Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

DP-GAN+B: A lightweight generative adversarial network based on depthwise separable convolutions for generating CT volumes

Xinlong Xing[a,b], Xiaosen Li[c], Chaoyi Wei[b], Zhantian Zhang[a,b], Ou Liu[b,*], Senmiao Xie[d], Haoman Chen[b], Shichao Quan[e], Cong Wang[f], Xin Yang[g], Xiaoming Jiang[b], Jianwei Shuai[b,**]

[a] *Postgraduate Training Base Alliance of Wenzhou Medical University, Wenzhou, Zhejiang, 325000, China*
[b] *Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, Zhejiang, 325000, China*
[c] *School of Artificial Intelligence, Guangxi Minzu University, Nanning, 530006, China*
[d] *Department of Radiology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, China*
[e] *Department of Big Data in Health Science, The First Affiliated Hospital of Wenzhou Medical University, China*
[f] *Department of Mathematics and Statistics, Carleton College, 300 N College St, Northfield, MN, 55057, USA*
[g] *School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China*

ARTICLE INFO

Keywords:
Bi-planar X-ray
Computed tomography
Generative adversarial networks (GAN)
Auxiliary diagnosis
Depthwise separable convolution

ABSTRACT

X-rays, commonly used in clinical settings, offer advantages such as low radiation and cost-efficiency. However, their limitation lies in the inability to distinctly visualize overlapping organs. In contrast, Computed Tomography (CT) scans provide a three-dimensional view, overcoming this drawback but at the expense of higher radiation doses and increased costs. Hence, from both the patient's and hospital's standpoints, there is substantial medical and practical value in attempting the reconstruction from two-dimensional X-ray images to three-dimensional CT images. In this paper, we introduce DP-GAN+B as a pioneering approach for transforming two-dimensional frontal and lateral lung X-rays into three-dimensional lung CT volumes. Our method innovatively employs depthwise separable convolutions instead of traditional convolutions and introduces vector and fusion loss for superior performance. Compared to prior models, DP-GAN+B significantly reduces the generator network parameters by 21.104 M and the discriminator network parameters by 10.82 M, resulting in a total reduction of 31.924 M (44.17%). Experimental results demonstrate that our network can effectively generate clinically relevant, high-quality CT images from X-ray data, presenting a promising solution for enhancing diagnostic imaging while mitigating cost and radiation concerns.

1. Introduction

After Wilhelm Rntgen's discovery in 1895, X-rays became the most commonly used image technique in clinical practice as it offers a non-invasive view of the internal structures of the human body. However, their two-dimensional nature inevitably leads to the overlapping of organs in images. Computed Tomography (CT), by providing a three-dimensional perspective, effectively resolves this issue of information overlap caused by organ superimposition. A major concern in the use of X-ray and CT imaging is the associated radiation dose, which has implications for patient health. Past research indicates that a chest X-ray exposes the body to approximately 0.1 millisieverts (mSv) of radiation, while a standard CT chest scan administers about 7 mSv [1]. Thus, the radiation from X-ray procedures is significantly lower than that from CT scans, underscoring the practical importance of developing methods to reconstruct three-dimensional CT scans from two-dimensional X-ray images.

In recent years, artificial intelligence has been widely used in many fields, including image processing [2] and bionics [3]. It has also provided great opportunities and achieved promising results in biomedical applications, such as medical image segmentation [4–8], drug analysis [9], disease diagnosis and prediction [10–13], single-cell multi-omics data analysis [14], RNA-RNA interaction [15], RNA-protein interaction [16], proteomics research [17], and Gene/protein signaling networks [18]. However, reconstructing CT from X-rays remains a challenging task. The primary challenge in CT reconstruction from X-rays lies in the

[*] Corresponding author.
[**] Corresponding author.
E-mail addresses: oliu@ucas.ac.cn (O. Liu), shuaijw@wiucas.ac.cn (J. Shuai).

https://doi.org/10.1016/j.compbiomed.2024.108393
Received 6 February 2024; Received in revised form 17 March 2024; Accepted 1 April 2024
Available online 2 April 2024
0010-4825/© 2024 Elsevier Ltd. All rights reserved.

inherent lack of depth information. This limitation significantly increases the complexity of the reconstruction process, particularly when the number of available X-ray images is limited, thus presenting substantial challenges to the accuracy and efficacy of CT reconstruction methods. Recent studies have attempted to address this issue by incorporating prior knowledge of human anatomy with advanced deep-learning models [19–25].

Earlier approaches utilized a 2D encoder-3D decoder structure to extract features from a single X-ray for CT reconstruction [20,23,25]. Henzler et al. [20] successfully reconstructed a complete 3D skull using X-rays from a single skull. Shen et al. [23] introduced PatReconNet, aiming to learn the feature space transformation between a single lung X-ray and 3D CT, generating the 3D CT volume from the single X-ray. However, relying on a single X-ray often leads to incomplete three-dimensional information [26], resulting in a considerably blurry effect in the three-dimensional reconstruction. Methods using multiple images tend to yield clearer results, but they also raise questions about the authenticity of the generated CT, potentially limiting their clinical application.

Another challenge is the high computational cost due to complex encoder-decoder structures. Since the advent of deep convolutional neural networks, such as AlexNet [27], there has been a trend towards deeper networks for improved accuracy [28–30]. While enhancing the depth of a neural network facilitates more comprehensive feature extraction, the number of parameters of the deep learning model continues to expand as the depth of the network increases. This not only adds to the complexity of the model but also imposes significant memory demands during training. Consequently, it elevates the training complexity threshold, leading to substantial resource utilization and raising the risk of overfitting, which are critical considerations in the model development process.

In numerous medical applications, such as real-time diagnostics and telesurgery, there is a critical need for prompt task execution on platforms with limited computational capabilities. Overly complex network models are thus impractical for use in portable medical devices, which often face strict constraints in processing capacity. Consequently, there is a growing interest in the development of lightweight and efficient neural networks that can perform effectively within these computational limits. A notable innovation in this regard is the depthwise separable convolution, a technique employed in advanced models like Xception [31], MobileNets [32], and ShuffleNet [33]. This approach is instrumental in designing networks that are both powerful and resource-efficient, catering to the needs of modern medical applications. Depthwise separable convolutions divide standard convolutions into depthwise convolutions and pointwise convolutions. This separation method significantly reduces the number of parameters in the model, making the network more lightweight, reducing the consumption of computing resources, and making it easier to deploy in hospitals. At the same time, in the 3D CT image reconstruction task, depthwise separable convolution processes spatial information and channel information respectively, which helps to better capture the spatial characteristics and channel correlation of 2D X-ray images and 3D CT images. This helps improve the model's sensitivity to image structure and texture, thereby improving the accuracy of image reconstruction.

In this work, we employ depthwise separable convolution to develop an efficient network structure for converting two-dimensional X-rays into three-dimensional CT scans. This approach significantly reduces the parameters in the model by 31.924 M (44.17%) through separating spatial and channel processing. Our aim is to develop compact, low-latency models for seamless integration with medical applications. To improve the accuracy of the generated CT scans, we introduce fusion loss and vector loss, ensuring high fidelity and texture similarity to actual CT scans. This method aligns with the diagnostic needs of clinicians. Our contributions in this paper can be summarized as follows.

1)We propose DP-GAN+B, a generative adversarial network, to reconstruct 3D CT volumes from 2D X-ray image.
2)We employ depthwise separable convolution to achieve lightweight network.
3)We propose fusion loss and vector loss to enhance the fidelity of generated CT images.
4)We empirically show that DP-GAN+B can significantly improve the performance of 3D reconstruction. When compared against the baseline model, DP-GAN+B produces better results than SOTA methods on the LIDC_IDRI dataset and reduces the number of parameters by 44.17%.

This article is organized as follows: Section 2 reviews related work, Section 3 details our proposed model DP-GAN+B, Section 4 discusses the loss function, Section 5 presents experimental results on LIDC-IDRI datasets and a detailed discussion, Section 6 discusses clinical applications, and Section 7 concludes the article.

2. Related work

In this section, we introduce previous related work, including the 3D reconstruction from single images, the 3D reconstruction from X-rays and Lightweight network.

2.1. The 3D reconstruction from single images

The task of reconstructing 3D volumes from single images, while challenging, is crucial for various practical applications. Humans have the unique ability to intuit 3D shapes from a single image, drawing on their prior knowledge and visual experience of the 3D world. Research in this area is predominantly focused on voxel-based reconstruction and point cloud reconstruction methods.

There has been considerable research focused on voxel-based reconstruction [34–36]. In these studies, 2D convolutional neural networks are typically employed to encode shape knowledge into vector representations, while 3D convolutional neural networks decode these into 3D object shapes [34]. Neubert Boris et al. [35] explored density estimation of 2D image projections using voxel values for guiding 3D reconstruction. Similarly, Tulsiani et al. [36] worked on integrating attitude estimation networks with 3D object voxel prediction for effective 3D reconstruction. However, voxel-based methods encounter computational challenges at high resolutions. To solve this problem, Fan et al. [37] introduced a point cloud-based 3D reconstruction representation as an alternative approach. Yotam Livny et al. [38] used a series of global optimizations to reconstruct the skeleton structure from tree point clouds.

These methods, while adept at reconstructing object surfaces, often fall short in detailing internal structures, thus limiting their applicability in clinical diagnostics. Methods for extracting 3D models from X-rays, which can penetrate most objects and produce layered 2D images, vary significantly from voxel and point cloud techniques and hold more clinical relevance.

