regions, thereby improving global understanding and enhancing detail localization. A gradient penalty is introduced to ensure more stable training, and an additional loss term is added to the generator's objective function to capture discrepancies between the source and generated images.

Our model exhibits superior quantitative performance compared to prevailing unsupervised learning techniques, achieving state-of-the-art evaluation metrics across both datasets. Comprehensive ablation studies, detailed in Tables 3 and 4, consistently reveal the positive impact of our proposed modules on the model's overall efficacy. Of particular note, the DCN module stands out with a more substantial boost, attributed to its inherent local characteristics that are effectively modeled by the self-attention mechanism in ViT. This integration enables ViT to engage in both global and local information exchange, thereby enhancing its capabilities.

In visual assessments, we rigorously tested our model's superiority through extensive experiments. As depicted in Figs. 6 and 7, our model-generated images exhibit reduced noise and enhanced detail, closely resembling authentic CT scans. For the H & N dataset's first row, our sCT exhibits the closest resemblance to real CT at the conus region (green arrow), with ADCycleGAN and RegGAN also displaying comparable performance.

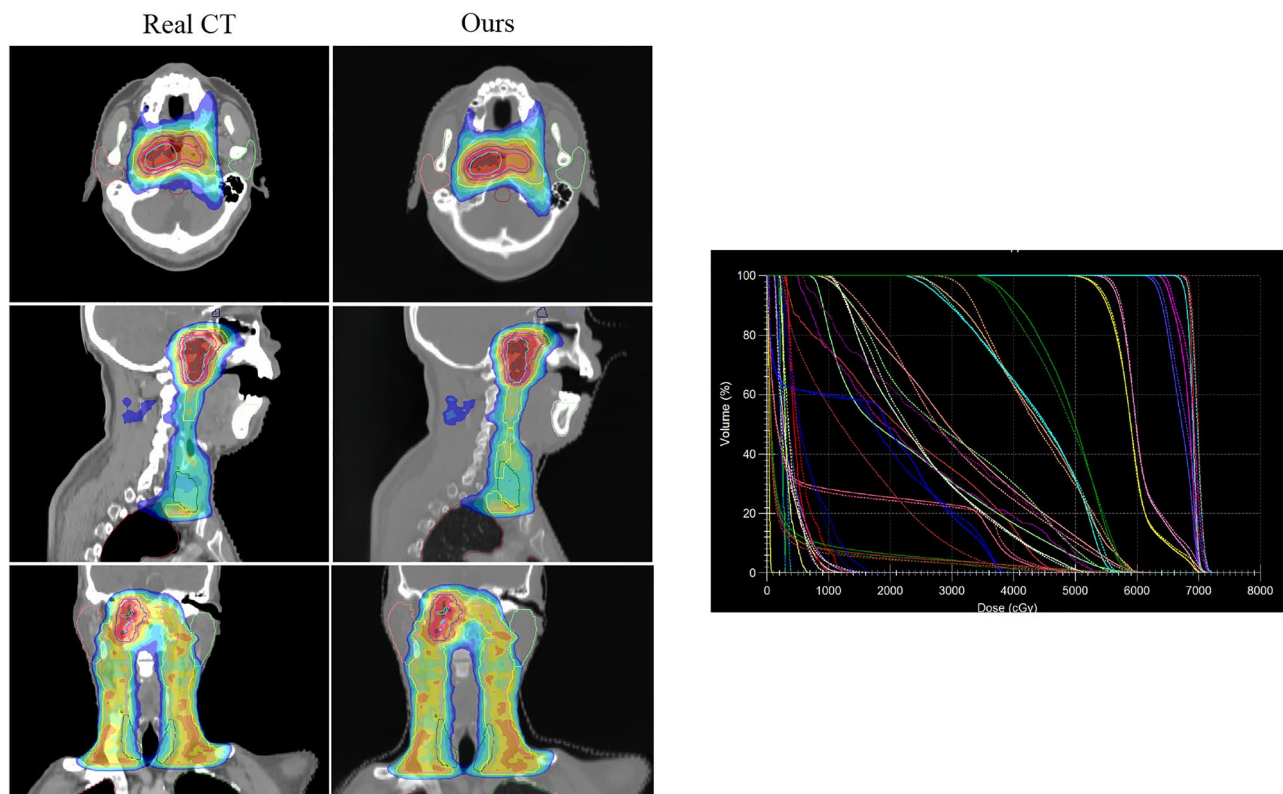


Figure 13. 3D reconstruction visualization of the H & N dataset. From top to bottom, representing axial, sagittal and coronal planes(left). Patient DVH differences, where the solid line is Real CT and the dashed line is sCT (right).

