Synthetic CT Generation from CBCT using MSG-GAN

Lu Bai[1], Chenyu[1], Chenqi[1], Shaobin Wang[1[0000]], Yi Du[2]

[1] MedMind Technology Co. Ltd., AB 1920 Techart Plaza, Beijing 100083, China
[2] Institute of Medical Technology, Peking University Health Science Center, Beijing 100191, China
wangshaobin@medicalmind.cn

Abstract. Since Cone-beam computed tomography (CBCT) requires shorter time and less exposure, it is used to obtain daily images for accurate dose calculations. However, CBCT images cannot provide accurate HU that describes the degree of x-ray attenuation by the tissue. Therefore, the gray level in CBCT should be converted to Hounsfield Unit (HU) in CT scan before dose calculation. This study is to synthesize CT image from CBCT input using a modified Multi-Scale Gradient Generative Adversarial Network (MSG-GAN) network. The results demonstrate that the distinction between soft tissue and bone structures in synthesized CT is close to that in the real CT. However, the MAE fluctuates from case to case. More attention should be paid to range of input CBCT image.

Keywords: Cone-beam computed tomography, CT Synthesis, Multi-Scale Gradient Generative Adversarial Network, .

1 Introduction

Cone-beam computed tomography (CBCT) uses a wild cone-beam of X-rays, therefore it can obtain the image more quickly with lower exposure [1]. It has been incorporated into radiotherapy devices to guide the patient positioning and to provide the up-to-date anatomical information in each radiation treatment. However, the quality of CBCT images suffer from X-ray scattering and truncated projections and the CBCT images cannot provide accurate HU. Therefore, CBCT cannot be directly used in dose calculation and treatment planning [1]. In order to obtain accurate dose calculation from CBCT, it is necessary to correct gray level in CBCT to HU value.

Motivated by the rapid development in deep learning in medical imaging, a variety of approaches to generated CT-like image with the same anatomical structure in CBCT have been proposed. Chen uses a deep U-net-based supervised approach to synthesize CT images. The averaged mean absolute error (MAE) between synthetic CT (sCT) and non-rigid registered real CT (rCT) on test data is reduced to 18.98 HU [1]. Thummerer compared the sCT of DCNN model to the corrected results given by deformable registration method and an analytical image-based correction method (AIC). Results demonstrated DCNN model gave the sCT with lowest MAE [2]. Chen proposed a unsupervised style-transfer-based method to generate synthetic CT from

CBCT images. The MAE of sCT reached 28.52 HU [3]. Xue tried CycleGAN, Pix2pix and UNet models to generate sCT. The average MAE was reduced to 15.4 HU for nasopharyngeal carcinoma patient. The 2D gamma index analysis were done on the different plans of sCT results [4]. Deng integrated the respath concept into cycleGAN to map CBCT to CT. The MAE reduced to 140.7 HU for head-and-neck (H&N) data and pelvic data [5]. There are many other GAN approaches such as GAN for unpaired data [6], conditional GAN [7].

GAN or cycleGAN is to optimize the generator network via confusing the discriminator instead of using supervised guidance of MAE. They cannot guarantee the structural consistency between synthetic and input images due to the lack of direct constraints between these two images. This study tries to use a modified Multi-Scale Gradient Generative Adversarial Network (MSG-GAN) [8] incorporating the AdaIN block used in StyleGAN [9] and convolutional operations to obtain sCT. The training strategy combines the advantages of MAE, MSE direct guidance as well as the discriminator loss to improve the quality of sCT. The training data are provided by SynthRAD 2023 challenge and no other data are used. The data contains 180 cases of brain data and 180 cases of pelvic data. The experiment results show that the MSG-GAN design can make the training more stable. The average MAE of sCT can reach 82.41 HU.

