EasyChair Preprint

№ 11239

Prior Information Enhanced Adversarial Learning for kVp Switching CT

Yizhong Wang, Ailong Cai, Ningning Liang, Shaoyu Wang, Junru Ren, Xinrui Zhang, Lei Li and Bin Yan

November 3, 2023

Prior information enhanced adversarial learning for kVp switching CT

Yizhong Wang[1], AiLong Cai[1], Ningning Liang[1], Shaoyu Wang[1], Junru Ren[1], Xinrui Zhang[1], Lei Li[1] and Bin Yan[1]

[1] Department of Henan Key Laboratory of Imaging and Intelligent Processing, PLA Strategic Support Force Information Engineering University, Zhengzhou, China

Abstract Dual energy computed tomography (DECT) can provide both structural and material information of the scanned object, and has been widely used in the medical field. However, patients may suffer from genetic damage and cancer under long-term high radiation dose of x-ray exposure. To reduce radiation dose and ensure optimal hardware cost. This work studies the switching technology based on the x-ray tube voltage (kVp). However, the kVp switching technology faces the problems of low sampling rate of each energy spectrum and the spatial misalignment of projection data of different energy spectrum. Thus, this study introduces an adversarial learning mechanism and proposes a Prior Information enhanced Projection data Inpainting Network (PINet). The experimental results show that the PINet framework is a promising approach for sparse-view angle DECT imaging.

1 Introduction

Different from traditional computed tomography (CT), dual energy computed tomography (DECT) can simultaneously provide the structural information and material information of the scanned object by obtaining the attenuation measurement of two different x-ray energy spectrum [1]. At present, DECT has been widely used in clinical diagnosis, such as virtual monoenergetic imaging [2],perfused blood volume imaging [3], and aortic disease diagnosis [4].

Increasing the radiation dose of x-ray is known to improve the quality of medical images. However, patients are likely to suffer from genetic damage and cancer under long-term high radiation dose of x-ray exposure. Therefore, lowering the radiation dose is also the focus of the medical imaging community. In order to reduce radiation dose and ensure optimal hardware cost. This paper study the switching technology based on x-ray tube voltage (kVp) as shown in Fig. 1. This technology only requires traditional energy integration detector and ray source, and the dose is about half of that of traditional DECT.

However, technology based on kVp switching not only faces the problem of low sampling rate of each energy spectrum, but also a common problem is that the projection of different energy spectrum is not aligned in space. Recently, deep learning has shown great potential in the field of medical image processing. Lee et al. proposed an interpolation method based on convolutional neural networks (CNN) to inpainting missing projection [5]. In 2022, Cao et al. developed a CNN framework for sparse-view projection completion and material decomposition [6]. Generative Adversarial Networks (GAN) also have great potential in the application of DECT. Kawahara et al. proposed an image synthesis framework based on GAN to material decomposition images of bone and fat scanned by DECT [7]. In 2022, Wang et al. designed a dual-way mapping GAN to mine the relationship between two different energy projection data, aiming at recovering the missing data[8].

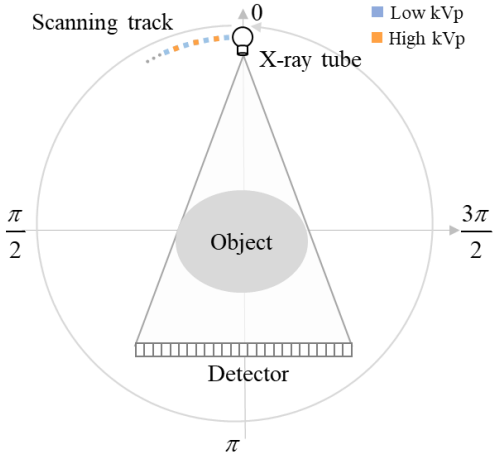

Figure 1: An illustration of the kVp switching technology.

