



Development of A deep Learning-based algorithm for High-Pitch helical computed tomography imaging

Xiaoman Duan ^a, Xiao Fan Ding ^a, Samira Khoz ^b, Xiongbiao Chen ^{a,b,*}, Ning Zhu ^{a,c,d,*}

^a Division of Biomedical Engineering, College of Engineering, University of Saskatchewan, Saskatoon, SK, S7N 5A9, Canada

^b Department of Mechanical Engineering, College of Engineering, University of Saskatchewan, Saskatoon, SK, S7N 5A9, Canada

^c Department of Chemical and Biological Engineering, College of Engineering, University of Saskatchewan, Saskatoon, SK, S7N 5A9, Canada

^d Canadian Light Source, Saskatoon, S7N 2V3, SK, Canada

ARTICLE INFO

Keywords:

Deep learning
high-pitch helical CT
Propagation-based imaging
Biomedical imaging

ABSTRACT

High-pitch X-ray helical computed tomography (HCT) imaging has been recently drawing considerable attention in biomedical fields due to its ability to reduce the scanning time and thus lower the radiation dose that objects (being imagined) may receive. However, the issue of compromised reconstruction quality caused by incomplete data in these high-pitch CT scans remains, thus limiting its applications. By addressing the aforementioned issue, this paper presents our study on the development of a novel deep learning (DL)-based algorithm, ViT-U, for high-pitch X-ray propagation-based imaging HCT (PBI-HCT) reconstruction. ViT-U consists of two key process modules of a vision transformer (ViT) and a convolutional neural network (i.e., U-Net), where ViT addresses the missing information in the data domain and U-Net enhances the post data-processing in the reconstruction domain. For verification, we designed and conducted simulations and experiments with both low-density-biomaterial samples and biological-tissue samples to exemplify the biomedical applications, and then examined the ViT-U performance with varying pitches of 3, 3.5, 4, and 4.5, respectively, for comparison in term of radiation does and reconstruction quality. Our results showed that the high-pitch PBI-HCT allowed for the dose reduction from 72% to 93%. Importantly, our results demonstrated that the ViT-U exhibited outstanding performance by effectively removing the missing wedge artifacts thus enhancing the reconstruction quality of high-pitch PBI-HCT imaging. Also, our results showed the superior capability of ViT-U to achieve high quality of reconstruction from the high-pitch images with the helical pitch value up to 4 (which allowed for the substantial reduction of radiation doses). Taken together, our DL-based ViT-U algorithm not only enables high-speed imaging with low radiation dose, but also maintains the high quality of imaging reconstruction, thereby offering significant potentials for biomedical imaging applications.

1. Introduction

Computed tomography (CT) has fundamentally transformed radiology and continues to advance in various biomedical fields. Nowadays, the CT systems are able to, for example, rapidly imagine the whole body of an object with isotropic resolution or the whole-heart perfusion in a single rotation and the temporal resolution. Notably, in these biomedical fields the radiation exposure or the dose that the object (being imagined) receives remains a critical issue, particularly in consideration of the trade-off of imaging quality and radiation dose. To address this issue, propagation-based imaging helical CT (PBI-HCT) has been emerging as an advanced imaging technique in biomedical fields, where the pitch is a

key trade-off parameter for the imaging quality and delivered radiation dose. Pitch in helical CT, is defined as the ratio of the vertical distance traveled per rotation to the height of the field of view (FOV). PBI-HCT presents a reliable solution to image low-density samples with fewer ring artifacts while maintaining the high image quality in terms of signal-to-noise ratio and spatial resolution (Duan, Li, et al., 2023). High-pitch PBI-HCT (i.e. the one with a pitch value exceeding 2) is highly desirable in many biomedical applications due to the benefits and/or merits, as highlighted below.

1) The surface entry radiation dose of PBI-HCT, $D_{\text{PBI-HCT}}$, is inversely proportional to the normalized helical pitch p for each single slice, as calculated in Equation (1). A larger p is usually preferred for radiation

* Corresponding authors.

E-mail addresses: xid896@usask.ca (X. Duan), xiaofan.ding@usask.ca (X. Fan Ding), samira.khoz@usask.ca (S. Khoz), xbc719@usask.ca (X. Chen), niz504@mail.usask.ca (N. Zhu).

dose reduction purposes.

$$D_{\text{PBI-HCT}} = \dot{D} \times \frac{N_p}{p} \times \Delta t \quad (1)$$

where \dot{D} is the measured surface entry radiation dose rate (Gy/s), Δt is the exposure time (s) for each projection, and $\frac{N_p}{p}$ is the effective projection number per pitch.

2) High-pitch imaging reduces the number of rotations or projections needed to cover the region of interest, increasing scanning speed. This shortens scan times, minimizing motion-induced blurring artifacts (Lapegue, et al., 2014; Lim, Ha, Hwang, & Lee, 2019). This attribute is particularly advantageous in procedures such as cardiac CT scans (Ahmed, et al., 2023; Hausleiter, et al., 2009), where it contributes to enhanced image quality.

3) High-pitch scanning with a small vertical FOV ensures uniform X-ray intensity and consistent image quality, aiding accurate data analysis and dose control in live animal imaging (Duan, Li, et al., 2023).

4) High-pitch scanning allows imaging of elongated samples despite limitations in detector height and X-ray flux, reducing artifacts and enabling consistent projections and rotations (Duan, Li, Chen, & Zhu, 2021).

However, when the pitch exceeds 2, the angle range of PBI-HCT imaging for each slice is truncated to less than 180° thus violating the Tuy's data sufficiency condition (Sun, Clackdoyle, Kim, Fulton, & Nuyts, 2017; Tuy, 1983) and results in missing wedge artifacts in images, as that in limited-view angle reconstruction problems (Anirudh, et al., 2018; Z. Chen, Jin, Li, & Wang, 2013). Specifically, for PBI-HCT imaging with pitches exceeding 2, the measured data experiences missing information, leading to the absence of some rows in sinograms, as shown in Figure 1 (A2)-(C2). As a result, missing wedge artifacts are observed in PBI-HCT images when the pitch exceeds 2 using the filtered-back projection (FBP) reconstruction algorithm, as shown in Figure 1 (A3)-(C3). Therefore, to enable fast speed, low- and uniform- dose, and high-quality PBI-HCT scans, it is critically important to address the scientific challenges in high-pitch PBI-HCT reconstruction problems.

The advancements in DL are opening up new avenues for addressing the challenges associated with high-pitch PBI-HCT. The high-pitch PBI-HCT imaging has been, usually, considered as a special CT condition with limited-angle scan (Hu, et al., 2021; Hu, Zhang, Liu, Luo, & Chen, 2022; J. Wang, Liang, Cheng, Guo, & Zeng, 2020; Q. Zhang, et al., 2020). Algorithms developed for limited-angle CT imaging can be applied to

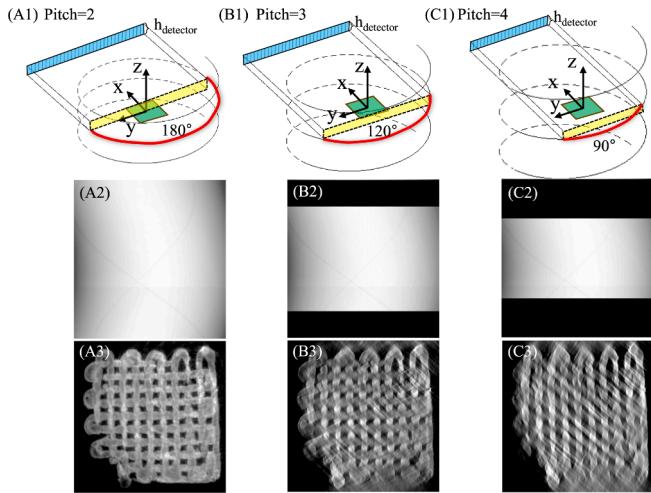


Fig. 1. Example of PBI-HCT imaging with various pitch scanning configurations. (A1)-(C1) illustration of PBI-HCT scan with vertical traveling distance same to FOV, i.e., active detector's height (h_{detector}), and the (A2)-(C2) converted sinograms from PBI-HCT imaging with various pitches and (A3)-(C3) corresponding reconstructed PBI-HCT images.

high-pitch PBI-HCT imaging. For example, sinogram-based deep learning image reconstruction (DLIR) (Hsieh, et al., 2019) (GE Healthcare, Waukesha, WI, USA) is able to learn the complex relationships between raw projection data and high-quality CT images; and to outperforms state-of-the-art iterative reconstruction algorithms, such as adaptive statistical iterative reconstruction-V (ASIR-V) (Fan, Yue, & Melnyk, 2014), thus offering superior image quality and faster reconstruction. Besides, hdNet (i.e., hybrid-domain deep convolutional neural network (CNN)) (Q. Zhang, et al., 2020) was introduced for limited-angle CT reconstruction, which comprises three key components: a CNN operating in sinogram domain, a domain transformation operation based on an analytical algorithm, and a CNN operating in CT image domain. The main contribution of the hdNet's is concurrent training across both sinogram and reconstructed CT image domains. Additionally, there are some studies tailored for high-pitch HCT imaging (Gong, Ren, McCollough, & Yu, 2020; Hayes, et al., 2021). For instance, UFP-net is a custom-designed network employing both local and non-local operators for ultra-fast-pitch image reconstruction in HCT, maintaining high image quality for simulation data with a mean structural similarity index measurement (SSIM) of 0.9 for a pitch of 3 (Gong, et al., 2020). Double-CNNs (Hayes, et al., 2021) is particularly developed for high-pitch HCT imaging to reduce artifacts along axial and coronal planes using a two-stage CNN.

