Unsupervised Semantic Image Synthesis for Medical Imaging

Research Thesis S1449

Wenwu Tang

ISS- University of Stuttgart

Content

- Motivation
- Method
- Datasets and Pre-processing
- Experiments and Results
  - CT Label-to-CT Scan
  - CT Label-to-MR Scan
  - Failure Cases
  - Ablation on 3D noise
- Conclusion
- Appendix

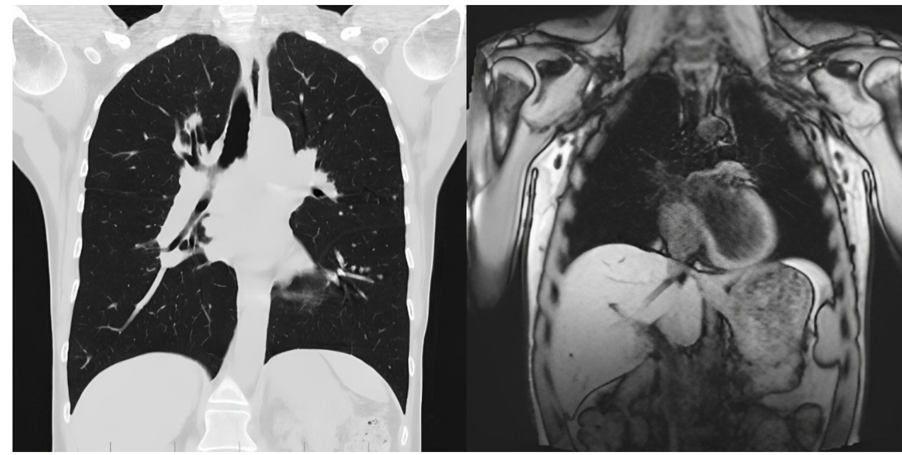
Content

- Motivation
- Method
- Datasets and Pre-processing
- Experiments and Results
  - CT Label-to-CT Scan
  - CT Label-to-MR Scan
  - Failure Cases
  - Ablation on 3D noise
- Conclusion
- Appendix

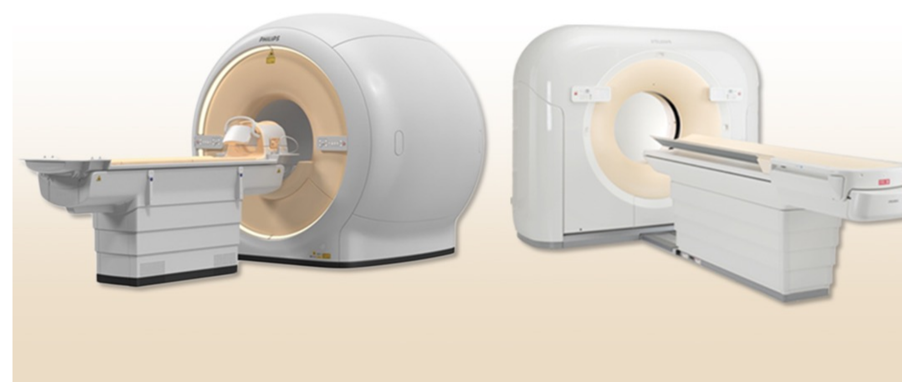
Motivation

Obtain MR-CT paired images without conducting both diagnostic procedures

https://www.starimagingindia.com/blog/difference-between-mri-scan-and-a-ct-scan/



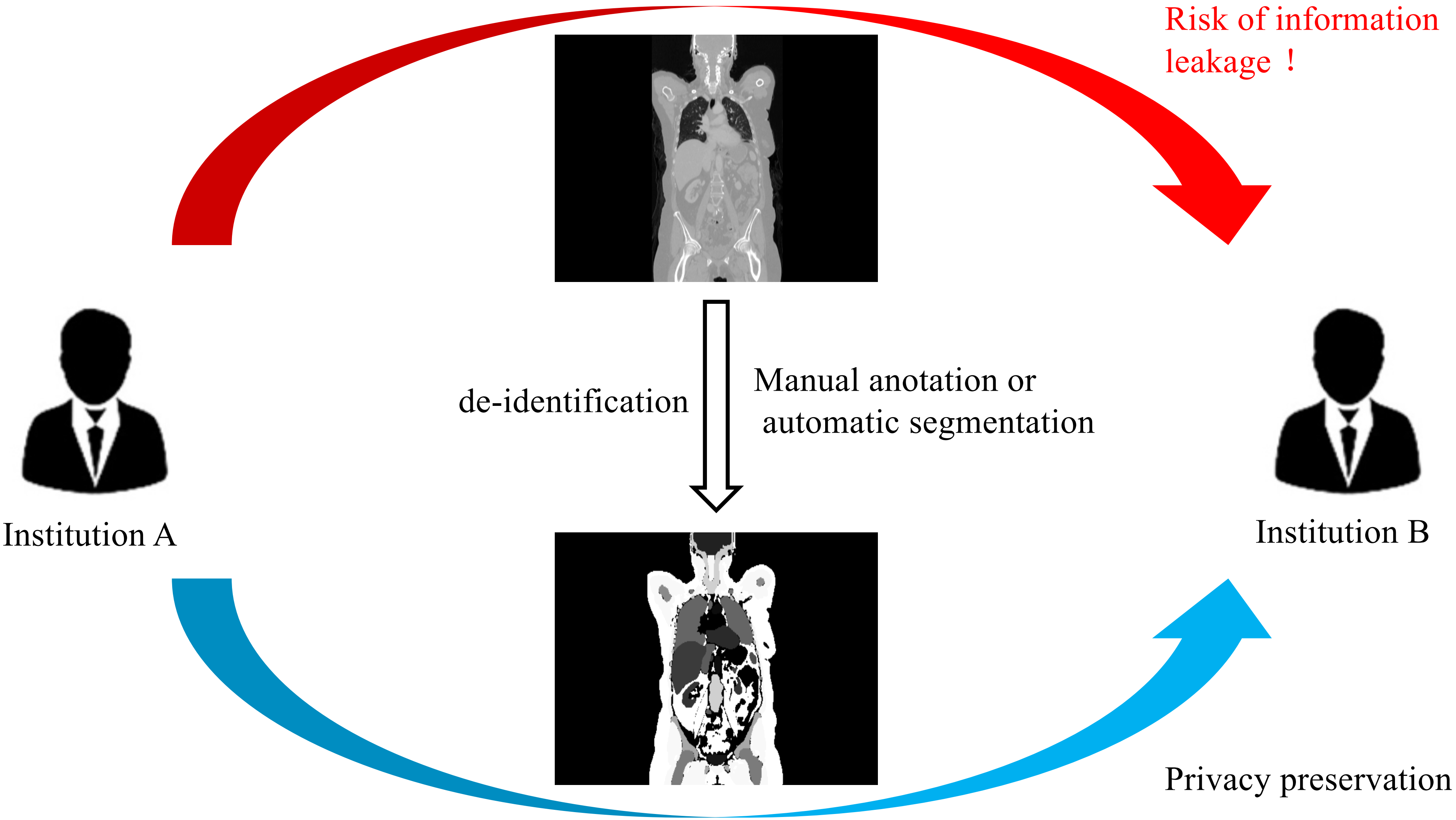
https://kiranpetct.com/ct-scans-vs-mri-scans-what-are-the-differences-between-them/



