Synthetic CT generation from PET images for attenuation correction with Deep Learning methods.

CS766 project mid-term report

Team members:
Xue Li (NetID: xli2245)
Ni Li (NetID: nli72)

1. Introduction and Motivation

Positron emission tomography (PET) and positron emission mammography (PEM) using 18F-fluorodeoxyglucose (FDG) have been shown to have high specificity for breast cancer detection. However, breast-specific PET is not routinely utilized; instead, magnetic resonance imaging (MRI) is used clinically for extent of disease evaluation and neoadjuvant therapy response assessment for patients with newly diagnosed breast cancer. The recent availability of simultaneous PET/MR scanners offers the benefit of combining the high sensitivity for the detection of breast cancer provided by contrast-enhanced MR imaging with functional imaging using PET.

However, there are two problems for MR images generated by clinical PET/MR system, the missing bone information and back truncation problem as shown in Fig 1. The former one is due to the technical challenges of obtaining positive MR signal in bone in clinical PET/MR system while the latter one is caused by position patient are placed during imaging to get better breast images. Both of two problems will lead to errors during breast cancer treatment, which is not desired. In current pipeline, MR images are used to generate synthetic CT (sCT) images for later attenuation correction of PET images. Therefore, the missing bone and back truncation problems will happen to sCT as well. As a result, inaccurate attenuation correction of PET images is calculated, and there is up to 20% error for the final treatment plan.

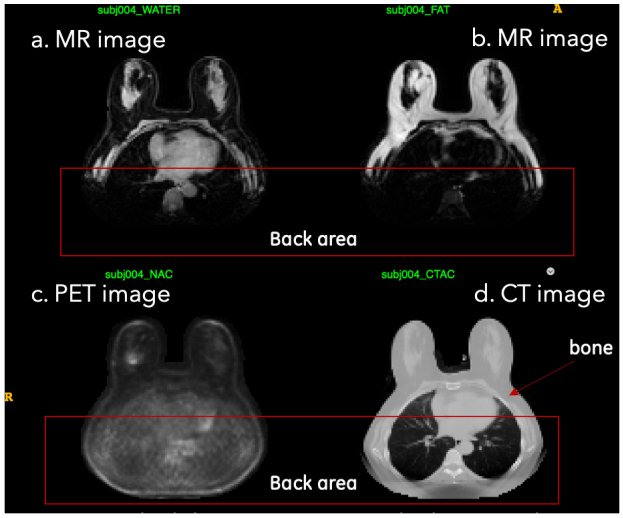


Figure 1. Comparison of MR images, PET image and CT image of the breast part. a) and b) are two MR images of water and fat components. Apparently the MR images have problems of missing bone and back truncation.

As the pipeline shows, the key to solve this problem is to get accurate sCT images with bone and back information. Recent advances in deep learning (DL) have demonstrated great success in accomplishing image processing tasks, like image transfer, converting images from one domain to anther domain. These methods have been applied to medical imaging successfully, like synthetic CT generation from MR images. Therefore, we want to train a DL model with breast

dataset to get accurate sCT images directly from PET images which don't have such problem mentioned.

2. Method and Related Works

Recently, a number of studies have used deep learning and convolutional neural networks to create sCT images[1-3]. Among the many kinds of convolutional neural networks, the Unet[4] has shown outstanding performance in medical image segmentation and synthesis. More recently, generative adversarial networks (GANs)[5] have become popular in creating realistic synthetic images. sCT images of pelvic, liver, brain, and head and neck regions have been produced by GANs and their variants[6-10]. However, almost all the studies got sCT from MR images, and very few studies show the transfer from PET images to sCT images [11]. Specifically, no research has been done to get sCT images from PET images in the breast tumor region which requires higher accuracy.

In this project, we will try to use Unet and UnetR[12] model to synthesis CT images directly from PET images. The sCT images should be accurate enough to have lower error than sCT generated from current pipeline. The official code for UnetR is pytorch, and it is used for medical image segmentation. We will play with it first, and then try to modify it to complete the PET-to-CT image generation task.

