

Final Report  
Cross Modality Medical Image Synthesis and  
Registration through Machine Learning  
600.456/656 Computer Integrated Surgery II

Ping-Cheng Ku (pku1@jh.edu)  
Mentors: Mehran Armand (Mehran.Armand@jhuapl.edu),  
Alejandro Martin-Gomez (alejandro.martin@jhu.edu)

May 2021

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Materials</b>	<b>4</b>
2.1	Datasets . . . . .	4
2.1.1	Arkansas Dataset . . . . .	4
2.1.2	New Mexico Decedent Image Database (NMDID) . . . . .	4
2.2	Dataset Generation . . . . .	5
2.2.1	Train/Test set split . . . . .	5
2.2.2	Data preprocessing . . . . .	5
<b>3</b>	<b>Methods</b>	<b>6</b>
3.1	CycleGAN Framework . . . . .	7
3.2	Training Loss . . . . .	7
3.2.1	Adversarial Loss ( $\mathcal{L}_{adv}$ ) . . . . .	7
3.2.2	Cycle-consistency Loss ( $\mathcal{L}_{cyc}$ ) . . . . .	8
3.2.3	Structural-consistency Loss ( $\mathcal{L}_{sc}$ ) . . . . .	8
3.2.4	Overall CycleGAN Loss . . . . .	8
<b>4</b>	<b>Experiments and Evaluation</b>	<b>8</b>
4.1	Evaluation metrics . . . . .	9
4.2	Ground truth CT . . . . .	10
4.3	Model parameters . . . . .	10
<b>5</b>	<b>Results and Discussion</b>	<b>11</b>
5.1	Training Results . . . . .	11
5.2	Evaluation Results . . . . .	12

<b>6</b>	<b>Conclusions and Future work</b>	<b>13</b>
<b>7</b>	<b>Project Management Summary</b>	<b>14</b>
7.1	Dependencies . . . . .	14
7.2	Project deliverables . . . . .	14
7.3	Project Significance . . . . .	15
7.4	Credits . . . . .	15
7.5	Lessons learned . . . . .	16
<b>8</b>	<b>Acknowledgements</b>	<b>16</b>

# 1 Introduction

Magnetic Resonance Imaging (MRI) is a commonly used imaging technology for the diagnosis of the hip osteonecrosis, which is a disease resulting in the death of bone cells. Early stage of osteonecrosis is often treated with core decompression, a commonly used surgical method for the removal of the osteonecrotic tissue (dead bone) from the femoral head. This surgical procedure involves drilling into the dead bone area, result in reduced pressure and increased blood flow.

Before the core-decompression procedure is performed, pre-operative MRI of the affected joints will be taken for the surgeons to identify the areas affected by the necrotic tissue as well as perform the planning of the drilling paths, as shown in Fig 1 Surgeons usually take intraoperative X-ray shots to monitor the procedure. However, there are currently no simple method that could easily translate the tool trajectories annotated on the MR scans to the intraoperative X-ray shots. The interpolation between the MR and X-ray volumes is essentially a registration problem between the two modalities.

Registration between different imaging modalities has been a popular topic considering its wide applications: helping surgeons during operations, monitoring the infected regions on the patients more easily, and reducing the number of scans that are required for patients to go through. Nevertheless, direct automatic registration between MR volumes and x-ray images has been challenging due to the lack of cross-modality information between the two modalities. Current MR-to-x-ray registration methods are mostly performed with assisting hardware or software model setup. Ashvin et al.[1] managed to calculate rigid transformation between MRI and x-ray by utilizing external fiducial markers that are visible under both imaging modalities. Many studies attempted to perform registration between breast MRI and x-ray mammography as they are the two main image modalities used for detection and diagnosis of breast diseases, and proposed approaches include the generation of Digitally Reconstructed Radiographs (DRR) from MRIs before registering to X-rays [2], as well as gradient-based methods with the construction of patient-specific biomechanical models before performing gradient projection and registration [3].

Recently, the registration approach between CT and x-rays [4, 5] or between CT and flouroscopy [6, 7] has been studied and implemented with very small error. Therefore, the aim of this project is to create image translation from MR volumes to CT volumes; once the synthesized CT volumes are acquired, the existing CT-to-x-ray registration methods could be implemented to interpolate any annotations or segmentation made in CT/MR domain to x-ray images.



Figure 1: Illustration of the annotated drilling path in both MRI (left) and X-ray (right) during core decompression surgery [8]

Various MR-to-CT synthesis algorithms has been proposed such as dictionary-based learning approaches [9, 10], random forest-based methods [11], and deep learning-based methods [12]. This project aims to achieve MR-to-CT synthesis using deep learning-based cycle generative adversarial networks (cycleGAN)[12]. The generation of synthesized CT volumes would be a crucial step when we move on to CT-to-x-ray registration using the synthesized CTs.

## 2 Materials

### 2.1 Datasets

Two datasets are used in this study: the Arkansas dataset and the New Mexico Decedent Image Database (NMDID).

#### 2.1.1 Arkansas Dataset

The Arkansas dataset consists of CT and MR volumes from 32 patients with osteonecrosis. The MRI sequence of the MR images include T1 and T2 weighted spin echo and short tau inversion recovery (STIR). To improve the consistency of the MR training images, only T1 weighted coronal MRI from 19 patients and 6 coronal CT scans are used in this project. The field of view (FOV) of MR volumes in Arkansas dataset ranges from 350mm to 400mm, the matrix size ranges from  $360 \times 360$  to  $528 \times 528$ , and the slice thickness ranges from 5.00mm to 8.40mm. The FOV of CT volumes in Arkansas dataset is 500mm, the matrix size is  $512 \times 512$ , and the slice thickness ranges from 2.00mm to 3.02mm.

Patients with CT volumes available also have their corresponding MR volumes available. Even though for the training of cycleGAN networks, the paired images are not required, the 2 paired MR-CT volumes would be used later for network evaluation purpose.