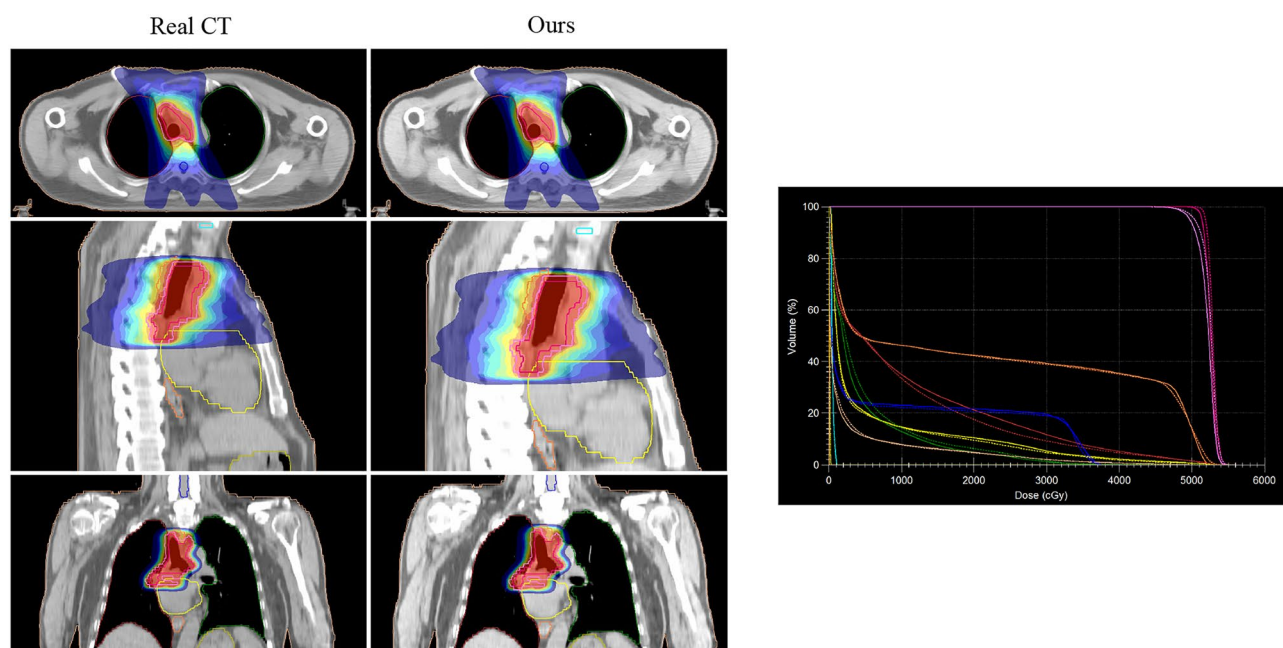


Figure 14. 3D reconstruction visualization of the Chest dataset. From top to bottom, representing axial, sagittal and coronal planes(left). Patient DVH differences, where the solid line is Real CT and the dashed line is sCT (right).