There is a certain correlation between low-energy and high-energy projection data of the same object for DECT. When DECT imaging is faced with a serious shortage of projection data, researchers usually use the correlation between energy spectrum data to achieve DECT imaging [9]. Therefore, this study fuses dual energy projection data as prior information, and proposes a Prior Information enhanced Projection data Inpainting Network (PINet) for sparse-view angle DECT. In the PINet framework, this work introduces an adversarial learning mechanism to generate results close to the real projection data. In order to make full use of the prior information and original information, the prior information and sparse-view dual energy projection data are sent to two separate encoders to extract and fuse useful features. Then, the two decoders perform differential learning on the projection data in different energy channels.

2 Methods

The PINet framework is shown in Fig. 2. First, the kVp switching technology is used to obtain two sparse-view projection data with different energy spectrum. Then, the generator G takes prior information and sparse-view projection data as input, extracts and uses various features to generate low-energy and high-energy full angle (360°) projection data. The prior information is the data after fusing the sparse-view projection data of two kinds of energy. Due to the different scanning angles of different energies, the data of prior information is twice as large as that of single energy sparse-view data. At the same time, the

discriminator D encourages the generator G to generate realistic results as much as possible. Once the PINet training is completed, the trained generator G can be used to generate the completed projection data.

$$G^*\left(\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Sparse}} \\ x_{\mathrm{H}}^{\mathrm{Sparse}} \end{pmatrix}, x_{\mathrm{Prior}}^{\mathrm{Fusion}}\right)=\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Pred}} \\ x_{\mathrm{H}}^{\mathrm{Pred}} \end{pmatrix}, \tag{1}$$

where G^* denotes the trained generator. $\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Sparse}} \\ x_{\mathrm{H}}^{\mathrm{Sparse}} \end{pmatrix}$ and $x_{\mathrm{Prior}}^{\mathrm{Fusion}}$ represent sparse-view dual energy projection data and prior information respectively, which are input into the generator G. $\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{pred}} \\ x_{\mathrm{H}}^{\mathrm{pred}} \end{pmatrix}$ represents the dual energy full angle projection data output by generator G. During training, the objective of PINet can be expressed as:

$$\begin{aligned}\mathcal{L}_{\mathrm{PINet}} = & \mathbb{E}_{\left(\begin{smallmatrix} x_{\mathrm{L}}^{\mathrm{Ref}} \\ x_{\mathrm{H}}^{\mathrm{Ref}} \end{smallmatrix}\right)}\left[\log D\left(\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Ref}} \\ x_{\mathrm{H}}^{\mathrm{Ref}} \end{pmatrix}\right)\right] \\ & + \mathbb{E}_{\left(\begin{smallmatrix} x_{\mathrm{L}}^{\mathrm{Pred}} \\ x_{\mathrm{H}}^{\mathrm{Pred}} \end{smallmatrix}\right)}\left[\log\left(1-D\left(\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Pred}} \\ x_{\mathrm{H}}^{\mathrm{Pred}} \end{pmatrix}\right)\right)\right] \\ & + \lambda\mathbb{E}_{\left(\begin{smallmatrix} x_{\mathrm{L}}^{\mathrm{Ref}} \\ x_{\mathrm{H}}^{\mathrm{Ref}} \end{smallmatrix}\right),\left(\begin{smallmatrix} x_{\mathrm{L}}^{\mathrm{Pred}} \\ x_{\mathrm{H}}^{\mathrm{Pred}} \end{smallmatrix}\right)}\left[\left\|\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Ref}} \\ x_{\mathrm{H}}^{\mathrm{Ref}} \end{pmatrix}-\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Pred}} \\ x_{\mathrm{H}}^{\mathrm{Pred}} \end{pmatrix}\right\|_1\right]\end{aligned} \tag{2}$$

where λ is the weight parameter. $\begin{pmatrix} x_{\mathrm{L}}^{\mathrm{Ref}} \\ x_{\mathrm{H}}^{\mathrm{Ref}} \end{pmatrix}$ is the label projection data of full angle. In the objective function, the mean absolute error (MAE) between the generated projection data and the label projection data is introduced to generate more realistic projection data. During network training, G tries to minimize the objective function, while D tries to maximize the objective function, i.e.,