However, these DL algorithms are designed for clinical HCT applications, and may not be directly used for high-pitch PBI-CT imaging. PBI-CT has relatively different image characteristics compared to clinical HCT imaging in terms of image resolution, noise, contrast, and artifacts distribution (Duan, Ding, et al., 2023; Duan, Li, et al., 2023). Therefore, the performance of directly employing these algorithms to high-pitch PBI-HCT data is not guaranteed, necessitating evaluation on experimental datasets. While it is possible to build a large model using a vast dataset of retrospective data in clinical helical practice and subsequently apply transfer learning (feature-based and/or model-based) to actual PBI-HCT data, this process can be complex and may not always yield successful outcomes, especially when the image features exhibit significant dissimilarities (S. J. Pan & Yang, 2009; W. Zhang, Deng, Zhang, & Wu, 2022). Therefore, there is a compelling need for the exploration and development of tailored DL methods that are specifically suitable to accommodate the distinctive attributes of high-pitch PBI-HCT data.

In recent years, the Transformer-based DL algorithms (Vaswani, et al., 2017), have garnered significant attention in the field of natural language processing and generation. Transformer is a neural network model used for processing sequence data, originally proposed to solve machine translation problems (Vaswani, et al., 2017). Its core is the self-attention mechanism, which captures long-distance dependencies (i.e., large effective receptive fields) by calculating the relationships between various positions in the input sequence, thereby better processing sequence data. Building on this success, researchers have tried to bring transformer encoder section into the vision domain, i.e., vision transformer (ViT) (Dosovitskiy, et al., 2020) to perform the image recognition task. Based on ViT, a series of networks (Cao, et al., 2022; Ze Liu, et al., 2021) have been developed, and has demonstrated the state-of-the-art performance in tasks such as image classification (C.-F. R. Chen, Fan, & Panda, 2021; Zhou, Dou, Su, Hu, & Zheng, 2023), object detection (Yanghao Li, Mao, Girshick, & He, 2022; D. Zhang, et al., 2023), denoising (D. Wang, et al., 2023), and semantic segmentation (Ando, et al., 2023; Strudel, Garcia, Laptev, & Schmid, 2021). Compared to convolutional neural networks (CNNs) with rather small convolution kernels (e.g., 3×3 or 5×5) (J. Li, et al., 2023), transformer-based DL algorithms are better at understanding nonlocal and global information. Besides, the advantages of the transformer model include better flexibility, interpretability, and generalization ability (B. Pan, et al., 2021; Shamshad, et al., 2023).

Motivated by the advantages of long-distance dependencies in ViT, this study is to leverage the innovative ViT architecture to address the

absence of information in the sinogram domain of high-pitch PBI-HCT imaging. As the pitch value increases, this challenge becomes more pronounced. Subsequently, we also employ a U-Net to further enhance the image quality of high-pitch PBI-HCT reconstruction in the CT image domain. The integrated algorithm is named as ViT-U. Our contributions can be summarized as follows: (i) This study developed ViT-U for high-pitch PBI-HCT imaging and the performance of ViT-U was compared with other state-of-the-art algorithms. (ii) The performance of ViT-U applied to high-pitch PBI-HCT imaging with different pitches was evaluated to identify the maximum pitch value and determine its maximum efficacy. (iii) The effectiveness of ViT-U was assessed in reducing radiation dose by applying it to high-pitch PBI-HCT imaging under different noise levels.

2. Materials and methods

2.1. Workflow of proposed ViT-U method

The architecture of ViT-U employed for high-pitch PBI-HCT imaging is depicted in the Figure 2, including sinogram filling module ViT network (Dosovitskiy, et al., 2020) and CT image post-processing module U-Net. Each part of the proposed ViT-U is introduced in following sections.

2.1.1. Sinogram filling module: ViT network

Sinogram filling module consists of patches creation and a classic ViT (a linear projection operation (head), transformer encoder, and another linear normalization operation (tail)).

1) Patches creation: Unlike traditional ViT, which uses square patches for input embedding, high-pitch PBI-HCT sinograms have

varying signal completeness across rows, with some rows blank and others containing usable ground truth. Cutting square patches disrupts data consistency. Modeling self-attention between views enhances the network's ability to learn critical features for sinogram completion (Yang, et al., 2022). Consequently, we create sinogram patches using data from each row of the measured sinogram, i.e., S^1, S^2, \dots, S^{N_p} , where N_p is the total number of projections (i.e., views or rows of sinogram). Notably, some rows within high-pitch PBI-HCT sinograms lack information and appear blank (i.e., grey value of 0). Nevertheless, in our algorithm, we retain these rows and utilize them as original input for ViT, hypothesizing that the information from these rows can be restored using the multi-head self-attention (MSA) mechanism in transformer encoder.

2) Linear projection: The aim of this step is to generate the token embedding (TE) to extract feature, which can be realized by using linear layer operator.

$$\langle TE^1, \dots, TE^{N_p} \rangle = \text{ReLU}(LN(S^1, S^2, \dots, S^{N_p})) \quad (2)$$

where LN is the linear normalization, and ReLU is the activate function. As this involves a regression problem aimed at filling the missing rows in images, the extra learnable classification token embedding in the original ViT is omitted here for simplification.

Before moving into encoder, we also add the same-quantity one-dimensional (1D) learnable position embedding $\langle PE^1, \dots, PE^{N_p} \rangle$ to maintain the position information (Dosovitskiy, et al., 2020). Position embedding allows the model to understand the order of the sequence, a crucial aspect for restoring the missing row information in our scenarios. Then $PE^i + TE^i, i = 1, \dots, N_p$ are the input into the transformer encoder directly.