https://www.starimagingindia.com/blog/difference-between-mri-scan-and-a-ct-scan/

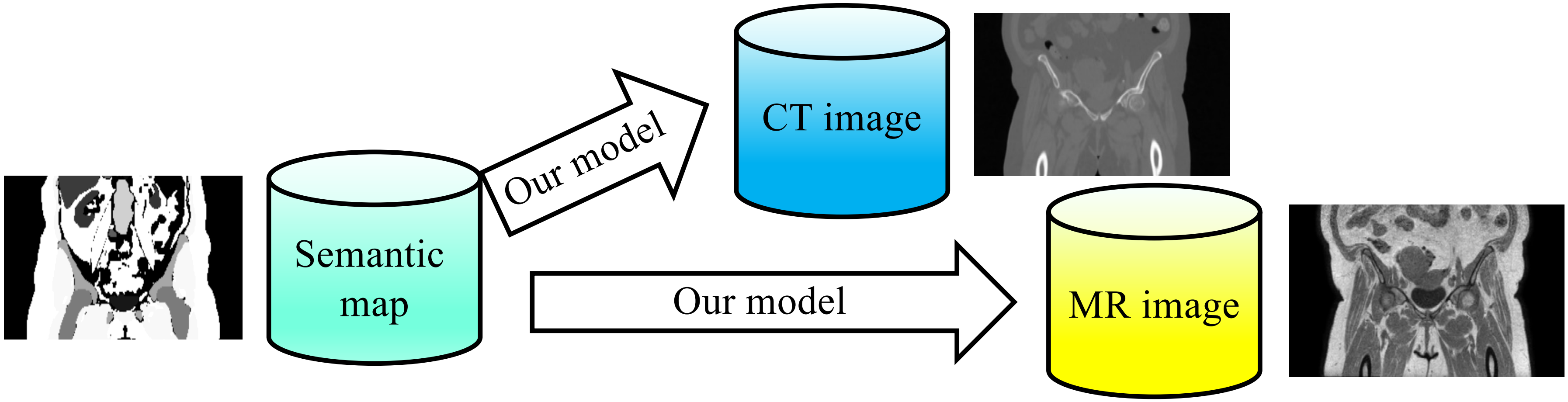
Motivation

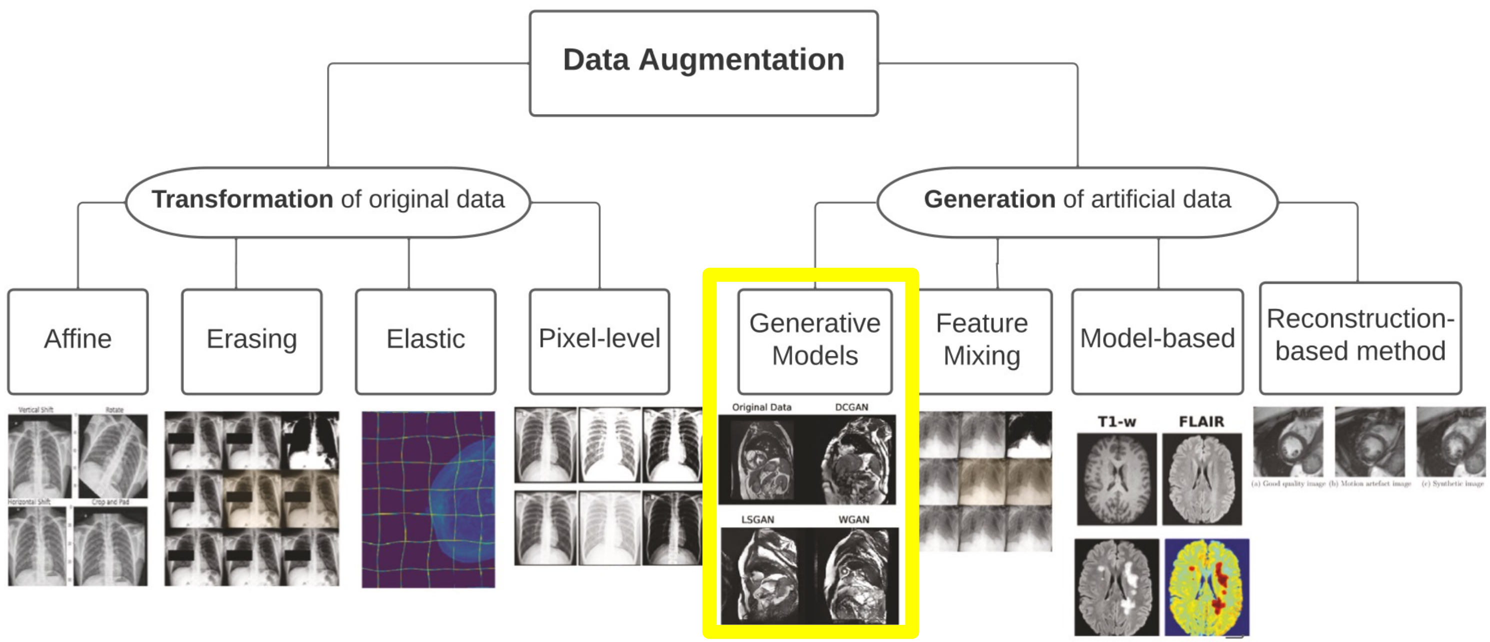
Privacy preservation by sharing anonymizing medical data