3. Change of Plan

Initially, we were going to re-implement a state-of-art model called UNETR proposed recently for 3D medical image segmentation[12]. However, after getting the feedback from TA and professor, we think we are not very clear about the project purpose and requirements. To figure that out, we had a talk with TA, and got some new ideas. Therefore, instead of just playing with the existing code and dataset, we want to try something new and interesting. As a result, we decide to synthesize CT images directly from PET images in the breast area with deep learning models to improve the accuracy for attenuation correction required for breast tumor treatment. We plan to try two models, classic UNet and UNETR. The former has shown outstanding performance in medical image segmentation and synthesis while the latter is a combination of UNet and Transformer which incorporates the capture of global context into local features.

Modifications to specific decoding layers and loss functions are necessary to adapt to the image synthesis task. The first step of our current plan is to get familiar with the Unet and UNETR models by training the models on segmentation dataset and compare their performance to confirm the improvement of UNETR on segmentation task. After that, we make modifications to the initial Unet and UNETR models and apply them to the task of CT images synthesis.

4. Evaluation and Comparison

4.1. Evaluation of models and sCT

Eight-fold-cross-validation was performed to evaluate model performance for both the UNet and UNETR models. The 23 patients were randomly divided into eight groups, seven groups of three patients and one group of two cases. For each cross-validation fold, seven groups were used for training and validation, and one group was left for testing. Finally, the synthesized sCT images were generated case by case and compared with the original CT images. Three measures were utilized to evaluate the difference and similarity between CT and sCT images: mean absolute error (MAE), peak-signal-to-noise-ratio (PSNR), and normalized cross-correlation (NCC). The formulas are shown as follows.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |sCT_i - CT_i| \tag{2}$$

$$PSNR = 20\log_{10}\frac{max(CT)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(sCT_i - CT_i)^2}} \tag{3}$$

$$NCC = \frac{\sum_{i=1}^{N}(CT_i - \frac{1}{N}\sum_{j=1}^{N} CT_j)(CT_i - \frac{1}{N}\sum_{j=1}^{N} CT_j)}{\sqrt{(\sum_{i=1}^{N}(CT_i - frac1N\sum_{j=1}^{N} CT_j)^2)(\sum_{i=1}^{N}(sCT_i - frac1N\sum_{j=1}^{N} sCT_j)^2)}} \tag{4}$$

4.2. Attenuation correction and Evaluation

After getting the sCT images, PET/CT based reconstruction will be run to get reconstructed PET. There are four kinds of them in total: one from CTAC, the ground truth, one from sCT generated by MR images in current pipeline, one from sCT generated by DL UNet model, and one from sCT generated by DL UNETR model. Once the reconstructed PET images are got, the treatment dose for each patient in each tumor area can be calculated. There are mainly tumor areas, including Breast, FGT, Liver, Blood, and LN. The average of absolute dose calculation percent error for each type of tumor will be used to evaluate each model. N is the number of patients.

$$percent\ error_{tumor} = \frac{1}{N}\sum_{i=1}^{N} |\frac{(dosecal.fromsCT_i - dosecal.fromCT_i)}{(dosecal.fromCT_i)}| \tag{5}$$

5. Progress

5.1 Implementing nnUNet for segmentation of the BTCV dataset

BTCV dataset is the dataset used by the authors of UNETR. This step is to get familiar with nnUNet [13] and UNETR models, make sure the UNETR works functionally, and validate the improvement of UNETR model on the segmentation performance.

5.1.1 BTCV dataset

The BTCV training dataset[14] contains 30 abdomen CT scans, and each scan is a 3D image file with only one image channel. 13 organs were manually labeled for segmentation training, and it's worth noting that the right kidney or gallbladder might be missing for some scans.
The 30 scans have various volume sizes (85 x 512 x 512 - 198 x 512 x 512), various in-plane spacings (0.54mm x 0.54mm - 0.98mm x 0.98mm) and slice spacings (2.5mm - 5mm).