2.2. The 3D reconstruction from X-rays

Recent studies have investigated the use of statistical models in conjunction with X-rays for 3D reconstruction. Aubert et al. [39] introduced a novel approach for the 3D reconstruction of rib cages from bi-planar radiographs using statistical parametric modeling. With some manual adjustments, the alignment between the generated results and original images can be further refined. Similarly, another study [40] employed a 3D statistical shape model of the rib cage, which is adapted to individual 2D projections. Lamecker et al. [41] developed a technique to reconstruct 3D shapes from X-rays, using3D statistical shape models combined with an algorithm that optimizes a similarity measure accessing the differences between projections of the reconstruction

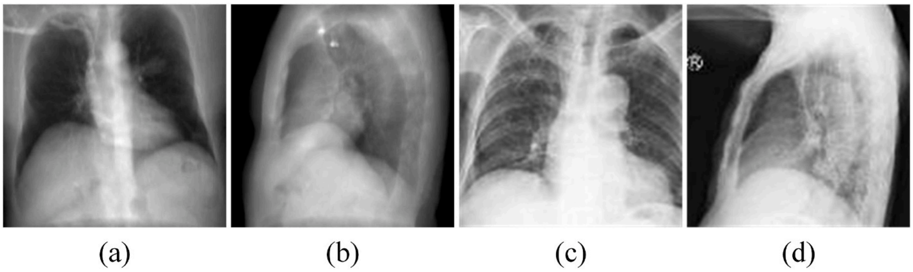

Fig. 1. Input of bi-planar X-ray images. (a) Frontal X-ray in public datasets. (b) Lateral X-ray in public datasets. (c) Frontal X-ray in hospital datasets. (d) Lateral X-ray in hospital datasets.

results and the actual X-ray images. These approaches utilize the known general shape of the rib cage and lungs for reconstruction. Koehler et al. [42] harnessed expert knowledge and anatomical insights to convert basic 3D templates into detailed models of these structures. Despite their innovation, such methods typically fall short of producing CT-like 3D volumes, which may limit their clinical utility. This highlights the need for techniques that can generate more detailed and clinically applicable 3D reconstructions.

To address existing challenges in medical imaging, the utilization of convolutional neural networks (CNNs) for three-dimensional reconstruction has gained prominence in the field [19–24]. Through convolution operations, convolutional neural networks adeptly extract and learn features from X-ray images. Notably, study [24] utilized an X-ray sinogram as input, a format not discernible to the human eye. Shen et al. [23] innovatively adjusted feature dimensions from X-rays to construct CT volumes. Henzler et al. [20] employed ResNet for extracting features from a single skull X-ray, achieving three-dimensional reconstruction. This approach used a single X-ray for CT reconstruction, employing traditional convolutions in both encoder and decoder components. X2Teeth [22] developed a novel ConvNet, reconstructing three-dimensional teeth from a single panoramic radiograph of cavities. This ConvNet comprises three CNN-based subnets. Further, Megumi Nalao et al. [43] combined CNNs for feature extraction from two-dimensional images with Graph neural networks [44] for learning mesh deformation, generating three-dimensional kidney meshes. Research [21] introduced an end-to-end CNN method for reconstructing knee bones from bi-planar X-rays, using a hop-connected network between the encoder and decoder. Tan et al. [45] proposed a lightweight CNN feature fusion network for lung CT reconstruction, demonstrating the broad applicability of CNNs in medical imaging.

In 2014, following the invention of generative adversarial networks (GAN) [46] by Goodfellow, GANs emerged as one of the most significant architectures in machine learning. Unlike CNNs, a GAN comprises two neural networks: a generator and a discriminator. The generator network uses random signals to create synthetic outputs, which are then assessed by the discriminator network against real data. GANs are extensively used in image processing, notably in converting low-resolution images to high-resolution [47,48], image-to-image translation [49], medical samples augmentation [50], medical image segmentation [51], disease diagnosis [52], etc. They also have significant applications in 3D X-ray reconstruction [53–55]. Projects like Oral-3D [54] and spine structure reconstructions [55] have utilized densely connected networks in the generator for feature extraction, with a hop-connected architecture linking the encoder and decoder. Studies [53] have demonstrated GANs' potential in improving abdominal anomaly predictions from orthogonal X-ray derived CT scans.

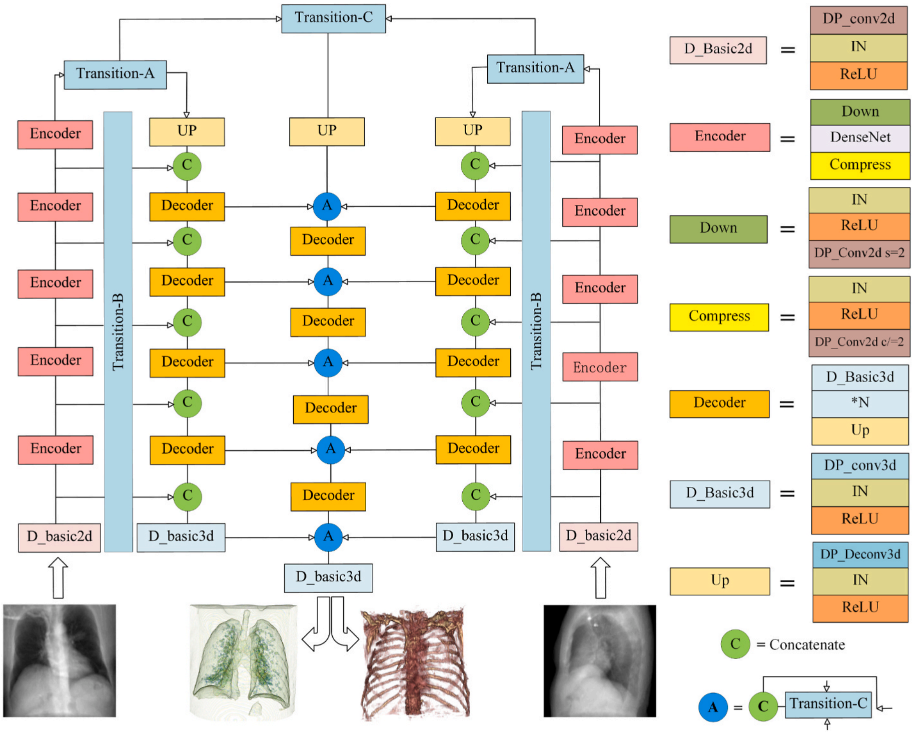

Fig. 2. Overview of DP-GAN+B 3D generator Architecture.

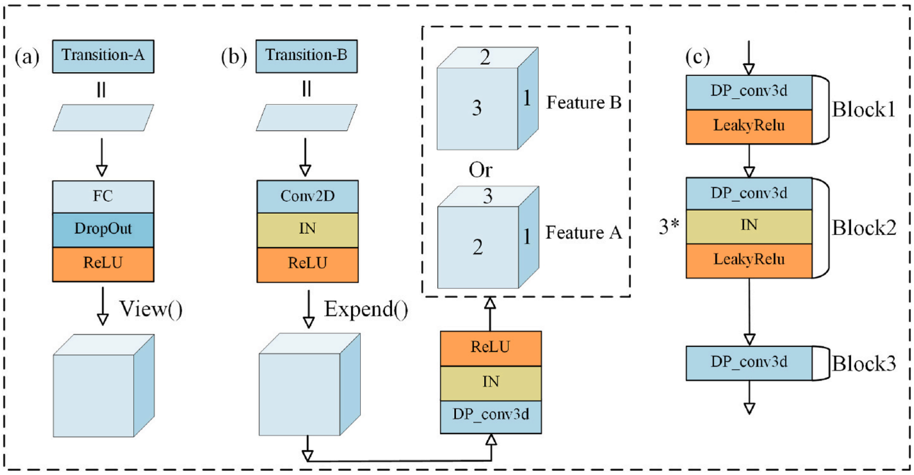

Fig. 3. Overview of DP-GAN+B transition architectures and 3D discriminator architecture. (a) Transition-A. (b) Transition-B. (c) 3D discriminator.

X2CT-GAN [25] innovatively applies this for lung reconstruction, using a 3D generator and discriminator for authenticating CT reconstructions. However, the reliance on deep networks for feature extraction can lead to redundancy in the model.

2.3. *Lightweight network*

Optimizing neural network structures for a balance between parameter efficiency and performance is a key research area. Since AlexNet [27], convolutional neural networks have become popular among researchers. However, increasing network depth often leads to challenges like gradient vanishing. Balancing parameter efficiency and performance in neural network structures is a crucial research focus. Laurent Sifre's invention of depthwise separable convolutions [56] was a breakthrough in addressing these issues. This innovation, inspired by the Transformation-invariant scattering theory by Sifre and Mallat [56, 57], was integrated into AlexNet during Sifre's internship at Google. It resulted in improved accuracy, enhanced speed, and reduced model size, offering a promising solution to the existing problems in deep neural networks.

MobileNetV1 [32] utilized depthwise separable convolutions to create a lightweight neural network, achieving only 1% reduction in accuracy compared to ImageNet with significantly fewer parameters. This approach was also adopted in Inception models [58,59] to reduce computations in initial layers. The Xception network [31] further refined this by replacing Inception modules with depthwise separable convolutions, maintaining parameter count while enhancing performance. This method reduces parameters by separating spatial filtering from feature generation, defined as depthwise convolution for spatial filtering and 1x1 pointwise convolution for feature generation.

In summary, the current 3D reconstruction technology based on neural networks still has problems such as model redundancy and ambiguous reconstruction results, which lead to low clinical feasibility. In order to solve these problems, we design our model based on the lightweight model. The specific model design is shown below.

3. DP-GAN+B: the depthwise separable convolution-based network for generating CT volumes

In this section, the network designed for 3D CT reconstruction from 2D bi-planar X-rays is presented. The input, as shown in Fig. 1, consists of bi-planar X-ray images.

Our model, drawing inspiration from X2CT-GAN+B [25], features both 3D generator and 3D discriminator architectures. These two components interact crucially in creating CT reconstructions. The process involves using two 2D X-ray projection images as input and aims to produce 3D CT volumes. The network follows an encoder-decoder structure, where extracted features are integrated and processed through a novel sampling decoder to yield the final 3D CT output.