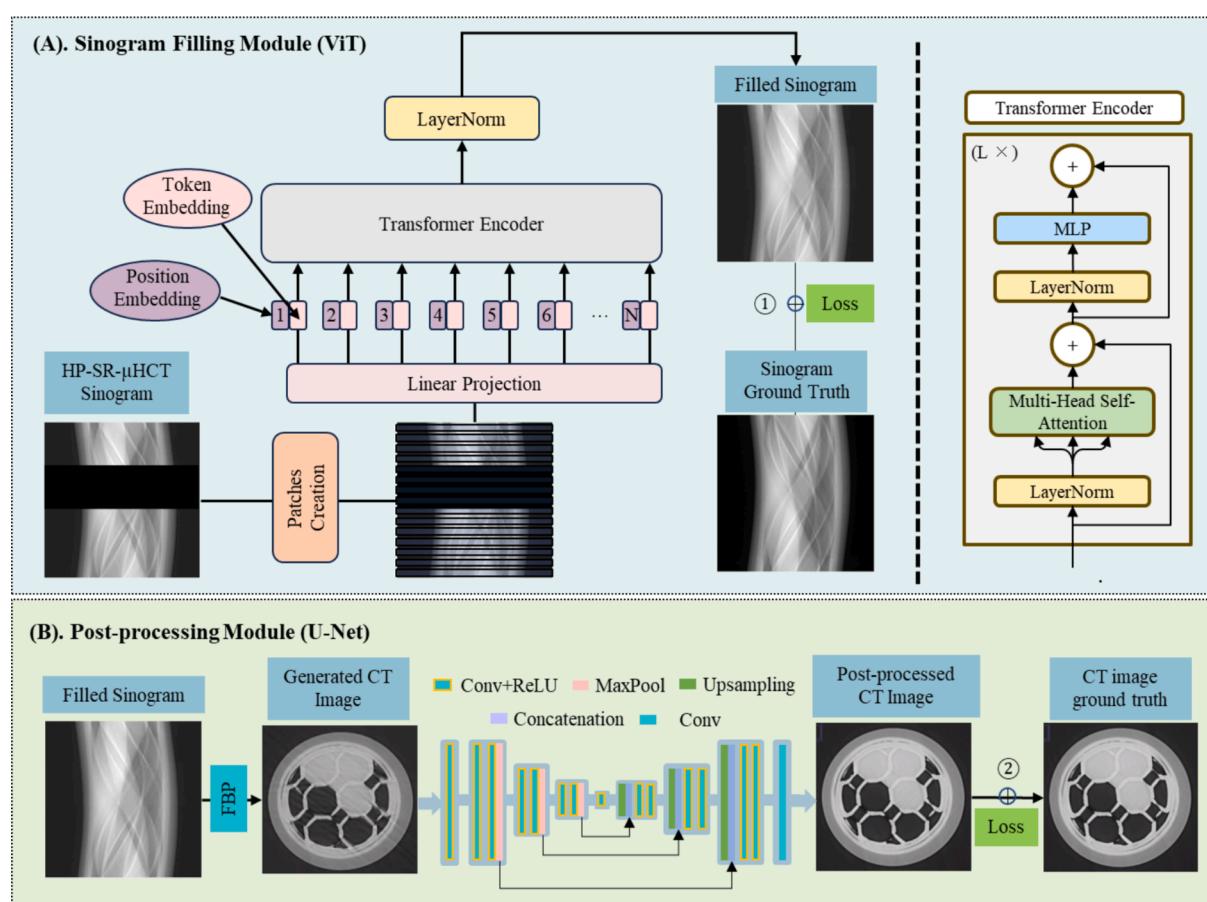


Fig. 2. Workflow of the proposed ViT-U method for high-pitch PBI-HCT imaging, including two modules, (a) sinogram filling module: ViT and (b) CT image post-processing module: U-Net.

3) Transformer encoder: In this part, the illustration of the transformer encoder was inspired by Dosovitskiy et al. (Dosovitskiy, et al., 2020). The transformer encoder consists of a *LN* and MSA which allows the model to jointly process information from different subspaces at different positions, and Multilayer Perceptron (MLP). The entire structure is shown as follows:

$$Z^0 = \langle PE^1 + TE^1, PE^2 + TE^2, \dots, PE^{N_p} + TE^{N_p} \rangle \quad (3)$$

$$Q^j = K^j = V^j = LN(Z^{j-1}) \quad (4)$$

where $j = 1, 2, \dots, L$. L is the repetition of the transformer encoder, which is taken as 6 in this work.

$$Z^j = MSA(Q^j, K^j, V^j) + Z^{j-1} \quad (5)$$

where MSA is expressed as,

$$MSA(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

The scaling factor $\frac{1}{\sqrt{d_k}}$ is based on the network depth. The output of transformer encoder is,

$$Z^j = MLP(LN(Z^j)) + Z^j \quad (7)$$

4) Layer normalization: After processed by the multi-blocks of transformer encoder, a final *LN* is applied to get the prediction results S_{pred} of the sinogram.

$$S_{pred} = LN(Z^j) \quad (8)$$

2.1.2. CT image post-processing module: U-net

After pre-training a robust model for predicting missing information in high-pitch PBI-HCT, the weights of the sinogram filling module are subsequently frozen. This trained module is then utilized as a pre-processing step, followed by a CT reconstruction (i.e., FBP with the Ultra-Fast-Online (UFO) (Vogelgesang, et al., 2016)) to obtain the CT-domain image. Subsequently, a U-Net (Shan, et al., 2018) with three down-sampling layers and three up-sampling layers (kernel size of 5 × 5) is applied as a post-processing step to further enhance the quality of the final reconstructed image.

2.2. Implementation of ViT-U and loss functions

To implement the ViT-U, the projections of high-pitch PBI-HCT were converted to virtual regular CT ones as per our previous study (Duan, Li, et al., 2023), and then the projection data were transposed to acquire the regular CT sinograms to be processed by ViT network. Before this, each sinogram row was transformed into a 1 × 1024 dimension by using linear projection. The transformer encoder consisted of 6 layers, each utilizing a 16-head attention mechanism based on the ablation study (see supplementary material). For optimization, we utilized ADAM with momentum, setting the initial learning rate to 10^{-3} , the first-order momentum to 0.9, and the second-order momentum to 0.999. Additionally, we employed a stepLR scheduler with a decay rate of 0.7 every 10 epochs. The loss function \mathcal{L}_{ViT} for this module is masked L1 loss for the rows that are blank for original high-pitch PBI-HCT data, which is proven to be able to keep the high-frequency structure for high-pitch PBI-HCT.

$$\mathcal{L}_{ViT} = Mask \bullet \|S_{pred} - S\|_1 \quad (9)$$

where the *mask* is binary, with a value of 1 indicating blank rows and 0 indicating normal rows with information. S_{pred} and S are the predicted and original rows from high-pitch PBI-HCT sinogram.

U-Net was trained using the Adam optimization algorithm with a mini-batch size of 64 and a learning rate of 10^{-4} . The loss function for U-Net is SSIM (Z. Wang, Bovik, Sheikh, & Simoncelli, 2004). Notably, the dataset required for training U-Net can be substantially reduced due to

relatively small module while still achieving high-quality prediction images compared with ViT module.

2.3. Simulations

The performance of ViT-U was firstly trained and tested on simulated data utilizing a training dataset comprising 10,000 sinograms, a validation dataset of 2,000 sinograms, and a test dataset containing 1,000 sinograms. Each sinogram was synthetically generated using the Radon Transform using ASTRA-toolbox (Van Aarle, et al., 2016), employing 1500 views extracted from the image slice of the object that measured 1024 × 1024 pixels. These objects were designed to encompass a diverse array of randomly generated objects, each exhibiting varying levels of contrast and intensity (Zhengchun Liu, Kettimuthu, & Foster, 2022). We generated simulated high-pitch PBI-HCT sinograms from PBI-CT sinograms based on experimental data instead of defining scanning trajectories. Missing row locations were identified using experimental high-pitch PBI-HCT data (see Section 2.4), and a 0–1 mask extracted relevant information for realistic sinogram generation. Different masks were created for varying pitches, with pitch 3 tested. The complete sinogram served as the training target. For reconstruction, we used the ViT module to fill the sinogram, applied the Inverse Radon Transform (“iRadon”) for FBP, and then trained the U-Net model to enhance image quality.

2.4. Imaging experiments

2.4.1. Biomaterial and biological samples

In this study, we utilized hydrogel tissue scaffolds (X. Chen, et al., 2023; Ning, et al., 2021) in water, representative of low-density biomaterials, and a rat’s hindlimb (male Sprague–Dawley rat) implanted with Polycaprolactone (PCL) stent and hydrogel scaffolds to aid in sciatic nerve regeneration, serving as the biological sample. These samples were employed to illustrate the biomedical visualization capabilities of high-pitch PBI-HCT imaging using the proposed ViT-U. The tissue scaffolds and rat’s hindlimb tissues utilized in this study were prepared following the method established in our prior research (Duan, Li, et al., 2023; Ning, et al., 2021).

2.4.2. Imaging setup

The high-pitch PBI-HCT imaging experiments were performed at the Biomedical Imaging and Therapy (BMIT) 05ID-2 beamline (Wysokinski, et al., 2007) at Canadian Light Source (CLS). All scans were performed at sample-to-detector distance of 1.5 m. The FOVs of the detector were 2 mm × 26.6 mm with pixel size of 13 μm for hydrogel tissue scaffolds sample in binning 1 × 1 mode and 2 mm × 26.6 mm with pixel size of 26 μm in binning 2 × 2 mode for rat’s hindlimb sample, respectively. The effective N_p was 1500 for 180° for both binning modes of high-pitch PBI-HCT imaging with pitch of 2 (as the training target), and high pitch of 3, 3.5, 4, and 4.5. The exposure time per projection was set at 30 ms. Phase retrieval algorithm (i.e., transport of intensity equation, TIE) converted edge-enhanced phase contrast to virtual absorption contrast for quantitative analysis. UFO (Vogelgesang, et al., 2016) performed phase retrieval on projections and high-pitch PBI-HCT reconstructions (FBP algorithm). The training dataset included 5,000 sinograms, with 1,000 for validation and 500 for testing. Training began with the ViT module pre-trained on simulated data.

