PARNet: Aortic Reconstruction from Orthogonal X-rays Using Pre-Trained Generative Adversarial Networks

Chengwei Cao[1][0009−0001−4988−368X], Jinhui Zhang[1][0000−0002−2405−894X], Yueyang Gao[1][0009−0002−2404−3809], and Zheng Li[1][0009−0004−6813−2293]

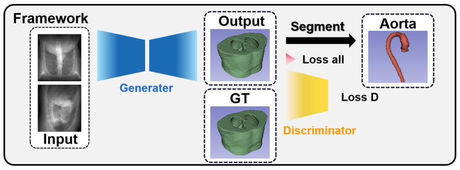
School of Automation, Beijing Institute of Technology, Beijing, China
caochw@bit.edu.cn
zhangjinh@bit.edu.cn

Abstract. The three-dimensional reconstruction of the aorta plays a crucial role in assisting minimally invasive vascular interventions to treat coronary artery disease, aiding surgeons in finding the optimal procedural angles for locating and delivering intervention devices. However, existing reconstruction methods face challenges such as weak imaging capability for low-density tissues in X-rays, limiting the accurate capture and reconstruction of the aorta and other blood vessels. To address these challenges, we propose PARNet, a deep-learning approach for 3D aortic reconstruction from orthogonal X-rays. PARNet leverages pre-training information to extract global and local features using Aortic Reconstruction with Background X-rays (ARB) module and Aortic Reconstruction with Mask X-rays (ARMask) module, respectively, thereby enhancing the model's reconstruction performance with more aortic details. Additionally, customized loss functions are introduced to adapt to the low-density characteristics of the aorta. The results demonstrate that our method outperforms existing approaches, producing results that are visually closest to the ground truth on mainstream datasets.

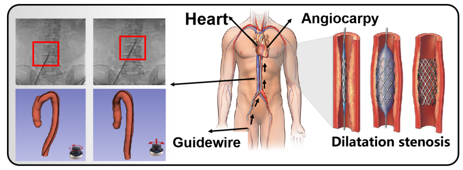
Keywords: 3D aortic reconstruction · GAN · Pre-trained.

1 Introduction

Cardiovascular disease has a high mortality rate [30], with coronary heart disease(CHD) being one of the most prevalent, primarily due to vascular stenosis or occlusion, which leads to myocardial ischemia, hypoxia, or necrosis. Patients with coronary heart disease typically undergo minimally invasive vascular intervention, involving percutaneous puncture of the thigh or arm, guiding devices through the aorta to the lesion site, where the stenosed vessel is dilated, and a stent is placed to restore normal blood flow. Therefore, the aorta, which can carry out internal blood delivery [27], is essential for the delivery of interventional devices. To assist surgeons in accurately locating interventional devices during surgery, the digital subtraction angiography (DSA) system is commonly



(a) Illustration of the PARNet for aorta reconstruction from X-rays to 3D visualization results



(b) 2D/3D aortic visualization results for interventional surgeries

Fig. 1: (a). Illustration of the proposed method for aorta reconstruction from X-rays to 3D visualization results. (b). The 2D/3D aortic visualization results for interventional surgeries. The first row shows that DSA systems can only produce 2D X-rays during the balloon catheter delivery. However, our method (see second row) can provide a 3D visualization of the aorta to simulate the guidewire delivery process and assist surgeons in delivering interventional devices. Surgeons can rotate to view multiple views of the aorta through the 3D Mouse, a six-dimensional manipulator in the bottom right corner.

used to generate X-rays (see Fig 1b), which map the devices' positions within the patient's blood vessels[10].

In vascular interventional surgery, X-rays are advantageous due to their fast imaging speed, lower radiation damage compared to CT [35], low cost, and convenient process [38]. As a result, surgeons can only use the two-dimensional X-ray images generated by DSA to guide interventional procedures with experience and imagination. In this case, providing surgeons with valuable 3D vascular information by developing a method to convert 2D aortic X-rays into 3D CT images holds great practical value. Due to the possibility of artifacts in the reconstruction of small vessels such as coronary arteries affecting the accuracy, this method cannot completely replace DSA. However, it can accurately reconstruct vascular structures like the aorta, providing clinically valuable 3D information during interventional surgery where CT imaging equipment is limited. 3D visualization of the aorta and other blood vessels can help surgeons locate the interventional devices and determine the optimal operating angle for their delivery (see Fig 1b), thereby improving the accuracy and safety of surgery and reducing radiation exposure for patients.

Due to the characteristics of X-rays, smaller anatomical tissues often lack distinct features and are occluded by denser surrounding tissues and organs, so the 3D reconstruction of the aorta faces great challenges [29, 23]. To assist surgeons in minimally invasive vascular interventional procedures, we designed a deep learning method called PARNet to convert 2D X-rays into 3D aortic CT images. We used digitally reconstructed X-ray images (DRRs) [28] to simulate X-rays and generate paired datasets. We also used CycleGAN [47] to learn the style of real X-rays to make up for the gap between DRR images and real X-rays [4, 5, 45]. Based on the central architecture of Generative Adversarial Networks

(GANs) [31], as shown in Fig 1a, PARNet takes the orthogonal X-rays of the human chest obtained by the above process as input and outputs 3D CT result containing the aorta. In addition, PARNet obtains the aortic X-ray map by pre-training as input information. Finally, using the automatic segmentation tools TotalSegmentator [37, 19] to obtain three-dimensional CT data of the aorta and visualization separately. Overall, this study enhances PARNet by incorporating pre-training information on aortic X-ray surface features and additional techniques to achieve a more accurate 3D reconstruction of the aorta. The specific contributions of this study are as follows.

1) A novel deep learning-based network, the PARNet, is designed, which utilizes aortic X-rays from pre-training information, enabling 3D aortic reconstruction through comprehensive orthogonal 2D X-rays, culminating in the acquisition of reconstructed aorta results using an automated segmentation tool.