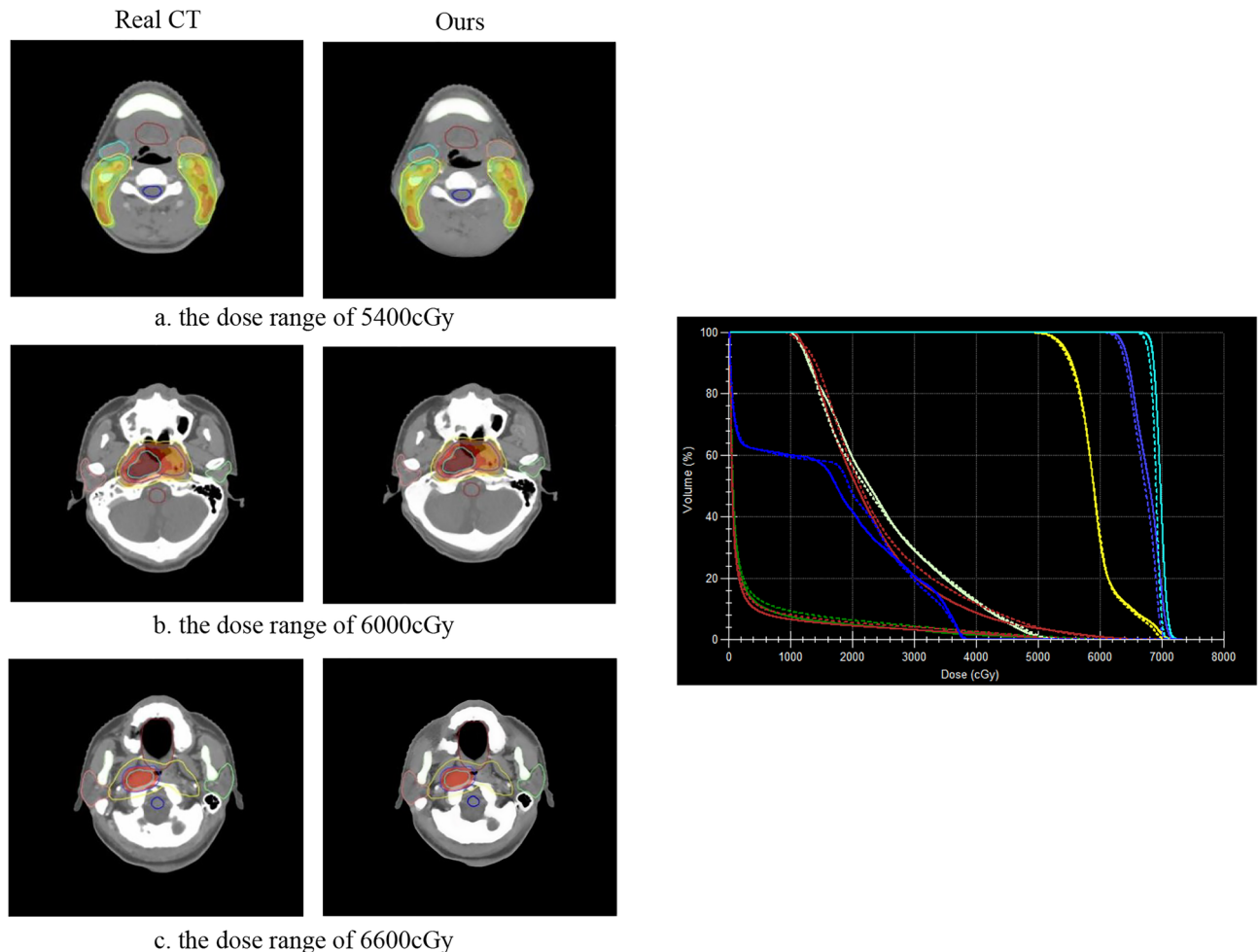


Figure 15. Differences between our method and Real CT for different dose distributions on the H & N dataset (left). Patient DVH differences, where the solid line is Real CT and the dashed line is sCT (right).