Fig. 2 provides a detailed overview of our generator network. This network is composed of encoder blocks, decoder blocks, transition blocks, and fusion blocks. In the diagram, the circle marked 'C' denotes the concatenate operation, and the circle marked 'A' represents a combination of concatenate operation and Transition-C. Our network's architecture is built around an encoder-transition-decoder structure. Initially, the encoder section takes in front and lateral X-ray images. Subsequent to the encoder, a series of blocks are employed to adjust feature dimensions, which then lead into the decoder blocks. The primary objective of our network is to effectively translate the 2D X-ray inputs in feature space into the desired 3D CT volume. By utilizing two parallel encoder-transition-decoder networks, our model incorporates a feature fusion module. This module is integral in reconstructing the 3D CT volume, as it synthesizes the bi-planar information from both encoder-decoder networks.

Fig. 3 illustrates the transition architectures and the discriminator network, including Transition-A, Transition-B, and 3D Discriminator. In the following section, we will introduce the generator (including encoder blocks, transition blocks, decoder blocks, and fusion blocks), and the discriminator.

3.1. *Encoder*

Each encoder block in our design comprises a down block, a DenseNet block, and a compressed block. Dense connections, a hallmark of DenseNet [60], are crucial for facilitating feature extraction across network layers. We have integrated a DenseNet architecture in the encoder to harness the full potential of information extracted from X-ray inputs. The encoder is structured into a 5-layer network, with each layer containing a specific number of depthwise separable layers (6, 12, 24, 16, and 6, respectively) and a progressive increase in channels by 32. The down block consists of an Instance Normalization (IN) layer, a Rectified Linear Unit (ReLU) layer, and a depthwise separable convolution layer with a stride of 2. In the compressed block, there is an IN layer, a ReLU layer, and a depthwise separable convolutional layer with halved channels. Moreover, our architecture adopts a cascading approach where features extracted layer by layer are transmitted to the decoder network through skip connections, ensuring efficient feature transmission to the decoder network and minimizing feature loss throughout the process.

3.2. Transition

In our network's Transition section, three distinct blocks are utilized to map 2D to 3D features. Initially, the Transition-A block is responsible for transforming 2D features into 3D. To connect with the encoder network effectively, the Transition-B block is then used for converting encoder features into inputs suitable for the decoder. Finally, the Transition-C block plays a crucial role in enabling feature fusion within the parallel encoder-transition-decoder structure, ensuring seamless integration of features for accurate 3D reconstruction. As demonstrated in Fig. 3 (a), the Transition-A block orchestrates the progression of two-dimensional features through a sequence comprising a Fully Connected layer, a Dropout layer, and a ReLU layer. This process refines the vector, which is then transformed into a three-dimensional form via the view() function in pytorch.

In the context of the Transition-B block, depicted in Fig. 3 (b), the procedure involves channeling the two-dimensional features through a Convolutional 2D layer with a kernel size of 1, an IN layer, and a ReLU layer. Following this, the expand() function in pytorch augments the two-dimensional vector into a three-dimensional structure. This enhanced three-dimensional feature subsequently undergoes processing through a depthwise convolution layer (kernel size=3), a pointwise convolution layer (kernel size=1), an IN layer, and a ReLU layer, which collectively prepare the feature for input into the decoder. The spatial distribution of the features yielded by the Transition-A and Transition-B blocks, termed Feature A, is contingent on the input being a frontal X-ray. Conversely, if a lateral X-ray is used as input, the resultant spatial feature distribution is referred to as Feature B.

In the Transition-C block, the parallel encoder-transition-decoder network necessitates the amalgamation of the extracted features. To facilitate this, the three-dimensional vectors are aligned into a uniform coordinate space. Following alignment, an averaging method is employed to fuse the three-dimensional features.

To encapsulate, the fundamental roles of Transition-A and Transition-B are to evolve two-dimensional features into a three-dimensional framework. Transition-C, on the other hand, is responsible for aligning Feature A and Feature B within the same coordinate space and subsequently averaging these features through an additive process.

3.3. Decoder

In the decoder block, the architecture incorporates an initial feature up block followed by four decoder blocks. This design mirrors the encoder network's approach, specifically in its application of depthwise separable convolution for effective feature extraction. Each decoder block is comprised of two distinct layers, both of which are three-dimensional depthwise separable convolution layers. Delving into the structural details, each three-dimensional depthwise separable convolution layer is constituted by a depthwise convolution layer with a kernel size of 3, a pointwise convolution layer with a kernel size of 1, an IN layer, and a ReLU layer. Furthermore, the termination of each decoder block is marked by an 'up' block. This up block is configured with a depthwise convolution layer (kernel size=3), a pointwise convolution layer (kernel size=1), an IN layer, and a ReLU layer, tasked with the upscaling of features. In alignment with the encoder network's framework, the extracted features at each layer are methodically channeled to the decoder network through the Transition-B block. The integration of features from the encoder with those continuously transmitted through the network is achieved using a concatenate operation. This operation ensures that post-merging, the amalgamated features persist in their propagation through the network's subsequent layers, maintaining a streamlined and efficient feature processing pipeline.

3.4. Feature fusion

It is crucial to acknowledge that features generated by distinct X-ray sources can display varying spatial arrangements. To address this variability and ensure uniformity in spatial arrangement prior to the fusion operation, we introduce connect A into our architecture. Connect A encompasses the Transition-C block followed by a concatenate operation. This configuration is instrumental in aligning the spatial arrangements of features from different sources, thereby facilitating a seamless integration. Parallel to this process, the decoder block executes a systematic up-sampling of the three-dimensional features on a layer-by-layer basis. This meticulous up-sampling is integral to the progression of the features through the decoder block, ultimately culminating in the generation of the final CT image.

3.5. Discriminator

In our 3D discriminator, we employ Phillip's 3DPatchDiscriminator [58] as the foundational model. The discriminator begins with Block1, which is composed of a depthwise convolution layer with a kernel size of 4, followed by a pointwise convolution layer with a kernel size of 1, and concluding with a Leaky Rectified Linear Unit (Leaky ReLU) layer. This initial block sets the stage for feature extraction and preliminary processing.

Following Block1, the discriminator's architecture includes a sequence of three Block2 structures. Each of these Block2 units is similarly configured, containing a depthwise convolution layer with a kernel size of 4, a pointwise convolution layer with a kernel size of 1, an IN layer, and a Leaky ReLU layer. The progression of the discriminator culminates in Block3. This final block is comprised of a depthwise convolution layer with a kernel size of 4, and a pointwise convolution layer with a kernel size of 1. The configuration of Block3 serves as the concluding stage of feature processing within the discriminator. The entire structure of the 3D discriminator, encompassing Block1, the three iterations of Block2, and Block3, is detailed and visualized in Fig. 3 (c).

In general, the encoder blocks in X-ray image processing are responsible for extracting features that will be used in subsequent processes. The Transition-A and Transition-B blocks convert the feature dimensions from two to three dimensions. Transition-C and Connect A implement feature fusion in a parallel network architecture. The decoder blocks generate the final three-dimensional volume by progressively upsampling the three-dimensional vector. All these blocks, including the encoders, transition blocks, fusion blocks, and decoders, constitute the generator. The discriminator uses a loss function to constrain the generated three-dimensional volume to ensure that the generated results are similar to the actual results.

4. Loss functions

In this section, we outline the loss functions utilized for constraining the proposed network. The composite loss function includes four components: adversarial loss, projection loss, and our specially developed fusion loss and vector loss. These components collectively guide and optimize the network's performance.

4.1. Adversarial loss

The learning process in a generative adversarial network (GAN) is analogous to the interaction between a counterfeiter, represented by the Generator D, and a policeman, represented by the Discriminator G. The Generator D aims to produce realistic fake data to fool the Discriminator G, while the Discriminator tries to distinguish between real and fake data. The goal is to reach a state of Nash Equilibrium, where the Generator creates data indistinguishable from real data, making it difficult for the Discriminator to differentiate. The mini-max game between generator D and discriminator G can be expressed mathematically by the following formula [46]:

$$\min_{G}\max_{D} V(\mathrm{G}, D) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim noise}[\log(1 - D(G(z)))] \tag{1}$$

where z is sampled from noise distribution.

The conventional GAN assumes the discriminator to be a classifier with a sigmoid cross-entropy loss function. However, this loss function may cause the vanishing gradient problem during the learning process. To overcome this issue, we use the Least Squares Generative Adversarial Network (LSGAN) [61] in this paper. LSGAN adopts the least squares loss function as the discriminator, which has two advantages over conventional GAN. First, LSGAN generates higher quality images. Second, it performs more stably during the learning process. LSGAN is deemed more suitable for our task.

We use the loss function of LSGAN with the following expressions:

$$L_{LSGAN}(D) = \frac{1}{2}\left[\mathbb{E}_{y\sim p(CT)}(D(y|x) - 1)^2 + \mathbb{E}_{x\sim p(Xray)}(D(G(x)|x) - 0)^2\right] \tag{2}$$

$$L_{LSGAN}(G) = \frac{1}{2}\left[\mathbb{E}_{x\sim p(Xray)}(D(G(x)|x) - 1)^2\right] \tag{3}$$

where x is the input of two synthetic X-rays and y is the reconstructed CT. Discriminator D and Generator G are alternately trained to compete with each other. The LSGAN replaces the logarithmic loss with a least-square loss, which helps to stabilize the training process.