5.1.2 Training of nnUNet model

nnUNet was selected to be compared with UNETR not only because it's one of the state-of-art model based on conventional U-Net model, but also because it automatically generate the configuration for any arbitrary dataset, which saves time for trial and test of various parameters. We performed training of both 2D U-Net configuration and 3D U-Net configuration. The spacings after automatic preprocessing for 2D U-Net training is 3mm × 0.76mm × 0.76mm while for 3D U-Net training is 3.22mm × 1.61mm × 1.61mm.

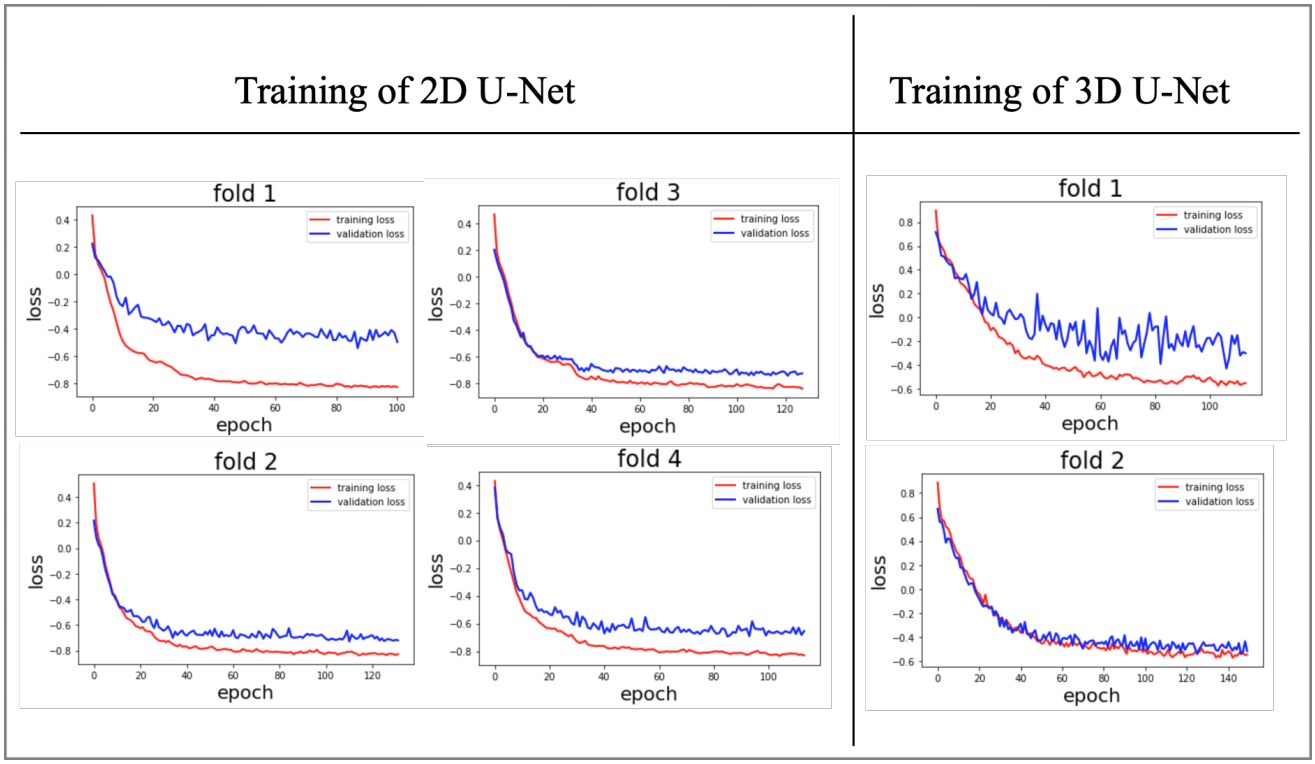


Figure 2. Training process of 2D an 3D U-Net models. Loss values of training and validation datasets are plotted as a function of epochs.

The training dataset was divided into 5 folds for cross-validation, where each fold contains 24 files for training and 6 files for validation. The training was performed on Google Colab with a GPU, and each fold was trained independently. One epoch of each fold takes ~10min for 3D configuration and ~4min for 2D configuration. As shown in Figure 1, the loss for both training and validation dataset kind of converges after 100 epochs, so we stopped training after ~100 epochs to save time.