2.5. Comparison algorithms

To comprehensively evaluate the feasibility and efficacy of the proposed ViT-U, the performance of two state-of-the-art algorithms, i.e., hdNet (Q. Zhang, et al., 2020) and Double-CNNs (Hayes, et al., 2021) were compared. The training dataset for both algorithms consisted of paired sinograms and corresponding high-pitch PBI-HCT with same amount to ViT-U. In addition, we also conducted an analysis to evaluate the filling performance by employing a trained U-Net model solely in the

sinogram domain for high-pitch PBI-HCT imaging, named sinogram-based U-Net. The purpose of these results is to compare the sinogram filling ability between ViT and U-Net.

3. Results

3.1. DL algorithm comparison based on simulations

The processed results of high-pitch PBI-HCT imaging with pitch of 3 by using different methods are shown in Figure 3. The pure U-Net model fails to restore large gaps in the sinogram, causing missing wedge artifacts in high-pitch PBI-HCT images. The hdNet method partially fills gaps but introduces artifacts and uniform shadowing. Double-CNNs cause blurring artifacts, leading to incomplete edges. In contrast, ViT-U shows significant improvement, with no blurring and effective edge restoration. While it lacks full sharpness compared to the reference with a pitch of 2, it maintains high quality by avoiding distortions and artifacts. We quantitatively assessed image quality and accuracy using peak signal-to-noise ratio (PSNR) and SSIM metrics (Duan, Ding, et al., 2023). Our analysis confirms that the ViT-U method consistently achieves superior SSIM and PSNR results, supporting our qualitative observations.

The figure above evaluates image quality using a single paired sinogram and high-pitch PBI-HCT image, which may not reflect dataset variability. For a more thorough examination, a statistical analysis of the 500 processed high-pitch PBI-HCT images in terms of SSIM and PSNR is presented in Figure 4. This analysis demonstrates the generalizability of algorithm performance across different high-pitch PBI-HCT image slices. All four algorithms show a significant enhancement in image quality ($p < 0.001$), with notable differences in their outcomes.

3.2. DL algorithm comparison based on biomaterial sample imaging

The processed sinograms and paired high-pitch PBI-HCT images of tissue scaffolds with a pitch of 3 are depicted in Figure 5. The figure (Figure 5 (A2)) illustrates that the original FBP reconstruction results with a pitch of 3 exhibit pronounced shadow and missing wedge artifacts throughout the images, leading to missing details and blurred edges. These findings underscore the limitations of analytical algorithms for tissue scaffolds imaging using high-pitch PBI-HCT. Upon inspection of the Figure 5 (B1), it is evident that only using the U-net in the sinogram domain fails to adequately restore missing information, particularly for positions distant from rows containing existing information. Consequently, the resulting high-pitch PBI-HCT image (Figure 5 (B2))

fails in artifact suppression and edge preservation, showing minimal improvement in image quality and remains afflicted by missing wedge artifacts. The scaffold images show distortion, obscuring the definition of individual strands and potentially causing segmentation errors. While hdNet restores missing information better, it roughly approximates the sample's structure and fails to preserve fine details, resulting in blurring and artifacts along strand edges in the high-pitch PBI-HCT image. Similarly, Double-CNNs improve missing sinogram restoration but struggle with fine detail reproduction, leaving noticeable blurring and residual missing wedge artifacts that affect the object's structural intricacies.

In contrast, ViT-U outperforms hdNet and Double-CNNs in restoring missing information with high fidelity. The high-pitch PBI-HCT image shows minimal artifacts and blurring, effectively reducing missing wedge artifacts and significantly improving image quality, closely approaching the ground truth at a pitch of 2. Quantitative assessments reveal a substantial increase in SSIM and PSNR, rising from 0.35 to 0.96 and from 8.63 to 26.59 after processing with ViT-U.

Figure 6 show the corresponding three-dimensional (3D) segmentation of reconstructed scaffolds in Figure 5. The original reconstructed results show significant noise, obscuring the scaffold strands in segmentation. Sinogram-based U-Net and hdNet reduce some noise but still yield cluttered results that hinder accurate microstructure analysis. Double-CNNs minimize noise but oversmooth, negatively affecting scaffold volume measurements. In contrast, ViT-U delivers the best segmentation results, closely resembling the ground truth, and while some differences remain, they are acceptable for quantitative analysis, achieving the Dice score (Ronneberger, Fischer, & Brox, 2015) of 0.95.

3.3. DL algorithm comparison based on biological sample imaging

Experimental high-pitch PBI-HCT images of rat's hindlimb with a pitch of 3 are shown in Figure 7. The figure shows shadow and missing wedge artifacts in original FBP reconstructions at a pitch of 3, with noticeable distortion in bone and soft tissues. Using only U-Net for sinogram restoration results in minimal improvement in high-pitch PBI-HCT image quality, making it difficult to identify the PCL stent and its structure for subsequent quantitative analysis. While hdNet and Double-CNNs provide better restoration, they still struggle with fine details, leading to blurring and artifacts. The PCL stent's clarity improves, but identifying scaffold regions remains challenging, and while the bone structure appears complete, it is still blurred. In contrast, ViT-U effectively restores missing information with minimal artifacts and blurring,

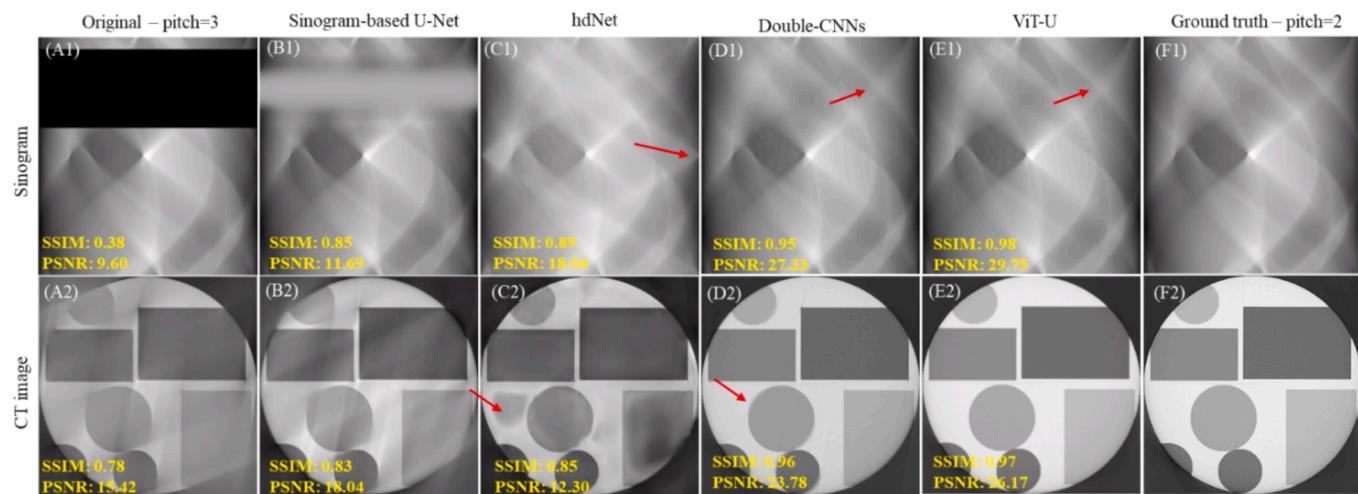


Fig. 3. Comparison of simulated sinograms and reconstructed images of high-pitch PBI-HCT imaging. (A1) Original sinogram with a pitch of 3, (B1) sinogram processed by U-Net, (C1) sinogram processed by hdNet, (D1) sinogram processed by Double-CNNs, (E1) sinogram processed by the proposed ViT-U, and (F1) ground truth sinogram (pitch of 2, with no missing information). (A2)-(F2) Reconstructed images from corresponding sinograms in (A1)-(F1).