4.2. Projection loss

To boost training efficiency, we utilize simpler shapes to facilitate the regularization process. Drawing on concepts from existing works [25,62], we implement volumetric projection loss. We use orthogonal projection to simplify process, this loss focuses on the consistency of the shape, producing stronger visual effects. Projection is carried out from three directions: the axial plane, the coronal plane, and the sagittal plane. The projection loss is defined as follows:

$$L_{PL} = \frac{1}{3}\left[\mathbb{E}_{x,y}\|P_{ax}(y) - P_{ax}(G(x))\|_1 + \mathbb{E}_{x,y}\|P_{co}(y) - P_{co}(G(x))\|_1 + \mathbb{E}_{x,y}\|P_{sa}(y) - P_{sa}(G(x))\|_1\right] \tag{4}$$

Where P_{ax}, P_{co}, P_{sa} represent the projection in the axial, coronal, and sagittal plane, respectively.

4.3. Fusion loss and vector loss

To tackle the accuracy issue observed before and after the three-dimensional vector fusion in the decoder network, we introduce two specific loss functions: fusion loss and vector loss. The primary objective of these losses is to minimize the information loss that occurs before and after the fusion of feature vectors. By implementing these losses, we aim to enhance both the authenticity and accuracy of the generated results from the network, ensuring that the critical information is effectively retained and accurately represented throughout the fusion process.

In order to minimize the loss before and after feature fusion, we design loss functions based on Manhattan distance and Euclidean distance respectively. Experimental results show that using Euclidean distance is more suitable for our task. Euclidean distance can be used to accurately measure the error between vectors to constrain the generator, making the generated CT closer to the real CT.

The fusion loss is defined as:

$$L_{FL} = \mathbb{E}_{x,y}\|T_y - T_x\|_2^2 \tag{5}$$

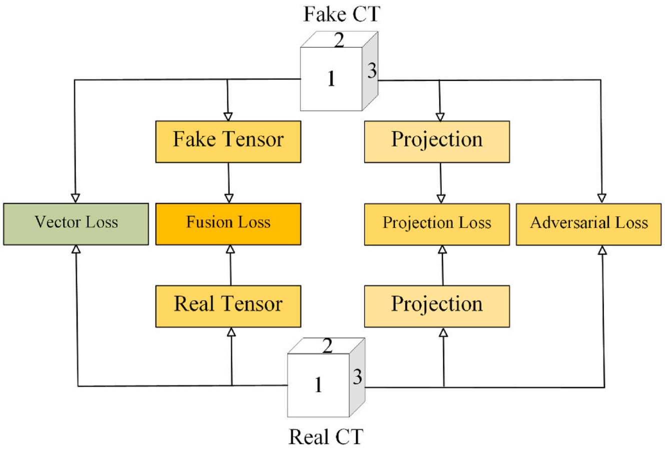

Fig. 4. The interaction of the loss functions.

where T_y represents the vector of the real CT and T_x represents the vector after fusion.

The vector loss is defined as:

$$L_{\mathrm{TL}} = \left[\mathbb{E}_{x1,y1}\left\|T_{y1} - T_{x1}\right\|_2^2 + \mathbb{E}_{x2,y2}\left\|T_{y2} - T_{x2}\right\|_2^2\right] \tag{6}$$

where T_{y1} represents the three-dimensional vector representing the front plane of the real CT and T_{x1} represents the three-dimensional vector converted from the front X-ray before fusion. T_{y2} represents the three-dimensional vector representing the side of the real CT and T_{x2} represents the three-dimensional vector converted from the side X-ray before fusion.

4.4. Total loss

Given the definitions of adversarial loss, projection loss, fusion loss, and vector loss, our total loss function is formulated as :

$$D^* = \operatorname*{argmin}_{D} \lambda_1 L_{\mathrm{LSGAN}}(D) \tag{7}$$

$$G^* = \operatorname*{argmin}_{G} [\lambda_1 L_{LSGAN}(D) + \lambda_2 L_{PL} + \lambda_3 L_{FL} + \lambda_4 L_{TL}] \tag{8}$$

Where λ_1 , λ_2 , λ_3 and λ_4 represent the importance of difference loss terms. In the reconstruction of 3D CT from X-ray projection, the adversarial loss is important to encourage local realism of the synthesized CT, but global shape consistency should be prioritized during the optimization process. Thus, we set $\lambda_1 = 0.1$, $\lambda_2 = 10$, $\lambda_3 = 10$ and $\lambda_4 = 10$ in our experiment. The interaction of the loss functions is shown in Fig. 4.

5. Simulation experiments

In this section, we delve into the experiments and results that were conducted to assess the performance of our proposed GAN model. To evaluate its efficacy, we compared the performance of our model with baseline models through a series of quantitative experiments and visualizations of the results. Additionally, structure ablation experiments were carried out to gain deeper insights into the model's architecture. These experiments played a crucial role in understanding the impact of different components and structures within the model. The basis for our experiments was a range of public datasets. These datasets provided a robust and reliable foundation for testing and validating the effectiveness of our model. Using public datasets also ensures that our results are reproducible and comparable with other studies. To comprehensively evaluate the proposed model, we employed several widely recognized metrics. These include the Mean Absolute Error (MAE), 3D Peak Signal-to-Noise Ratio (PSNR-3D), Structural Similarity (SSIM), and Cosine Similarity. Each of these metrics offers a different perspective on the model's performance, covering aspects from error measurement to image quality and structural similarity. The combination of these diverse metrics provides a holistic view of the model's capabilities and areas of improvement.

5.1. Datasets

For our experiments, we utilized the LIDC-IDRI datasets [63], which include 1018 chest CT scans. Due to the practical challenges in collecting real paired datasets, we used digitally reconstructed radiograph (DRRs) technology to create synthetic X-rays from the CT volumes. This resulted in 1018 synthesized frontal and lateral X-ray images.

To account for pixel and noise disturbances, we employed Ten-fold cross-validation in our experiment. In this method, the dataset is divided into ten parts, with a training to testing ratio of 9:1, resulting in 916 training images and 102 testing images. The final results are obtained by averaging the quantitative outcomes of the same metrics across all ten experiments, with values reported to two decimal places.

To verify the compatibility of the model with real data, we use the hospital datasets to validate the model. We use private hospital X-ray and paired CT datasets containing 1 chest CT scan and frontal and lateral X-ray images. We perform preprocessing operations on the data to meet the model input conditions, including adjusting X-rays to 8bit and resampling the resolution to 256*256 pixels, and resampling CT scans to 256*256*256 voxels.

5.2. Metrics

MAE [64] is a commonly used metric for measuring the difference between predicted and ground-truth images in image reconstruction. It calculates the L1-norm discrepancy between the reconstructed and actual images. We computed MAE across the complete 3D volume in our experiments, which provided comprehensive evaluation of the model's accuracy in image reconstruction. This measure is widely adopted and helpful in assessing the fidelity of our predictions, providing insights into the overall performance of our 3D reconstruction methodology.

PSNR-3D [65] is an improved version of the PSNR metric, designed specifically for three-dimensional image. It measures the accuracy of volumetric reconstructions by comparing the peak signal strength to the noise level, providing a comprehensive evaluation of 3D image quality. Using PSNR-3D in our methodology ensures a robust and accurate assessment of the reconstruction results, which is essential for determining the effectiveness of our 3D reconstruction approach.

SSIM [66] serves as a comprehensive index that includes factors such as brightness, contrast, and structural similarity between two images. Unlike other indicators, SSIM is closer to human subjective evaluation. In our method, SSIM plays a crucial role in assessing the fidelity of the reconstructed 3D images, ensuring that the reconstructed images are convenient for clinicians to judge, thereby enhancing the clinical utility of our 3D reconstruction method.

Cosine Similarity [67] evaluates the cosine of the angle formed between the vectors of the predicted CT image and the vectors of the real CT image, providing a similarity score from −1 to 1. This metric serves as a valuable measure of similarity, capturing the directional relationship between predicted and real image vectors. In our approach, cosine similarity helps quantify the consistency between predicted and actual CT images, providing a robust assessment of directional alignment, which is crucial to ensure the accuracy of the 3D image reconstruction process.

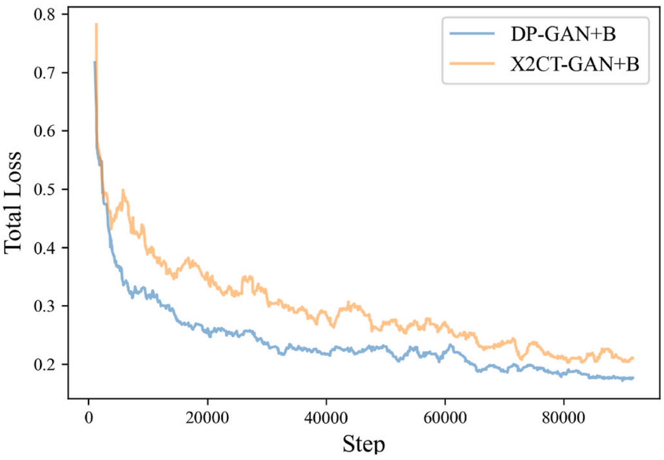

Fig. 5. The training loss curves.

Table 1
Comparison of metrics of different networks.

Model	PSNR-3D(dB)↑	SSIM ↑	MAE ↓	Cosine Similarity ↑
2DCNN	23.10	0.461	–	–
X2CT-GAN+B	26.19	0.656	93.17	0.955
DP-GAN+B	26.39	0.672	87.48	0.959

Note: '-' means we cannot find the specific parameter value.

Table 2
Comparison of parameters of different networks.

Method	Parameters_G (M)	Parameters_D (M)	Parameters_T (M)
2DCNN	–	–	9.068
X2CT-GAN+B	61.740	11.055	72.795
DP-GAN+B	40.636	0.235	40.871

5.3. Training and inference details

In our training methodology, the generator and discriminator of the GAN were trained alternately, adhering to the standard process. We utilized the Adam optimizer for this purpose, starting with an initial learning rate of 2e-4 and setting the momentum parameters at $\beta1 = 0.5$ and $\beta2 = 0.99$. The training regimen spanned over 100 epochs. During the first 50 epochs, we maintained the initial learning rate, and then implemented a linear decay strategy for the learning rate, gradually bringing it down to zero over the remaining 50 epochs. All the experiments detailed in this paper were conducted using the PyTorch framework. For the computational resources, we used Nvidia A100 GPUs, each with a memory size of 80 GB, to carry out the experiments. The training loss curves for DP-GAN+B and X2CT-GAN+B can be seen in Fig. 5. In these visualizations, the loss curve of DP-GAN+B is represented by a blue line, while the loss curve for X2CT-GAN+B is depicted in orange. These curves provide a graphical representation of the loss trends over the course of the training epochs.

