Tackling the Synthetic Computerized Tomography Problem

Yorick Chern[1]

[1]Bioengineering, University of California, Berkeley.

Contributing authors: yorick_c@berkeley.edu;

Abstract

The synthetic computerized tomography (sCT) problem describes a general goal of synthesizing CT scans with some type of input, whether it be magnetic resonance images (MRI) or other types of medical images. In this paper, I attempted to create and train a model that takes an MR image as an input and outputs the corresponding CT scan. This project will also focus on the brain. Two types of models were proposed, the naive U-Net and the pix2pix General Adversarial Network (GAN). In the end, the GAN model produced much clearer synthetic CT's than the naive U-Net. Note: this report was written using the Springer Nature report's template.

1 Introduction

In medicine, it is very standard practice for patients, healthy or ill, to receive periodic magnetic resonance imaging (MRI) checkups. Prior to MRI's, X-Rays have always been the primary diagnostic and treatment tool for oncological patients. Then, with computerized tomography (CT) being developed, CT scans have become the primary imaging modality in radiotherapy (RT) for cancerous patients. Radiologists would refer to CT scans to calculate dose delivery and determine patient organ geometry with the goal of maximizing treatment efficiency. When the MRI technology is developed, physicians began to refer to MR images due to its superb soft-tissue contrast. In modern radiotherapy for patients undergoing cancer treatments, MRI's are usually registered to CT's to benefit from the various advantages each imaging technology has to offer. However, this process has shown to increase workload, increase patients' diagnostic-to-treatment time, and increased hospital bills. Moreover, for younger patients, the extra exposure to ionization and radiation may cause more harm than

good. Thus, the development of an "MR-only RT" will benefit both the providers and the patients in multiple ways: reduced treatment costs, increased efficiency for patients and physicians, and decreased exposure of radiation to fragile patient populations such as children.

With the increase in data availability through the common use of electronic health systems by hospitals around the world and the rapid advancements in artificial intelligence (AI) along with computing power, the sCT problem has began to garner interest in the field of medicine and computer science. More specifically, with new generative AI's being developed every other day, researchers have been optimistic and creative with the use of these generative AI's to tackle the sCT problem along with other medical imaging problems.

One of the most successful and fundamental image generation algorithm was the creation of the U-Net [1]. This model has shown great promise in segmentation and bounding box generation and is one of the earlier image generation methods utilizing convolutional neural networks. The second proposed method was using the pix2pix GAN [2], which uses U-Nets as the backbone of the image generation mechanism but involves many other logical steps such as training a discriminator network that helps produce sharper images.

2 Results

Setting the U-Net as the baseline model, we will compare the performance of the pix2pix GAN and against that of the basic U-Net model.

2.1 U-Net

For our U-Net, we introduce a very light-weight U-Net with only kernels of the dimensions $(H, W, C) = (3, 3, 256)$ or less. Here are some images generated by this model (fig. 1-3).

It is quite apparent that the U-Net is capable of capturing the overall shape and structure of the MRI input, but it is unable to produce clear and distinct edges. This is a common symptom when using the mean squared error loss function in image synthesis algorithms.

2.2 pix2pix GAN

Now, we will move on to the pix2pix GAN. This algorithm, like other GAN's, includes a generator network and a discriminator network. It does not use mean squared error as the metric, but it is visually much better than that of the images generated by the U-Net. Here are some examples (fig 4-6):