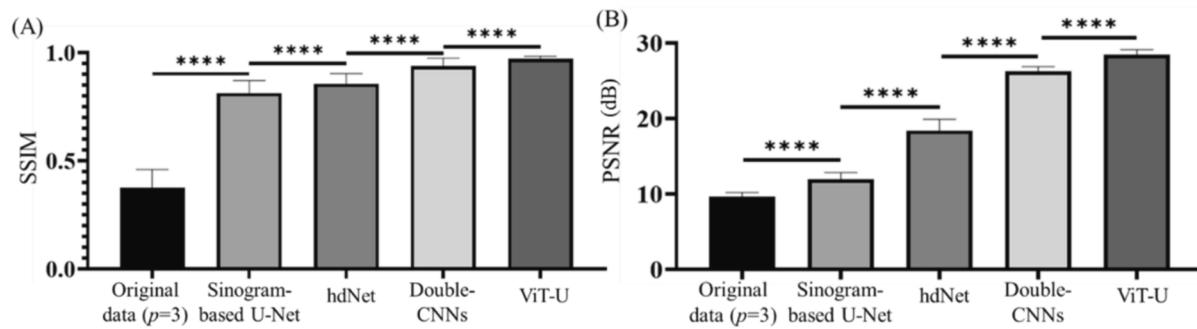


Fig. 4. Statistical comparison of simulated reconstructed high-pitch PBI-HCT original images and processed images by using four different algorithms in terms of (A) SSIM and (B) PSNR. Paired *t*-test was utilized with “****” representing $p < 0.0001$.

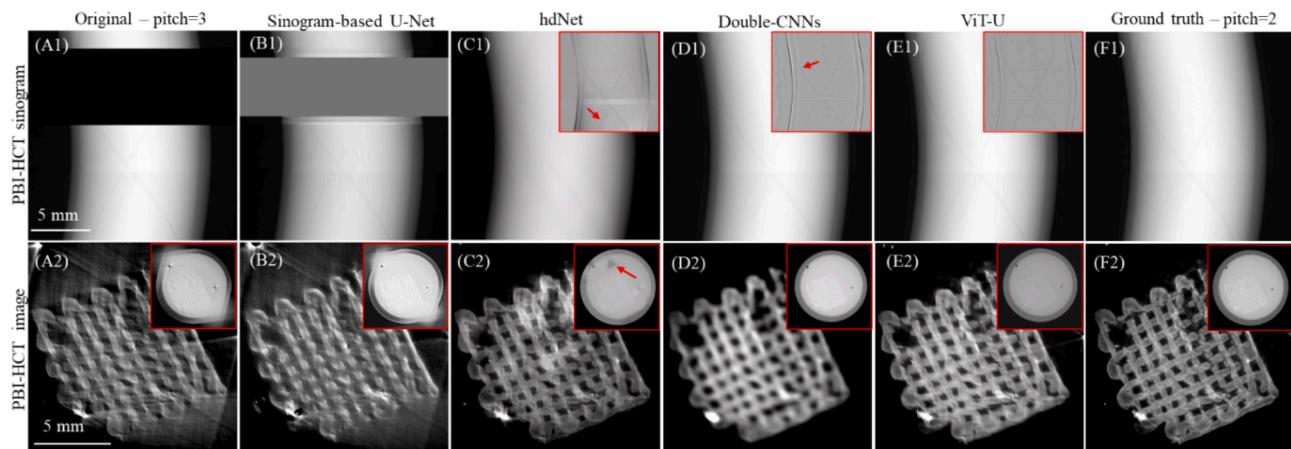


Fig. 5. Comparison of sinograms and reconstructed images (tissue scaffolds) of high-pitch PBI-HCT imaging after phase retrieval ($\delta/\beta = 1000$), (A1) Original sinogram with a pitch of 3, (B1) sinogram processed by U-Net, (C1) sinogram processed by hdNet, (D1) sinogram processed by Double-CNNs, (E1) sinogram processed by the proposed ViT-U, and (F1) ground truth sinogram (pitch of 2, with no missing information). (A2)-(F2) Corresponding reconstructed images from the sinograms in (A1)-(F1). The images highlighted within the red rectangle in (C1)-(E1) represent the sinogram differences when compared with the ground truth. The images enclosed within the red rectangle in (A2)-(F2) depict the entirety of the images. These images provide clearer boundary information for analysis. The red arrows indicate areas of weaker performance by the respective algorithms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

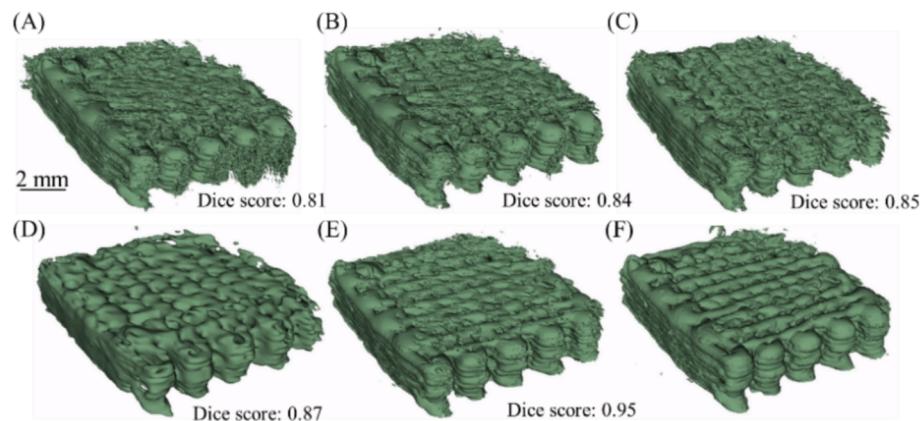


Fig. 6. Comparison of reconstructed scaffolds segmentation presented in 3D rendering using PBI-HCT imaging with a pitch of 3, processed by (A) original FBP reconstructed images and processed results obtained through various methods, including (B) sinogram-based U-Net, (C) hdNet, (D) Double-CNNs, and (E) our proposed ViT-U, along with (F) the ground truth of PBI-HCT imaging with a pitch of 2.

significantly enhancing image quality and clearly revealing both the PCL stent and scaffold regions for potential quantitative analysis. Some streaking artifacts in soft tissue are present but minimally impact bone detail analysis, as shown in Figure 7 (A2)-(F2).

Quantitative assessments show significant improvements in SSIM

and PSNR after processing with ViT-U, increasing from 0.34 to 0.94 and from 8.62 to 26.31, respectively. Although some missing wedge artifacts persist near bone edges due to contrast requirements for soft tissues, the majority of image components remain largely unaffected and can be accurately assessed. This indicates that the overall integrity and



Fig. 7. Comparison of sinograms and reconstructed images (rat's hindlimb) of high-pitch PBI-HCT imaging after phase retrieval, $\delta/\beta = 200$. (A1) Original sinogram with a pitch of 3, (B1) sinogram processed by U-Net, (C1) sinogram processed by hdNet, (D1) sinogram processed by Double-CNNs, (E1) sinogram processed by the proposed ViT-U, and (F1) ground truth sinogram (pitch of 2, with no missing information). (A2)-(F2) Corresponding reconstructed images from the sinograms in (A1)-(F1). The images highlighted within the red rectangle in (C1)-(E1) represent the sinogram differences when compared with the ground truth. The images enclosed within the red rectangle in (A2)-(F2) depict additional bone details shown with appropriate contrast. The red arrows highlight areas of the PCL stent and yellow arrows indicate areas of the degraded and swollen scaffolds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

usability of the image are maintained despite the artifacts.

4. Performance evaluations of ViT-U for varied imaging conditions

4.1. Evaluation of ViT-U's performance with different pitches

We explored the application of ViT-U in high-pitch PBI-HCT imaging across different scanning pitches. Our goal was to assess the performance of ViT-U in relation to helical pitch and identify the optimal pitch for future high-pitch PBI-HCT applications, taking into account factors like radiation dose and image quality.

The high-pitch PBI-HCT image results of tissue scaffolds obtained with various pitches (i.e., 3, 3.5, 4, and 4.5) are depicted in Figure 8. Compared to a pitch of 3, pitches of 3.5 and 4 show increased distortion and blurring, hindering strand identification. While the strand boundaries remain discernible, the structure becomes incomplete. Higher pitches allow for a reduction in radiation dose by factors of 1.75 (pitch of 3.5) and 2 (pitch of 4), maintaining high image quality compared to the

baseline pitch of 2. However, when image quality is critical and radiation dose is less of a concern, ViT-U may not perform well at pitches of 3.5 and 4. Results from high-pitch PBI-HCT imaging at a pitch of 4.5 (reducing radiation dose by a factor of 2.25) show artifacts that distort strand shapes, making boundary identification difficult, indicating that ViT-U is ineffective in this case.