2) The ARB and ARMask modules are proposed to improve model performance. PARNet uses the ARMask module to extract pixel-level features as local features from pre-training aortic X-rays and performs feature fusion with the global spatial feature information extracted from chest orthogonal X-rays by the ARB module. This approach enhances PARNet's perception of aortic features, avoids the loss of aortic feature information, and improves reconstruction accuracy.

3) To enhance PARNet's 3D reconstruction capabilities, a customized loss function is implemented to help accurately reconstruct low-density soft tissues like the aorta, preventing the loss of reconstruction details.

2 RELATED WORKS

To provide valuable preoperative information for surgeons, it is imperative to develop 3D visualizations of the aorta and other blood vessels. While certain studies have successfully automated the meticulous reconstruction of 3D vascular structures, including the aorta, from chest CT scans [42, 15], these methods rely on three-dimensional CT data for vascular reconstruction. However, during surgical procedures, it is more feasible to obtain X-rays of patients. The aforementioned methods are incapable of directly deriving the necessary 3D reconstruction from X-rays, thus failing to meet the visualization requirements of vascular interventional surgeries.

Other studies have focused on visualizing and reconstructing the aorta and other vasculature based on thoracic and pulmonary reconstructions [16, 22, 40], but these methods do not generate images comparable to three-dimensional CT scans. However, due to the lower radiation exposure [35] and widespread use of X-rays for vascular imaging, the 3D reconstruction of anatomical structures from 2D X-rays has long been a significant research focus in the field of medical imaging.

Recent advancements in 3D reconstruction from X-ray images have predominantly relied on deep learning techniques, with numerous research teams pioneering innovative approaches employing convolutional neural networks (CNNs), GANs, and transformer models to attain high-accuracy 3D anatomical reconstruction from single or dual-view X-ray perspectives [25, 2]. Other studies are concerned with 3D reconstruction based on sparse or limited view X-rays and require a large number of simulations of multi-view projection results [7, 6, 43, 41, 9]. Henzler et al. [16] developed a deep CNN method that initially learns a coarse fixed-resolution volume and then fuses it with input X-rays to produce a high-resolution volume, ultimately generating the full 3D skull. Shen et al. [34] utilized deep learning to map projected X-rays of patients to their corresponding 3D anatomical structures, generating volumetric X-ray tomography images from a single projection view to reconstruct regions such as the upper abdomen, lungs, and head and neck. Kasten et al. [21] designed an end-to-end convolutional neural network method capable of directly reconstructing the 3D structure of knee joints from orthogonal X-rays, effectively learning the shape distribution of bones from training images. In the context of sparse view CT reconstruction, Wang et al. [36] proposed a deep learning model called DIDR-Net, which achieves 3D reconstruction and detail recovery through an iterative reconstruction sub-network and a detail recovery sub-network.

Ying et al. [39] were the first to leverage the advantages of GANs to reconstruct CT volumes from orthogonal X-rays. The proposed X2CT-GAN benefits from the additional views, enabling high-quality image synthesis and extracting finer anatomical structures, such as blood vessels within the lung. Since then, most related research has been carried out based on CNNs and GANs, which are also the main architectures of our proposed PARNet. Saravi et al. [32] proposed an improved method of X2CT-GAN and Gao et al. [12] proposed 3DSRNet, demonstrating the potential of GANs for synthesizing orthogonal X-rays for 3D spine reconstruction. Saravi et al.. effectively utilized feature fusion techniques to combine information from multiple views. Gao et al.. employed the CNN-transformer framework and detail extraction method for pixel-level reconstruction, which inspired the design of PARNet, particularly in integrating local and global features to enhance the accuracy of aortic reconstruction. Zhang et al. [44] proposed XTransCT, a super-fast volume CT reconstruction method using a transformer framework, achieving significant progress in reconstruction speed.

Despite the potential adaptability of these methods for aortic reconstruction, they still face significant hurdles in effectively handling low-density vascular tissue details in 2D X-rays. Existing techniques heavily rely on CNNs and transformers to extract features from orthogonal 2D X-rays of the human chest, often falling short in capturing sufficient aortic feature information. Moreover, the presence of high-density organs like bones, which can obscure aortic imaging, exacerbates the challenge. Crucial low-level features may be lost in deep convolutions, impeding the extraction of precise high-level details from the extensive sensory field. This, in turn, leads to blurring and loss of image information, undermining their efficacy in practice. Therefore, the accuracy of existing methods for reconstructing 3D geometry from 2D X-rays of the aorta still requires improvement.

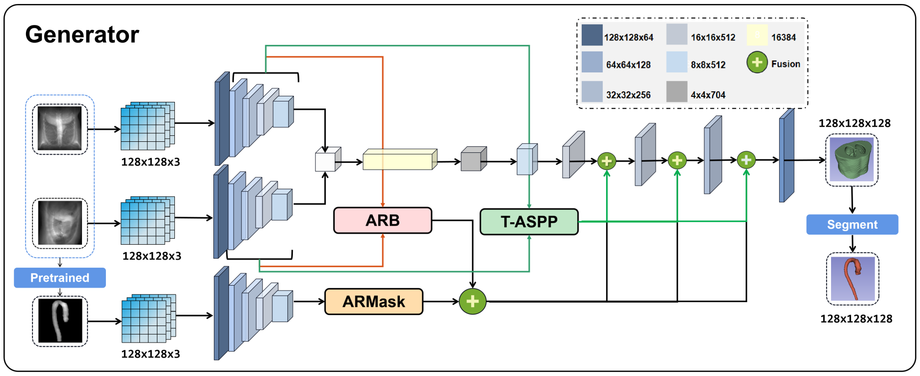
3 Method

In this paper, we introduce PARNet, an orthogonal 2D X-ray-based network that leverages pre-training information within a generative adversarial framework for 3D CT image reconstruction of the aorta. Consistent with other 3D GAN architectures, our network comprises a 3D generator and a 3D discriminator. In this cross-modal application scenario, our method facilitates competition between the generator and discriminator during the training phase, ingeniously recreating the rich 3D details of the aorta through a self-learning mechanism. The overall architecture of the reconstructed network is depicted in Fig 1a. Within the human chest, low-density anatomical tissues such as the aorta exhibit different imaging details compared to high-density tissues like bone, with these features varying across different regions. Building on this understanding, this paper employs aortic X-rays obtained through pre-training as extra information input.