5.4. Results compared with other models

In this section, we focus on the metric enhancement of our proposed method. To quantitatively evaluate the outcomes of our methods, we employ a set of metrics: PSNR-3D, SSIM, MAE, Cosine Similarity, and

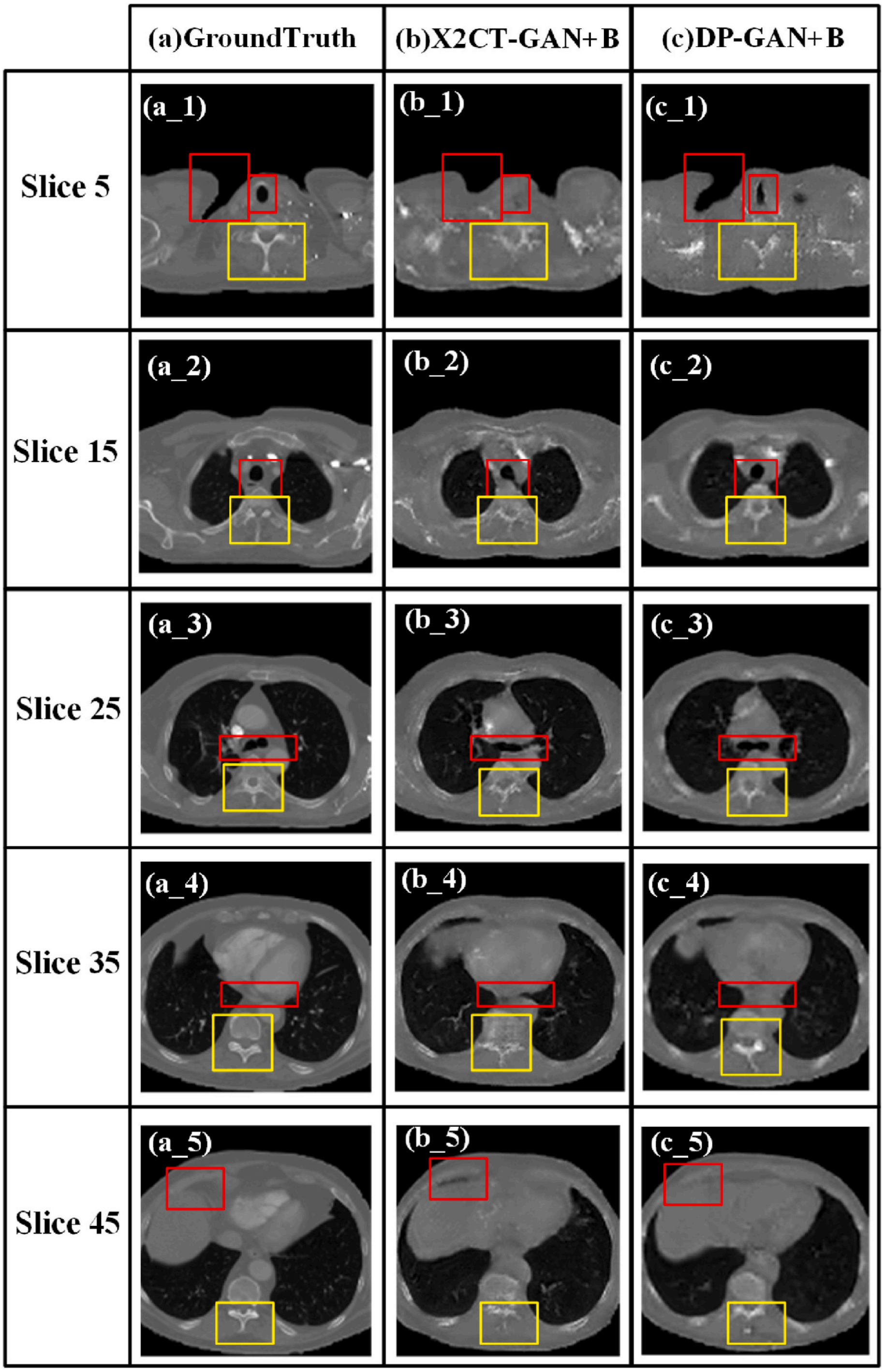

Fig. 6. Reconstructed CT scans from different approaches. (a) GroundTruth: the real CT scans. (b) X2CT-GAN+B: the baseline model. (c) DP-GAN+B: our proposed model.

network parameters. The results of these evaluations are presented in Table 1 and Table 2. In Table 1, we compare different models. ’2DCNN’ refers to the X2CT-GAN+B reproduced baseline model, which uses only a single X-ray input. ’X2CT-GAN+B′ is identified as our baseline network. ’DP-GAN+B′ represents our proposed model, where ’+B′ indicates the use of bi-planar X-rays input. Table 2 provides insights into the network parameters. ’Parameters_G′ denotes the parameters of the Generator network. ’Parameters_D′ refers to the parameters of the Discriminator network. ’Parameters_T′ signifies the parameters of the total network. These tables collectively offer a comprehensive view of the performance and complexity of each model under consideration.

Our analysis reveals that 3D reconstruction using the GAN network outperforms traditional CNN methods. The use of dual-view inputs, which contain more information, contributes to higher reconstruction accuracy. Specifically, our model, DP-GAN+B, shows a PSNR-3D that is 3.29 dB higher and an SSIM that is 0.211 higher than the 2DCNN model. Compared to X2CT-GAN+B, DP-GAN+B demonstrates an improvement of 0.2 dB in PSNR and 0.016 in SSIM, indicating better image quality and structural similarity. Notably, DP-GAN+B achieves a significant reduction in parameters by 31.925 M (44.17%), with the generator network streamlined to 0.235 M. This reduction enhances model performance, as evidenced by a decrease of 5.69 in MAE and an increase of 0.004 in

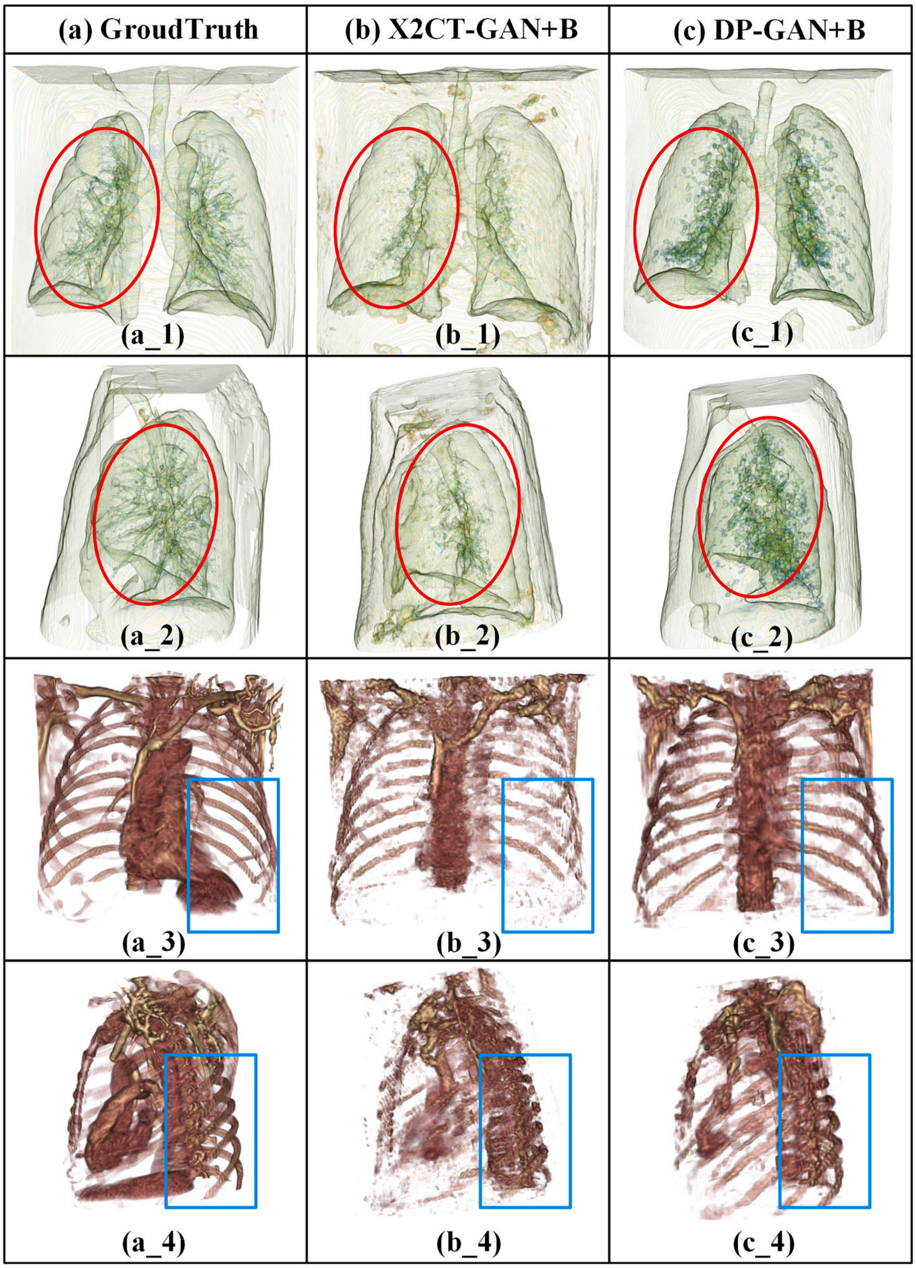

Fig. 7. Reconstructed CT volumes from different approaches. (a) Groundtruth, the real lung and chest rib volumes. (b) X2CT-GAN+B, the baseline model. (c) DP-GAN+B, our proposed model.