In addition to qualitative image quality analysis, we also conducted statistical quantitative analysis using SSIM and PSNR for the entire test dataset, as depicted in Figure 9. As depicted in the figure, the SSIM and PSNR exhibit relatively low values across pitches of 3, 3.5, 4, and 4.5 in the absence of utilizing ViT-U for high-pitch PBI-HCT imaging. This suggests a notable degradation in image quality across these pitch values, emphasizing the need for additional enhancement techniques like ViT-U to improve imaging outcomes. By using ViT-U, a notable trend emerges, demonstrating a substantial decrease in both SSIM and metrics with increasing pitch values. This reduction becomes particularly pronounced at a pitch of 4.5, indicating a significant decline in image quality compared to lower pitch values. This observation aligns with our conclusions from Figure 8.

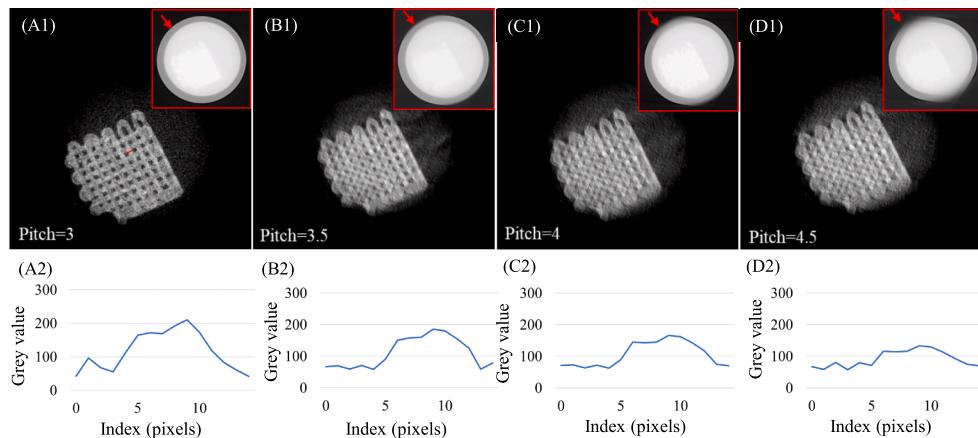


Fig. 8. Comparison of tissue scaffolds reconstructed images of high-pitch PBI-HCT imaging (after phase retrieval, $\delta/\beta = 1000$) using ViT-U, examining with a pitch of (A1) 3, (B1) 3.5, (C1) 4, and (D1) 4.5. (A2)-(D2) Corresponding grey value at position indicated by red lines in (A1)-(D1). The images enclosed within the red rectangle are displayed with a different contrast. The red arrows highlight areas where the respective algorithms demonstrate less effective performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

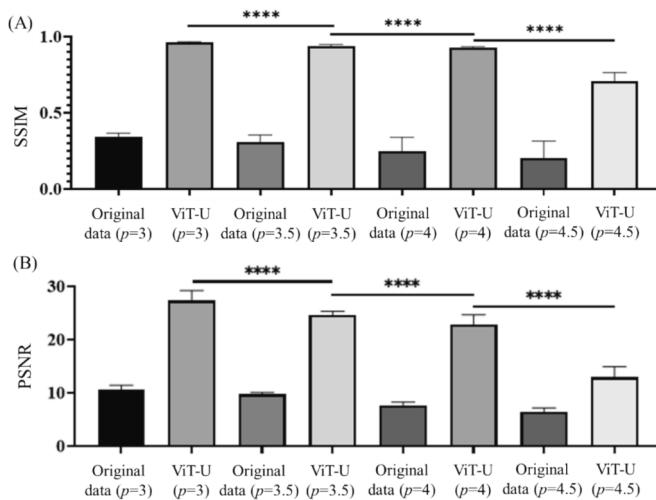


Fig. 9. Statistical comparison of reconstructed high-pitch PBI-HCT images of tissue scaffolds with four different pitches before and after using ViT-U in terms of (A) SSIM and (B) PSNR. Paired t-test was employed, with “****” indicating $p < 0.0001$.

In addition, the 3D segmentation accuracy of the hydrogel scaffolds using high-pitch PBI-HCT imaging before and after the use of ViT-U with different pitches was evaluated, represented by the Dice score, as shown in Figure 10. The accuracy of 3D segmentation is crucial for comprehensive quantitative analysis of tissue scaffolds, such as porosity characterization or degradation behavior visualization, offering a deeper understanding of the tissue regeneration mechanism.

4.2. Evaluation of ViT-U's performance for low-dose conditions

The evaluation of ViT-U's performance for the high-pitch PBI-HCT imaging in low dose condition (all with a pitch of 3) was illustrated in Figure 11. We adjusted photon flex to have 72 % and 93 % dose reduction by utilization of neutral density filters with thicknesses of 40 mm and 80 mm, respectively. The results indicate that the restoration of missing information is unaffected by noise, compared with noise-free high-pitch PBI-HCT image. Despite the presence of noise impacting image quality, the scaffold information can still be effectively restored, with relatively clear and sharp edges observed. This resilience to noise underscores the robustness of the restoration process of ViT-U, ensuring that critical structural details are accurately preserved despite environmental disturbances and low-flux scanning.

Additionally, with ViT-U, we performed a statistical comparison of clean (0 % dose reduction), and noisy (72 % and 93 % dose reduction) reconstructed high-pitch PBI-HCT images of tissue scaffolds in terms of

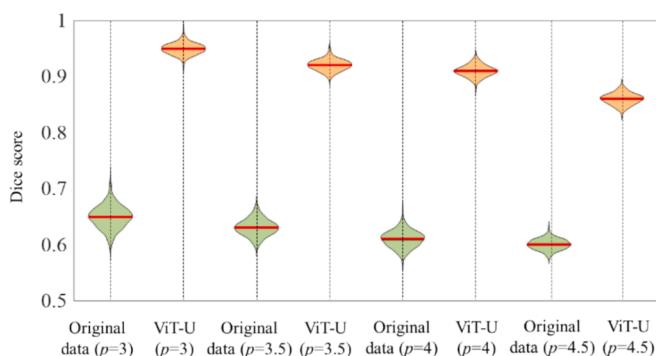


Fig. 10. Comparison of reconstructed high-pitch PBI-HCT images of tissue scaffolds with four different pitches before and after using ViT-U in terms of Dice score.

SSIM and PSNR, depicted in Figure 12.

When assessed using SSIM, ViT-U consistently performs well across both clean and noisy images. However, significant differences in PSNR values are observed between clean and noisy images ($p < 0.1$ for a 73 % dose reduction and $p < 0.01$ for a 93 % dose reduction, respectively). Nonetheless, when considering $p < 0.001$ as the significance threshold, ViT-U demonstrates consistent restoration performance across all clean and noisy images, underscoring its robustness and adaptability across diverse imaging scenarios.

5. Discussions

This work developed ViT-U, a novel DL-based algorithm for high-pitch PBI-HCT imaging to correct the missing wedge artifacts, ultimately providing superior imaging quality compared to existing algorithms. ViT-U's performance is evaluated across various pitch imaging conditions and in processing low-dose imaging data, showcasing its potential in low-dose imaging. This innovative algorithm enables rapid and precise imaging at low doses, making it promising for live animal imaging and longitudinal research applications.

The results demonstrate that ViT-U exhibits outstanding performance in removing missing wedge artifacts, as evidenced by both simulation and experimental data, including hydrogel tissue scaffolds and rat's hindlimb samples. Compared to biomaterial samples used in this work which only have two components (i.e., water and hydrogel), the practical applications of biological samples (rat's hindlimb with implants) are more complex, and improving image quality is also more difficult. In particular, all algorithms evaluated demonstrate efficacy in high-pitch PBI-HCT imaging of tissue scaffolds samples, with ViT-U exhibiting the most superior performance in image restoration. However, the missing wedge artifacts observed in Figure 7 for rat's hindlimb samples persist using ViT-U. This may result from the discontinuities of grayscale values between rows with missing information and rows with existing information in the sinogram domain. This issue becomes more pronounced when complex components with significantly different contrast coexist within the same slice. Existing algorithms are unable to effectively mitigate this problem. One potential solution involves training separate models for different components with varying contrast levels or training a model specifically for small regions. Such an algorithm could substantially reduce the complexity of biological samples and potentially improve artifact reduction.

It's worth noting that ViT-U heavily depends on precise geometry calibration in high-pitch PBI-HCT imaging. Alignment errors, especially in roll and pitch, can cause streaking artifacts and degrade ViT-U's performance. Mechanical accuracy in the rotation stage, including straightness and flatness, is critical. In high-pitch imaging, even small mechanical errors can introduce artifacts, particularly in high-resolution scenarios, impacting ViT-U results.