Motivation

Data augmentation methods produce synthetic data to enhance performance in downstream tasks.





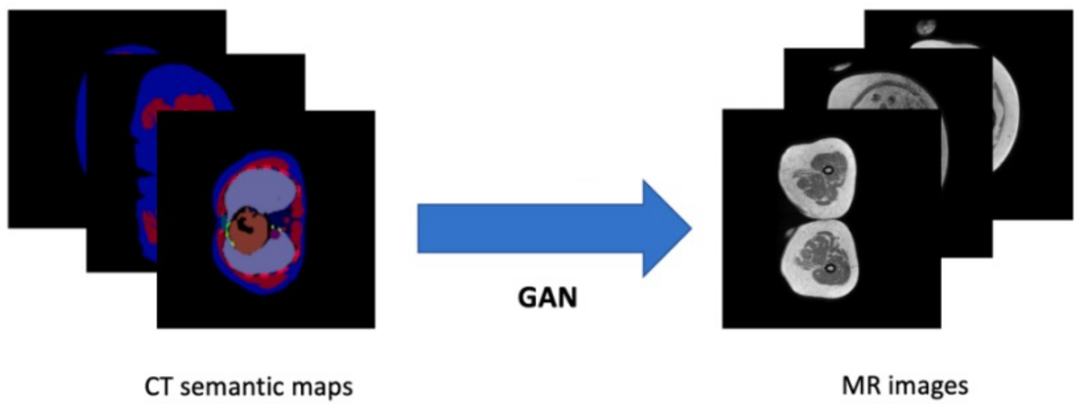
1. Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: A systematic literature review. Comput Biol Med. 2023 Jan;152:106391. doi: 10.1016/j.compbiomed.2022.106391. Epub 2022 Dec 9. PMID: 36549032.

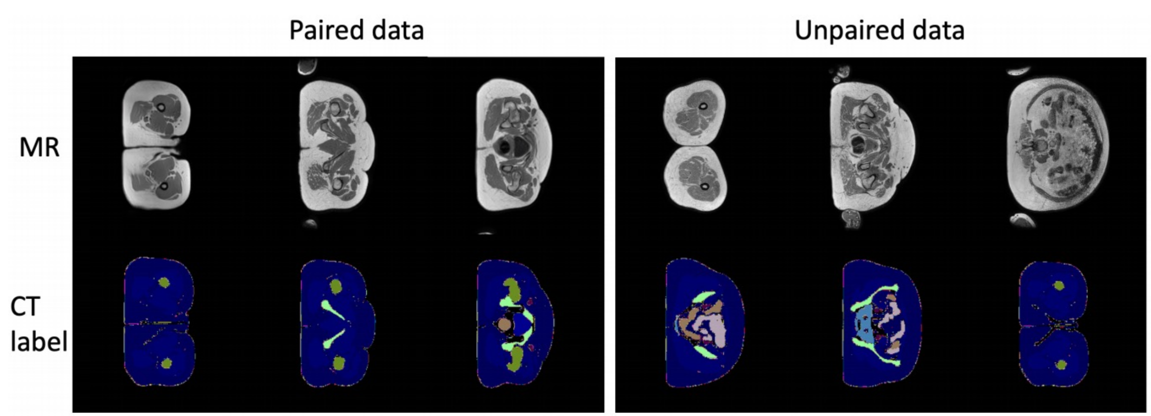
Motivation

Problem definition

Our goal is the unsupervised synthesis of **2D MR images** $\mathbf{x}_{MR}$ based on **CT semantic maps** $\mathbf{m}_{CT}$. Semantic map can be either manually annotated or even automatically generated by segmentation tools