Up to now, we performed training of 2D U-Net on 4 folds of data and 3D U-Net on 2 folds of data. The soft Dice score (defined in our proposal) is used as our evaluation metric, and was computed and recorded for each epoch. The evolution of evaluaction metric as a function of epoch is plotted in Figure 2. Except for fold 1, the dice score reaches about 0.8 for both 2D and 3D U-Net configurations. The result is consistent with what Hatamizadeh, Ali, *et al.* [12] obtained.

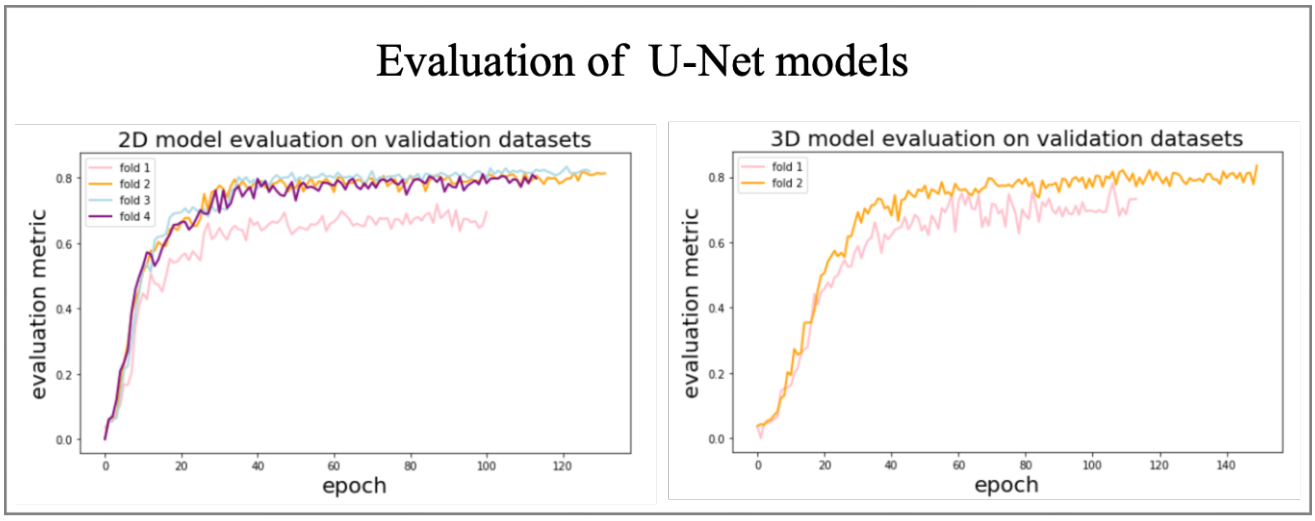


Figure 3. Evaluation of U-Net models by plotting the Dice score as a function as epochs for different folds of validation data. The evaluation metric on the y axis represent the Dice score.

5.2. Image systhesis

5.2.1 Breast Dataset

The breast dataset contains 23 subjects in total, including CT images and PET images. CT images have the size of 192x192x47 with spacing information (3.646mm, 3.646mm, 3.27mm) while PET images have the size of 192x192x89 with spacing information (3.125mm, 3.125mm, 3.78mm).

5.2.2 Data process

The noise of the CT images and PET images were removed firstly. Then, PET images were registered to CT images to have the same data size and spacing information. The HU value of CT images is cropped into the range of [-1000, 2000] according to the meaning of HU value. Both CT and PET images are normalized. The normalization for CT is shown as follows:

$$x_{new} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \tag{1}$$

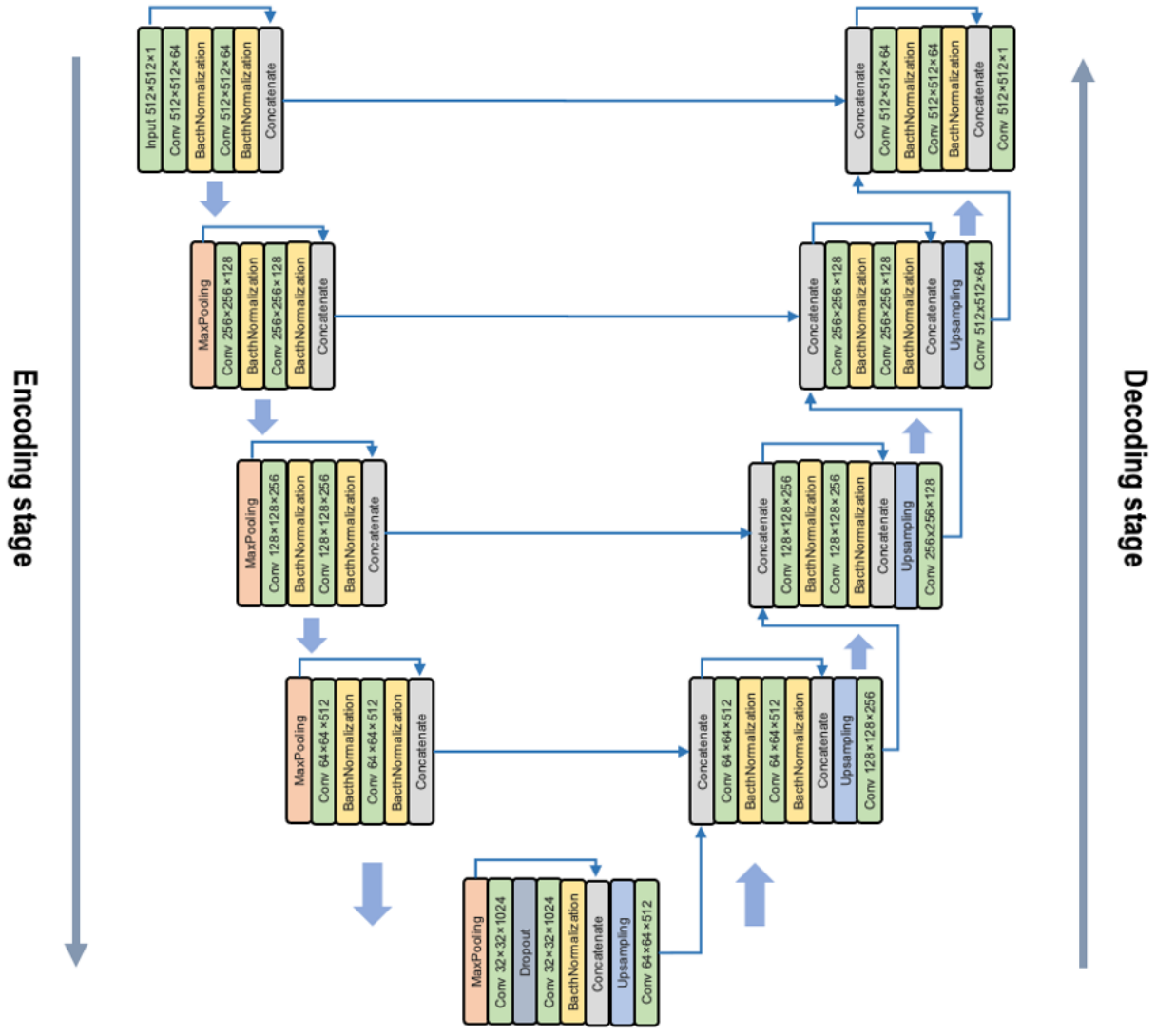


Figure 4. model of U-Net based model for sCT image generation.

The normalization for PET is simple. Just divide by 1500. Since not all the CT images have perfect back images, we cropped the imperfect back area for both CT and PET images.