6. Future directions

Further to the present study, the following directions for future research would be interesting:

- 1) **Fairness-aware learning:** Fairness-aware learning is great to be incorporated in such imaging algorithms as the one developed in the present study. The capability of fairness-aware learning is able to ensure that artifacts are uniformly removed across different samples, thus preventing biases in image reconstruction for specific regions or sample types. Also, it allows for consistent image quality across both fine details and broader areas. Furthermore, by applying fairness-aware principles universal models could be developed to enhance generalizability, thus ensuring consistent and unbiased artifact removal and image quality across different helical pitches and dose reduction levels. These universal models would be used to support equitable and unbiased biomedical applications.

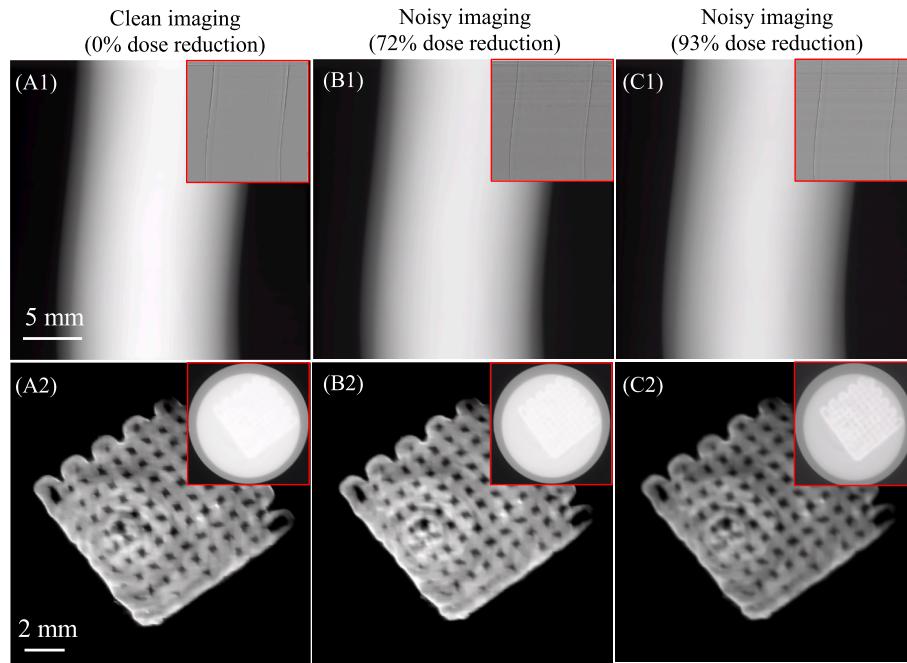


Fig. 11. Results comparison (pitch of 3) among processed sinograms and high-pitch PBI-HCT images of tissue scaffolds (after phase retrieval, $\delta/\beta = 1000$) using ViT-U with dose reduction of (A1)-(A2) 0 %, (B1)-(B2) 72 %, and (C1)-(C2) 93 %. The images highlighted within the red rectangle in (A1)-(C1) represent the sinogram differences when compared with their own ground truth. The images enclosed within the red rectangle in (A2)-(C2) are shown in a different contrast. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

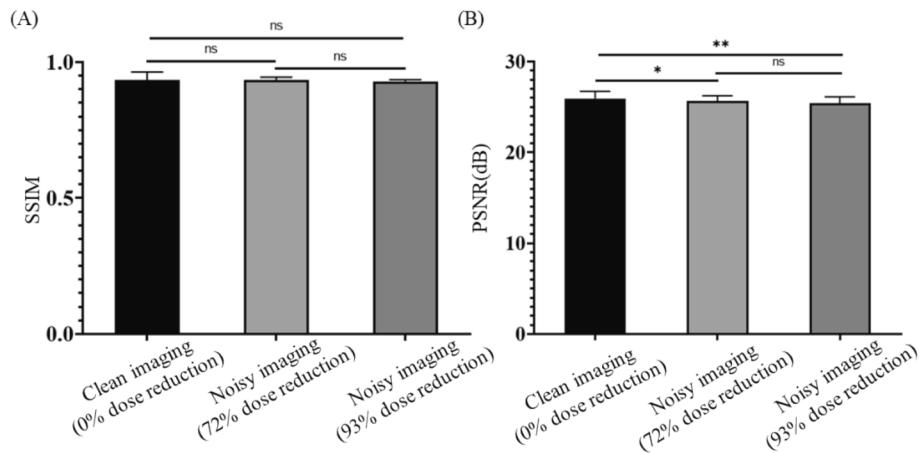


Fig. 12. Statistical comparison of reconstructed high-pitch PBI-HCT clean images and noisy images (72 % and 93 % dose reduction) of tissue scaffolds using ViT-U in terms of (A) SSIM and (B) PSNR. Paired *t*-test was employed, “ns” denoting not significant; “**” indicating $p < 0.1$, and “***” representing $p < 0.01$.

- 2) **Unified framework:** The current approach based on separate training provides flexibility for handling diverse samples with varying density and features. Future work could involve training ViT in the sinogram domain and U-Net in the CT image domain simultaneously, and/or exploring DL models that replace both FBP reconstruction and post-processing in a unified framework. Another direction of interesting would be using ViT alone to directly generate high-pitch PBI-HCT reconstructions, leveraging the relationship between sinogram and image pixels. While this could streamline the process, it requires significant computational resources, especially for large transformer-based models. Hardware advances and optimized models would enable this approach to become practical, potentially revolutionizing high-pitch PBI-HCT imaging.
- 3) **Computational efficiency:** Given that the current approach involves substantial computational resources, it might be interesting to investigate or optimize model complexity and efficiency,

especially for large-scale transformer-based models. On this regard, Swin-transformer (Parakh, et al., 2021) and LocalViT (Yawei Li, Zhang, Cao, Timofte, & Van Gool, 2021) would be of options to reduce the training time without losing or even enhanced image quality.

- 4) **Low-dose high-pitch PBI-HCT imaging:** High-pitch PBI-HCT imaging, as a low-dose strategy, allows for reduced projections per rotation, thus making it compatible with low-dose CT reconstruction algorithms, like the sinogram-based DLIR approach, for further reducing radiation dose but retaining the image quality. Also, reducing the time needed for each projection would be of a great value to further minimize the radiation dose. On this regard, it is urged to develop advanced methods or algorithms, similar to Sparse2Noise (Duan, Ding, et al., 2023), for reducing radiation dose without compromising image quality or resolution, and to integrate these advanced methods or algorithms with high-pitch PBI-HCT to

further address the trade-off issue of radiation dose and imaging quality.

7. Conclusions

High-pitch PBI-HCT techniques offer flexibility and low radiation doses but can lead to missing wedge artifacts at pitches above 2, compromising image quality. This work introduces the ViT-U algorithm, which effectively mitigates these artifacts, enabling imaging with pitches exceeding 2. Simulation and experimental results show that ViT-U outperforms sinogram-based U-Net, hdNet, and Double-CNNs, preserving object details and reconstructing clearer boundaries. We achieved accurate reconstructions up to a pitch of 4 with minimal artifacts and demonstrated ViT-U's effectiveness in low-dose noisy conditions, achieving a 93 % dose reduction, highlighting its potential for imaging radiation-sensitive samples.

Declaration of generative AI and AI-assisted technologies in the writing process

The author(s) used ChatGPT to correct grammatical errors. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Xiaoman Duan: Conceptualization, Methodology, Data curation, Validation, Writing – original draft. **Xiao Fan Ding:** Data curation. **Samira Khoz:** Data curation. **Xiongbiao Chen:** Supervision, Resources, Funding acquisition, Writing – review & editing. **Ning Zhu:** Supervision, Methodology, Data curation, Funding acquisition, Writing – review & editing.

Declaration of competing interest

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

Acknowledgements

This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC, No. RGPIN 06007-2019 and No. RGPIN 06396-2019). Experimental research described in this paper were performed at the Canadian Light Source, which is a national research facility supported by the Canada Foundation for Innovation (CFI), NSERC, National Research Council (NRC), Canadian Institutes of Health Research (CIHR), Government of Saskatchewan, and the University of Saskatchewan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eswa.2024.125663>.

Data availability

Data will be made available on request.