#### 2.1.2 New Mexico Decedent Image Database (NMDID)

The NMDID consists of full body CT scans, and in this work a total of 12 coronal torso volumes are used. The FOV is 800mm, the matrix size is  $512 \times 512$ , and the slice thickness is 3.00mm.

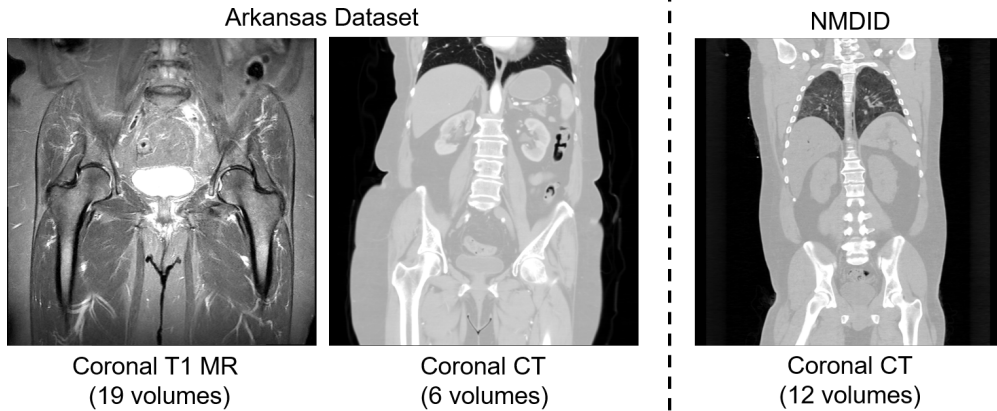


Figure 2: Datasets used in this work for CycleGAN network training

## 2.2 Dataset Generation

### 2.2.1 Train/Test set split

Combining the Arkansas Dataset and NMDID, there are a total of unpaired 19 MR volumes and 18 CT volumes. To ensure that no paired information are used during the training phase of cycleGAN, 5 MR volumes and 3 CT volumes in Arkansas Dataset are used preserved as the evaluation dataset, which includes the paired MR and CT volumes from 2 patients. Two additional CT volumes in NMDID are also added to the evaluation dataset. The remaining 14 MR volumes and 13 CT volumes are used for training.

### 2.2.2 Data preprocessing

In this study, only coronal MR and CT slices are used. Each MR volume in the Arkansas Dataset contains about 20 coronal slices, and each CT volume in Arkansas Dataset and NMDID contains about 100 coronal slices. Due to the different FOV between the CT scans in Arkansas Dataset and NMDID as described in Section 2.1, a region selection process is performed so that the all unpaired MR and CT images are roughly aligned to each other. The region selection process is achieved by cropping the original coronal slices so that the new volume centers around the hip region with roughly the same FOV. The volumes are then resized and padded to  $290 \times 290$  for MR volumes and  $320 \times 320$  for CT volumes.

Since we do not have sufficient training volumes available, offline data augmentation is performed to enhance the size and quality of our training dataset. The transform modules from Medical Open Network for AI (MONAI) [13] framework is used in this project to provide data preprocessing as well as augmentation support for the 3D data. In this project, the following transforms are used:

- **Intensity Thresholding:** The intensity range of both MR and CT volumes are thresholded so that the range becomes  $[0, 400]$  for MR volumes and  $[-1000, 1400]$  Hounsfield unit (HU) for CT volumes. This is mainly to remove unreasonable values within the volumes.

- Intensity Normalization: Since the MR volumes in the Arkansas dataset are collected under various settings and different FOV, the inconsistency between the intensity range has been quite large. Studies has shown that for image synthesis tasks using MR images, performing intensity normalization in pre-processing step could substantially improve the result [14]. Therefore, the Z-score intensity normalization is carried out. Denoting the input MR image as  $\mathcal{I}(\mathbf{x})$  (pixel position  $\mathbf{x} \subset \mathbb{N}^3$ ), the intensity normalized image  $\mathcal{I}_{IN}(\mathbf{x})$  is obtained as follow:

$$\mathcal{I}_{IN}(\mathbf{x}) = \frac{\mathcal{I}(\mathbf{x}) - \mu}{\sigma}$$

where  $\mu$  is the mean intensity and  $\sigma$  is the intensity standard deviation of the input volume.

- Intensity re-scaling: The intensity of both MR and CT volumes is re-scaled to [0, 255].
- Augmentation - random zoom: The volumes are zoomed within the range of [0.9, 1.1] along the anteroposterior axis to interpolate the pixel values. Values of zeros are added around the volume while zooming out.
- Augmentation - Random rotate: The volumes are rotated by [-0.1, 0.1] radians along the anteroposterior axis. Values of zeros are added to fill in empty values during the operation.
- Augmentation - Random spatial crop: Random cropping is performed to crop both the CT and MR volumes to  $256 \times 256 \times 3$  so that the output volume shape aligns with the cycleGAN input image shape.

For the generation training dataset, all the transforms above are performed; while for the generation of the test dataset, the augmentation functions are skipped. In addition of the generation of testing images (with output size of  $256 \times 256 \times 3$ , additional testing volumes are also generated with output shape of  $256 \times 256 \times n_{cor}$ , preserving the original number of coronal slice count  $n_{cor}$ . The entire testing volumes are used later for network performance evaluation.

### 3 Methods

The proposed MR-to-CT translation training workflow is shown in Fig 3 below. With the pre-processed training dataset established, the cycle generative adversarial networks (cycleGANs) is used for network training.