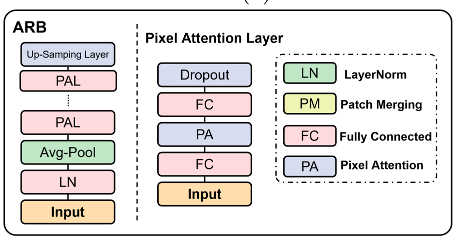
The ARB and ARMask algorithms are proposed to enhance reconstruction ability by utilizing both the global structural information of chest X-rays and the local details of the aorta obtained from pre-training information respectively and fuse these features effectively. Specifically, the ARB module extensively extracts the global structural information of orthogonal chest X-rays as background information, while the ARMask module processes the pre-training features from aortic X-rays as focused local attention features to enhance the model's visual field perception of the aorta. The results obtained by these two modules are fused and output to the generator's decoding module to enhance the aorta's detailed features. Additionally, PARNet integrates ASPP from 3DSRNet [48, 12] with transformer and other modules, with the proposed T-ASPP maximizing the utilization of global features and low-level detail information to enhance reconstruction ability. Together, these methods contribute to PARNet's efficacy for 3D reconstruction of aortic vessels. These methods are described in detail below.

3.1 Generator Architecture

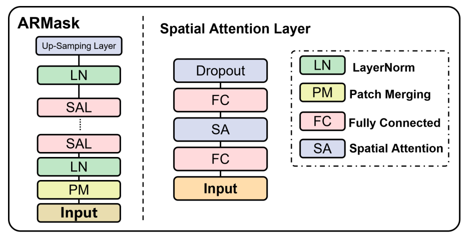
The overall architecture of the generator network is shown in Fig 2a. The generator network comprises two encoder-decoder networks with identical architectures for posteroanterior (PA) and lateral (LA) X-ray inputs, respectively. By introducing convolutional layers for feature extraction, the encoder-decoder network learns a mapping in the feature space from the input 2D X-rays to the target 3D CT images. Parallel skip connections are utilized for feature extraction to augment the original detail information. Additionally, a separate set of encoders is specifically designed for extracting feature information from pre-training aortic X-rays. After fusing these feature representations with the feature output extracted by the encoder from the orthogonal X-ray input, the fusion result is fed into the feature fusion network of the generator. This feature fusion network captures and backpropagates any structural inconsistencies between the outputs of the two decoders, ensuring that the visual constraints of both views



(a) Architectural details of the generator



(b) Network architecture of the ARB module



(c) Network architecture of the ARMask module

Fig. 2: (a). The framework of the generator. Two parallel encoders extract features from two orthogonal input images. ARB in (b) and ARMask in (c) are two modules that can obtain more features of aortic images.

are enforced. The decoder part consists of 3D deconvolution layers, which are responsible for generating the 3D CT image and ultimately producing and visualizing the 3D reconstruction result of the aorta using segmentation tools. In addition, a feature map dimension transformation module is introduced between the encoder and decoder parts of the network to expand the 2D feature map to a pseudo-3D feature map. This proposed method preserves rich low-level information, enhancing the strong correlation between input and output regarding shape and appearance.

The generator network uses a pyramid-like convolutional neural network encoder-decoder architecture to aggregate local features hierarchically. Although CNNs can effectively extract 2D features through the above multiple pyramid-like convolutional layers, these features alone are not sufficient for 3D aortic reconstruction. Therefore, we adopt the T-ASPP module [12], which integrates atrous spatial pyramid pooling (ASPP) and transformer algorithm as skip connections to improve the accuracy and completeness of reconstruction. The generator feature fusion network is connected, and the fusion feature results are inputted. This enables the generator network to focus on both high-level attributes and low-level details of the image to produce the final 3D output and enhance the effectiveness of 3D aortic reconstruction. In summary, PARNet can achieve richer and more detailed 3D aortic structure reconstruction.

3.2 Pre-training Process

In the 3D reconstruction of the aorta, mitigating the impact of high-density organs like bones on aortic feature extraction, aiding the model in extracting sufficient high-level features from 2D X-ray images, and enhancing the ability to capture key aortic features from the extensive sensory field accurately is paramount. Drawing inspiration from the pre-trained model [24], our research embarks on a similar path. PARNet leverages pre-training information through the pre-training process to improve the specificity of the generative adversarial network in the 3D aortic CT reconstruction task and enhance reconstruction performance. Unlike the typical pre-training process of general deep learning models, PARNet does not retain hyperparameters such as weights after pre-training. Instead, it uses the X-rays of the aorta as input information for the model. During the model pre-training process, the generator utilizes only chest orthogonal X-rays as input information, without separate aortic X-rays as additional pre-training information. In this state, where the ARB and ARMask modules are not yet employed for feature detail extraction and fusion, the model is pre-trained to obtain 3D CT reconstruction results of the entire chest. The output of the pre-trained model is then processed using a segmentation tool to isolate the 3D CT data of the aorta. Finally, digitally reconstructed radiograph (DRR) technology [28] generates 2D X-rays of the aorta as the final pre-training results, providing richer pre-training information for subsequent training. Incorporating the pre-trained 2D X-ray of the aorta into the model as input for the ARMask module enhances the input information for the generator, improving the model's perception and attention to aortic feature details. Subsequent experiments demonstrate that this method effectively improves the accuracy and detail performance of 3D aortic reconstruction.