$$G^*, D^* = \arg\min_G \max_D \mathcal{L}_{\mathrm{PINet}} \tag{3}$$

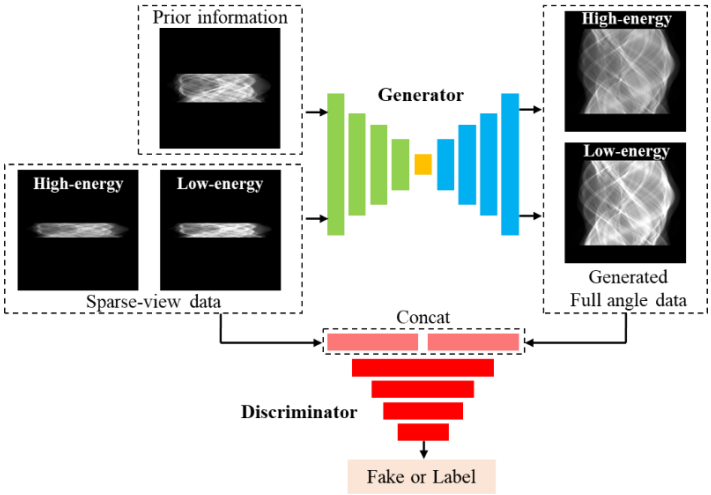

Figure 2: Schematic of the PINet framework.

The PINet consists of a generator G and a discriminator D. The generator G is improved from the classic U-Net [10] as shown in Fig.3. To make full use of prior information and match two energy channels, the structure of G is extended to a dual input and dual output network structure. The network structure consists of three parts: encoding module, fusion module and decoding module. The encoding module contains two encoding channels, which process prior information and sparse-view projection data respectively. The sparse-view projection data channel is used as the main channel, and the initial number of channels is set to 32. The prior information channel is used as an auxiliary component, and the initial channel number is set to 8. Then, the fusion module aims to achieve the fusion of the features extracted from the two encoding channels. Considering that the difference between different energy projection data is the key to material identification, the decoding module uses two decoding channels to process the fused feature information to generate DECT data. During decoding, the shallow extracted features of the encoding module will be copied and connected to the low-energy and high-energy decoding channels.

The structure of the discriminator D is a CNN, and its input is paired sparse-view angle projection data and full angle projection data (generated or label). The discriminator D has five layers. The first layers contain convolution, batch norm (BN) and Rectified Linear Unit (ReLU) operations. The size of the convolution kernel is 3×3, the stride is 2, and the number of channels is 32. The second and third layers contain 3×3 convolution with stride 2, 3×3 convolution with stride 1, BN and ReLU operations, and the number of channels is 64 and 128 respectively. The last layer includes global average pooling, full connection and sigmoid operations The output of the discriminator D is true or false to match the projection data pair, which is equivalent to 0-1 classification.

3 Experimental Results

The experimental dataset was established from real clinical dataset, and the DECT images were obtained using the SOMATOM Definition Flash DECT scanner (Siemens Healthcare, Germany). 80 kVp and 140 kVp spectra are used for low-energy and high-energy scanning, respectively. The dataset includes 1491 cranial cavity images of 6 patients. The size of the image is 512 × 512. The projection dataset used to train PINet is generated using 1000 images of 5 patients, while 100 images of another patient are used as the test dataset of the network.