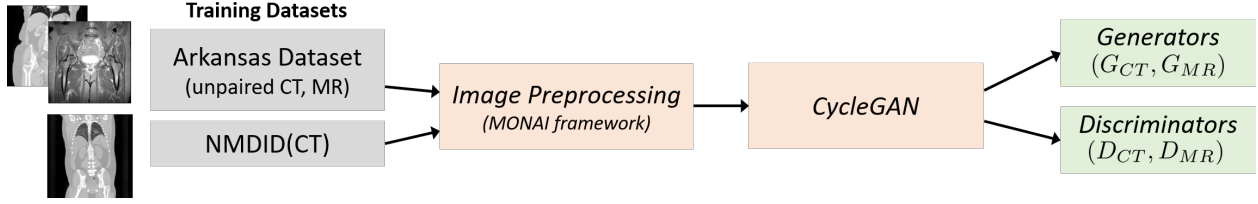


Figure 3: MR-to-CT translation training flowchart

### 3.1 CycleGAN Framework

The approach to perform image-to-image translation is to utilize the cycleGAN network proposed by Zhu et al. [12]. In a CycleGAN network, the main workflow to translate an image from MR domain to CT domain as well as the corresponding loss functions are shown in Fig 4 below, and the same workflow structure is applied to the case of translation from CT to MR domain.

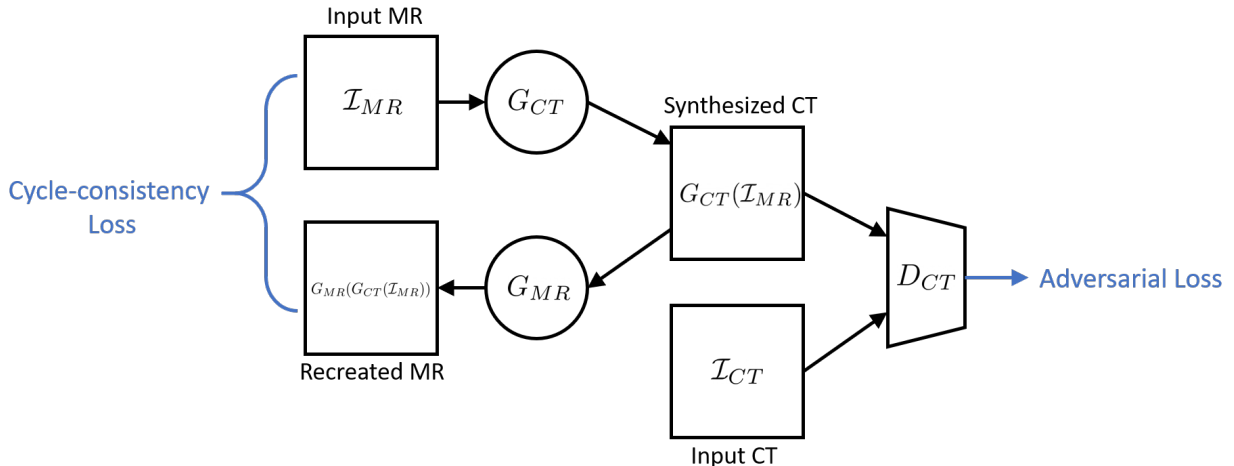


Figure 4: Workflow of image translation from MR domain to CT domain in a CycleGAN network

### 3.2 Training Loss

#### 3.2.1 Adversarial Loss ( $\mathcal{L}_{adv}$ )

The goal in our cycleGAN network is to learn the mapping functions between the MR domain to the CT domain. To achieve this objective, the CT generator  $G_{CT}$  generates synthetic CT given input MR images  $\mathcal{I}_{MR}$ , while the CT discriminator attempts to distinguish the synthesized CT images from real CT images  $\mathcal{I}_{MR}$ . Therefore, the adversarial loss for the translation from MR to CT domain is formulated as

$$\mathcal{L}_{adv}(G_{CT}, D_{CT}) = D_{CT}(G_{CT}(\mathcal{I}_{MR}))^2 + (1 - D_{CT}(\mathcal{I}_{CT}))^2$$

Similarly, the adversarial loss for the translation from CT to MR domain (with MR generator  $G_{MR}$  and MR discriminator  $D_{MR}$ ) is formulated as

$$\mathcal{L}_{adv}(G_{MR}, D_{MR}) = D_{MR}(G_{MR}(\mathcal{I}_{CT}))^2 + (1 - D_{MR}(\mathcal{I}_{MR}))^2$$

### 3.2.2 Cycle-consistency Loss ( $\mathcal{L}_{cyc}$ )

Cycle-consistency loss calculates the difference between the input real image and the reconstructed image, which is the inverse mapping of the synthesized image. Reducing the cycle-consistency loss encourages the network to produce reconstructed images that are identical to the real images. The cycle-consistency loss is formulated as follow

$$\mathcal{L}_{cyc}(G_{CT}, G_{MR}) = \|G_{CT}(G_{MR}(\mathcal{I}_{CT})) - \mathcal{I}_{CT}\|_1 + \|G_{MR}(G_{CT}(\mathcal{I}_{MR})) - \mathcal{I}_{MR}\|_1$$

### 3.2.3 Structural-consistency Loss ( $\mathcal{L}_{sc}$ )

MR-to-CT synthesis through the original cycleGAN proposed by Zhu et al. [12] often result in inconsistent structures between input and synthetic images. Previous works that attempt to perform cross-modality image synthesis often address this problem by implementing additional loss functions. Haisa et al. [15] implemented gradient consistency loss base on the gradient correlation (GC) between the input and output image of the generators; while Tanner et al. [16] and Yang et al. [17] implemented Modality Independent Neighborhood Descriptor (MIND) loss to ensure the consistency instead.

In this project, the MIND loss was implemented that calculates the structural consistency of the generators ( $\mathcal{L}_{sc}(G_{CT}, G_{MR})$ ) and the result is compared to the one without this loss function implemented.