2 Method

Generative Adversarial Network (GANs) have gained huge successes in image synthesis task. Therefore, this study uses a GAN model to predict sCT. However, GAN is unstable during training and sensitive to hyperparameters [8]. The reason for these disadvantages is that gradients passing from the discriminator to generator become uninformative. To addressing the problem, MSG-GAN proposed to use the glow of gradients from the discriminator to the generator at multiple scales [8]. Inspired from MSG-GAN, a modified 2D CT synthesis generator with multi-scale outputs is adopted in this study. Its discriminator takes multi-scale images as input. Therefore, the multi-scale connection between the generator and discriminator are build to make the gradient more informative.

As shown in Fig. 1, the left encoder is revised from ResNet-50 model. It extracts the features at five different scale using the basic block with bottleneck operation and residual shortcut [10]. The decoder uses the similar basic block in encoder part. The decoder and encoder are connected like UNet [11]. The skip connection between encoder and decoder are modified by the AdaIN block which is proposed by the StyleGAN model [9]. The AdaIN block rescales the features from encoder part to presents the same mean and variation as those in decoder part. After the AdaIN block, the rescaled encoder feature are concatenated with the features from decoder part to do further processing. The decoder features at different scales give a respective sCT output.

Fig. 2 is the illustration of multi-scale discrimination model. It takes the sCT or rCT in different scales as input and combines the features from different scales to generate the feature vectors. The label result is obtained after the processing of three fully connected layers (FC).

The integration of multi-scale processing is to build more direct connection between generator and discrimination. Then the gradient passing from discriminator to generator would be more informative and the training would became more stable. The integration of AdaIN block is to reduce the gap between the encoder features and decoder features.

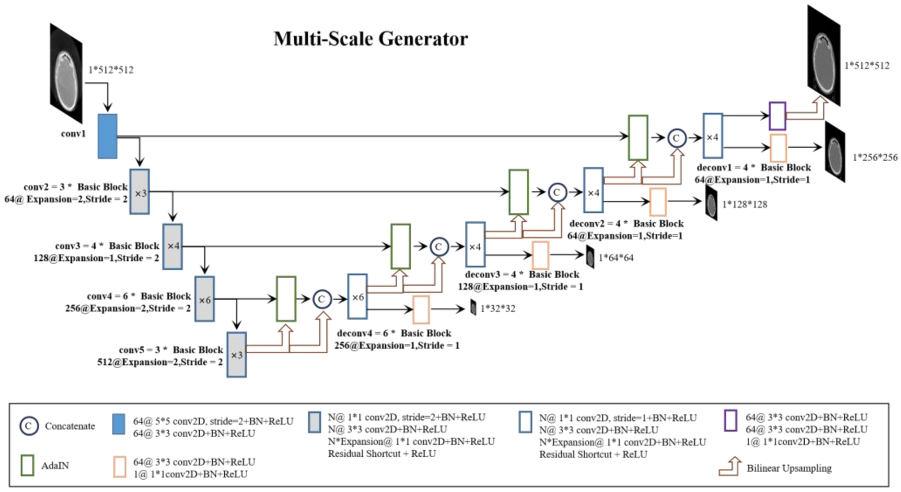


Fig. 1. Illustration of architectures of generator network in our MSG-GAN.

The training processing of the proposed MSG-GAN is shown in Fig. 3. The gradient flow is divided into two step. The first step is to optimize the parameters in discriminator network to improve the label prediction accuracy. The second step is to update parameters in generator to give predictions similar to rCT. The optimization objects in two steps are different.

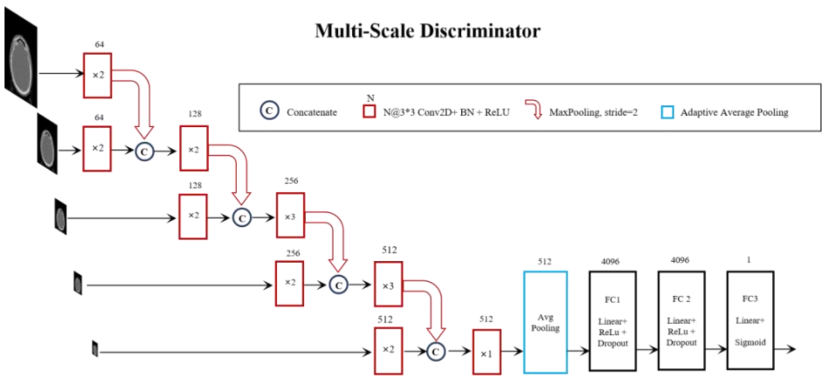


Fig. 2. Illustration of architectures of discriminator network in our MSG-GAN.