Cosine Similarity, reflecting lower average absolute error and improved feature vector similarity.

The results from training loss and experimental comparisons indicate that incorporating depthwise separable convolution in the adversarial network is more effective for learning tasks, adapts better to training data, and significantly reduces the number of model parameters while improving overall model performance.

5.5. Result visualization

Fig. 6 illustrates the enhanced accuracy in the reconstruction achieved by our model, DP-GAN+B, in comparison to the baseline model, X2CT-GAN+B. CT slices in our experiment are sequentially numbered starting from 1. For analysis, we extracted CT slices at a sampling interval of 10, beginning with slice number 5. The reconstructions are highlighted in Fig. 6, where the red box draws attention to the enhanced shape details in our model's reconstruction, DP-GAN+B. This area demonstrates a more accurate and precise representation of the CT structure, particularly in terms of tissue slice structure and shape, compared to the baseline model X2CT-GAN+B. Additionally, the yellow-boxed area in the figure focuses on the reconstruction results of the vertebrae. Notably, the reconstructions of the vertebrae region by our DP-GAN+B model exhibit a closer resemblance to the ground truth when compared to those generated by X2CT-GAN+B.

In summary, our DP-GAN+B model demonstrates a consistent and accurate reconstruction of major anatomical structures, maintaining a high level of detail. The model is particularly effective in capturing bone tissue details, with a special emphasis on the accuracy in the vertebrae region. These results underscore the superiority of our approach in generating high-fidelity CT reconstructions.

Fig. 7 presents the reconstructed three-dimensional (3D) computed tomography (CT) volumes of lung structures alongside the skeletal

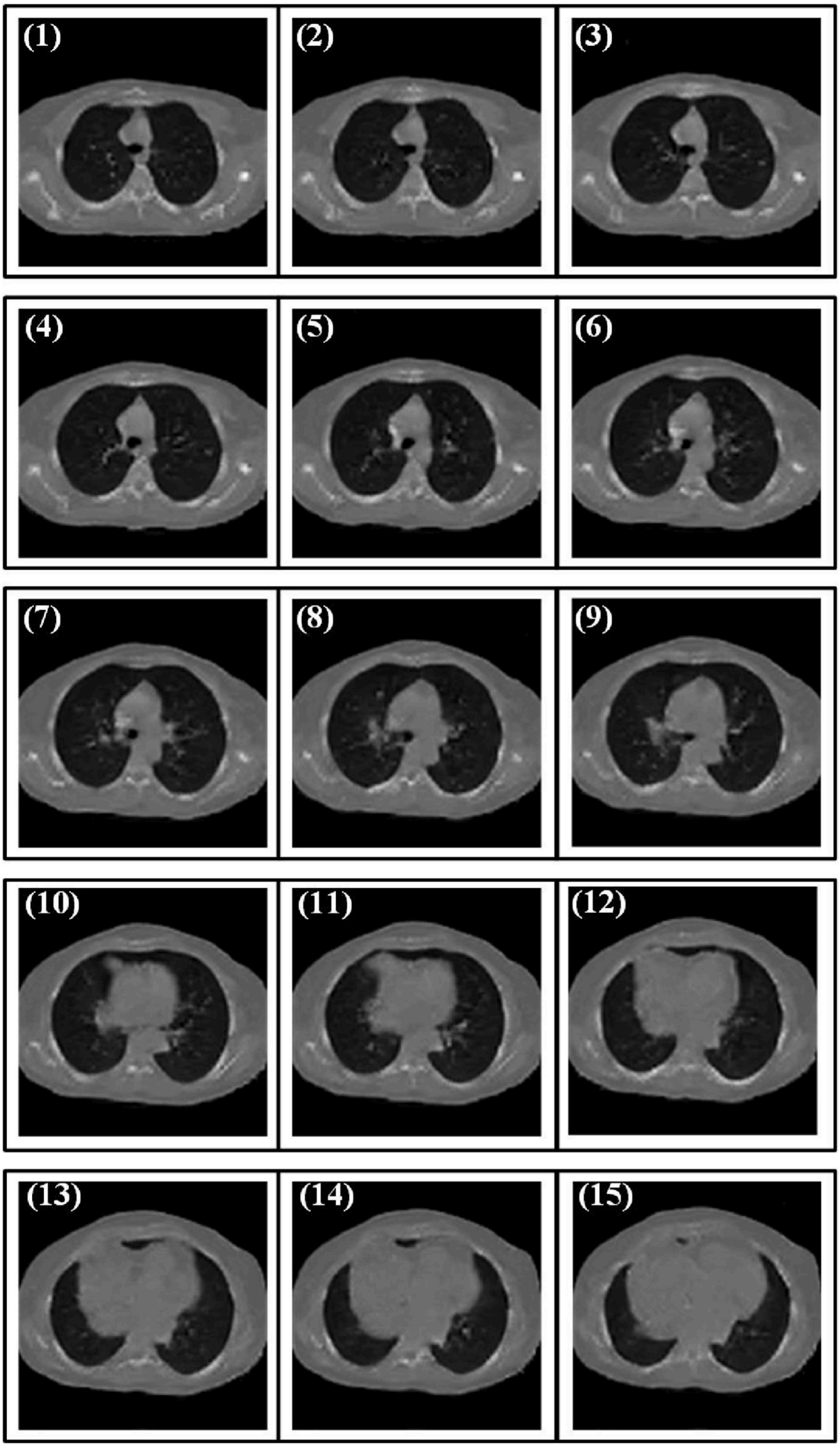

Fig. 8. Reconstructed CT scans from hospital datasets.

framework of the chest. In this figure, the 3D lung reconstructions are depicted as green volumes, while the 3D reconstructions of chest ribs are indicated in red. Our model demonstrates a notable improvement in alignment with the ground truth, especially when compared to X2CT-GAN+B. This enhanced alignment is particularly evident within the red oval, which highlights our model's superior visualization of the lung field's extent and the detailed trajectory of the small bronchi. Furthermore, as delineated within the blue box, our method exhibits increased accuracy and completeness in the reconstruction of skeletal structures, with a specific emphasis on the ribs, when contrasted with the results obtained from X2CT-GAN+B.

In summary, the performance of our reconstruction method is

Table 3
Evaluation of different settings in the base GAN network.

Combination			Metrics			
DP	TL	FL	PSNR-3D	SSIM	MAE	Cosine Similarity Params
			26.19	0.656	93.17	0.955 72.796
✓			26.25	0.669	89.11	0.957 40.871
✓	✓		26.30	0.668	88.38	0.958 40.871
✓	✓	✓	26.39	0.672	87.48	0.959 40.871

exemplary across multiple dimensions, including 2D slice views, 3D lung CT volumes, and 3D lung skeleton structures. It is crucial to acknowledge that, despite the significant advancements our reconstructed CT images represent over traditional X-ray images, they do not completely replace real CT scans, primarily due to the ultra-low radiation levels utilized in our X-ray imaging technique. However, our reconstructed CT from X-ray images substantially mitigates the challenges associated with overlapping tissues and organs in conventional X-rays, thereby significantly enhancing the visibility and differentiation of specific tissues and organs. As a result, our proposed reconstruction method has considerable potential for application in clinical settings. This includes uses in pre-operative planning, intra-operative guidance for minimally invasive procedures, and aiding in the diagnosis of fractures and other organic pathologies.

To verify the transferability of the model, we use hospital dataset to validate the model. Fig. 8 shows the CT slices generated using the hospital's datasets input model. We select the middle 15 CT slices in order. The result show that the generated slices have very accurate tissue structure and details, which is consistent with anatomical knowledge. It also proved that our model has strong transferability and has very important clinical value.

5.6. Ablation experiments

To assess the impact and effectiveness of depthwise separable convolution, tensor loss, and fusion loss within the DP-GAN+B framework, we executed a comprehensive ablation study. The outcomes of these ablation experiments are presented in Table 3. In this table, ’DP’ is used to denote depthwise separable convolution, while ’TL’ and ’FL’ correspond to tensor loss and fusion loss, respectively.

The integration of depthwise separable convolution within our model has resulted in a substantial reduction of parameters by 44.17%. This modification has led to significant performance enhancements, evidenced by a 0.06 dB increase in PSNR-3D, a 0.013 enhancement in SSIM, and a decrease in MAE from 93.17 to 89.11.

Furthermore, the incorporation of specialized loss functions has been instrumental in further improving the model's efficacy. Notably, the addition of fusion loss to the network has yielded the most significant improvements. While tensor loss also contributes to performance enhancement, its impact is comparatively modest. The combined application of these two loss functions has cumulatively increased the PSNR by 0.14 dB and improved the SSIM by 0.004. Additionally, there has been a noticeable reduction in MAE by 0.9 and an increase in cosine similarity by 0.002. The empirical results of these improvements are visually represented in Fig. 9. Specifically, Fig. 9(a) illustrates the PSNR-3D improvements, while Fig. 9(b) displays the enhancements in SSIM.

6. Discussion

In this section, we provide a certain clinical discussion of the proposed model, including its clinical application value, clinical application potential, and how to integrate this technology into actual clinical workflows.

The algorithm for generating chest CT based on frontal and lateral chest radiographs through neural adversarial network training can simulate or predict high-quality 3D CT images through low-dose, low-cost X-rays, which has significant clinical application value.