### 3.2.4 Overall CycleGAN Loss

The overall cycleGAN loss ( $\mathcal{L}$ ) is defined as

$$\mathcal{L}(G_{CT}, G_{MR}, D_{CT}, D_{MR}) = \mathcal{L}_{adv}(G_{CT}, D_{CT}) + \mathcal{L}_{adv}(G_{MR}, D_{MR}) + \lambda_{cyc}\mathcal{L}_{cyc}(G_{CT}, G_{MR}) + \lambda_{str}\mathcal{L}_{sc}(G_{CT}, G_{MR})$$

where the cycle consistency weight  $\lambda_{cyc}$  and structural consistency weight  $\lambda_{sc}$  control the importance of the corresponding loss terms.

## 4 Experiments and Evaluation

To evaluate the performance of the network, several experiments has been carried out. The evaluation workflow is shown in Fig 5



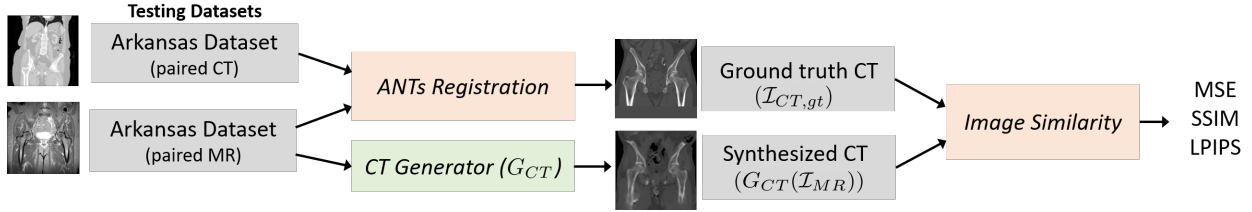


Figure 5: Workflow of performance evaluation on the trained network

## 4.1 Evaluation metrics

The quantitative evaluation of the cycleGAN network performance is achieved by comparing the mean absolute error (MAE), structural similarity index measure (SSIM)[18], and learned Perceptual Image Patch Similarity (LPIPS) metric [19].

Given two input images (volumes)  $\mathcal{I}$  and  $\hat{\mathcal{I}}$  with identical shape, each evaluation metric is introduced as follow

- Mean absolute error (MAE):

$$MSE(\mathcal{I}, \hat{\mathcal{I}}) = \frac{1}{N} \sum_{i=1}^N (\mathcal{I}_i - \hat{\mathcal{I}}_i)^2 \quad (1)$$

where  $N$  represents the total number of data values in each image.

- Structural similarity index measure (SSIM):

Structural similarity index is first proposed by Wang et al.[18]. SSIM models any image distortion by combining the correlation loss, the distortion of luminance, and distortion of contrast. The SSIM is defined as

$$SSIM(\mathcal{I}, \hat{\mathcal{I}}) = [l(\mathcal{I}, \hat{\mathcal{I}})]^\alpha \cdot [c(\mathcal{I}, \hat{\mathcal{I}})]^\beta \cdot [s(\mathcal{I}, \hat{\mathcal{I}})]^\gamma$$

where

$$l(\mathcal{I}, \hat{\mathcal{I}}) = \frac{2\mu_{\mathcal{I}}\mu_{\hat{\mathcal{I}}} + C_1}{\mu_{\mathcal{I}}^2 + \mu_{\hat{\mathcal{I}}}^2 + C_1} \quad c(\mathcal{I}, \hat{\mathcal{I}}) = \frac{2\sigma_{\mathcal{I}}\sigma_{\hat{\mathcal{I}}} + C_2}{\sigma_{\mathcal{I}}^2 + \sigma_{\hat{\mathcal{I}}}^2 + C_2} \quad s(\mathcal{I}, \hat{\mathcal{I}}) = \frac{\sigma_{\mathcal{I}\hat{\mathcal{I}}} + C_3}{\sigma_{\mathcal{I}} + \sigma_{\hat{\mathcal{I}}} + C_3}$$

$$\mu_{\mathcal{I}} = \frac{1}{N} \sum_{i=1}^N \mathcal{I}_i \quad \sigma_{\mathcal{I}}^2 = \frac{1}{N} \sum_{i=1}^N (\mathcal{I}_{ij} - \mu_{\mathcal{I}})^2 \quad \sigma_{\mathcal{I}\hat{\mathcal{I}}} = \frac{1}{N} \sum_{i=1}^N (\mathcal{I}_i - \mu_{\mathcal{I}})(\hat{\mathcal{I}}_i - \mu_{\hat{\mathcal{I}}})$$

In this work, the weights  $\alpha, \beta, \gamma$  are all set to 1. A high SSIM score represents strong structural similarity between the two input images and vice versa.

- Perceptual similarity metric (LPIPS): [19] One of the drawbacks of the considered criteria is their low correlation with the visual perception of a human being. For example, low MSE value can be obtained if there are noticeable local distortions. LPIPS, a similarity metric based on convolutional neural networks (CNN) that approximates

the process of perception of visual distortion by a human being has been developed. The authors provided LPIPS implementation on architectures on AlexNet [20], VGG16 [21], and SqueezeNet [22]. We are using pretrained AlexNet for the LPIPS metric in this work.

## 4.2 Ground truth CT

To evaluate the synthesized CT volumes from the CT generator  $G_{CT}$ , the ground truth CT has to be obtained. In the Arkansas dataset, the two subjects with both MR and CT images available will be used in the evaluation phase.

To perform robust MR-to-CT registration, the registration toolkit from Advanced Normalization Tools (ANTs) [23] were used. 20 landmarks on the pelvis and the femurs are manually labelled in both the paired MR and CT volumes and are exported as FCSV files. The ANTs registration software then computes the rigid transformation between the two sets of input data. The transformation collected is then imported in 3D Slicer, where the CT volumes are warped using the transformation and exported as the ground truth CT. Fig 6 illustrates the MR-CT registration result from the same patient after the calculated transformation was applied on the CT volume.