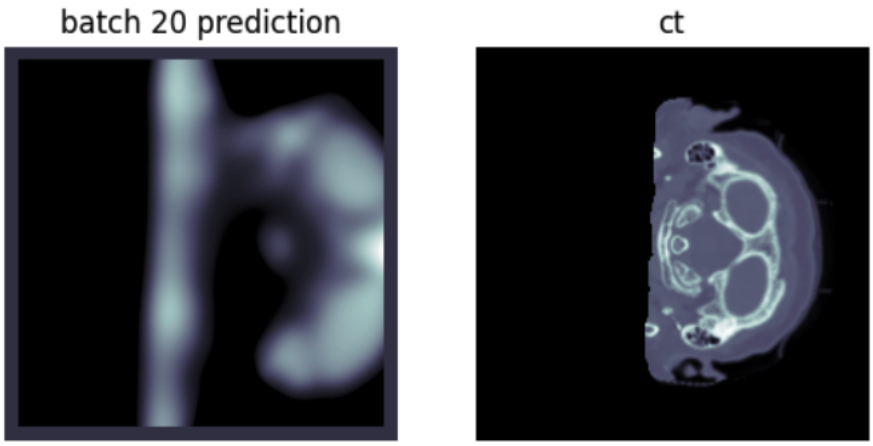

Fig. 1 The image on the left is generated by the U-Net upon the 20th batch. The image on the right is the ground truth CT scan. The mean-squared-error (MSE) between these two images is 0.79.

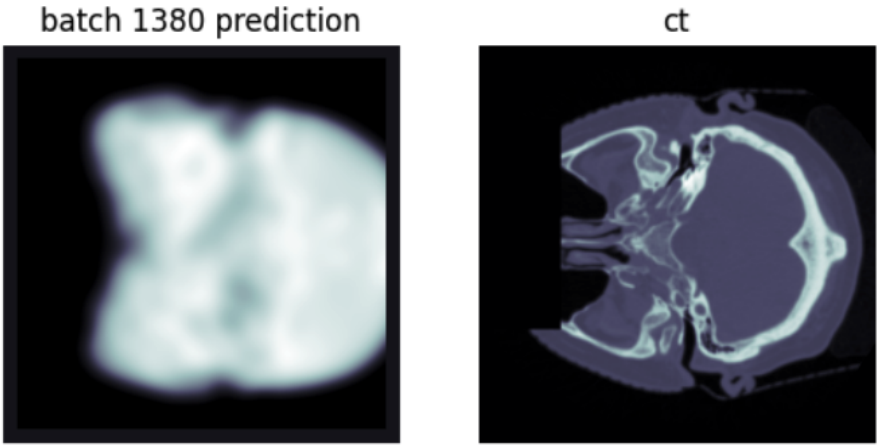

Fig. 2 The image on the left is generated by the U-Net after training through 1380 batches. The MSE between these two images is 0.026.

3 Methods

3.1 Data processing

The data was obtained by SynthRAD2023 [3]. The dataset included MR images, ground-truth CT scans, and the mask of the MR image. Each set of images came from a patient, and the facial structure of the images has been masked out to comply with confidential reasons. A script in PyTorch was created in order to load the dataset efficiently into GPU memory. Reshape and recrop in the Dataset Class was enforced for the U-Net models and the pix2pix GAN to run smoothly.

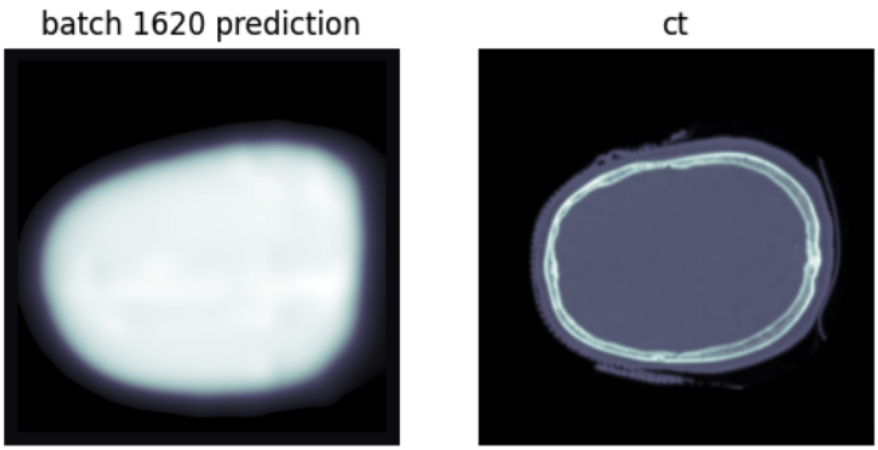

Fig. 3 The image on the left is generated by the U-Net after training through 1620 batches. The MSE between these two images is 0.016. Although it has a lower MSE than the previous image in figure 2, we can see that the CT scan is much less convoluted than that of figure 2.

