2025년도 한국통신학회 동계종합학술발표회

S-XR2CT: CT Reconstruction from Single X-Ray Image using GAN-based Approach

[1]Siekny Hor, [1]Ehwa Yang, [2]Jae-Hun Kim

[1]Research Institute for Future Medicine, Samsung Medical Center, Seoul, Korea,

[2]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea.

GAN 기반 접근 방식을 활용한 단일 X-Ray 에서 CT 영상의 생성

[1] 호식니, [1] 양이화, [2] 김재훈

[1] 삼성서울병원 미래의학연구원, [2] 삼성서울병원 영상의학과, 성균관대학교 의과대학

Abstract

CT scans are more essential for accurate disease diagnosis, offering detailed 3D information. However, they are significantly more expensive than X-ray machines, making them less accessible in clinical settings. In this work, we propose a method called S-XR2CT to reconstruct CT images from a single X-ray using a GAN-based approach. Our method includes a 2D mapping network and a 3D encoder-decoder. Through qualitative and quantitative evaluations, our model demonstrates improved structural accuracy compared to existing methods.

Ⅰ. Introduction

X-ray and CT imaging are commonly used in hospitals. CT scans offer a significant advantage by presenting tissues in 3D, effectively resolving the tissue of overlapping structures seen in 2D imaging. However, CT scans are expensive and require extensive infrastructure, making them inaccessible in resource-limited or underserved areas. In contrast, X-rays are cheaper and more widely available imaging modality, but providing limited 2D anatomical information, which can obstruct accurate diagnosis. In this work, we propose a method to reconstruct a 3D CT volume from a single X-ray captured from an orthogonal plane. The major challenge in reconstruction is the ambiguity inherent in X-ray imaging, multiple different 3D CT volume can produce the same 2D X-ray when projected. Existing method, such as X2CT-GAN[1], used generative adversarial networks (GANs) for CT reconstruction, demonstrating promising results. However, it loses high-frequency details and produce loss-resolution outputs due to the challenges of inverse imaging problems. To address these issues and enhance reconstruction performance, we introduce a method called S-XR2CT: Reconstruction CT from a Single X-Ray view with GAN. Our method uses a single frontal X-ray image as input and incorporates an initial template CT to guide the 3D network. The generated 3D CT volumes are then passed into the discriminator for improved fidelity.

Ⅱ. Method

Let X_i denotes the frontal X-ray view for patient i, captured using a specific acquisition geometry. ($X_i \in R^{1\times128\times128}$). The patient's ground truth CT scan is represented as $Y_i \in R^{1\times128\times128\times128}$. Our architecture, illustrated in Figure 1, consists of a generator and a discriminator. The generator produces a synthetic CT image $Y_i' \in R^{1\times128\times128\times128}$, and then then generated CT and the real CT are passed into a discriminator to determine whether the CT volume is real or fake.

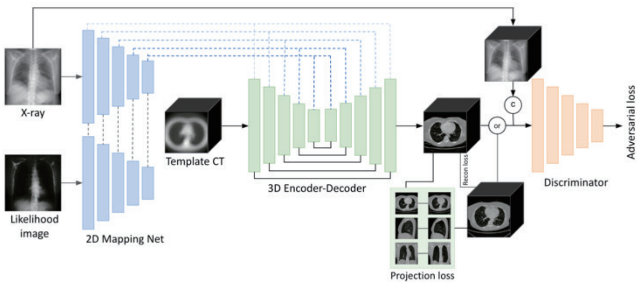

Figure 1. Overview of our method, consisting of a generator and a discriminator. The generator is composed of a 2D mapping network and a 3D Encoder-Decoder.

1. Generator

2D Mapping Network: For each X-ray, we compute a likelihood image using the mean and standard deviation of pixel intensities as described in the equation below. We extract multi-scale 2D features from X-ray using a ConvNeXt[2] network. The likelihood image is downsampled and concatenated with the X-ray features, which are then fed into a 3D network to generate the final results.

$$L = \log\sqrt{2\pi\sigma^2} + \frac{(X-\mu)^2}{2\sigma^2}$$

3D Encoder: The 3D encoder takes a template CT image as a module parameter and the X-ray style features as inputs, concatenating them at each layer. The 3D decoder then upsamples the features to reconstruct the CT volume. To enhance the reconstruction quality, we incorporate self-attention following the Swin UNEt-TRansformer (Swin UNETR) model[3] after the first layer in the decoder.