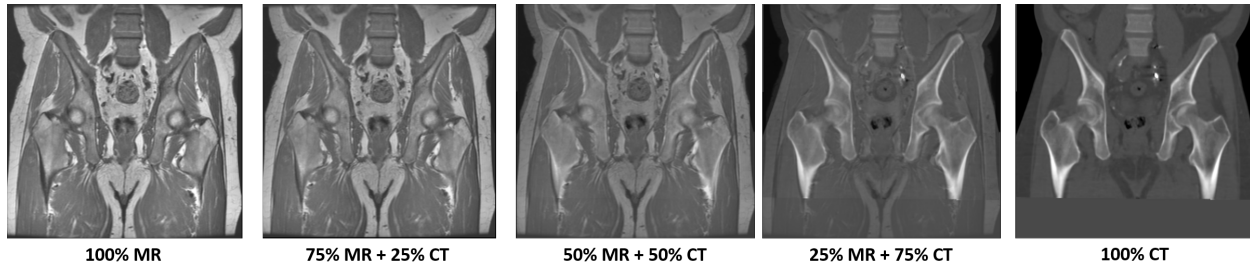


Figure 6: Visualization of ANTs registration result in 3D Slicer through MR/CT volume blending

An important observation in Fig 6 is that most parts of femoral shafts and the tissue around them are missing due to the different FOV in MR and CT when the images were originally taken. The fact that around  $\frac{1}{6}$  of our ground truth CT is empty will have a significant impact when we attempt to compare it with the synthesized CT later.

## 4.3 Model parameters

To quantitatively evaluate the performance of the network and the dependency on the training parameters, the following cycleGAN training experiments have been carried out: (1) Training with dataset size of 4000 MR and 4000 CT images, and with dataset size of 1000 MR and 1000 CT images (2) Training with and without the implementation of structural consistency loss  $\mathcal{L}_{sc}$ . In the case where structural consistency loss is included, the weight  $\lambda_{sc}$  in the loss function is set to 0.2.

Some model parameters remain constant throughout the experiment. The initial learning rate of cycleGAN training is set to 0.2. The weight of cycle-consistency loss ( $\mathcal{L}_{cyc}$ ) is set

to 10. The majority of the models are trained using the 12GB NVIDIA Tesla K80 GPU on Google Cloud. Each run terminates when reaching epoch number 50, and the best performing generator weights will be used and evaluated in the evaluation step. The procedure will be re-run two additional times under the same settings.

## 5 Results and Discussion

### 5.1 Training Results

The training results from the cycleGAN generators  $G_{CT}$ ,  $G_{MR}$  are shown in Fig 7 and Fig 8, respectively. Both synthesized CT and MR images appear to have reasonable image intensity with proper anatomical structures highlighted. However, as mentioned in the previous section, very few CT volumes in our training dataset captures the femoral shaft. In Fig 7, it can be observed that the synthesized CTs show good image translation result on bony structures like pelvis femoral heads, but performs poorly translating the femoral shafts from MR volumes.

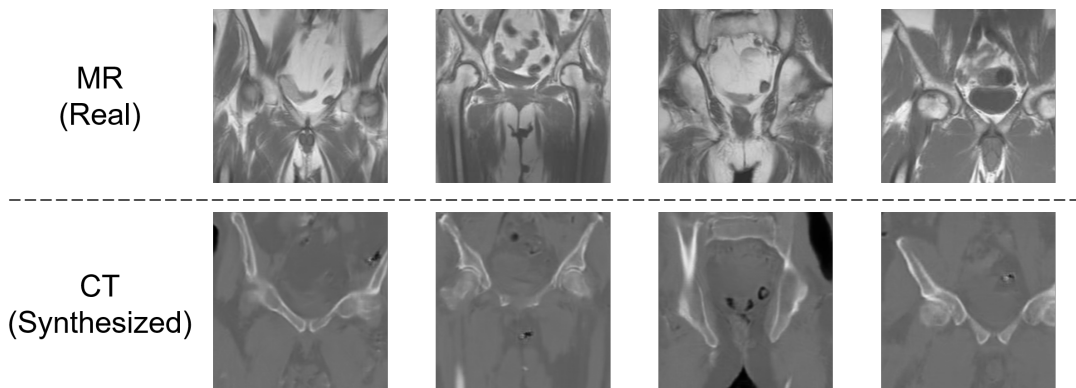


Figure 7: CycleGAN CT generator ( $G_{CT}$ ) training results (training size: 4000, with implementation of structural consistency loss  $\mathcal{L}_{str}$ )

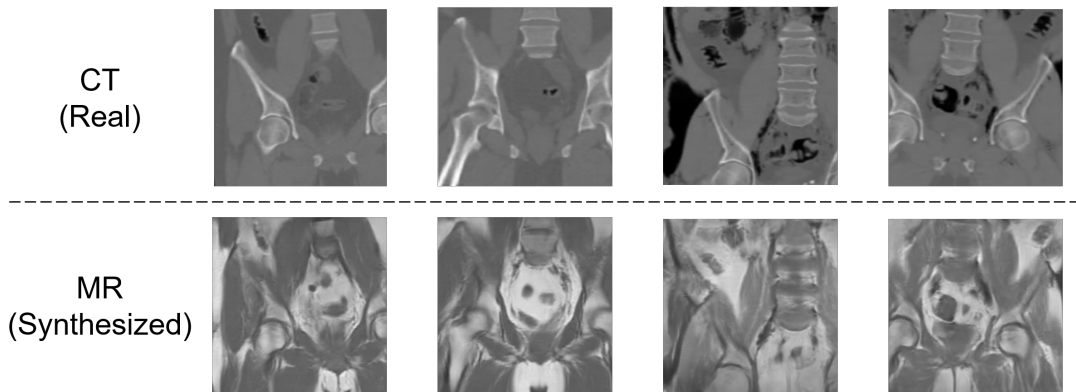


Figure 8: CycleGAN CT generator ( $G_{MR}$ ) training results (training size: 4000, with implementation of structural consistency loss  $\mathcal{L}_{sc}$ )

## 5.2 Evaluation Results

Table 1 below shows the mean squared error (MSE), structural similarity index measure (SSIM), and perceptual similarity metric (LPIPS) result between the synthesized CT volumes and the ground truth CT volumes we obtained earlier through ANTs registration. The results are also visualized in the box plot in Fig 9 below.