5.2.3 Model

A 4-layer 3D UNet model was used for the training. The input is PET images with cube size (64x64x32), and the ground truth is the corresponding CT images with same cube size. The network

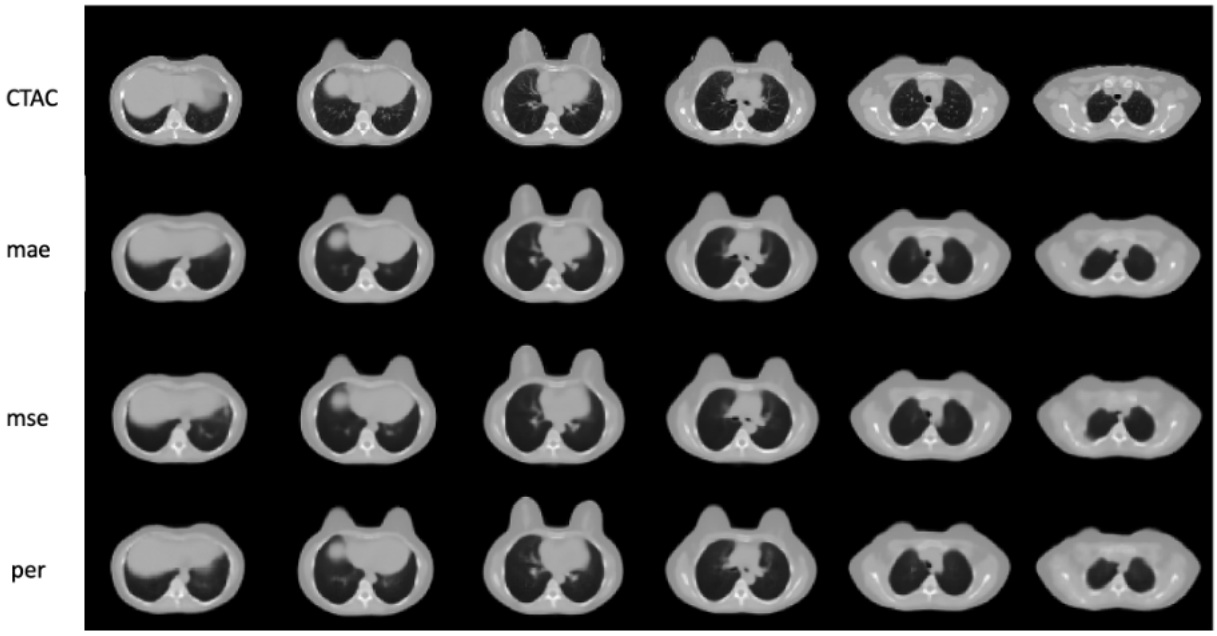
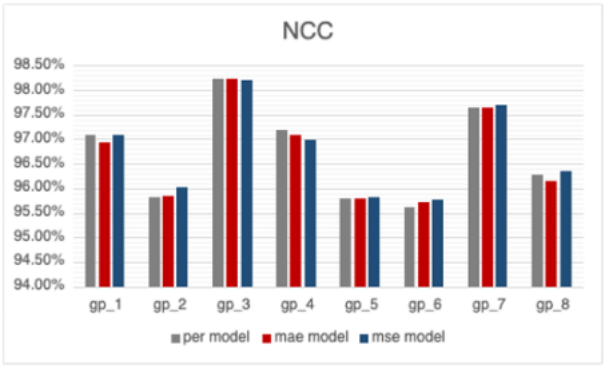


Figure 5. Generated sCT images by different models.

Cross validation

	Per model	Mae model	Mse model
mae	16.54	16.24	16.60
psnr	22.525	22.525	22.624
ncc	96.71%	96.67%	96.75%



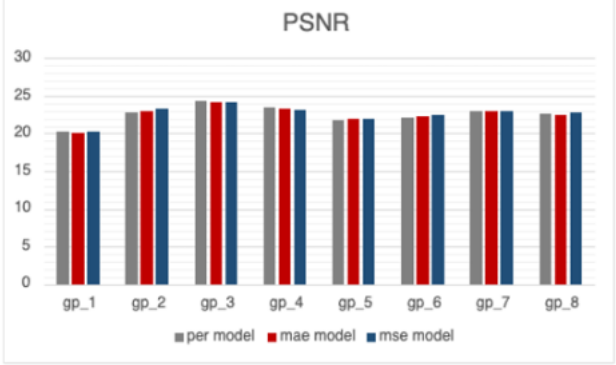


Figure 6. Cross validation results of various models.