m_{CT} x_{MR}





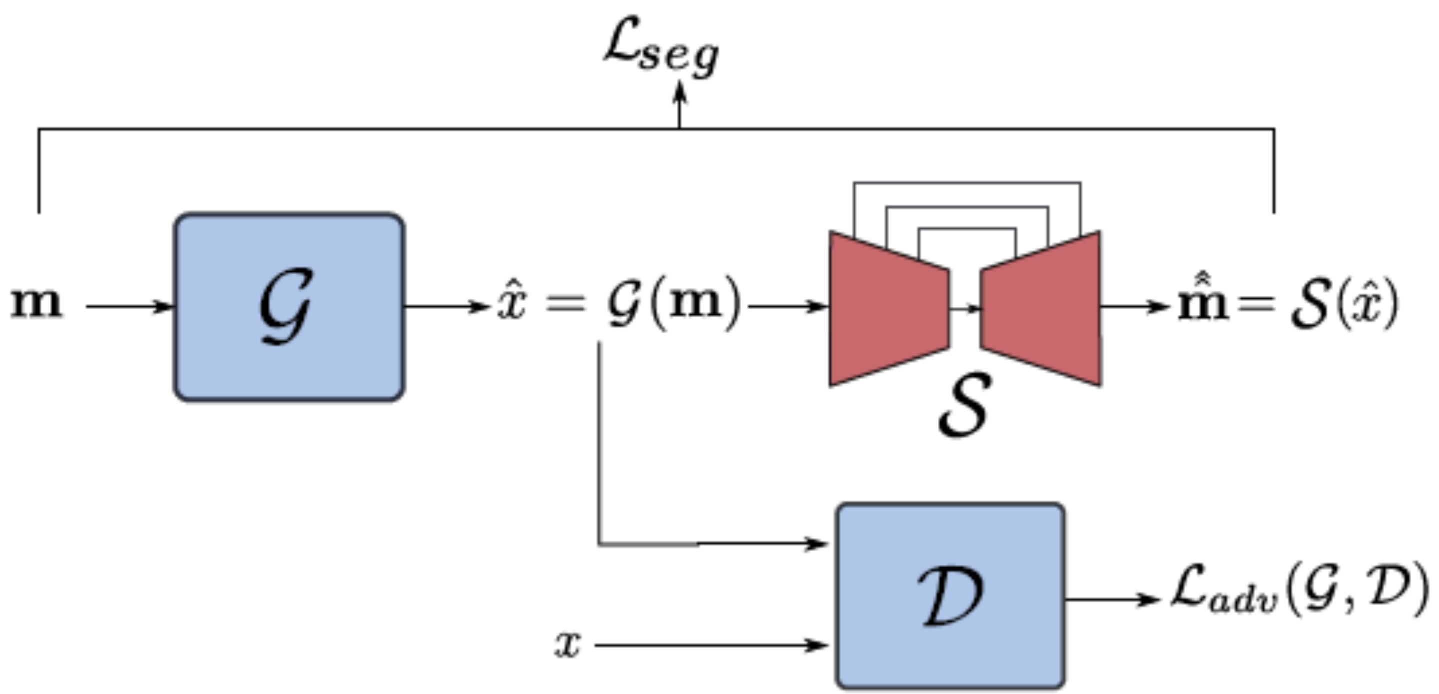
Content

- Motivation
- Method
- Datasets and pre-processing
- Experiments and Results
  - CT Label-to-CT Scan
  - CT Label-to-MR Scan
  - Failure Cases
  - Ablation on 3D noise
- Conclusion

Method: Baseline model

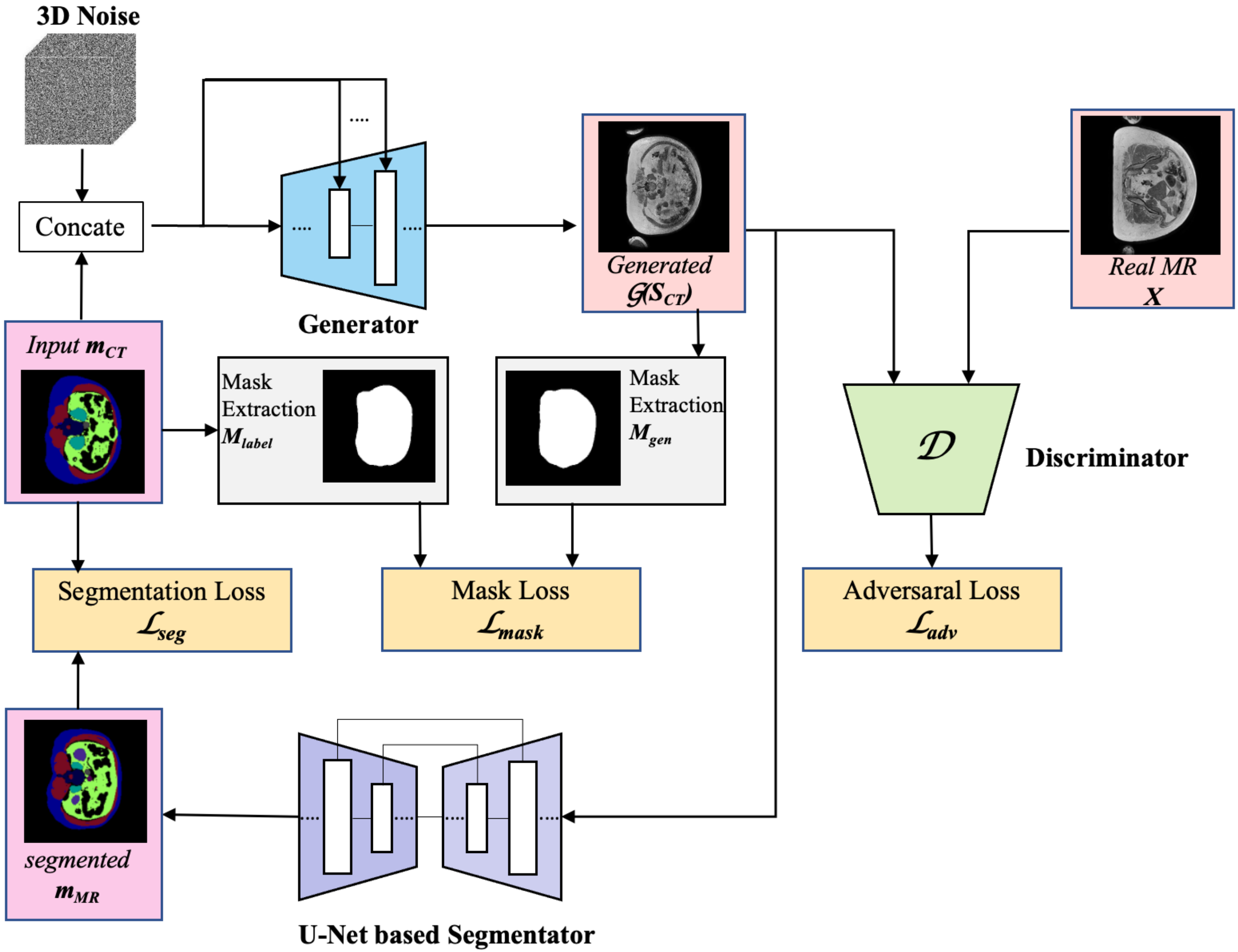
Baseline model: USIS-Unsupervised Semantic Image Synthesis