Metric	Test volume	1000 training images		4000 training images	
		without $\mathcal{L}_{sc}$	with $\mathcal{L}_{sc}$	without $\mathcal{L}_{sc}$	with $\mathcal{L}_{sc}$
MSE	run_1, Subject_1	0.1279	0.1447	0.1425	0.1441
	run_1, Subject_2	0.1167	0.1473	0.1144	0.0846
	run_2, Subject_1	0.1416	0.0955	0.0790	0.0846
	run_2, Subject_2	0.1147	0.1147	0.0775	0.0816
	run_3, Subject_1	0.1168	0.1442	0.1060	0.0921
	run_3, Subject_2	0.1335	0.1188	0.1119	0.1014
	Average $\pm$ std	$0.1252 \pm 0.0100$	$0.1293 \pm 0.0207$	$0.1052 \pm 0.0223$	$0.0981 \pm 0.0216$
SSIM	run_1, Subject_1	0.2744	0.1897	0.1637	0.1385
	run_1, Subject_2	0.2192	0.2274	0.2801	0.4487
	run_2, Subject_1	0.1151	0.3547	0.4082	0.3075
	run_2, Subject_2	0.3515	0.3515	0.5403	0.5393
	run_3, Subject_1	0.2352	0.2601	0.2341	0.3546
	run_3, Subject_2	0.2507	0.2272	0.3323	0.3448
	Average $\pm$ std	$0.2410 \pm 0.0705$	$0.2767 \pm 0.0663$	$0.3265 \pm 0.1222$	$0.3556 \pm 0.1238$
LPIPS	run_1, Subject_1	0.4688	0.4594	0.4355	0.4227
	run_1, Subject_2	0.4223	0.4664	0.3984	0.3787
	run_2, Subject_1	0.4447	0.4233	0.3667	0.4541
	run_2, Subject_2	0.4879	0.4879	0.3549	0.3419
	run_3, Subject_1	0.4109	0.4356	0.3633	0.3836
	run_3, Subject_2	0.4169	0.3909	0.3943	0.3862
	Average $\pm$ std	$0.4419 \pm 0.0284$	$0.4545 \pm 0.0229$	$0.3855 \pm 0.0275$	$0.3945 \pm 0.0355$

Table 1: Mean squared error (MSE), structural similarity index measure (SSIM), and perceptual similarity metric (LPIPS) result between the synthesized and the ground truth CT volumes

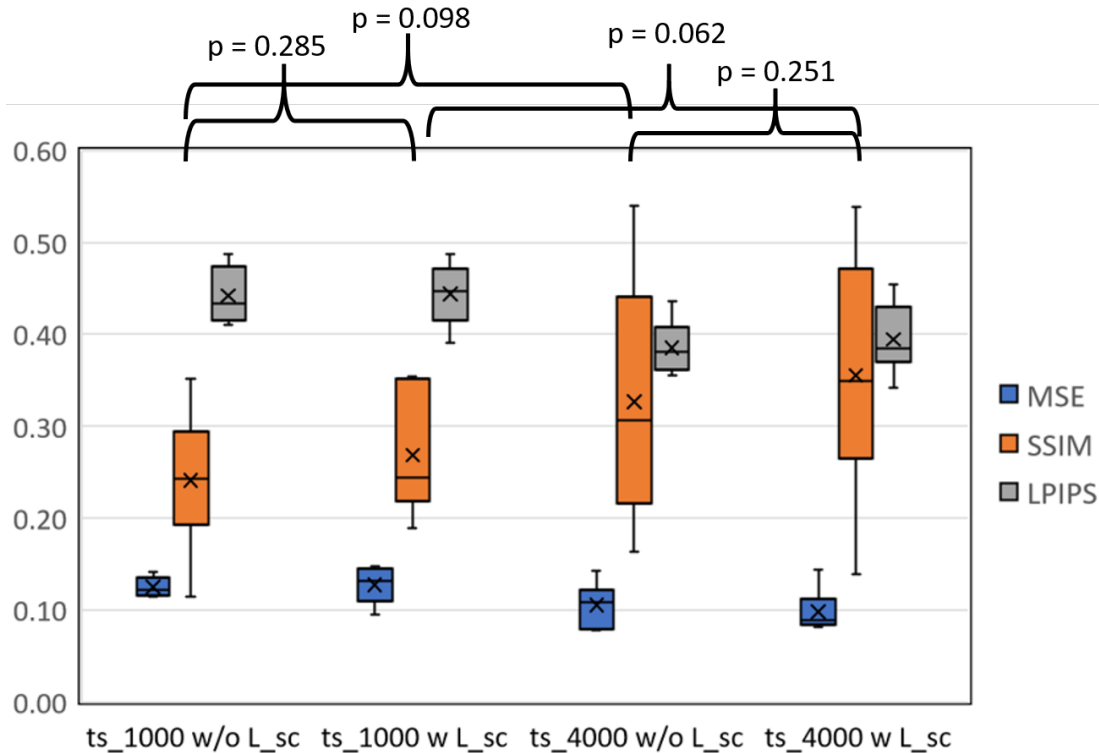


Figure 9: Evaluation of image similarity between ground truth and synthesized CT volumes under different experiment settings. (y-axis formatting: ts\_(training size)\_(with or without structural consistency loss))

It is shown in Table 1 that the average structural similarity index measure (SSIM) values under different runs and different settings ranges between 0.24 to 0.36, which is a relatively low value, indicating the synthesized CT and the ground truth CT do not have similar structural appearance. One reason is due to the missing femoral shaft information in our CT volumes, resulting a great portion of the ground truth CT to be filled with empty values.