The discriminator is to separate the sCT and rCT input correctly. For input i, the real label is y_i and the predict label is $\hat{y}_i$ .Therefore, the objective of first optimization step is defined as the Binary Cross Entropy Loss (BCE):

$$\mathrm{L1}(y,\hat{y}) = BCE(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot \log \hat{y}_i + (1-y_i)\cdot \log(1-\hat{y}_i)\right) \tag{1}$$
$$\hat{y}_i = \mathrm{Discriminator}(sCT, sCT_1, sCT_2, sCT_3, sCT_4)$$

The generator is to generate sCT that are so close to rCT to confuse the discriminator. The objective of generation optimization is defined as a combination of four parts:

$$\mathrm{L2} = \mathrm{MAE}(sCT, rCT) + \alpha \cdot \mathrm{MSE}(sCT, rCT) + \beta \sum_{i=1}^{4} \mathrm{MAE}(sCT_i, rCT_i) - \gamma \cdot \mathrm{BCE}(y, \hat{y}_i) \tag{2}$$

The first two parts is to constrain the sCT in original scale directly. The third part is to constrain the sCT in different scales. The final part is to confuse the discriminator. Parameter α, β and γ are real number to control the contribution of each part to the total loss. In this study, α is set to 0.002, β is set as 0.125 and γ is set to 20. The MAE and MSE are calculated within the valid area given by the mask.

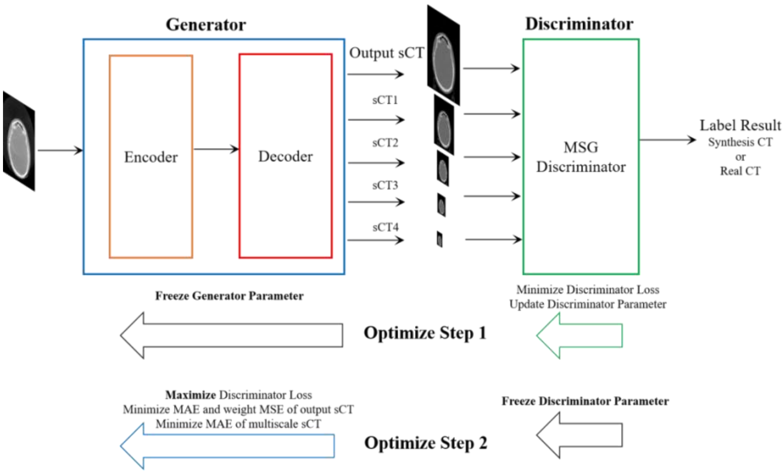


Fig. 3. Illustration of training process of our MSG-GAN.

3 Experiments

The experiment data are provided by SynthRAD 2023 challenge committee. The training data consist of 360 cases of CBCT, rigid registered CT images and mask information. Half the training data cover the brain part. The left data are from the pelvic part. The validation data contain 60 cases of paired data. They are from the brain and the pelvic part.

The size of input CBCT data are different. During training processing, the slices from each case are resampled to 512 * 512. In order to make good use of the data, the input CBCT is randomly rotated and randomly applied translation in small range when the training data are loaded. The registered rCT and mask are augmented in the same way. The training computer is equipped with two GPU (NVIDIA RTX A6000 with 48G) and a CPU (Intel Core i9-10900X CPU@3.70GHz). Its operating system is Ubuntu 18.04.1 LTS.