3.3 Aortic Reconstruction with Background and Mask X-rays Modules

In the aortic reconstruction task, we aim to fully leverage the global structural information of the human chest and the local detail information of the aortic vessels from the pre-training information to enhance reconstruction capability. To achieve this, we propose the ARB algorithm module and the ARMask algorithm module shown in Fig 2b and Fig 2c, inspired by [8], building on CNN feature extraction. These modules use a branch feature extraction and fusion method to meet our objectives better. The ARB algorithm module is designed to extract the global structural information from the human chest orthogonal X-rays as background information, while the ARMask algorithm module focuses on the aortic X-ray features from the pre-training information. The feature extraction results from both module branches are then fused and input into the feature fusion network of the generator to provide detailed features. Detailed descriptions of these two modules are provided below.

The human chest orthogonal X-ray is processed through the encoder's layered convolution to obtain the feature map. Specifically, the feature maps from layers 2 to 5 of the encoder are selected as the input for the ARB algorithm module, allowing for a more comprehensive acquisition of the global receptive field of the chest X-ray. The ARB module employs multiple convolutional layers, normalization layers, and activation layers to perform transformer operations effectively. This enhances the intermediate layers of PARNet with a larger receptive field, enabling the full extraction of key features from the global structural information of the chest X-rays. Additionally, compared to methods employing inter-channel attention and pixel attention, spatial attention used in the ARB module offers a favorable trade-off between transmission speed and spatial resolution performance [8]. It effectively balances the trade-off between resolution accuracy and computational resource consumption. We employ average pooling within the spatial attention module to compress features into two dimensions, generating two-dimensional attention masks. This process injects global information into the output features, thereby enhancing the CNN network training process. The specific steps of ARB are as follows.

$$I_0 = G_{\mathrm{CNN}}(X_0), \quad I_1 = G_{\text{Trans-Spatial}}\left(I_0'\right) \tag{1}$$

$$G_{\text{Trans-Spatial}}\left(I_0'\right) = \text{Spatial Attention}(Q, K, V). \tag{2}$$

where I_0 represents the feature maps generated by the middle layers of the CNN from the input image X_0. Meanwhile, I_0' denotes the intermediate features derived from I_0, and I_1 represents the predicted feature maps after being processed by the spatial self-attention mechanism with query (Q), key (K), and value (V) from I_0', along with the dimension (d).

To provide more feature details for 3D aorta reconstruction, the ARMask algorithm module is specifically designed to treat the pre-training information in Fig 3.2, primarily focusing on extracting detailed features from 2D aortic X-rays. Much like ARB, ARMask employs feature extraction techniques that facilitate independent extraction of underlying image information and multi-scale detail information from aortic X-rays within the pre-training information embedded in the generator network. These extracted features are subsequently fused with background features extracted by the ARB module branch. Following multi-layer feature extraction of the input through CNN, the feature map from layer 2 to layer 5 in the middle part of the encoder serves as input to the ARMask branch. While these shallow features are inputted into the transformer branch of ARMask, a distinct attention module is employed. ARMask utilizes Pixel Attention (PA) [46] to discern X-ray color and shape (edges). Through this process, PA generates 3D attention masks without any pooling or downsampling, ensuring that the output feature maps retain rich local information to be focused on. Furthermore, PA boasts fewer parameters than spatial attention (SA), enabling it to achieve more precise attention features. The specific steps of ARMask are as follows.

$$I_2 = G_{\text{CNN}}(X_2), \quad I_3 = G_{\text{Trans-Pixel}}\left(I_2'\right) \tag{3}$$

$$G_{\text{Trans-Pixel}}\left(I_0'\right) = \text{Pixel Attention}(Q, K, V) \tag{4}$$

where I_2 represents the feature maps generated by the middle layers of the CNN from the input image X_2, while I_2' denotes the intermediate features derived from I_2, and I_3 represents the predicted feature maps processed by the pixel self-attention mechanism.

Drawing inspiration from extraction and fusion of feature branches [11], our PARNet employs a similar hybrid fusion of feature branches from the ARB and ARMask as follows,

$$I_O = l(\alpha) * I_1 + (1 - l(\alpha)) * I_3 \tag{5}$$

where I_1 and I_3 represent the output images of the spatial and pixel branch networks, respectively, and the parameter $l(\alpha)$ is a learnable parameter of the activation function, constrained between 0 and 1.

3.4 Loss Functions

To ensure stable and efficient convergence during the training of the aortic reconstruction model within the GAN framework, we have devised a customized loss function comprising multiple components. Firstly, to ensure that the generated 3D results align with the semantic information provided by the input 2D X-rays, it is crucial to enforce a constraint that minimizes the discrepancy between the generated CT image and the ground truth in terms of voxel values. Traditionally, this is achieved by introducing the Mean Squared Error (MSE) loss function, formulated as follows,

$$\mathcal{L}_{GAN}(G) = \mathbb{E}_{MSE}\|y - G(x)\|_2^2 \tag{6}$$

where x is the input of the 2D X-ray image and y is the 3D CT volume.

Besides, motivated by the challenges posed by Federated Learning (FL) [13], we introduce the Weighted Gradient Smoothing (WGS) loss function. As shown below, the WGS loss function aims to reduce noise and discontinuities in generated images, thereby preventing the generator from overfitting the training data and improving the model's generalization ability.

$$\begin{aligned}\mathcal{L}_{WGS} = \sum_{i,j,k} \Bigg(& \frac{1 - l(\beta)}{2} \Big((G(x)_{i,j+1,k} - G(x)_{i,j,k})^2 \\ & + (G(x)_{i+1,j,k} - G(x)_{i,j,k})^2 \Big) + l(\beta) \Big((G(x)_{i,j,k+1} - G(x)_{i,j,k})^2 \Big) \Bigg)\end{aligned} \tag{7}$$

where $G(x)$ represents the pixel value and β is the weighting coefficient for the three directions, which we adjust to fit the input aortic data.