The results in Fig 9 shows that by increasing the number of training data from 1000 to 4000, the data yielded statistically marginal significant improvement ( $p < 0.1$ ) according to the paired t-test results. However, the implementation of structural consistency loss does not seem to have a statistically meaningful impact on the final results ( $p > 0.1$ ) even though the mean SSIM seem to slightly increase.

## 6 Conclusions and Future work

It has been shown in both the training and the evaluation results that the inconsistent field of view between the MR and CT volumes in our dataset is a problem that significantly impacts the evaluation performance score. I did not notice the effect of this issue early enough to make corresponding plans to mitigate the negative effect of this problem. I believe that being able to access MR and CT datasets with more aligned FOV may help alleviate this issue. And I will be able to make more meaningful comparisons between the ground-truth

and the synthesized CT volumes. An additional approach would be to manually inspect and crop each MR and CT volumes to ensure all images in both image modalities share similar regional FOVs. However, the trained CT generator ( $G_{CT}$ ) does seem to be properly trained in Fig 7 neglecting the lower part of the image affected by the missing femoral shafts in our CT volumes and has the potential to be used for further applications once the issue is resolved.

In this project, I did not manage to implement the state-of-the-art cycleGAN-based networks for evaluation performance benchmarking in time. Some cycleGAN-based algorithms that show great potentials include the work from Hiasa et al.[15] where gradient correlation (GC) is used for cycleGAN loss calculation, and the structure-constrained cycleGAN (sc-cycleGAN) developed by Yang et al.[17] where spectral normalization and self-attention blocks are implemented to the cycleGAN network. I believe this is a crucial step to take as it would allow meaningful comparisons between the similarity metric values collected from our trained network and the best-performing networks, which is a crucial step while moving on towards research publications.

## 7 Project Management Summary

### 7.1 Dependencies

The dependencies of this project are shown in Fig.10 below, with the plans and the estimated resolving time. Alternative plans are proposed for each dependency to mitigate the risk of original plans failing. All dependencies were addressed and resolved in the early stage of this project.




Dependency	Plan	Estimated time	Current Status
GPU resource access for network training 	Talk to Dr. Armand to get access to MARCC	2/26 3/25	Resolved on 3/16. Allocation to MARCC has been acquired.
	(alternative) Lab GPU resource (thin6), Google cloud	Resolved	Access to thin6 has been acquired.
Access to CT dataset for additional training images 	Request access to New Mexico Database	2/26	Access to images of 24 additional patients submitted on 2/22, acquired on 2/26
	(alternative) Raise the issue during the 3/2 weekly meeting with Dr. Armand.	3/2	N/A
Acquire tool path annotations in MR images 	Check with Alejandro if we can get collect annotated MRs from surgeons	3/18	Discussed during 3/1 meeting with Alejandro. Going for the alternative approach below.
	(alternative) Generate simulated tool paths ourselves	3/20	I have access to Imfusion Suite which allows simple path annotations to be made (two points that forms a line). The method to extract the paths out from the software still needs to be figured out.

Figure 10: Project dependencies

### 7.2 Project deliverables

The project deliverables are listed in Fig ?? below. Multiple project deliverables have been significantly adjusted during the course of this project. The maximum deliverable that aims

to perform synthesized CT to X-ray registration has mostly been removed from this project and are put in the future work of this project that I will continue be working on. Most expected deliverables, including the MR-to-CT image translation algorithm as well as the evaluation workflow have both been implemented and tested.

The code documentation of this work is fully accessible on Github ([https://github.com/AxDante/MR\\_to\\_synCT](https://github.com/AxDante/MR_to_synCT))

	Task	Date	Status	
<b>Mini mum</b>	Generate reasonable synthesized CT images from unpaired MR images.	3/15	Completed by 3/15	✓
	Validate the performance of registration between synthesized CT and ground truth CTs.	<del>4/6</del> 4/20	Completed by 4/20	✓
	Proper code documentation and repository publishment.	5/6	All code available on Github, will keep it maintained and updated in the future	✓
<b>Expe cted</b>	Code optimization and generation of best trained network models and ideal parameters.	<del>4/15</del> 5/6	Trained with different parameters, will be running the training for improvements	✓
	<b>(NEW)</b> Implementation of similarity metrics for network performance validation	3/30	Completed by 3/30	✓
	<b>(NEW)</b> Manual registration between existing MR and CT images, and validation of the robustness of cycleGAN generator performance.	4/11	Manual registration completed by 4/11, evaluation completed by 4/20	✓
	<del>Validate the registration workflow by registering manually annotated paths from MR to x-ray images.</del>	4/20	Manually annotated MRs completed, but the registration to X-ray is <b>removed</b> .	
<b>Maxi mum</b>	<del>Improve workflow and combine networks for loss optimization.</del>	4/29	<b>Removed</b>	
	<del>Develop a direct MR to x-ray registration method.</del>	4/29	<b>Removed</b>	
	<del>Annotation projection from MR to X-rays</del>	5/6	<b>Removed</b>	

Figure 11: Project deliverables

### 7.3 Project Significance

- This project provides the MR-to-CT translation pipeline using cycleGAN network. Including training and evaluation workflows.
- The flexible preprocessing package developed in this project is capable of reading input data from different datasets and different imaging modalities
- This project implemented MIND-based structural loss function in cycleGAN, which could be optimized in the future to further boost the network performance.
- The image similarity package developed for this project accepts nifty volumes as input instead of only accepting 2D images, which would be a lot more useful for medical imaging evaluation purposes.

### 7.4 Credits

Ping-Cheng Ku is in charge of the entire implementation of the project.

## 7.5 Lessons learned

I have learned that keeping track of the training parameters that are being adjusted is very crucial. Lots of training time were in fact wasted training using incorrect parameter settings. I have also learned a lot while establishing the network evaluation workflow. Proper evaluation methods and proper statistical analysis of the obtained results are crucial to create strong and convincing publications.