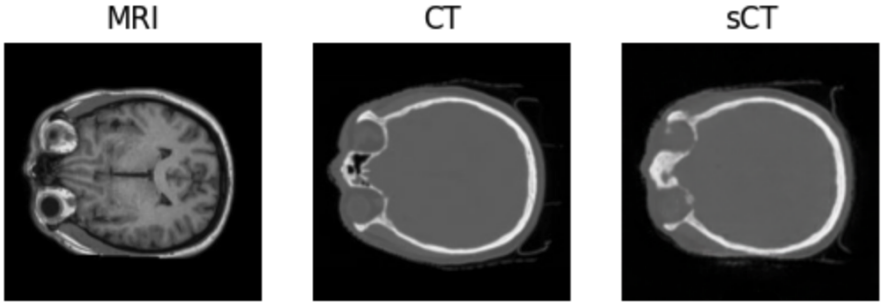

Fig. 4 The image on the left is generated by the U-Net upon the 20th batch. The image on the right is the ground truth CT scan. The mean-squared-error (MSE) between these two images is 0.79.

3.2 U-Net

The U-Net tested in this project is a very lightweight model that was created in-house using PyTorch. It contained convolutional layers with dimensions no larger than $(H, W, C) = (3, 3, 256)$. No regularization techniques were applied and the loss function was simply the mean-squared-error function comparing the generated image and the ground-truth CT image. This model was trained with less than 1 epoch but with $1680 * 10 = 16,800$ images (batch size of 10 and 1680 batches were trained). The AdamW optimizer was used with a learning rate of $1e-4$. The model was trained for 3 hours with a single Nvidia GPU on Google Colab. In the end, training was forced to stop because the model plateaud at an MSE of 0.0155.

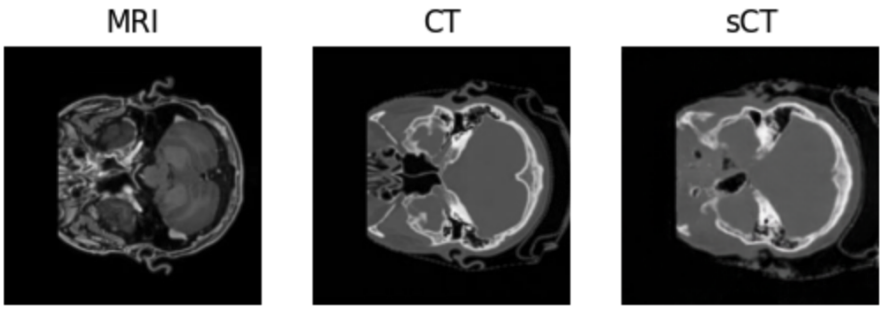

Fig. 5

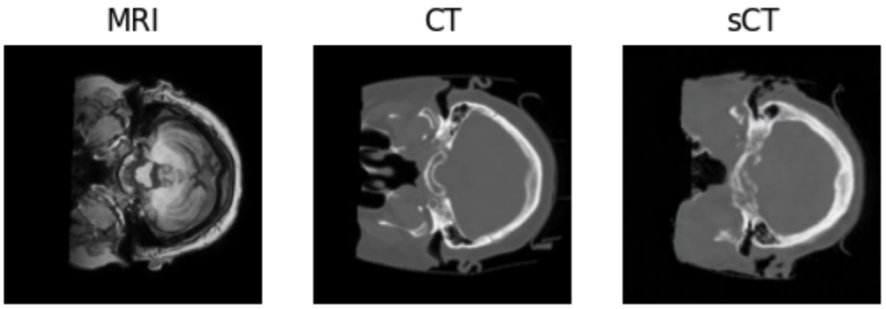

Fig. 6 Even with more complex internal organ structures, we can still see the GAN model producing clear edges and capturing the structures with fine details.