- Generator
- Semantic Consistency Network
- Unconditional Discriminator



2. G. Eskandar, M. Abdelsamad, K. Armanious and B. Yang, “Usis: Unsupervised semantic image synthesis,” Computers & Graphics, vol. 111, pp. 14–23, 2023.

Med-USIS Framework



Method: Overall Training and Optimization

- Loss function for Generator

$$\mathcal{L}_{Gen} = \lambda_{seg}\mathcal{L}_{seg}(m, \mathcal{S}(\mathcal{G}(m))) + \mathcal{L}_{advG}(\mathcal{D}(\mathcal{G}(m))) + \lambda_{mask}\mathcal{L}_{mask}(m, \mathcal{G}(m))$$

- Loss function for Discriminator

$$\mathcal{L}_{Dis} = \mathcal{L}_{advD}(\mathcal{D}(x), \mathcal{D}(\mathcal{G}(m)))$$

- Loss function for U-Net

$$\mathcal{L}_{seg} = -\mathbb{E}_m\left[\sum_{c=1}^{C}\alpha_c\sum_{i=1}^{H}\sum_{j=1}^{W} m_{c,i,j}\log(\mathcal{S}(\mathcal{G}(m))_{c,i,j})\right] \qquad \alpha_c = \frac{H \times W}{\sum_{i,j}^{H\times W}\mathbb{E}_m\left[1[m_{c,i,j}]\right]}$$

- Mask Consistency Loss

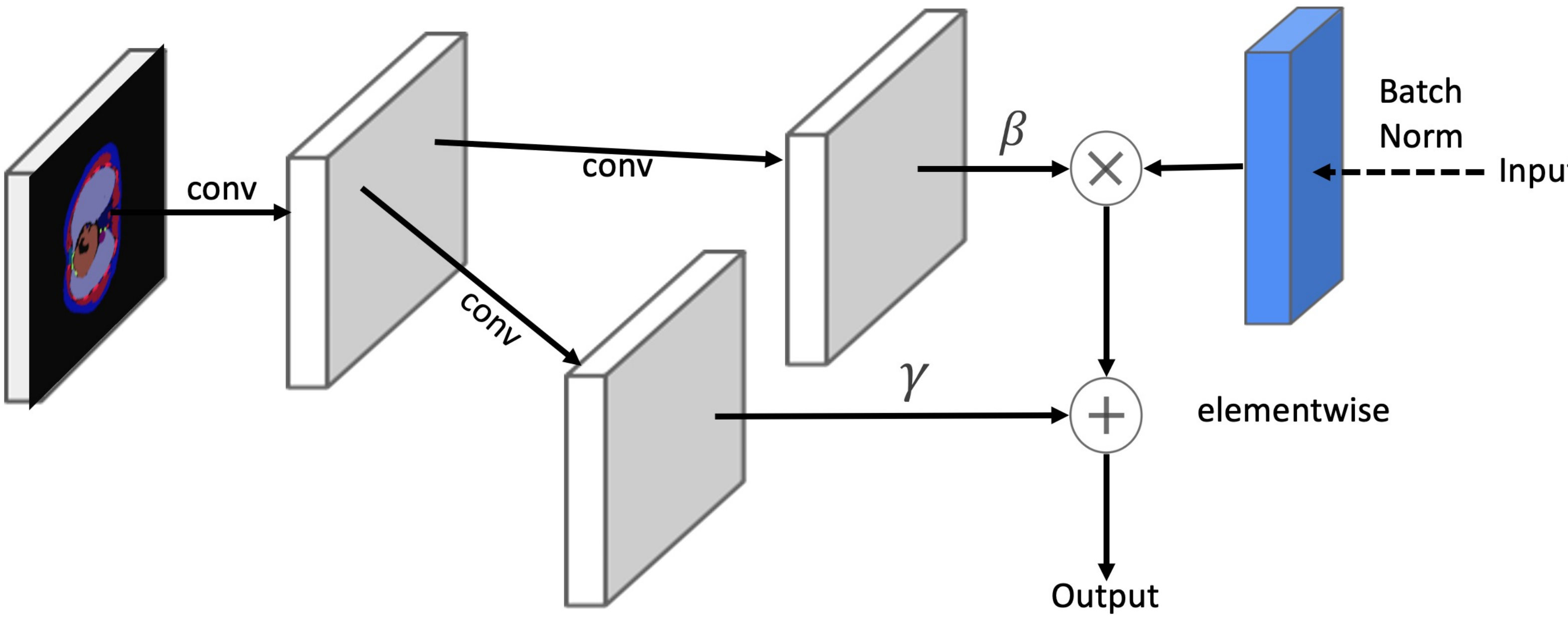
$$\mathcal{L}_{\text{mask}} = \|\mathcal{E}(m) - \mathcal{E}(\mathcal{G}(m))\|_2^2$$