## 8 Acknowledgements

Thank you to Dr. Armand Mehran, Robert Grupp, Alejandro Martin-Gomez for the continuous support, assistance and encouragement throughout this project.

## References

- [1] Ashvin K. George et al. “Robust automatic RIGID registration of MRI and x-ray using External fiducial markers for Xfm-guided interventional procedures”. In: *Medical Physics* 38.1 (2010), pp. 125–141. DOI: [10.1118/1.3523621](https://doi.org/10.1118/1.3523621).
- [2] Thomy Mertzaniidou et al. “MRI to X-ray mammography registration using a volume-preserving affine transformation”. In: *Medical Image Analysis* 16.5 (2012), pp. 966–975. DOI: [10.1016/j.media.2012.03.001](https://doi.org/10.1016/j.media.2012.03.001).
- [3] Eloy García et al. “Breast MRI and X-ray mammography registration using gradient values”. In: *Medical Image Analysis* 54 (2019), pp. 76–87. DOI: [10.1016/j.media.2019.02.013](https://doi.org/10.1016/j.media.2019.02.013).
- [4] Haofu Liao et al. “Multiview 2D/3D Rigid Registration via a Point-Of-Interest Network for Tracking and Triangulation”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). DOI: [10.1109/cvpr.2019.01292](https://doi.org/10.1109/cvpr.2019.01292).
- [5] Gautham Nandakumar et al. “Fully automated CT to x-ray registration of infected lung regions for COVID-19 patient monitoring”. In: *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications* (2021). DOI: [10.1117/12.2582163](https://doi.org/10.1117/12.2582163).
- [6] Robert B. Grupp et al. “Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration”. In: *International Journal of Computer Assisted Radiology and Surgery* 15.5 (2020), pp. 759–769. DOI: [10.1007/s11548-020-02162-7](https://doi.org/10.1007/s11548-020-02162-7).
- [7] Po Su et al. “Fast CT-CT fluoroscopy registration with respiratory motion compensation for image-guided lung intervention”. In: *Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling* (2012). DOI: [10.1117/12.910822](https://doi.org/10.1117/12.910822).
- [8] *Osteonecrosis of the Hip - OrthoInfo - AAOS*. URL: <https://orthoinfo.aaos.org/en/diseases--conditions/osteonecrosis-of-the-hip>.
- [9] A. Torrado-Carvajal et al. “Fast Patch-Based Pseudo-CT Synthesis from T1-Weighted MR Images for PET/MR Attenuation Correction in Brain Studies”. In: *Journal of Nuclear Medicine* 57.1 (2015), pp. 136–143. DOI: [10.2967/jnumed.115.156299](https://doi.org/10.2967/jnumed.115.156299).



- [10] Lei Yang et al. “Magnetic resonance imaging-based pseudo computed tomography using anatomic signature and joint dictionary learning”. In: *Journal of Medical Imaging* 5.03 (2018), p. 1. DOI: [10.1117/1.jmi.5.3.034001](https://doi.org/10.1117/1.jmi.5.3.034001).
- [11] Yang Lei et al. “MRI-based synthetic CT generation using semantic random forest with iterative refinement”. In: *Physics in Medicine amp; Biology* 64.8 (2019), p. 085001. DOI: [10.1088/1361-6560/ab0b66](https://doi.org/10.1088/1361-6560/ab0b66).
- [12] Jun-Yan Zhu et al. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks”. In: *2017 IEEE International Conference on Computer Vision (ICCV)* (2017). DOI: [10.1109/iccv.2017.244](https://doi.org/10.1109/iccv.2017.244).
- [13] Project-MONAI. *Project-monai/monai*. URL: <https://github.com/Project-MONAI/MONAI>.
- [14] Jacob C. Reinhold et al. “Evaluating the impact of intensity normalization on MR image synthesis”. In: *Medical Imaging 2019: Image Processing* (2019). DOI: [10.1117/12.2513089](https://doi.org/10.1117/12.2513089).
- [15] Yuta Hiasa et al. “Cross-Modality Image Synthesis from Unpaired Data Using CycleGAN”. In: *Simulation and Synthesis in Medical Imaging* (2018), pp. 31–41. DOI: [10.1007/978-3-030-00536-8\\_4](https://doi.org/10.1007/978-3-030-00536-8_4).
- [16] Christine Tanner et al. *Generative Adversarial Networks for MR-CT Deformable Image Registration*. 2018. eprint: [arXiv:1807.07349](https://arxiv.org/abs/1807.07349).
- [17] Heran Yang et al. “Unsupervised MR-to-CT Synthesis Using Structure-Constrained CycleGAN”. In: *IEEE Transactions on Medical Imaging* 39.12 (2020), pp. 4249–4261. DOI: [10.1109/tmi.2020.3015379](https://doi.org/10.1109/tmi.2020.3015379).
- [18] Z. Wang et al. “Image Quality Assessment: From Error Visibility to Structural Similarity”. In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. DOI: [10.1109/tip.2003.819861](https://doi.org/10.1109/tip.2003.819861).
- [19] Richard Zhang et al. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018). DOI: [10.1109/cvpr.2018.00068](https://doi.org/10.1109/cvpr.2018.00068).
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet classification with deep convolutional neural networks”. In: *Communications of the ACM* 60.6 (2017), pp. 84–90. DOI: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [21] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. eprint: [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [22] Forrest N. Iandola et al. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and j0.5MB model size*. 2016. eprint: [arXiv:1602.07360](https://arxiv.org/abs/1602.07360).
- [23] Brian B. Avants et al. “The Insight ToolKit image registration framework”. In: *Frontiers in Neuroinformatics* 8 (2014). DOI: [10.3389/fninf.2014.00044](https://doi.org/10.3389/fninf.2014.00044).