The algorithm is implemented in Python 3.7 environment. The training required approximately 120 minutes per epoch and the batch size is 12. The initial learning rate is set to 0.0006 and Adam optimizer is used. The total 106 epoch cost about 212 hours. The first 46 epoch is to pre-train the model using small learning rate 0.0001. The left 60 epoch is the formal training process. As shown in Fig. 4, the changes in train MAE and validate MAE are very similar. Therefore, there no obvious occurrence of overfitting or underfitting. The training process is sensitive to the learning rate. That is why the MAE does not decrease in the same pace.

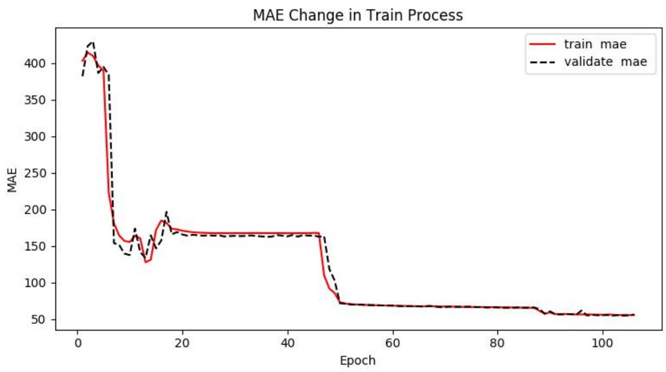
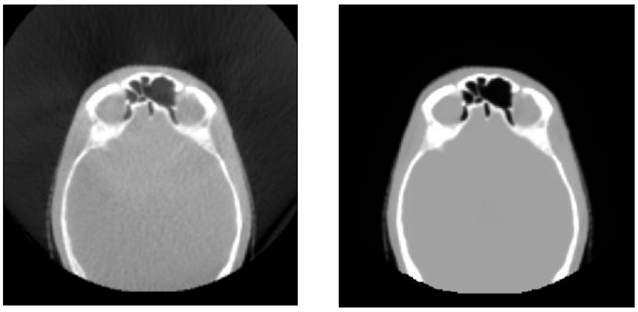


Fig. 4. Train MAE and Validate MAE Change During Training Process.

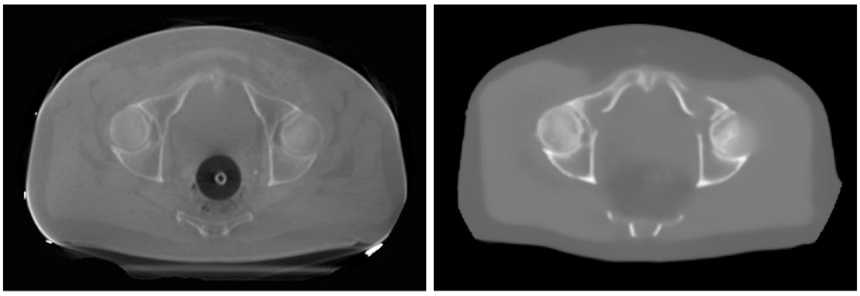
The generator gives five synthesis CT in different scale but only the sCT with the largest size are needed. The sCT should be resampled to original shape. The HU value of sCT in valid area shown in the mask would be kept while the others would be set to invalid value -1000. The final sCT results of the validation data are shown in Fig. 5 and Fig. 6. Fig. 5 is an example of sCT results of brain part. As shown in the result, the noises in the center part of the CBCT is suppressed and there are not obvious fluctuation in sCT result. Fig. 6 gives the synthesis CT result of pelvic part. Compared to the CBCT image, the sCT fails to keep the fine structures of the bone and soft tissue. The pelvic result needs further optimization.

The metric statistics of total 60 validation cases are presented in Fig. 6. The box chart of MAE, PSNR and SSIM metrics are analyzed. Seen from the chart, PSNR and SSIM result are more stable in different validation cases. However, the MAE result changes from 42.97 HU to 172.81 HU. The variation of the MAE result is large. Further efforts should be done to reduce the variation of MAE.