In this paper, Sidden's ray tracing algorithm [11] is used to simulate the geometry of the fan beam. The distances from the x-ray source to the object and the detector are set to 1000 mm and 1500 mm respectively. Both low-energy and high-energy projections collect 360 frames of projection data within the 360 degree scanning range. The projection data

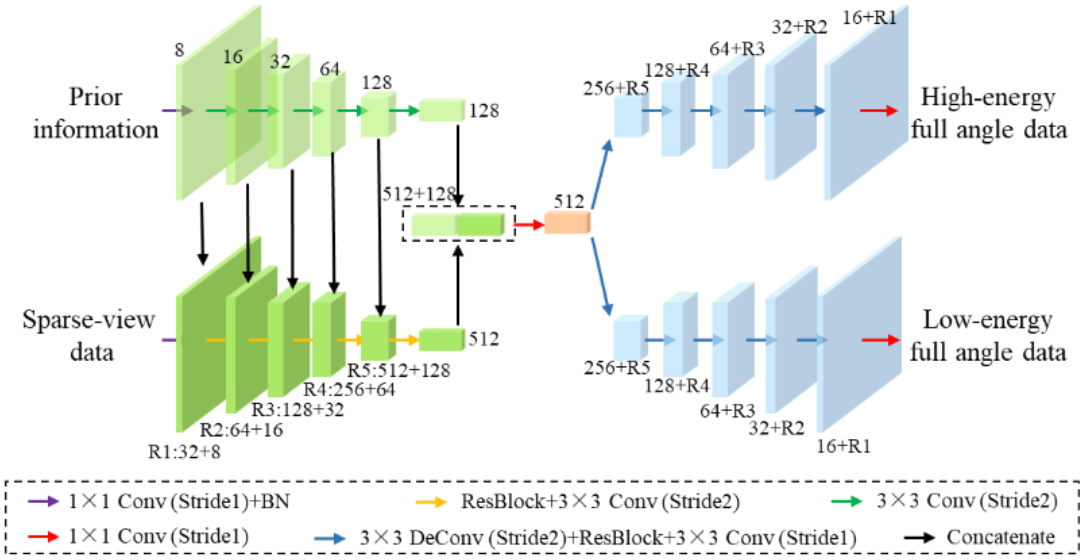

Figure 3: Network structure for the generator G.

of each frame is collected by a linear detector, which consists of 512 bins. The size of the generated full angle projection data is 360 ×512, and then two 76 ×512 size all zero matrices are added to the generated data. Finally, the size of full angle projection data is 512 × 512 obtained as the label of PINet training. In this work, low-energy and high-energy sparse-view projection data are obtained by kVp switching in every 3 °rotation range. The size of each energy projection data is 60 × 512. The size of prior projection data is 120 ×512. Then, using operations similar to label data generation, the size of sparse projection data and prior projection data are 512 × 512 input into PINet. Peak Signal to Noise Ratio (PSNR), Root Mean Square Error (RMSE) and Structure Similarity Index (SSIM) are used to evaluate the reconstructed images. In order to evaluate the performance of the proposed method, it is compared with SC-CNN [6].

Fig. 4 shows the results of projection data inpainting and images reconstruction by the PINet method and comparison method under sparse-view angle scanning. Then, we use the full angle projection data generated by the network to reconstruct the CT image. It can be observed that the image reconstructed by SC-CNN still has obvious artifacts, and the proposed method can effectively reduce the serious artifacts caused by the missing projection data. Furthermore, this study also compares PSNR, RMSE and SSIM of different methods as shown in Tabel 1. Compared with SC-CNN method, the PSNR of high-energy and low-energy images obtained by the proposed method is improved by 1.5413 dB and 1.2953 dB, and SSIM of proposed methods also has significant advantages. The RMSE of SC-CNN is higher than 0.0183, while the RMSE of proposed methods is lower than 0.0173. Numerical results show that the proposed method has some advantages in noise suppression and structure preservation.