3.3 pix2pix GAN

According to the authors, pix2pix GAN works most effectively when the specified input and the output have high similarity in visual contents. In this case, the input being an MR image and the output being a CT scan is the perfect pair. Both imaging modalities are volumetric, and the same location of the same organ will have similar structures and outlines in an MR image and a CT scan, as shown in Fig. 7. The pix2pix GAN used in this project was not created in-house. Instead, the architecture was imported from the official Github of the paper [2]. After the data has been processed using the preprocessing script mentioned above, the data was further tailored to fit the format of the pix2pix GAN requirement. The model that produced the images shown in fig. 4-6 were trained for 200 epochs on Google Colab with 1 Nvidia GPU. The training set consisted of only 1,000 MRI/CT pairs. The generator network consisted of 54.414 million parameters and the discriminator network consisted of 2.769 million parameters. The total training time took 3 hours. The AdamW optimizer was used with a batch size of 1. Additionally, the default learning rate scheduler was applied during training.

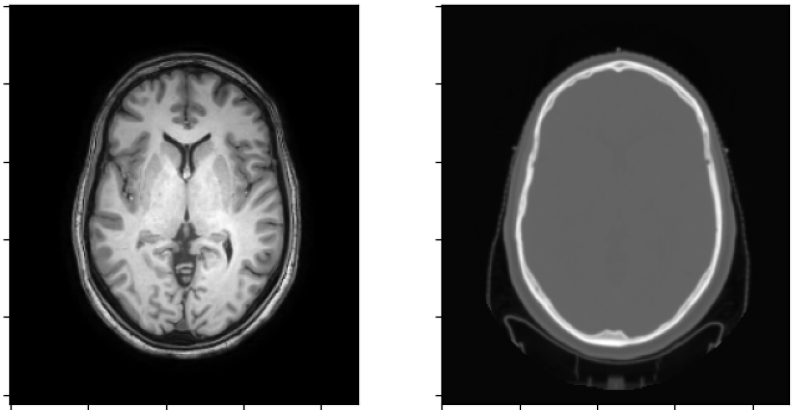

Fig. 7 On the left is the input, the MR image, and on the right is the ground-truth, the CT.

4 Discussion

With the dataset providing over 30,000 MRI/CT pairs, the success of training pix2pix GAN on only 1,000 images shows high potential. With more computational resources and larger training size, it is no doubt that pix2pix GAN would synthesize even more realistic and accurate CT scans. For future implementation, I would like to also try to tackle this problem using diffusion models. The most state-of-the-art diffusion models, such as stable diffusion [4] has been able to produce hyper-realistic images given a prompt (text-to-image synthesis). It is, then, not hard to imagine a diffusion model generating highly realistic and accurate CT scans based on an MRI input. Although the pix2pix GAN was able to produce images that are visually similar to the CT scans, there remain several limitations and concerns of the problem. Firstly, pathology preservation needs to be ensured. For example, a fracture in a skull shown in the MRI needs to be reflected in the CT scan synthesized by the model. The pathology lost in translation poses a great threat to the overall health of the patient and may result in poor diagnoses and treatment. Adding on to the previous point, this is a diagnostic AI, meaning that the AI's performance is one of the factors that will decide a patient's health outcome, and so the evaluation and training of said AI needs to be handled with much care and caution.

5 Conclusion

In this paper, we have demonstrated the success of tackling the synthetic CT using pix2pix GAN. With a training size of 1,000 MR images, the model was able to generate images that are highly similar to the ground-truth images, and this shows a high potential of tackling sCT and other medical image generations using pix2pix GAN or algorithms alike.

References

[1] Olaf Ronneberger, T.B. Philipp Fischer: U-net: Convolutional networks for

biomedical image segmentation. arXiv (2015)

[2] Philip Isola, A.A.E.: Image-to-image translation with conditional adversarial networks. Computer Vision Foundation (2016)

[3] SynthRAD2023: Synthesizing computed tomography for radiotherapy. https://synthrad2023.grand-challenge.org/SynthRAD2023/ (2023)

[4] Robin Rombach, B.O.: High-Resolution Image Synthesis with Latent Diffusion Models, (2022)