References

- Ahmed, Z., Campeau, D., Gong, H., Rajendran, K., Rajiah, P., McCollough, C., & Leng, S. (2023). High-pitch, high temporal resolution, multi-energy cardiac imaging on a dual-source photon-counting-detector CT. *Medical physics*, 50, 1428–1435.
- Ando, A., Gidaris, S., Bursuc, A., Puy, G., Boulch, A., & Marlet, R. (2023). RangeViT: Towards Vision Transformers for 3D Semantic Segmentation in Autonomous Driving. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5240–5250).
- Anirudh, R., Kim, H., Thiagarajan, J. J., Mohan, K. A., Champley, K., & Bremer, T. (2018). Lose the views: Limited angle CT reconstruction via implicit sinogram completion. In *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6343–6352).
- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022). In *Swin-unet: Unet-like pure transformer for medical image segmentation* (pp. 205–218). Springer.
- Chen, C.-F.-R., Fan, Q., & Panda, R. (2021). Crossvit: Cross-attention multi-scale vision transformer for image classification. In *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 357–366).
- Chen, X., Anvari-Yazdi, A. F., Duan, X., Zimmerling, A., Gharraei, R., Sharma, N., Sweilem, S., & Ning, L. (2023). *Biomaterials/bioinks and extrusion bioprinting*. *Bioactive Materials*, 28, 511–536.
- Chen, Z., Jin, X., Li, L., & Wang, G. (2013). A limited-angle CT reconstruction method based on anisotropic TV minimization. *Physics in Medicine & Biology*, 58, 2119.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., & Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:1612.08008*.
- Duan, X., Li, N., Chen, X., & Zhu, N. (2021). Characterization of Tissue Scaffolds Using Synchrotron Radiation Microcomputed Tomography Imaging. *Tissue Engineering Part C: Methods*, 27, 573–588.
- Fan, J., Yue, M., & Melnyk, R. (2014). *Benefits of ASIR-V reconstruction for reducing patient radiation dose and preserving diagnostic quality in CT exams*. GE Healthcare: White paper.
- Gong, H., Ren, L., McCollough, C. H., & Yu, L. (2020). *Ultra-fast-pitch acquisition and reconstruction in helical CT* (Vol. 11312, pp. 51–57).
- Hausleiter, J., Bischoff, B., Hein, F., Meyer, T., Hadamitzky, M., Thierfelder, C., Allmendinger, T., Flohr, T. G., Schömöig, A., & Martinoff, S. (2009). Feasibility of dual-source cardiac CT angiography with high-pitch scan protocols. *Journal of cardiovascular computed tomography*, 3, 236–242.
- Hayes, J. W., Montoya, J., Budde, A., Zhang, C., Li, Y., Li, K., Hsieh, J., & Chen, G.-H. (2021). High pitch helical CT reconstruction. *IEEE Transactions on medical imaging*, 40, 3077–3088.
- Hsieh, J., Liu, E., Nett, B., Tang, J., Thibault, J.-B., & Sahney, S. (2019). *A new era of image reconstruction: TrueFidelityTM*. GE Healthcare: White Paper.
- Hu, D., Zhang, Y., Liu, J., Du, C., Zhang, J., Luo, S., Quan, G., Liu, Q., Chen, Y., & Luo, L. (2021). SPECIAL: Single-shot projection error correction integrated adversarial learning for limited-angle CT. *IEEE Transactions on Computational Imaging*, 7, 734–746.
- Hu, D., Zhang, Y., Liu, J., Luo, S., & Chen, Y. (2022). DIOR: Deep iterative optimization-based residual-learning for limited-angle CT reconstruction. *IEEE Transactions on Medical Imaging*, 41, 1778–1790.
- Lapegue, F., Faruch-Bilfeld, M., Demondion, X., Apredoaie, C., Bayol, M., Artico, H., Chiavassa-Gandois, H., Railhc, J.-J., & Sans, N. (2014). Ultrasoundography of the brachial plexus, normal appearance and practical applications. *Diagnostic and interventional imaging*, 95, 259–275.
- Li, J., Chen, J., Tang, Y., Wang, C., Landman, B. A., & Zhou, S. K. (2023). Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. *Medical image analysis*, 85, Article 102762.
- Li, Y., Mao, H., Girshick, R., & He, K. (2022). In *Exploring plain vision transformer backbones for object detection* (pp. 280–296). Springer.
- Lim, H. K., Ha, H. I., Hwang, H. J., & Lee, K. (2019). High-pitch, 120 kVp/30 mAs, low-dose dual-source chest CT with iterative reconstruction: Prospective evaluation of radiation dose reduction and image quality compared with those of standard-pitch low-dose chest CT in healthy adult volunteers. *Plos one*, 14, e0211097.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012–10022).
- Ning, L., Zhu, N., Smith, A., Rajaram, A., Hou, H., Srinivasan, S., Mohabatpour, F., He, L., McLnnes, A., Serpooshan, V., Papagarakis, P., & Chen, X. (2021). Noninvasive Three-Dimensional In Situ and In Vivo Characterization of Bioprinted Hydrogel Scaffolds Using the X-ray Propagation-Based Imaging Technique. *ACS Applied Materials & Interfaces*, 13, 25611–25623.
- Pan, B., Panda, R., Jiang, Y., Wang, Z., Feris, R., & Oliva, A. (2021). IA-RED \$2\$: Interpretability-Aware Redundancy Reduction for Vision Transformers. *Advances in neural information processing systems*, 34, 24898–24911.
- Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22, 1345–1359.
- Parakh, A., Cao, J., Pierce, T. T., Blake, M. A., Savage, C. A., & Kambadakone, A. R. (2021). Sinogram-based deep learning image reconstruction technique in abdominal CT: Image quality considerations. *European radiology*, 31, 8342–8353.
- Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2023). Transformers in medical imaging: A survey. *Medical image analysis*, 102802.
- Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M. K., Sun, L., Cong, W., & Wang, G. J. I. t. o. m. i. (2018). 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. *37*, 1522–1534.
- Strudel, R., Garcia, R., Laptev, I., & Schmid, C. (2021). Segmenter: Transformer for semantic segmentation. In *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 7262–7272).
- Sun, T., Clackdoyle, R., Kim, J.-H., Fulton, R., & Nyuyt, J. (2017). Estimation of local data-insufficiency in motion-corrected helical CT. *IEEE transactions on radiation and plasma medical sciences*, 1, 346–357.

- Tuy, H. K. (1983). An inversion formula for cone-beam reconstruction. *SIAM Journal on Applied Mathematics*, 43, 546–552.
- Van Aarle, W., Palenstijn, W. J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Batenburg, K. J., & Sijbers, J. (2016). Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics express*, 24, 25129–25147.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vogelgesang, M., Farago, T., Morgeneyer, T. F., Helfen, L., dos Santos Rolo, T., Myagotin, A., & Baumbach, T. (2016). Real-time image-content-based beamline control for smart 4D X-ray imaging. *Journal of synchrotron radiation*, 23, 1254–1263.
- Wang, D., Fan, F., Wu, Z., Liu, R., Wang, F., & Yu, H. (2023). CTformer: Convolution-free Token2Token dilated vision transformer for low-dose CT denoising. *Physics in Medicine & Biology*, 68, Article 065012.
- Wang, J., Liang, J., Cheng, J., Guo, Y., & Zeng, L. (2020). Deep learning based image reconstruction algorithm for limited-angle translational computed tomography. *Plos one*, 15, e0226963.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE transactions on image processing*, 13, 600–612.
- Wysokinski, T. W., Chapman, D., Adams, G., Renier, M., Suortti, P., & Thomlinson, W. (2007). Beamlines of the biomedical imaging and therapy facility at the Canadian light source—Part 1. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 582, 73–76.
- Yang, L., Li, Z., Ge, R., Zhao, J., Si, H., & Zhang, D. (2022). Low-dose ct denoising via sinogram inner-structure transformer. *IEEE Transactions on medical imaging*, 42, 910–921.
- Zhang, D., Liang, D., Zou, Z., Li, J., Ye, X., Liu, Z., Tan, X., & Bai, X. (2023). A simple vision transformer for weakly semi-supervised 3d object detection. In *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8373–8383).
- Zhang, Q., Hu, Z., Jiang, C., Zheng, H., Ge, Y., & Liang, D. (2020). Artifact removal using a hybrid-domain convolutional neural network for limited-angle computed tomography imaging. *Physics in Medicine & Biology*, 65, Article 155010.
- Zhang, W., Deng, L., Zhang, L., & Wu, D. (2022). A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica*, 10, 305–329.
- Zhou, W., Dou, P., Su, T., Hu, H., & Zheng, Z. (2023). Feature learning network with transformer for multi-label image classification. *Pattern Recognition*, 136, Article 109203.