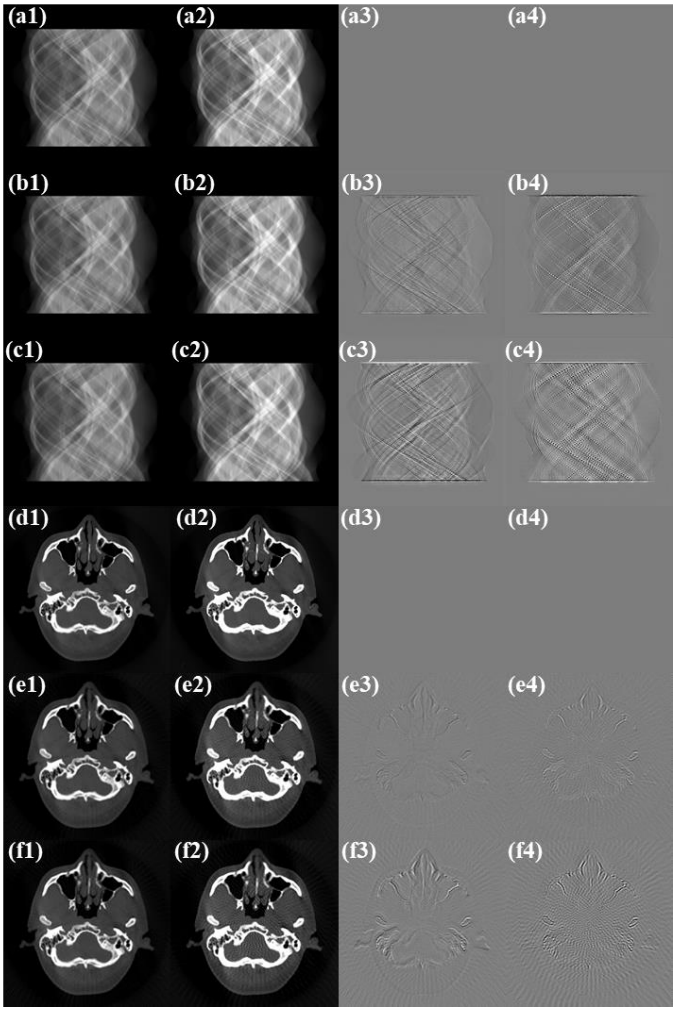

Figure 4: Results of inpainting projection data and reconstructed images from sparse-view angle scanning. (a) and (d) represent label projection data and reconstructed images. (b) and (e) represent projection data generated by PINet and reconstructed images. (c) and (f) represent projection data generated by SC-CNN and reconstructed images. (1) and (2) represent high-energy and low-energy, (3) and (4) represent corresponding error maps. Display windows of projection data, reconstructed images, error maps of projection data and error maps of reconstructed images are [0, 2.5], [0, 0.04], [-0.1, 0.1] and [-0.02, 0.02], respectively.

	avg. PSNR	avg. RMSE	avg. SSIM
PINet(H)	**36.2473**	**0.0154**	**0.9030**
PINet(L)	**35.2881**	**0.0172**	**0.8890**
SC-CNN(H)	34.7060	0.0184	0.8629
SC-CNN(L)	33.9928	0.0200	0.8535

Tabel 1: Quantitative results (PSNR: Peak Signal to Noise Ratio; RMSE: Root Mean Square Error; SSIM: Structure Similarity Index). Averaged over 100 test samples.

In order to verify the performance of the proposed method, the reconstructed images of PINet and SC-CNN are further decomposed to obtain the decomposition results of tissues and bone materials, as shown in Fig. 5. It can be seen that the basis material decomposed by PINet from the reconstructed image is closer to the ground truth, and the decomposition accuracy is higher.

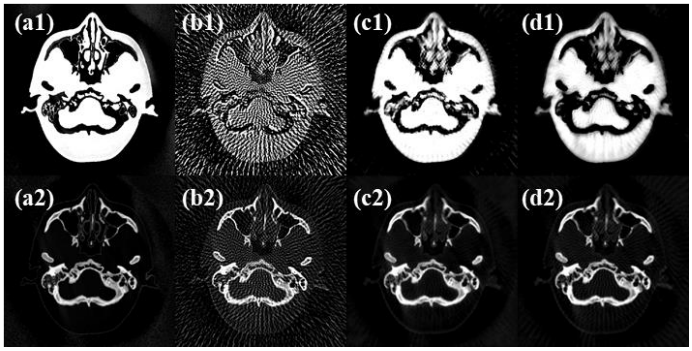

Figure 5: Decomposition results of different methods. (a), (b), (c) and (d) represent ground truth, raw sparse data, PINet and SC-CNN decomposition results, respectively. (1) and (2) represent tissue and bone materials. All display windows are [0, 1].