However, in the nasal cavity, ADCycleGAN and RegGAN differ significantly from the real CT in shape. CycleGAN, DualGAN, and AttentionGAN exhibit larger discrepancies, characterized by blurred details and excessive noise in the conus area. In the nasopharynx (blue arrow) of the second row, ADCycleGAN, RegGAN, CycleGAN, and DualGAN present similar shapes with minor differences from the real CT, such as blurred boundaries and missing details. Our sCT stands out with clear details, while AttentionGAN performs the least favorably. In the pituitary region (yellow arrow) of the third row, our sCT most closely matches the real CT, with distinct boundaries and minimal shape variations. ADCycleGAN, RegGAN, and CycleGAN lose some details, and eye distortion is prominent. DualGAN and AttentionGAN generally underperform. In the eye region (orange arrow) of the fourth row, ADCycleGAN, RegGAN, and CycleGAN exhibit minimal differences from the real CT, but brain tissue distortion is severe. Our sCT excels, whereas DualGAN and AttentionGAN lag behind. For the Chest dataset, in the bronchial bifurcation (green arrow), ADCycleGAN, RegGAN, and CycleGAN exhibit smaller differences from the real CT, but overall image detail is lacking. Our sCT stands out, while DualGAN and AttentionGAN falter. In the conus region (blue arrow) of the second row, ADCycleGAN and RegGAN show less shape deviation compared to CycleGAN, DualGAN, and AttentionGAN, which exhibit the greatest discrepancies. Our sCT is the closest to the real CT in this region. In the lung tumor area (yellow dashed circle), ADCycleGAN and RegGAN have similar shapes to the real CT, but they lack information in the heart and conus regions. Our sCT outperforms others, with CycleGAN, DualGAN, and AttentionGAN performing the worst. In the lung tumor area of the fourth row (orange dashed circle), ADCycleGAN and RegGAN have slightly less shape deviation from the real CT, but they severely lack heart information. Our sCT demonstrates superior clarity with minimal distortion, while CycleGAN, DualGAN, and AttentionGAN remain inferior. The HU value histograms in Fig. 8 reveal that our method's curve shape and peak are closer to the real CT, indicating a certain level of fidelity. Figures 9 and 10 illustrate the difference plots between CBCT, our method, and CT for both datasets. The difference plots reveal that our method exhibits significantly smaller discrepancies compared to CBCT, indicating that the synthetic sCT generated by our approach approaches the CT standard to a certain extent. In summary, our model, trained on unpaired data, is capable of extracting and preserving crucial features and fine-grained details, automatically focusing on various image regions, and enhancing the understanding of

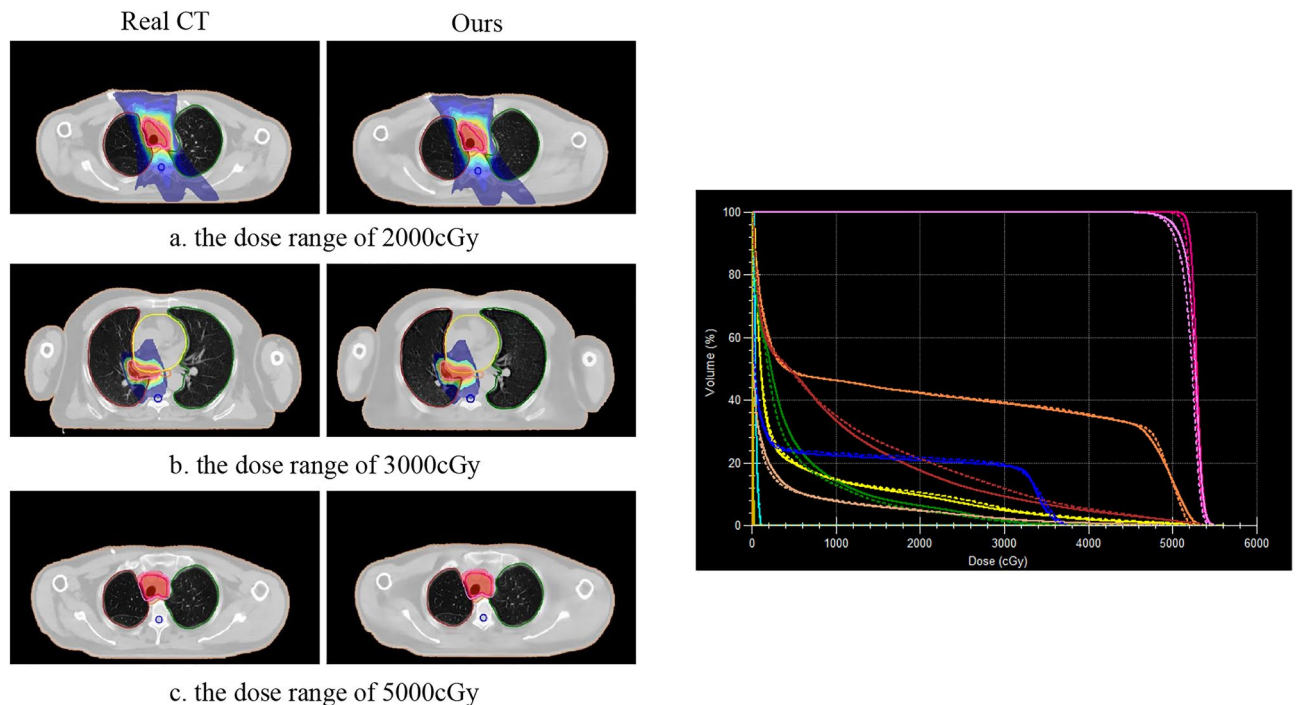


Figure 16. Differences between our method and Real CT for different dose distributions on the Chest dataset (left). Patient DVH differences, where the solid line is Real CT and the dashed line is sCT (right).