In addition to the WGS loss function, in the frequency domain, the following adjustments to the function [20] are necessary to accommodate the characteristics of low-density aortic structures.

$$G(a,b,c) = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \sum_{z=0}^{Z-1} I(x,y,z) f(e) \tag{8}$$

$$f(e) = e^{-i2\pi\left(\frac{ax}{X} + \frac{by}{Y} + \frac{cz}{Z}\right)} \tag{9}$$

where $I(x,y,z)$ represents the pixel values of the aortic data in the space. We redefine the weighting function $W(u,v,w)$ using a three-dimensional Gaussian function to reflect the characteristics of low-density structures. The 3D loss function to constrain the generated structure in the frequency domain is as follows,

$$\begin{aligned} W(p,q,r) = \mathcal{G}(p,q,r) = G^a, Z = \frac{1}{XYZ} \\ G^m = |G_r(p,q,r) - G_f(p,q,r)|^m \end{aligned} \tag{10}$$

$$\mathcal{L}_{\text{3D}} = Z \times \sum_{p=0}^{X-1} \sum_{q=0}^{Y-1} \sum_{r=0}^{Z-1} W(p,q,r) G^2 \tag{11}$$

where $\mathcal{G}(u,v,w)$ represents the 3D Gaussian function, and the dynamic scaling factor a is an integer not less than 1 to meet the special flexibility requirements of the aortic structure.

We combined various loss components to integrate the above information into a unified loss function within the GAN framework and obtain the following final loss function for aortic reconstruction.

$$\begin{aligned} \mathcal{L}_{\text{all}} = \lambda_{\text{GAN}} \cdot \mathcal{L}_{\text{GAN}}(G) + \lambda_{\text{WGS}} \cdot \mathcal{L}_{\text{WGS}} \\ + \lambda_{3D} \cdot \mathcal{L}_{\text{3D}} \end{aligned} \tag{12}$$

where λ_{GAN}, λ_{WGS}, and λ_{3D} are parameters that control the relative importance of each loss function. It is necessary to focus on the local feature information of the aorta. Therefore, we let $\lambda_{GAN} = 0.3$, $\lambda_{WGS} = 0.4$, and $\lambda_{3D} = 0.3$.

4 Experiments

To evaluate the performance of PARNet, we selected several public human chest datasets, including LIDC-IDRI [3], VerSe '20 [33], VerSe '19 [33], and LungCT-Diagnosis [14], totaling 1200 CT datasets for this study. The voxel spacing of all CT data and the image size of input X-rays were resampled and adjusted to the same value. To validate the effectiveness of our method and its improvements, we conducted comparative experiments with four mainstream algorithms: PSR [34], X2CT [39], XTransCT [44] and 3DSRNet [12]. To be fair, we only compared algorithms based on X-ray-CT paired learning. PSR takes a single X-ray as input, while X2CT, 3DSRNet, and XTransCT are based on the GAN framework and use orthogonal X-rays as input data, which is consistent with our method. We will conduct a fair comparison and comprehensive analysis of our proposed method against other mainstream methods based on multiple metrics, including Peak Signal-to-Noise Ratio (PSNR) [17], Cosine Similarity (CS), Mean Absolute Error(MAE) and Mean Squared Error (MSE) [1].

4.1 Datasets

To validate our proposed CT reconstruction method, ideally, we need to construct a large dataset containing paired X-ray projections and corresponding CT reconstructions. However, in practical scenarios, such datasets are often scarce and costly to collect. Therefore, we adopt the DRR technique to simulate corresponding X-rays from actual CT volumes, aiming to reduce costs while enriching the dataset. Furthermore, our utilized reconstruction algorithm dataset comprises 1200 pairs of data, with 1000 for training, 100 for validation, and 100 for testing. Since multiple datasets were utilized in this study, the imaging modalities of the data differ, having distinct capture ranges and resolutions. Therefore, a fixed cubic region of $320 \times 320 \times 320$ $(\mathrm{mm})^3$ was cropped from each CT scan, and the image size of the input X-rays was adjusted to 128×128 pixels.

4.2 Tools for Training and Implementation Details

The specific segmentation tool to be employed in the pre-training process of Fig 3.2 is TotalSegmentator [37, 19]. It can provide efficient segmentation in a wide range of images, ensuring robust results. It is a powerful tool used to isolate the aorta from the reconstruction results for pre-training and evaluation, removing other tissues and organs, which is widely used in previous work. To bridge the disparity between simulated synthetic X-rays from DRRs and real X-rays, we employ CycleGAN [47] to learn the characteristics of real X-rays onto simulated X-rays and establish the mapping relationship. Specifically, our approach involves using 100 real X-rays from CheXpert [18] and an equal number of randomly selected synthetic X-rays from our study's training dataset for the CycleGAN training process.

We used the Stochastic Gradient Descent (SGD) optimizer [26] with a decay weight of 0.01 and momentum of 0.99 to train PARNet over a total of 300 epochs, validating every ten epochs. The experiments were conducted on an NVIDIA 4090 GPU using the PyTorch framework, with a batch size of 2 and a learning rate starting at 0.01 on the GPU.

5 Results

In this study, we utilized a consistent segmentation tool to assess the overall efficacy of aortic reconstruction. Following established practices in related research

[39, 32, 12, 44], our experimental findings exclusively present averages across all datasets to assess the generalization ability and robustness of our method, objectively quantifying performance disparities and enhancements.