4 Discussion and Conclusion

DECT has great potential in medical field. To reduce the radiation dose and ensure the best hardware cost, this paper studies the kVp switching technology. In addition, aiming at the problem of missing projection data faced by this technology, this study uses the correlation between the projection data to fuse the sparse-view projection data under two different energies, and introduces the adversarial learning mechanism to propose a PINet framework with prior information. In the clinical data experiment, the feasibility of the proposed method is demonstrated. In the future work, we will extend this work to the application of spectral CT, and design corresponding image post-processing module and material decomposition module based on deep learning to achieve high-resolution spectral CT imaging.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 62101596), the National Key Research and Development Project of China (Grant No. 2020YFC1522002), the National Natural Science Foundation of China (Grant No. 62201616) and the China Postdoctoral Science Foundation (Grant No. 2019M663996).

References

[1] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher. “Dual-and multienergy ct: principles, technical approaches, and clinical applications”. *Radiology* 276.3 (2015), pp. 637–653. DOI: 10.1148/radiol.2015142631

[2] S. Leng, L. Yu, J.G. Fletcher, et al. “Maximizing iodine contrast-to-noise ratios in abdominal CT imaging through use of energy domain noise reduction and virtual monoenergetic dual-energy CT”. *Radiology* 276.2 (2015), pp. 562. DOI: 10.1148/radiol.2015140857.

[3] K. Li, Y. Li, Z. Qi, et al. “Quantitative lung perfusion blood volume using dual energy ct–based effective atomic number (zeff) imaging”. *Medical physics* 48.11 (2021), pp. 6658–6672. DOI: 10.1002/mp.15227.

[4] B. Ruzsics, H. Lee, P.L. Zwerner, et al. “Dual-energy CT of the heart for diagnosing coronary artery stenosis and myocardial ischemia-initial experience”. *Eur Radiol* 18.11 (2008), pp. 2414–2424. DOI: 10.1007/s00330-008-1022-x.

[5] H. Lee, J. Lee, S. Cho. “View-interpolation of sparsely sampled sinogram using convolutional neural network”. *Medical Imaging 2017: Image Processing* 10133 (2017), pp. 1013328. DOI: 10.1117/12.2254244.

[6] W. Cao, N. Shapira, A. Maidment, et al. “Hepatic dual-contrast CT imaging: slow triple kVp switching CT with CNN-based sinogram completion and material decomposition”. *Journal of Medical Imaging* 9.1 (2022), pp. 014003. DOI: 10.1117/1.JMI.9.1.014003.

[7] D. Kawahara, A. Saito, S. Ozawa, and Y . Nagata, “Image synthesis with deep convolutional generative adversarial networks for material decomposition in dual-energy ct from a kilovoltage ct”. *Computers in Biology and Medicine* 128 (2021), pp. 104111, 2021. DOI: 10.1016/j.compbiomed.2020.104111.

[8] Y. Z. Wang, A. L. Cai, N. N. Liang, et al. “One half-scan dual-energy CT imaging using the Dual-domain Dual-way Estimated Network (DoDa-Net) model”. *Quantitative Imaging in Medicine and Surgery* 12.1 (2022), pp. 653. DOI: 10.21037/qims-21-441.

[9] H. Y. Zhang, Y. X. Xing. “Reconstruction of limited-angle dual-energy CT using mutual learning and cross-estimation (MLCE)”. *Medical Imaging 2016: Physics of Medical Imaging* 2016. DOI: 10.1117/12.2211224.

[10] O. Ronneberger, P . Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. *Proc. Med. Image Comput. Comput.-Assist. Intervent. Springer* (2015), pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[11] R. L. Siddon. “Fast calculation of the exact radiological path for a three‐dimensional CT array”. *Medical physics* 12.2 (1985), pp. 252-255. DOI: 10.1118/1.595715.