- Inference phase: Only the Generator part is used

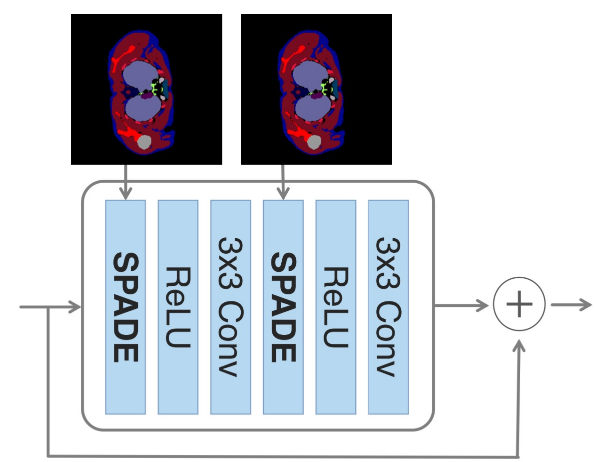
Method: Generator

Generator: based on SPADE structure

Problem: normalization layers tend to *"**wash away**"* semantic information



Solution: Implementing ***spatially-adaptive normalization*** **(SPADE)** after batch normalization

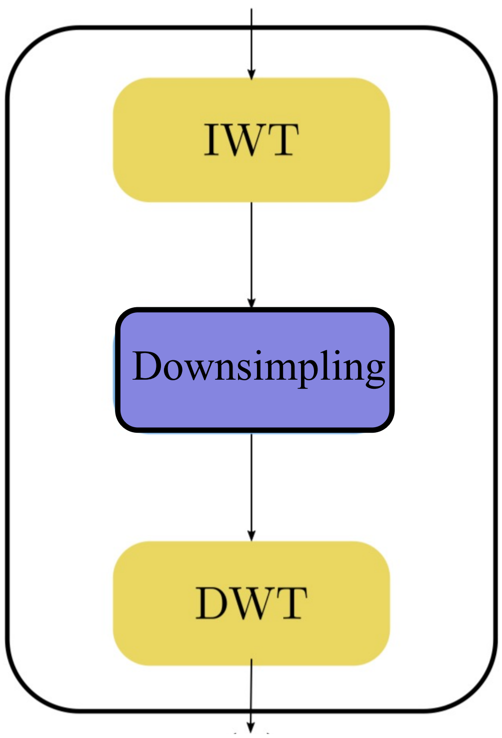


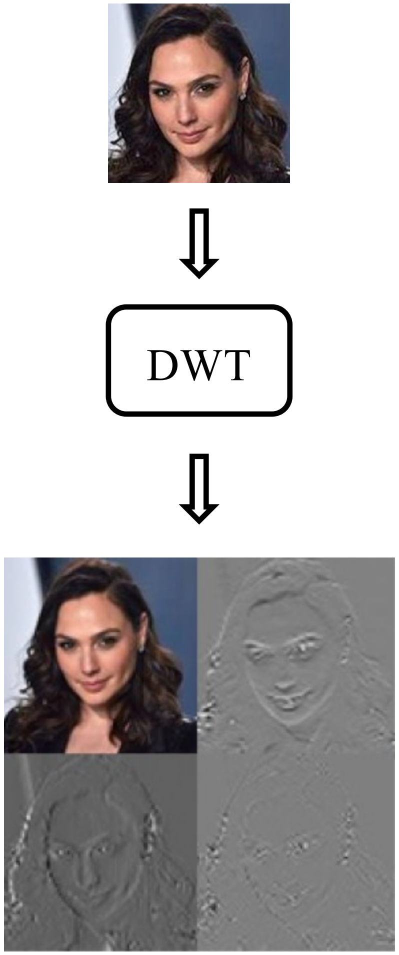
SPADE ResBlk

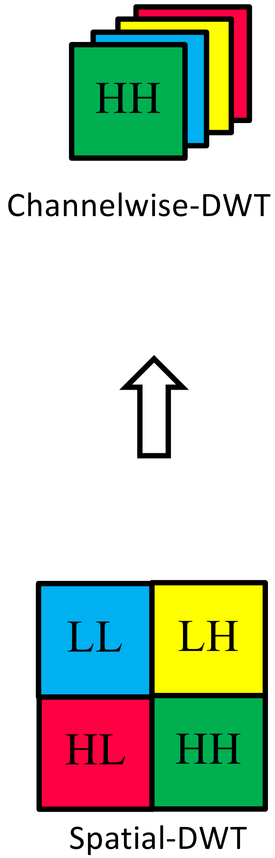
3. T. Park, M. Liu, T. Wang and J. Zhu, "Semantic image synthesis with spatially-adaptive normalization," CoRR, vol. abs/1903.07291, 2019.