5.1 Quantitative Results

Based on the characteristics of aortic 3D reconstruction and the needs of vascular interventional surgery, we aim to reconstruct and restore the details of the aorta precisely. Data in Table 1 show that PARNet outperforms other 3D reconstruction methods across all evaluation metrics.

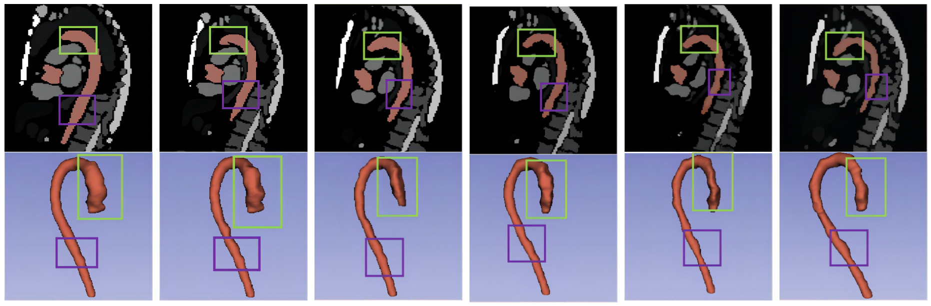
Table 1: Average Quantitative Results of Different Methods for Aortic Reconstruction.

Method	PSNR ↑	SSIM ↑	CS ↑	MAE ↓	MSE ↓
PSR [34]	37.633	0.743	0.555	35.320	14308.362
X2CT [39]	42.355	0.849	0.704	26.358	10543.356
XTransCT [44]	38.453	0.802	0.578	30.336	13728.213
3DSRNet [12]	43.263	0.862	0.712	24.384	8356.254
Ours(PARNet)	**45.153**	**0.881**	**0.741**	**19.378**	**3269.211**

PARNet achieved a PSNR of 45.153, SSIM of 0.881, and CS of 0.741. Additionally, PARNet recorded the lowest MAE and MSE values at 19.378 and 3269.211, respectively. These results demonstrate that PARNet has significant advantages in image reconstruction quality and accuracy, particularly in producing excellent aortic CT results. Compared to the single-view method PSR, PARNet, and other approaches using dual-view X-rays can capture more geometric constraints and reconstruction information, significantly enhancing reconstruction accuracy and capability. Similar to X2CT, XTransCT, and 3DSRNet, PARNet also employs a dual-view orthogonal GAN framework. Experimental comparisons indicate that the ARB and ARMask algorithm modules integrated into PARNet, along with additional pre-training information, effectively extracted more reconstruction details, thereby improving the accuracy and precision of aortic reconstruction. These findings underscore the potential of our proposed method in delivering detailed and accurate image reconstructions, which is essential for supporting surgeons in vascular interventional surgery.

5.2 Qualitative Results

Fig 3 visually compared the aortic reconstruction results of our proposed PARNet method with those of other methods qualitatively. PARNet reconstructed the ascending aorta, aortic arch, and abdominal aorta most closely to the ground truth compared with other methods. The sagittal view clearly illustrated the advantage of PARNet in terms of overall reconstruction completeness.



(a) GT (b) PARNet (c) 3DSRNet (d) XTransCT (e) X2CT (f) PSR

Fig. 3: 3D visualization of qualitative results. These results demonstrate that PARNet performs well in aortic reconstruction. Significant differences are highlighted in the boxes within the sagittal views in the first row and the 3D views in the second row.

Table 2: Ablation results of PARNet. Each module was removed in turn for the experiment, and ✓ indicated the presence of the module.

Combination			Metrics				
ARB + ARMsak	WGS Loss	3D Loss	PSNR ↑	SSIM ↑	CS ↑	MAE ↓	MSE ↓
✓	✓	✓	**<u>45.153</u>**	**<u>0.881</u>**	**<u>0.741</u>**	**<u>19.378</u>**	**<u>3269.211</u>**
	✓	✓	43.218	0.855	0.721	24.897	8305.419
		✓	42.355	0.849	0.714	26.358	10543.356
			38.601	0.723	0.654	33.725	15264.377

5.3 Ablation Study

To assess the effectiveness of the proposed innovative module, we conducted a series of ablation experiments on PARNet with various settings. We deactivated these modules in turn in four experiments and the 3D visualization results are shown in Fig 4 (b)-(e).

Through the ablation experiment results in the Table 2, we can observe the impact of different modules on the performance of PARNet. These ablation experiment results highlight the effectiveness of each module in enhancing reconstruction performance.

5.4 Evaluation on Real-World Data

In addition to experimenting on the dataset, we also conducted experiments on clinically acquired real X-ray data CheXpert [18], verifying the effectiveness of the PARNet algorithm, as shown in Fig 5.

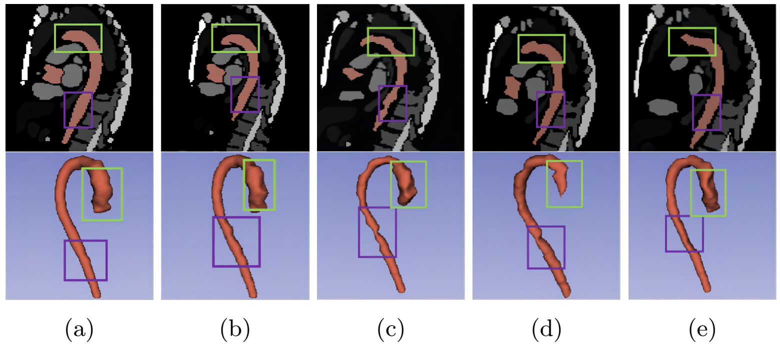


Fig. 4: 3D reconstruction results of ablation study. (a) is the ground truth, and (b) - (e) are the results of the four ablation experiments.