the overall image structure. Through extensive experimentation on both datasets, our model outperforms other competing algorithms in terms of results.

Furthermore, in addition to visual presentations, this paper also evaluates the clinical significance through expert analysis. Figures 13 and 14 illustrate the 3D reconstructions, demonstrating that sCT preserves the original anatomical structures and maintains a certain continuity in internal organ tissues. Figures 15 and 16 showcase the distribution of our sCT under different doses compared to the real clinical dose. The results indicate that the dose difference between our sCT and the real CT is minimal, confirming the clinical applicability of our method.

While our approach outperforms other unsupervised learning models, the results on the Chest dataset remain relatively average. For future work, we plan to experiment with the state-of-the-art diffusion models in the image generation domain and further investigate 3D image generation capabilities.

Conclusions

This study proposes an unsupervised learning model, IViT-CycleGAN, aiming to synthesize sCT from CBCT for future clinical practice. IViT-CycleGAN presents a U-net framework, which is built upon the ViT architecture within the generator. This framework leverages the U-net structure to effectively extract and retain crucial features and intricate details. Moreover, we enhance the ViT model by integrating a deep convolutional network and the self-attention mechanism of Transformer into the feed-forward neural network. The objective is to automatically prioritize information from various image locations during the image generation process, leading to a better comprehension of the overall image structure and emphasizing regions with finer details. Consequently, the generated images exhibit enhanced clarity and realism. To enhance the stability of model training, a gradient penalty is introduced to ensure minimal variations in network weights for minor changes in the model's input. Additionally, an additional loss term is included in the generator loss to reinforce the consistency between the generated and source images by capturing their differences. The results demonstrate that IViT-CycleGAN outperforms other unsupervised learning models in terms of generating sCT, thus validating the clinical applicability and robustness of our model. In future clinical practice, this method can assist clinicians in developing radiotherapy treatment plans.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 16 January 2024; Accepted: 6 May 2024

Published online: 20 May 2024

References

- Jaffray, D. A. Image-guided radiotherapy: from current concept to future perspectives. *Nat. Rev. Clin. Oncol.* **9**, 688–699. <https://doi.org/10.1038/nrclinonc.2012.194> (2012).