mean	mae	mse	per	petmr
SUV mean Liver	6.46%	6.92%	7.71%	8.98%
SUV max Liver	5.64%	5.74%	6.87%	7.93%
SUV min Liver	13.43%	14.38%	14.87%	13.87%
SUV mean Blood	13.82%	14.52%	15.23%	17.63%
SUV max Blood	14.02%	14.53%	14.93%	16.91%
SUV min Blood	13.86%	14.73%	15.97%	18.62%
SUV mean FGT	3.73%	3.90%	4.45%	3.73%
SUV max FGT	3.80%	4.01%	4.38%	3.63%
SUV min FGT	3.92%	4.29%	4.58%	3.84%
SUV mean BR	5.21%	5.47%	5.79%	4.96%
SUV max BR	5.15%	5.57%	5.59%	4.80%
SUV min BR	4.61%	5.40%	5.29%	4.80%
SUV mean LN	4.91%	4.86%	6.08%	6.40%
SUV max LN	5.04%	4.52%	6.35%	6.05%
SUV min Ln	6.99%	7.31%	8.15%	6.66%

Table 1. Absolute percent error of each tumor.

structure is shown in Fig 4.

5.2.4 results

We tried UNet with three kinds of loss functions, including mae, mse, perceptual loss. The sCT images generated by each model is shown as Fig. 5. All those sCT images look good, and the UNet with mae loss model produces better tissue contrast than the other two models.
The eight-fold cross validation results are shown in Fig. 6. No obvious difference were found between them. The absolute percent error for each tumor is shown as Table. 1.
UNet with mae model out-performs the current pipeline (petmr) in the Liver and Blood area, but for the breast and FGT area, current pipeline produced better results.

Difficulties and Problems

1. New to Pytorch

Both team members are not familiar with PyTorch, but the codes provided by the authors are in PyTorch. Although one member is familiar with Keras, it takes time to understand how does PyTorch work and do coding with PyTorch.

2. Limited computing speed

We have no access to fast servers. The RAM of GPU that we have access to is very limited and the training time is very long compared to expected.

Milestones

Timeline	
02/24	Submit proposal
4/5	Submit midterm report
4/12	Finish the development of UNETR model for image synthesis
4/19	Finish the Cross-Validation and Evaluation of UNETR model
04/25	Finish final webpage, collect codes, presentation preparation
If extra time	If we have extra time, we'll try to modify swing U-Net model for image Synthesis

References

[1]. Dinkla AM, Florkow MC, Maspero M, et al. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. *Med Phys.* 2019;46:4095-4104.

[2]. Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. *J Appl Clin Med Phys*. 2019;20:105-114.

[3]. Bird D, Scarsbrook AF, Jonathan Sykes J, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. *BMC Cancer.* 2015;15:844.

[4]. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. *Springer*; 2015:234-241.

[5]. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. arXiv:1406.2661 [stat.ML]. June 2014. Available at: https://arxiv.org/abs/1406.2661. Accessed: April 14, 2020.

[6]. Qi M, Li Y, Wu A, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. *Med Phys.* 2020;47:1880-1894.

[7]. Kazemifar S, McGuire S, Timmerman R, et al. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother Oncol.* 2019;136:56-63.

[8]. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46:3565-3581.

[9]. Liu Y, Lei Y, Wang Y, et al. MRI-based treatment planning for proton radiotherapy: Dosimetric validation of a deep learning-based liver synthetic CT generation method. *Phys Med Biol.* 2019;64: 145015.

[10]. Liu Y, Lei Y, Wang Y, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. *Phys Med Biol.* 2019;64: 205022.
[11]. X. Dong et al., "Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging," *Phys. Med. Biol.*, vol. 64, no. 21, p. 215016, Nov. 2019, doi: 10.1088/1361-6560/ab4eb7.
[12]. Hatamizadeh, Ali, et al. Unetr: Transformers for 3d medical image segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
[13]. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods,* 18(2), 203-211.
[14]. Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge - syn3193805 - Wiki