2. Discriminator

We adapt a similar architecture from X2CT-GAN, conditioning on expanded input x-ray to guide the training process effectively.

3. Loss function

We follow the loss functions used in X2CT-GAN including adversarial loss (L_G, L_D), reconstruction loss (L_{rec}), and projection loss (L_{pl}). Additionally, we adopt perceptual loss (L_p) to promote realistic reconstruction details. For the adversarial loss, we follow the loss in the X2CT-GAN.

$$D = \arg min \lambda_1 L_D,$$

$$G = \arg\min(\lambda_1 L_G + \lambda_2 L_{rec} + \lambda_2 L_{pl} + \lambda_3 L_p),$$

where, $\lambda_1 = 0.1$, $\lambda_2 = 10$, $\lambda_3 = 1$.

4. Dataset and Preprocessing

To train and validate the proposed reconstruction CT from X-ray approach, we need a dataset with paired X-ray and corresponding CT images. We use LIDC-IDRI dataset [4], which contains 1018 chest scans. We first resample the CT scans to the $1\times1\times1mm^3$, then cropping into a $320\times320\times320$ resolution. To obtain the paired X-ray from corresponding CT for each subject, the digitally reconstructed radiographs (DRR) is adopted to project the X-ray image. We randomly divide the dataset into a training set (868 samples), a validation set (50 samples), and a testing set (100 samples). Input images to the model are resized to 128×128 and the output is set to $128\times128\times128$.

5. Results

To demonstrate the advantages of our methods, we compare our approach with X2CT-GAN single-view CT reconstruction method as shown in Table 1. We also conduct a qualitative assessment of the reconstruction performance compare with X2CT-GAN, as shown in Figure 2. our method generates more accurate and shows small anatomical structure.

Methods	SSIM ↑	PNSR ↑	LIPIP ↓
X2CT-GAN	0.5133	23.3972	0.3768
S-XR2CT	**0.5226**	**23.5115**	**0.3216**

Table 1. Quantitative comparison results between X2CT-GAN and ours.

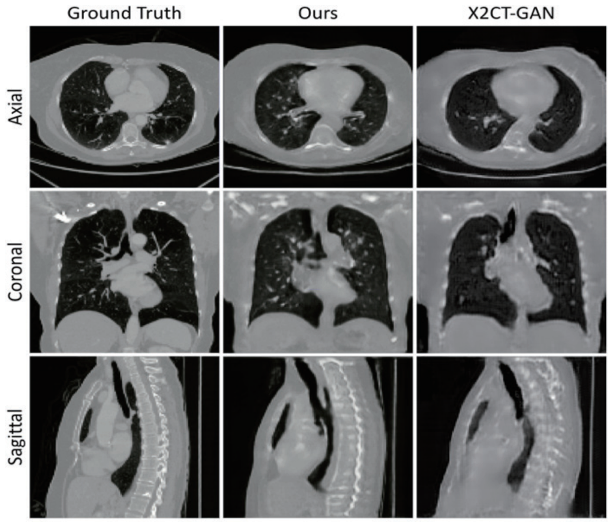

Figure 2. Visualization comparison between X2CT-GAN and ours.

Ⅲ. Conclusion

In this paper, we present a method for reconstructing 3D CT scans from a single 2D X-ray using a GAN framework. Our qualitative and quantitative experiments shows that our approach produces more accurate and detailed structures compared to existing methods, highlighting the strong generalization capability of our proposed model.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Kore (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C1004516).

REFERENCES

[1] Xingde Ying, Heng Guo, Kai Ma, Jian Wu, Zhengxin Weng, and Yefeng Zheng. X2CT-GAN. Reconstructing CT from biplanar X-rays with generative adversarial networks. In CVPR, 2019.

[2] huang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In CVPR, 2022.

[3] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. InInternational MICCAI brainlesion workshop. Springer, 2021.

[4] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNittGray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. MP, 38(2):915–931, 2011.