2. Dawson, L. A. & Sharpe, M. B. Image-guided radiotherapy: rationale, benefits, and limitations. *Lancet Oncol.* **7**, 848–858. [https://doi.org/10.1016/S1470-2045\(06\)70904-4](https://doi.org/10.1016/S1470-2045(06)70904-4) (2006).
3. Boda-Heggemann, J., Lohr, F., Wenz, F., Flentje, M. & Guckenberger, M. kv-cone-beam-ct-basierte bildgeführte strahlentherapie-ein klinischer überblick: A clinical review. *Strahlenther. Onkol.* **187**, 284–291. <https://doi.org/10.1007/s00066-011-2236-4> (2011).
4. Harms, J. *et al.* Paired cycle-gan-based image correction for quantitative cone-beam computed tomography. *Med. Phys.* **46**, 3998–4009. <https://doi.org/10.1002/mp.13656> (2019).
5. Liu, Y. *et al.* Mri-based treatment planning for proton radiotherapy: Dosimetric validation of a deep learning-based liver synthetic ct generation method. *Phys. Med. Biol.* **64**, 145015. <https://doi.org/10.1088/1361-6560/ab25bc> (2019).
6. Hunter, A. K. & McDavid, W. Characterization and correction of cupping effect artefacts in cone beam ct. *Dentomaxillofacial Radiol.* **41**, 217–223. <https://doi.org/10.1259/dmfr/19015946> (2012).
7. Zou, Y., Sidky, E. Y. & Pan, X. Partial volume and aliasing artefacts in helical cone-beam ct. *Phys. Med. Biol.* **49**, 2365. <https://doi.org/10.1088/0031-9155/49/11/017> (2004).
8. Richter, A. *et al.* Initial results for patient setup verification using transperineal ultrasound and cone beam ct in external beam radiation therapy of prostate cancer. *Radiat. Oncol.* **11**, 1–7. <https://doi.org/10.1186/s13014-016-0722-7> (2016).
9. Long, J., Shelhamer, E., & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3431–3440. <https://doi.org/10.1109/cvpr.2015.7298965> (2015).
10. Nie, D., Cao, X., Gao, Y., Wang, L. & Shen, D. Estimating ct image from MRI data using 3d fully convolutional networks. In *Proceedings of Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016*, 170–178. https://doi.org/10.1007/978-3-319-46976-8_18 (Springer, 2016).
11. Chen, S., Qin, A., Zhou, D. & Yan, D. U-net-generated synthetic ct images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Med. Phys.* **45**, 5659–5665. <https://doi.org/10.1002/mp.13247> (2018).
12. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning* 1597–1607 (PMLR, 2020).
13. Dinkla, A. M. *et al.* Dosimetric evaluation of synthetic ct for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. *Med. Phys.* **46**, 4095–4104. <https://doi.org/10.1002/mp.13663> (2019).
14. Kida, S. *et al.* Cone beam computed tomography image quality improvement using a deep convolutional neural network. *Cureus* <https://doi.org/10.7759/cureus.2548> (2018).
15. Li, Y. *et al.* A preliminary study of using a deep convolution neural network to generate synthesized ct images based on cbct for adaptive radiotherapy of nasopharyngeal carcinoma. *Phys. Med. Biol.* **64**, 145010. <https://doi.org/10.1088/1361-6560/ab2770> (2019).
16. Xie, S., Liang, Y., Yang, T. & Song, Z. Contextual loss based artifact removal method on CBCT image. *J. Appl. Clin. Med. Phys.* **21**, 166–177. <https://doi.org/10.1002/acm2.13084> (2020).
17. Liu, Y. *et al.* A two-step method to improve image quality of cbct with phantom-based supervised and patient-based unsupervised learning strategies. *Phys. Med. Biol.* **67**, 084001. <https://doi.org/10.1088/1361-6560/ac6289> (2022).
18. Choi, K., Kim, S. H. & Kim, S. Self-supervised denoising of projection data for low-dose cone-beam ct. *Med. Phys.* **50**, 6319–6333. <https://doi.org/10.1002/mp.16421> (2023).
19. Park, T., Efros, A. A., Zhang, R., Zhu, J.-Y. Contrastive and learning for unpaired image-to-image translation. In *Proceedings of Computer Vision-ECCV, 16th European Conference, Part IX, Glasgow, UK, August 23–28*, vol 16, 319–345. https://doi.org/10.1007/978-3-030-58545-7_19 (Springer, 2020).
20. Wang, T. *et al.* An unsupervised dual contrastive learning framework for scatter correction in cone-beam ct image. *Comput. Biol. Med.* **165**, 107377. <https://doi.org/10.1016/j.compbiomed.2023.107377> (2023).
21. Wang, T. *et al.* Ring artifacts correction for computed tomography image using unsupervised contrastive learning. *Phys. Med. Biol.* **68**, 205008. <https://doi.org/10.1088/1361-6560/acfa60> (2023).
22. Liu, X., Liang, X., Deng, L., Tan, S. & Xie, Y. Learning low-dose ct degradation from unpaired data with flow-based model. *Med. Phys.* **49**, 7516–7530. <https://doi.org/10.1002/mp.15886> (2022).