Method: Discriminator

Discriminator: based on Wavelet-transform





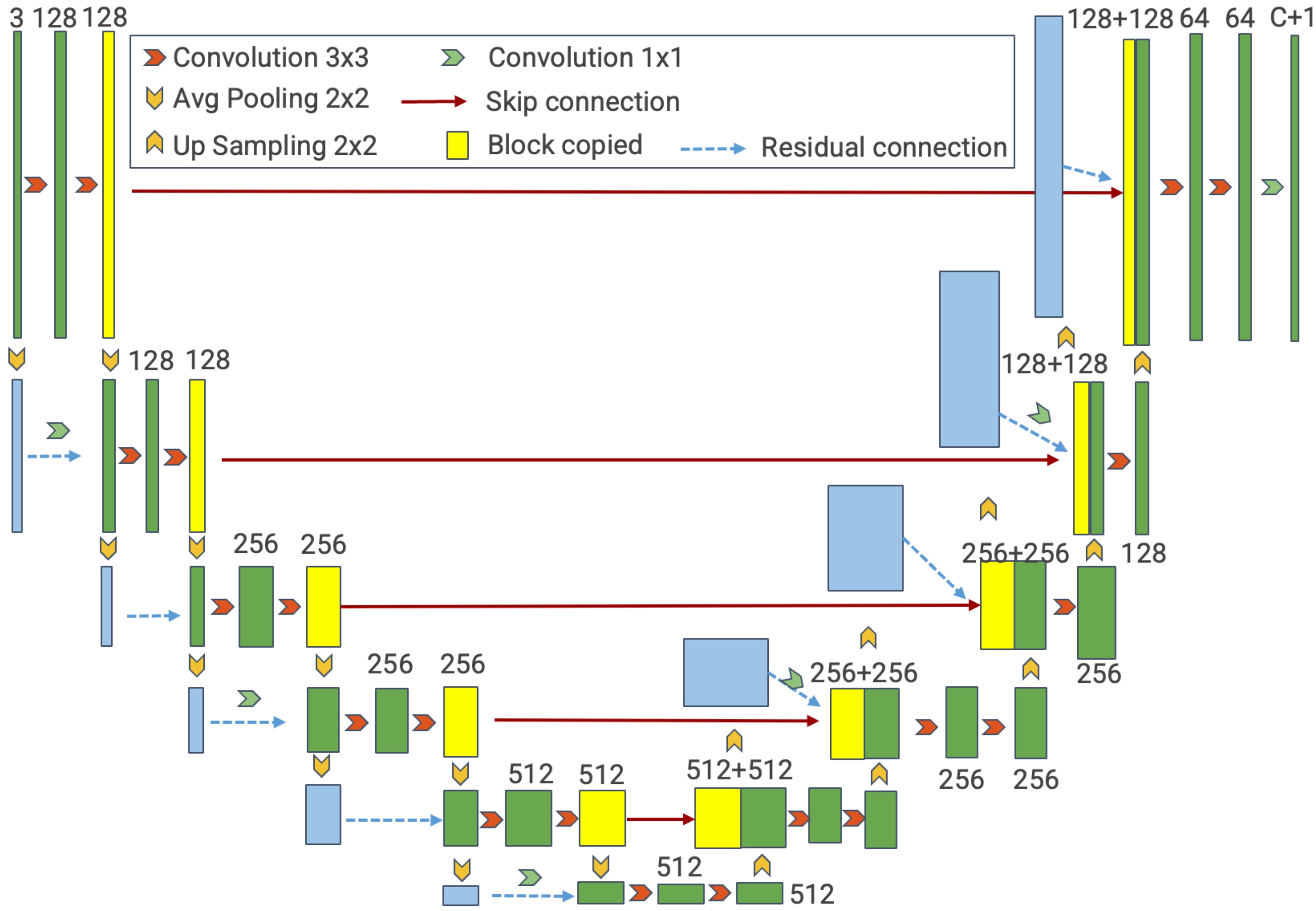


T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila, “Analyzing and improving the image quality of stylegan,” 2020

Method: U-Net

U-Net based residual segmentation network

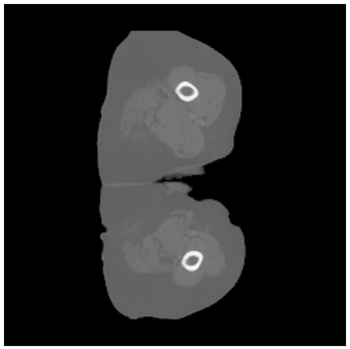
- Enforces semantic alignment
- Preserves relationship between input pixels



Method: Shape consistency loss

Shape consistency loss (Mask loss)

Generated image



Semantic map

Traditional image processing methods

Binarization
Normalization
Binary opening
Connected component analysis

Mask from label
Use it as pseudo-groundtruth

L2-Loss

Method: Metrics for evaluation

- Fréchet Inception Distance (FID)

$$\mathrm{FID}(X, \hat{X}) = \|\mu_x - \mu_{\hat{x}}\|_2^2 + \mathrm{tr}(\Sigma_x + \Sigma_{\hat{x}} - 2(\Sigma_x \Sigma_{\hat{x}})^{\frac{1}{2}})$$

- Learned Perceptual Image Patch Similarity (LPIPS)

$$\mathrm{LPIPS} = d(\phi(x), \phi(\hat{x}_i))$$

- Structural Similarity Index (SSIM)

$$\mathrm{SSIM}(x, \hat{x}) = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}$$

- Peak Signal-to-Noise Ratio (PSNR)