6 Conclusion

This study proposed a deep learning method, PARNet, for three-dimensional aortic reconstruction based on orthogonal X-rays. PARNet utilized pre-trained generated aortic X-rays as pre-training information to enhance its accurate perception of low-density tissues like the aorta, enriching the final details of aortic reconstruction and improving the model's reconstruction performance. Finally, experiments verified the effectiveness of each innovative module of the proposed method, demonstrating its superiority over existing methods. The model's effectiveness was further validated using clinical real-world data, demonstrating its capability to furnish surgeons with valuable three-dimensional vascular information.

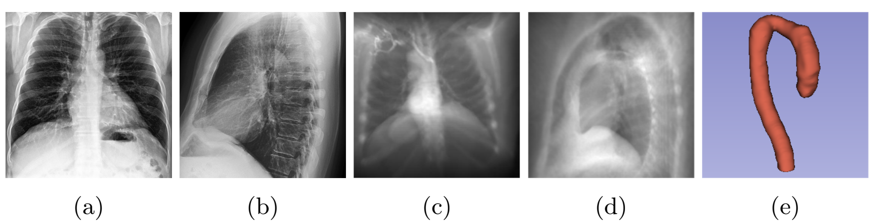


Fig. 5: 3D reconstruction result of clinical real data samples. (a) and (b) are clinically acquired real orthogonal X-rays, (c) and (d) are the corresponding simulated generated orthogonal X-rays, and (e) is the 3D reconstruction result based on (a) and (b).

Acknowledgments. This work was supported in part by Beijing Natural Science Foundation under Grant Z220016.

References

1. Allen, D.M.: Mean square error of prediction as a criterion for selecting variables. Technometrics **13**(3), 469–475 (Aug 1971)
2. Anirudh, R., Kim, H., Thiagarajan, J.J., Mohan, K.A., Champley, K., Bremer, T.: Lose the views: Limited angle ct reconstruction via implicit sinogram completion. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). pp. 6343–6352 (2018)
3. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical Physics **38**(2), 915–931 (2011)
4. Bahrami, K., Shi, F., Rekik, I., Shen, D.: Convolutional neural network for reconstruction of 7t-like images from 3t mri using appearance and anatomical features. In: Deep learning and data labeling for medical applications, pp. 39–47. Springer, Berlin, Germany (2016)
5. Burgos, N., et al.: Robust ct synthesis for radiotherapy planning: Application to the head and neck region. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. pp. 476–484. Springer, Cham, Switzerland (2015)
6. Cai, Y., Wang, J., Yuille, A., Zhou, Z., Wang, A.: Structure-aware sparse-view x-ray 3d reconstruction. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR) (2024)
7. Chen, X., Wu, Q., Wang, S.: Research on 3d reconstruction based on multiple views. In: Proc. 13th Int. Conf. Comput. Sci. Educ. (ICCSE). pp. 1–5 (Aug 2018)
8. Chen, Y., Zhang, C., Chen, B., Huang, Y., Sun, Y., Wang, C., Fu, X., Dai, Y., Qin, F., Peng, Y., Gao, Y.: Accurate leukocyte detection based on deformable-detr and multi-level feature fusion for aiding diagnosis of blood diseases. Computers in Biology and Medicine **170**, 107917 (2024)
9. Chung, H., Ryu, D., Mccann, M.T., Klasky, M.L., Ye, J.C.: Solving 3d inverse problems using pre-trained 2d diffusion models. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22542–22551 (2023)
10. Doenst, T., Haverich, A., Serruys, P., et al.: Pci and cabg for treating stable coronary artery disease: Jacc review topic of the week. Journal of the American College of Cardiology **73**(8), 964–976 (2019)
11. Fan, C.M., Liu, T.J., Liu, K.H.: Compound multi-branch feature fusion for image deraindrop. In: 2023 IEEE International Conference on Image Processing (ICIP). pp. 3399–3403 (2023)
12. Gao, Y., Tang, H., Ge, R., Liu, J., Chen, X., Xi, Y., Ji, X., Shu, H., Zhu, J., Coatrieux, G., Coatrieux, J.L., Chen, Y.: 3dsrnet: 3-d spine reconstruction network using 2-d orthogonal x-ray images based on deep learning. IEEE Transactions on Instrumentation and Measurement **72**, 4506214 (2023)
13. Ghosh, A., Chung, J., Yin, D., Ramchandran, K.: An efficient framework for clustered federated learning. IEEE Transactions on Information Theory **68**(12), 8076–8091 (Dec 2022)
14. Grove, O., Berglund, A.E., Schabath, M.B., Aerts, H.J.W.L., Dekker, A., Wang, H., Velazquez, E.R., Lambin, P., Gu, Y., Balagurunathan, Y., Eikman, E., Gatenby, R.A., Eschrich, S., Gillies, R.J.: Data from: Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma (2015)