23. Li, Y. *et al.* Zero-shot medical image translation via frequency-guided diffusion models. *IEEE Trans. Med. Imaging* <https://doi.org/10.1109/TMI.2023.3325703> (2023).
24. Liu, X. *et al.* Diffusion probabilistic priors for zero-shot low-dose ct image denoising. <https://doi.org/10.48550/arXiv.2305.15887> (2023).
25. Özbey, M. *et al.* Unsupervised medical image translation with adversarial diffusion models. *IEEE Trans. Med. Imaging* <https://doi.org/10.1109/TMI.2023.3290149> (2023).
26. Peng, J. *et al.* Cbct-based synthetic ct image generation using conditional denoising diffusion probabilistic model. *Med. Phys.* **51**, 1847–1859. <https://doi.org/10.1002/mp.16704> (2024).
27. James, G., Witten, D., Hastie, T., Tibshirani, R. *et al.* An introduction to statistical learning, Vol. 112 (Springer, 2013).
28. Almahairi, A., Rajeshwar, S., Sordani, A., Bachman, P. & Courville, A. Augmented cyclegan: Learning many-to-many mappings from unpaired data. In *International Conference on Machine Learning* 195–204 (PMLR, 2018).
29. Liu, Y. *et al.* Cbct-based synthetic ct generation using deep-attention cyclegan for pancreatic adaptive radiotherapy. *Med. Phys.* **47**, 2472–2483. <https://doi.org/10.1002/mp.14121> (2020).
30. Lei, Y. *et al.* Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks. *Med. Phys.* **46**, 3565–3581. <https://doi.org/10.1002/mp.13617> (2019).
31. Lei, Y. *et al.* Learning-based cbct correction using alternating random forest based on auto-context model. *Med. Phys.* **46**, 601–618. <https://doi.org/10.1002/mp.13295> (2019).
32. Liang, X. *et al.* Generating synthesized computed tomography (ct) from cone-beam computed tomography (cbct) using cyclegan for adaptive radiation therapy. *Phys. Med. Biol.* **64**, 125002. <https://doi.org/10.1088/1361-6560/ab22f9> (2019).
33. Liu, F., Yadav, P., Baschnagel, A. M. & McMillan, A. B. Mr-based treatment planning in radiation therapy using a deep learning approach. *J. Appl. Clin. Med. Phys.* **20**, 105–114. <https://doi.org/10.1002/acm2.12554> (2019).
34. Zeng, G., Zheng, G. Hybrid and generative adversarial networks for deep mr to ct synthesis using unpaired data. In *Proceedings of 22nd International Conference Medical Image Computing and Computer Assisted Intervention-MICCAI, Shenzhen, China, October 13–17, 2019. Part IV*, vol. 22, 759–767. https://doi.org/10.1007/978-3-030-32251-9_83 (2019).
35. Wang, J., Wu, Q. J. & Pourpanah, F. An attentive-based generative model for medical image synthesis. *Int. J. Mach. Learn. Cybern.* <https://doi.org/10.1007/s13042-023-01871-0> (2023).
36. Torbunov, D. *et al.* Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* 702–712. [arXiv:2203.02557](https://arxiv.org/abs/2203.02557) (2023).
37. Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision* 2223–2232. <https://doi.org/10.48550/arXiv.1703.10593> (2017).
38. Li, Y. *et al.* Localvit: Bringing locality to vision transformers. In *2023 IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS)* 9598–9605. <https://doi.org/10.1109/IROS55552.2023.10342025> (2023).
39. Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. <https://doi.org/10.48550/arXiv.1710.10196> (2017).
40. Torbunov, D. *et al.* Rethinking cyclegan: Improving quality of gans for unpaired image-to-image translation. <https://doi.org/10.48550/arXiv.2303.16280>. [arXiv:2303.16280](https://arxiv.org/abs/2303.16280) (2023).

41. Yi, Z., Zhang, H., Tan, P. & Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision* 2849–2857, <https://doi.org/10.1109/ICCV.2017.310> (2017).
42. Kong, L. *et al.* Breaking the dilemma of medical image-to-image translation. *Adv. Neural. Inf. Process. Syst.* **34**, 1964–1978. <https://doi.org/10.48550/arXiv.2110.06465> (2021).

Acknowledgements

This research was funded by the Jiangsu Provincial Key Research and Development Program (BE2020714).

Author contributions

Conceptualization, H.Z., C.L. and C.H.; methodology, C.H.; software, C.H.; validation, C.H. and N.C.; formal analysis C.H. and N.C.; resources C.H. and N.C.; writing-original draft preparation, C.H.; writing-review and editing, C.H., Y.H.; funding acquisition, N.C. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.L. or C.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024