Firstly, the risk of radiation exposure can be reduced. Traditional CT scans have a higher radiation dose than ordinary chest. Thus, our algorithm can protect patients from potential long-term health effects. Secondly, it enhances diagnostic accuracy. Although X-rays are common screening methods, their two-dimensional spatial characteristics limit the assessment of the three-dimensional position relationship of lesions. By generating 3D CT images, doctors can obtain richer anatomical information, facilitating more accurate lesion localization, mass volume measurement, observation of the spatial relationship between the mediastinum and lung tissue, and improving the diagnostic accuracy of complex diseases. Thirdly, it contributes to reducing medical costs and enhancing medical efficiency. In areas with limited medical resources, obtaining three-dimensional information similar to CT from X-rays can save a lot of CT equipment purchase and operating costs, shorten patient waiting time, and improve the efficiency of medical services. Fourthly, monitor the progression of the disease. For patients with chronic respiratory diseases, consecutive time-point CTs generated by the model can be used to observe disease evolution, assisting clinicians in evaluating treatment responses and adjusting treatment plans. Lastly, it facilitates quick decision-making in emergency situations. During emergencies when immediate CT examinations are not feasible, the

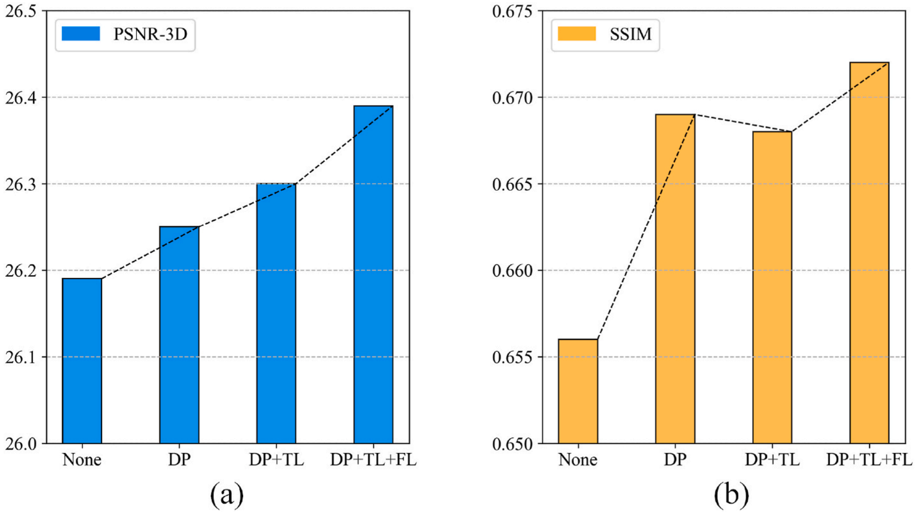

Fig. 9. Visualization of ablation experiment results. (a) PSNR-3D. (b) SSIM.

model can quickly generate preliminary 3D images from existing X-rays, providing clinicians with timely information to guide initial treatment plans.

Our technology addresses the challenge of converting two-dimensional to three-dimensional imaging, a limitation in current technology. While X-ray images offer only two-dimensional plane information and CT scans provide three-dimensional spatial data, our deep learning algorithms effectively bridge this gap, enabling transformation from two-dimensional to three-dimensional space. This innovation fills a critical technical void in medical image technology. At the same time, it can also reduce hospitals' over-reliance on expensive equipment. More precise imaging techniques require expensive equipment to support them. Our technology helps relieve the pressure on CT equipment in large medical institutions and reduces hospitals' reliance on expensive equipment.

In order to use this technology in actual clinical work, we can start from these aspects. First of all, we can use lots of real data to train the model, and conduct strict testing and verification through a large number of X-rays and corresponding CT images under different pathological conditions to ensure that the generated CT images are reliable enough in terms of clinical accuracy. Secondly, the model is packaged into easy-to-use software and integrated into the existing medical imaging information system (PACS), so that clinicians can directly use the model to output corresponding CT images when viewing X-rays. Then, technical training is provided to medical staff to educate them on the advantages and limitations of this technology and how to use the software to ensure that clinicals can accurately judge the patient's condition from the generated results. And establish a real-time monitoring and feedback mechanism. When the model is used in clinical practice, clinicals' actual usage and opinions are collected, and the model is continuously optimized and updated based on real-world diagnostic results. Under the conditions of legal compliance and ethics, through some novel and effective patient privacy protection [68] and data retrieval methods [69], ensure that this technology can truly serve patients and improve the efficiency of diagnosis and treatment.

7. Conclusion

In this paper, we apply depthwise separable convolution to construct a lightweight model, addressing the redundancy in existing models and improving performance. Our experimental results indicate that replacing traditional convolutions with depthwise separable convolutions not only reduces the model's parameters but also significantly improves its performance. Importantly, our method demonstrates a superior ability to reconstruct CT volumes with enhanced visual quality, capturing more accurate anatomical structures and shapes. Compared with Xception, MobileNets, ShuffleNet, we pioneered the application of depthwise separable convolution on GANs for generative tasks, while previous work only applied it on AlexNet, VGG and so on for classification, semantic segmentation and object detection tasks.

At the same time, we aim to develop compact, low-latency models for real-time inference on hospital mobile devices or deployment in PACS. Therefore, building a model based on depthwise separable convolution is a good choice. Depthwise separable convolution divides standard convolution into two steps: depth convolution and point-wise convolution. The depth convolution stage uses a convolution kernel to process each input channel, while the pointwise convolution stage uses a 1x1 convolution kernel for blending between channels. This separation method significantly reduces the number of parameters in the model, making the network more lightweight, reducing the consumption of computing resources, and facilitating deployment in hospitals.

Although the use of depthwise separable convolution is usually to reduce the cost of calculation and parameters, for some specific situations or specific tasks, it can show excellent performance. The reason why the model accuracy does not decrease but increases may be that in the 3D CT image reconstruction task, depthwise separable convolution processes spatial information and channel information respectively, which helps to better capture the spatial features and channel correlations of 2D X-ray and 3D CT images. This feature helps improve the model's sensitivity to image structure and texture, thereby improving the accuracy of image reconstruction.

The DP-GAN+B network addresses several clinical challenges. It shows promise in early pneumonia diagnosis, assessing post-treatment recovery, and providing precise measurements of chest nodule sizes. One of the notable benefits of our algorithm is its facilitation of rapid examinations. This is particularly advantageous in scenarios involving uncooperative pediatric patients, leading to improved detection rates and reduced waiting times for CT scans. The efficiency of our method saves crucial time for both healthcare providers and patients. Furthermore, it ensures patient safety by utilizing the significantly lower radiation doses associated with X-ray imaging compared to traditional CT scans.

Our work has three main limitations. First of all, currently only the lungs can be reconstructed. In theory, a variety of organs and tissues can be reconstructed as long as they comply with the model input. Secondly, only frontal and lateral X-rays are allowed as input to the model. It is known that there are not only frontal and side X-ray images, but also left-side and right-side X-ray images in clinical settings. If we could feed all of these images into the model, the accuracy of the results would be significantly enhanced and the outcomes would be more relevant from clinical standpoint. Finally, during the training process, we found that the training of GAN is very unstable and difficult to achieve convergence. Perhaps we can consider using a stable diffusion model to replace GAN to improve training efficiency and stability.

Looking forward, our research will pursue addressing the above limitations. Firstly, we aim to expand the reconstruction capabilities of our network to include different organs. This expansion will encompass the lumbar spine, liver, heart, and the application to pediatric chest X-rays for detecting conditions such as mycoplasma infections or pneumonia. Secondly, instead of being limited to frontal and side inputs, we can expend the range of input images, such as left-side and right-side images, so that the model can learn more features to reconstruct a more accurate model. Finally, we plan to incorporate diffusion models into our framework for 3D organ reconstruction. The diffusion model, known for its training stability and ability to generate more realistic visual effects compared to GAN networks, is expected to produce medical images that are more conducive for clinical diagnosis.

Code availability

All code used for dataset preparation and image processing will be available on request.

Funding

This work is supported by the Ministry of Science and Technology of the People's Republic of China (STI2030-Major Projects2021ZD0201900), the National Natural Science Foundation of China under Grant 12090052, and Wenzhou Institute University of Chinese Academy of Sciences (WIUCASQD2022010).

CRediT authorship contribution statement

Xinlong Xing: Methodology, Software, Writing – original draft. **Xiaosen Li:** Conceptualization, Methodology. **Chaoyi Wei:** Conceptualization, Methodology. **ZhanTian Zhang:** Data curation, Validation, Visualization. **Ou Liu:** Funding acquisition, Methodology, Supervision, Writing – review & editing. **Senmiao Xie:** Data curation, Resources. **Haoman Chen:** Data curation, Resources. **Shichao Quan:** Data curation, Resources. **Cong Wang:** Writing – review & editing. **Xin Yang:** Software. **Xiaoming Jiang:** Conceptualization, Methodology. **Jianwei Shuai:** Writing – review & editing, Supervision, Funding acquisition, Methodology.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] P. Safety, Radiation Dose in X-Ray and CT Exams, American College of Radiology and Radiological Society of North America, 2012.

[2] X. Zhang, T. Wang, J. Wang, et al., Pyramid channel-based feature attention network for image dehazing, in: Computer Vision and Image Understanding, 197, 2020 103003.

[3] H. Liang, H. Bai, K. Hu, et al., Bioinspired polarized skylight orientation determination artificial neural network, JBE 20 (2023) 1141–1152.

[4] X. Li, X. Qin, C. Huang, et al., SUnet: a multi-organ segmentation network based on multiple attention, Comput. Biol. Med. 167 (2023) 107596.

[5] K. Hu, L. Zhao, S. Feng, et al., Colorectal polyp region extraction using saliency detection network with neutrosophic enhancement, Comput. Biol. Med. 147 (2022) 105760.

[6] X. Jiang, Y. Ding, M. Liu, et al., BiFTransNet: a unified and simultaneous segmentation network for gastrointestinal images of CT & MRI, Comput. Biol. Med. 165 (2023) 107326.