15. He, J., et al.: Learning hybrid representations for automatic 3d vessel centerline extraction. In: Martel, A., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Lecture Notes in Computer Science. vol. 12266. Springer, Cham (2020)
16. Henzler, P., Rasche, V., Ropinski, T., Ritschel, T.: Single-image tomography: 3d volumes from 2d cranial x-rays. Comput. Graph. Forum **37**(2), 377–388 (2018)
17. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of psnr in image/video quality assessment. Electron. Lett. **44**(13), 800–801 (Jun 2008)
18. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R.L., Shpanskaya, K., Seekins, J., Mong, D.A., Halabi, S.S., Sandberg, J.K., Jones, R., Larson, D.B., Langlotz, C.P., Patel, B.N., Lungren, M.P., Ng, A.Y.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. CoRR **abs/1901.07031** (2019)
19. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
20. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV). pp. 13899–13909 (Oct 2021)
21. Kasten, Y., Doktofsky, D., Kovler, I.: End-to-end convolutional neural network for 3d reconstruction of knee bones from bi-planar x-ray images. In: Machine Learning for Medical Image Reconstruction. pp. 123–133. Springer, Cham, Switzerland (2020)
22. Koehler, C., Wischgoll, T.: 3-d reconstruction of the human ribcage based on chest x-ray images and template models. IEEE MultiMedia **17**(3), 46–53 (2010)
23. Kuppuswamy, R.R.: Recent progress in coaxial electrospinning: New parameters, various structures, and wide applications. Journal of Materials Science **54**(12), 8530–8545 (Jun 2019)
24. Lee, S., Chung, H., Park, M., Park, J., Ryu, W.S., Ye, J.C.: Improving 3d imaging with pre-trained perpendicular 2d diffusion models. 2023 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 10676–10686 (2023)
25. Lin, Y., Luo, Z., Zhao, W., Li, X.: Learning deep intensity field for extremely sparse-view cbct reconstruction. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 13–23. Springer Nature Switzerland, Cham (2023)
26. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. arXiv: Learning (2016)
27. Maton, A., Hopkins, J., McLaughlin, C.W., Johnson, S., Warner, M.Q., LaHart, D., Wright, J.D.: Human Biology Health. Prentice Hall, Englewood Cliffs, NJ, USA (1995)
28. Milickovic, N., Baltas, D., Giannouli, S., Lahanas, M., Zamboglou, N.: Ct imaging based digitally reconstructed radiographs and their application in brachytherapy. Physics in Medicine & Biology **45**(10), 2787–2800 (2000)
29. Müller, R., et al.: Morphometric analysis of human bone biopsies: A quantitative structural comparison of histological sections and micro-computed tomography. Bone **23**(1), 59–66 (Jul 1998)
30. Ndumele, C.E., Rangaswami, J., Chow, S.L., Neeland, I.J., Tuttle, K.R., Khan, S.S., Coresh, J., Mathew, R.O., Baker-Smith, C.M., Carnethon, M.R., Despres, J.P., Ho, J.E., Joseph, J.J., Kernan, W.N., Khera, A., Kosiborod, M.N., Lekavich,

C.L., Lewis, E.F., Lo, K.B., Ozkan, B., Palaniappan, L.P., Patel, S.S., Pencina, M.J., Powell-Wiley, T.M., Sperling, L.S., Virani, S.S., Wright, J.T., Singh, R.R., Elkind, M.S.V., Assoc, A.H.: Cardiovascular-kidney-metabolic health: A presidential advisory from the american heart association. Circulation **148**(20), 1606–1635 (Nov 2023)
31. Nie, D., et al.: Medical image synthesis with context-aware generative adversarial networks. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. pp. 417–425. Springer, Cham, Switzerland (2017)
32. Saravi, B., Guzel, H.E., Zink, A., Ülkümen, S., Couillard-Despres, S., Wollborn, J., Lang, G., Hassel, F.: Synthetic 3d spinal vertebrae reconstruction from biplanar x-rays utilizing generative adversarial networks. J. Pers. Med. **13**(12), 1642 (2023)
33. Sekuboyina, A., et al.: VERSE: A vertebrae labelling and segmentation benchmark for multi-detector CT images. Medical Image Analysis **73**, 102166 (Oct 2021)
34. Shen, L., Zhao, W., Xing, L.: Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. Nature Biomed. Eng. **3**(11), 880–888 (Oct 2019)
35. Vo, C.D., Jiang, B., Azad, T.D., Crawford, N.R., Bydon, A., Theodore, N.: Robotic spine surgery: Current state in minimally invasive surgery. Global Spine Journal **10**, 34–40 (Apr 2020)
36. Wang, X., He, J., Liu, Y., Zhang, P., Cheng, W.W., Wang, B., Gui, Z.: Didr-net: a sparse-view ct deep iterative reconstruction network with an independent detail recovery network. Journal of Instrumentation **18** (2023)
37. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence (2023)
38. World Health Organization: Baseline Country Survey on Medical Devices 2010. World Health Organization (2011)
39. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2ct-gan: Reconstructing ct from biplanar x-rays with generative adversarial networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10611–10620 (2019)
40. Yuan, X., Liu, C., Feng, F., Zhu, Y., Wang, Y.: Slice-mask based 3d cardiac shape reconstruction from ct volume. In: Proceedings of the Asian Conference on Computer Vision (ACCV). pp. 1909–1925 (December 2022)
41. Zang, G., Idoughi, R., Li, R., Wonka, P., Heidrich, W.: Intratomo: Self-supervised learning-based tomography via sinogram synthesis and prediction. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1940–1950 (2021)
42. Zeng, Q., Zhou, J., Ji, Y., Wang, H.: A semiparametric gaussian mixture model for chest ct-based 3d blood vessel reconstruction. Biostatistics **2024**(kxae013) (April 2024)
43. "Zha, R., Zhang, Y., Li, H.: Naf: Neural attenuation fields for sparse-view cbct reconstruction. In: 2022 Medical Image Computing and Computer Assisted Intervention(MICCAI). pp. 442–452. Springer Nature Switzerland, Cham (2022)
44. Zhang, C., Liu, L., Dai, J., Liu, X., He, W., Chan, Y., Xie, Y., Chi, F., Liang, X.: Xtransct: ultra-fast volumetric ct reconstruction using two orthogonal x-ray projections for image-guided radiation therapy via a transformer network. Phys. Med. Biol. **69**(085010) (2024)

45. Zhang, Z., Yang, L., Zheng, Y.: Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition(CVPR). pp. 9242–9251 (2018)
46. Zhao, H., Kong, X., He, J., Qiao, Y., Dong, C.: Efficient image super-resolution using pixel attention. In: ECCV. pp. 56–72. Springer (2020)
47. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition(CVPR). pp. 2242–2251 (2017)
48. Zhu, L., Ji, D., Zhu, S., Gan, W., Wu, W., Yan, J.: Learning statistical texture for semantic segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12532–12541 (2021)