(a) Input CBCT (b) Output sCT

Fig. 5. Synthesis CT Result for Brain Part (2BA016).



(a) Input CBCT (b) Output sCT

Fig. 6. Synthesis CT Result for Pelvic Part.

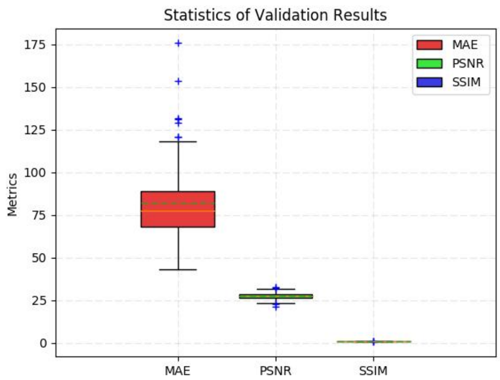


Fig. 7. Box Chart of Statistics of Validation Results.

4 Discussion

A MSG-GAN approach is proposed to synthesis CT like image from CBCT image. As demonstrated by the analysis of validation results, the average MAE is about 82.41 HU. MAE fluctuates a lot. After the analysis of the training process, we found that the training CBCT have different distributions. Some of them have the minimum value 0 and the left of them have minimum value -1000. In our approach, the input CBCT are truncated to range [0, 2048]. That is why the performance of results changes a lot. Since the regression training is very sensitive to the distribution of input and output, more attention should be paid to check the input type.

Although the metric of sCT results are not good as expect, the training process is stable. There are no obvious fluctuation of MAE and training loss in the training stage.The MSG design used in generation and discriminator works well to make the training more stable.

References

1. Chen L, Liang X, Shen C, Jiang S, Wang J.: Synthetic CT generation from CBCT images via deep learning. Med Phys **47**(3), 1115-1125(2020).
2. Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, Langendijk JA, Both S, Spadea MF, Knopf AC.: Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol. 65(9), 095002(2020).
3. Liyuan Chen, Xiao Liang, Chenyang Shen, Dan Nguyen, Steve Jiang, Jing Wang.: Synthetic CT generation from CBCT images via unsupervised deep learning. Phys. Med. Biol. 66, 115019(2021)
4. Xue X, Ding Y, Shi J, Hao X, Li X, Li D, Wu Y, An H, Jiang M, Wei W, Wang X.: Cone Beam CT (CBCT) Based Synthetic CT Generation Using Deep Learning Methods for Dose Calculation of Nasopharyngeal Carcinoma Radiotherapy. Technol Cancer Res Treat 20, 15330338211062415(2021)
5. Liwei Deng, Jie Hu, Jing Wang, Sijuan Huang, Xin Yang. : Synthetic CT generation based on CBCT using respath-cycleGAN. Med Phys 49(8), 5317-5329(2022)
6. Wolterink, J.M., Dinkla, A.M., Savenije, M.H.F., Seevinck, P.R., van den Berg, C.A.T., Išgum, I.: Deep MR to CT Synthesis Using Unpaired Data. Simulation and Synthesis in Medical Imaging. SASHIMI 10557, 14-23 (2017)
7. Zhao B, Cheng T, Zhang X, Wang J, Zhu H, Zhao R, Li D, Zhang Z, Yu G.: CT synthesis from MR in the pelvic area using Residual Transformer Conditional GAN. Comput Med Imaging Graph 103, 102150(2023)
8. A. Karnewar, O. Wang.: MSG-GAN: Multi-Scale Gradients for Generative Adversarial Networks. Conference on Computer Vision and Pattern Recognition (CVPR), 796-7805 (2020).
9. T. Karras, S. Laine, T. Aila.: A Style-Based Generator Architecture for Generative Adversarial Networks. Conference on Computer Vision and Pattern Recognition (CVPR), 4396-4405(2019)
10. K. He, X. Zhang, S. Ren and J. Sun.: Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition, 770-778(2016)

11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention, 9315(2015)