[7] A. Qi, D. Zhao, F. Yu, et al., Directional mutation and crossover boosted ant colony optimization with application to COVID-19 X-ray image segmentation, Comput. Biol. Med. 148 (2022) 105810.

[8] H. Su, D. Zhao, H. Elmannai, et al., Multilevel threshold image segmentation for COVID-19 chest radiography: a framework using horizontal and vertical multiverse optimization, Comput. Biol. Med. 146 (2022) 105618.

[9] W. Liu, Y. Du, G. Fang, et al., Efficient Gaussian sample specific network marker discovery and drug enrichment analysis validation, Comput. Biol. Chem. 83 (2019) 107139.

[10] H. Chen, H. Ye, F. Chen, et al., Revolutionizing infection risk scoring: an ensemble "from weak to strong" deduction strategy and enhanced point-of-care testing tools, Adv. Intell. Syst. 5 (2023) 2300224.

[11] H. Gao, J. Sun, Y. Wang, et al., Predicting metabolite–disease associations based on auto-encoder and non-negative matrix factorization, Briefings Bioinf. 24 (2023) bbad259.

[12] C. Wei, X. Xiang, X. Zhou, et al., Development and validation of an interpretable radiomic nomogram for severe radiation proctitis prediction in postoperative cervical cancer patients, Front. Microbiol. 13 (2023) 1090770.

[13] F. Zhu, J. Ding, X. Li, et al., MEAs-Filter: a novel filter framework utilizing evolutionary algorithms for cardiovascular diseases diagnosis, Health Inf. Sci. Syst. 12 (2024) 8.

[14] H. Hu, Z. Feng, H. Lin, et al., Gene function and cell surface protein association analysis based on single-cell multiomics data, Comput. Biol. Med. 157 (2023) 106733.

[15] W. Wang, L. Zhang, J. Sun, et al., Predicting the potential human lncRNA–miRNA interactions based on graph convolution network with conditional random field, Briefings Bioinf. 23 (2022) bbac463.

[16] J. Zhao, J. Sun, S.C. Shuai, et al., Predicting potential interactions between lncRNAs and proteins via combined graph auto-encoder methods, Briefings Bioinf. 24 (2023) bbac527.

[17] Q. He, C.-Q. Zhong, X. Li, et al., Dear-DIAXMBD: deep autoencoder enables deconvolution of data-independent acquisition, Proteomics, Research 6 (2023) 179.

[18] X. Chen, R. Zhu, J. Zhong, et al., Mosaic composition of RIP1–RIP3 signalling hub and its role in regulating cell death, Nat. Cell Biol. 24 (2022) 471–482.

[19] K. Hammernik, T. Würfl, T. Pock, et al., A deep learning architecture for limited-angle computed tomography reconstruction, Bildverarbeitung für die Medizin 2017, Springer, 2017, pp. 92–97.

[20] P. Henzler, V. Rasche, T. Ropinski, et al., Single-image tomography: 3D volumes from 2D cranial x-rays, Comput. Graph. Forum 37 (2018) 377–388.

[21] Y. Kasten, D. Doktofsky, I. Kovler, 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, Springer, 2020, pp. 123–133.

[22] Y. Liang, W. Song, J. Yang, et al., X2teeth: 3d teeth reconstruction from a single panoramic radiograph, in: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Springer, 2020, pp. 400–409.

[23] L. Shen, W. Zhao, L. Xing, Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning, Nat. Biomed. Eng. 3 (2019) 880–888.

[24] T. Würfl, F.C. Ghesu, V. Christlein, et al., Deep learning computed tomography, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2016, pp. 432–440.

[25] X. Ying, H. Guo, K. Ma, et al., X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10619–10628.

[26] M. Bertero, T.A. Poggio, V. Torre, Ill-posed problems in early vision, Proc. IEEE 76 (1988) 869–889.

[27] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90.

[28] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[29] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.

[30] C. Szegedy, V. Vanhoucke, S. Ioffe, et al., Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826.

[31] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1251–1258.

[32] A.G. Howard, M. Zhu, B. Chen, et al., Mobilenets: efficient convolutional neural networks for mobile vision applications, in: arXiv Preprint arXiv:1704.04861, 2017.

[33] X. Zhang, X. Zhou, M. Lin, et al., Shufflenet: an extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6848–6856.

[34] C.B. Choy, D. Xu, J. Gwak, et al., 3d-r2n2: a unified approach for single and multi-view 3d object reconstruction, in: European Conference on Computer Vision (ECCV), Springer, 2016, pp. 628–644.

[35] B. Neubert, T. Franken, O. Deussen, Approximate image-based tree-modeling using particle flows, ACM Trans. Graph. 26 (2007) 88.

[36] S. Tulsiani, A.A. Efros, J. Malik, Multi-view consistency as supervisory signal for learning shape and pose prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 2897–2905.

[37] H. Fan, H. Su, L.J. Guibas, A point set generation network for 3d object reconstruction from a single image, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 605–613.

[38] Y. Livny, F. Yan, M. Olson, et al., Automatic reconstruction of tree skeletal structures from point clouds, Proc. of ACM SIGGRAPH Asia (2010) 1–8.

[39] B. Aubert, C. Vergari, B. Ilharreborde, et al., 3D reconstruction of rib cage geometry from biplanar radiographs using a statistical parametric model approach, Comput Methods Biomech Biomed Eng Imaging Vis 4 (2016) 281–295.

[40] J. Dworzak, H. Lamecker, J. von Berg, et al., 3D reconstruction of the human rib cage from 2D projection images using a statistical shape model, Int. J. Comput. Assist. Radiol. Surg. 5 (2010) 111–124.

[41] H. Lamecker, T.H. Wenckebach, H.-C. Hege, Atlas-based 3D-shape reconstruction from X-ray images, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2006, pp. 371–374.

[42] C. Koehler, T. Wischgoll, Knowledge-assisted reconstruction of the human rib cage and lungs, IEEE Comput Graph Appl 30 (2010) 17–29.

[43] M. Nakao, F. Tong, M. Nakamura, et al., Image-to-graph convolutional network for deformable shape reconstruction from a single projection image, in: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Springer, 2021, pp. 259–268.

[44] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907 (2016).

[45] Z. Tan, J. Li, H. Tao, et al., XctNet: reconstruction network of volumetric images from a single X-ray image, Comput. Med. Imag. Graph. 98 (2022) 102067.

[46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial networks, Commun. ACM 63 (2020) 139–144.

[47] C. Ledig, L. Theis, F. Huszár, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4681–4690.

[48] B. Wu, H. Duan, Z. Liu, et al., SRPGAN: perceptual generative adversarial network for single image super resolution, in: arXiv Preprint arXiv:1712.05927, 2017.

[49] J.-Y. Zhu, T. Park, P. Isola, et al., Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV), 2017, pp. 2223–2232.

[50] B. Jahanyar, H. Tabatabaee, A. Rowhanimanesh, MS-ACGAN: a modified auxiliary classifier generative adversarial network for schizophrenia's samples augmentation based on microarray gene expression data, Comput. Biol. Med. 162 (2023) 107024.

[51] H. Zhao, X. Qiu, W. Lu, et al., High-quality retinal vessel segmentation using generative adversarial network with a large receptive field, Int. J. Imag. Syst. Technol. 30 (2020) 828–842.

[52] Z. Wang, S. Stavrakis, B. Yao, Hierarchical deep learning with Generative Adversarial Network for automatic cardiac diagnosis from ECG signals, Comput. Biol. Med. 155 (2023) 106641.

[53] S. Bhattacharya, A. Bhattacharya, S. Shahnawaz, Generating synthetic computed tomography (CT) images to improve the performance of machine learning model for pediatric abdominal anomaly detection, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV), 2023, pp. 3865–3873.

[54] W. Song, Y. Liang, J. Yang, et al., Oral-3d: reconstructing the 3d structure of oral cavity from panoramic x-ray, Proc. AAAI Conf. Artif. Intell. (2021) 566–573.

[55] C.J. Yang, C.L. Lin, C.K. Wang, et al., Generative adversarial network (GAN) for automatic reconstruction of the 3D spine structure by using simulated Bi-planar X-ray images, Diagnostics 12 (2022) 1121.

[56] L. Sifre, S. Mallat, Rigid-motion Scattering for Texture Classification, 2014 arXiv preprint arXiv:1403.1687.

[57] L. Sifre, S. Mallat, Rotation, scaling and deformation invariant scattering for texture discrimination, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1233–1240.

[58] C. Szegedy, W. Liu, Y. Jia, et al., Going deeper with convolutions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.

[59] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 2015, pp. 448–456.

[60] G. Huang, Z. Liu, L. Van Der Maaten, et al., Densely connected convolutional networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[61] X. Mao, Q. Li, H. Xie, et al., Least squares generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV), 2017, pp. 2794–2802.

[62] L. Jiang, S. Shi, X. Qi, et al., Gal: geometric adversarial loss for single-view 3d-object reconstruction, in: Proceedings of the European Conference on Computer Vision, ECCV), 2018, pp. 802–816.

[63] S.G. Armato 3rd, G. McLennan, L. Bidaut, et al., The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans, Med. Phys. 38 (2011) 915–931.

[64] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, Geosci, Model Develop 7 (2014) 1247–1250.

[65] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 2366–2369.

[66] Z. Wang, A.C. Bovik, H.R. Sheikh, et al., Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.

[67] F. Rahutomo, T. Kitasuka, M. Aritsugi, Semantic cosine similarity, in: The 7th International Student Conference on Advanced Science and Technology (ICAST), 2012, p. 1.

[68] Z. Wu, H. Liu, J. Xie, et al., An effective method for the protection of user health topic privacy for health information services, World Wide Web (2023) 1–23.

[69] Z. Mei, J. Yu, C. Zhang, et al., Secure multi-dimensional data retrieval with access control and range query in the cloud, Inf. Syst. 122 (2024) 102343.