$$\mathrm{PSNR} = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

- Root Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2}$$

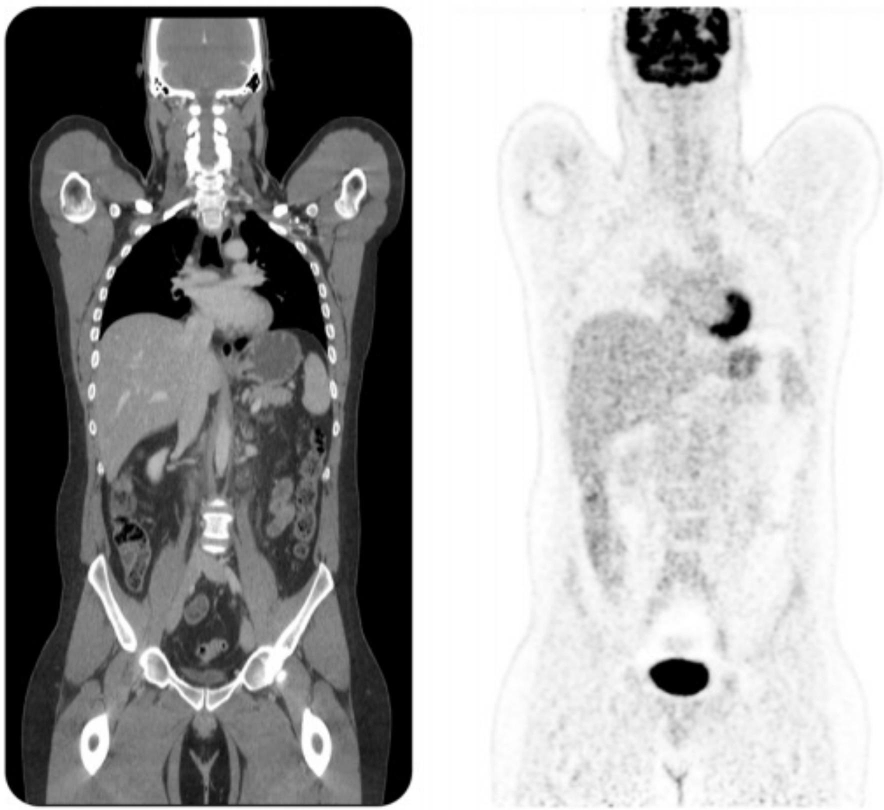
Content

- Motivation
- Method
- Datasets and Pre-processing
- Experiments and Results
  - CT Label-to-CT Scan
  - CT Label-to-MR Scan
  - Failure Cases
  - Ablation on 3D noise
- Conclusion

Datasets and Pre-processing

AutoPET Dataset

AutoPET dataset is a comprehensive collection of CT and PET imaging from1,014 studies (900 patients), It contains **37 annotated classes**, which are highly imbalanced.



CT classes of AutoPET dataset

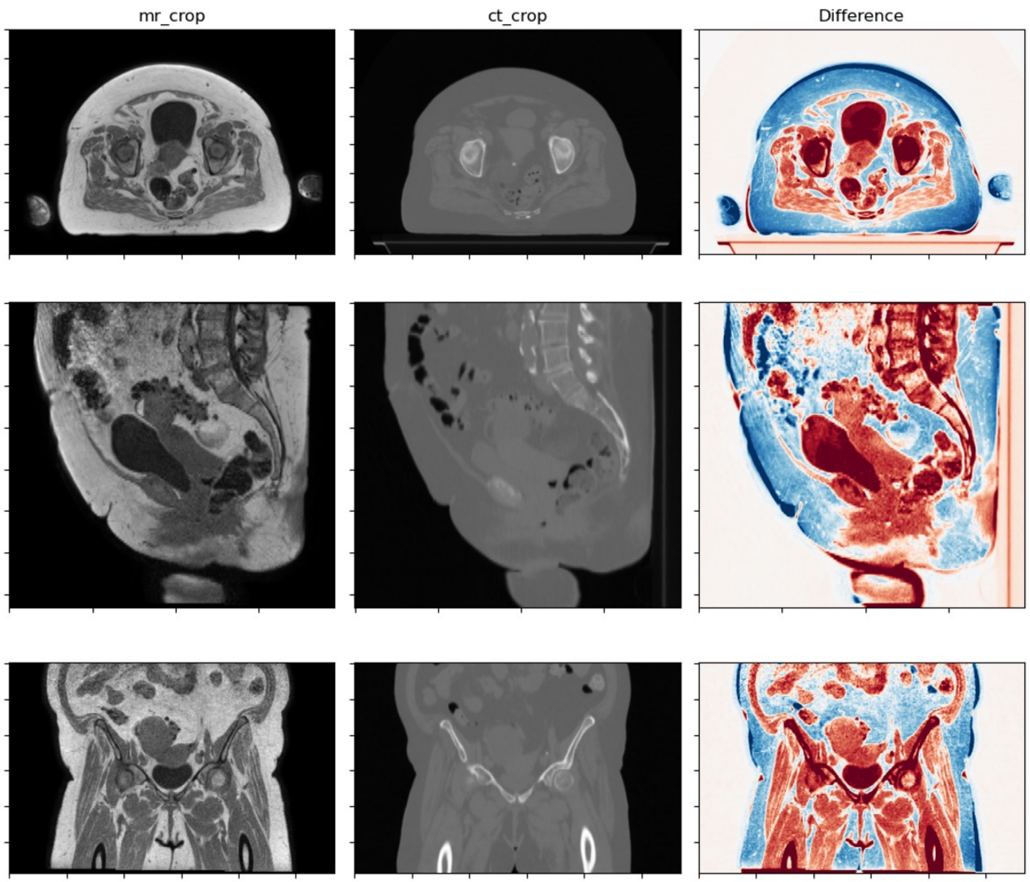
ID	Percent/%	Class ID	Percent/%	Class ID	Percent/%
0	**81.4155**	13	0.0016	26	0.2103
1	0.0025	14	0.0252	27	0.4839
2	0.0827	15	0.2208	28	0.0245
3	0.0425	16	0.0073	29	0.0180
4	0.3732	17	0.1037	30	0.0397
5	0.2138	18	0.0020	31	0.0014
6	0.1233	19	0.0063	32	0.0062
7	0.6116	20	0.0025	33	**5.7686**
8	0.0214	21	0.2879	34	**6.5854**
9	0.0907	22	0.0009	35	1.6175
10	0.0067	23	0.0098	36	0.1043
11	0.0286	24	0.1551		
12	1.2245	25	0.0803		

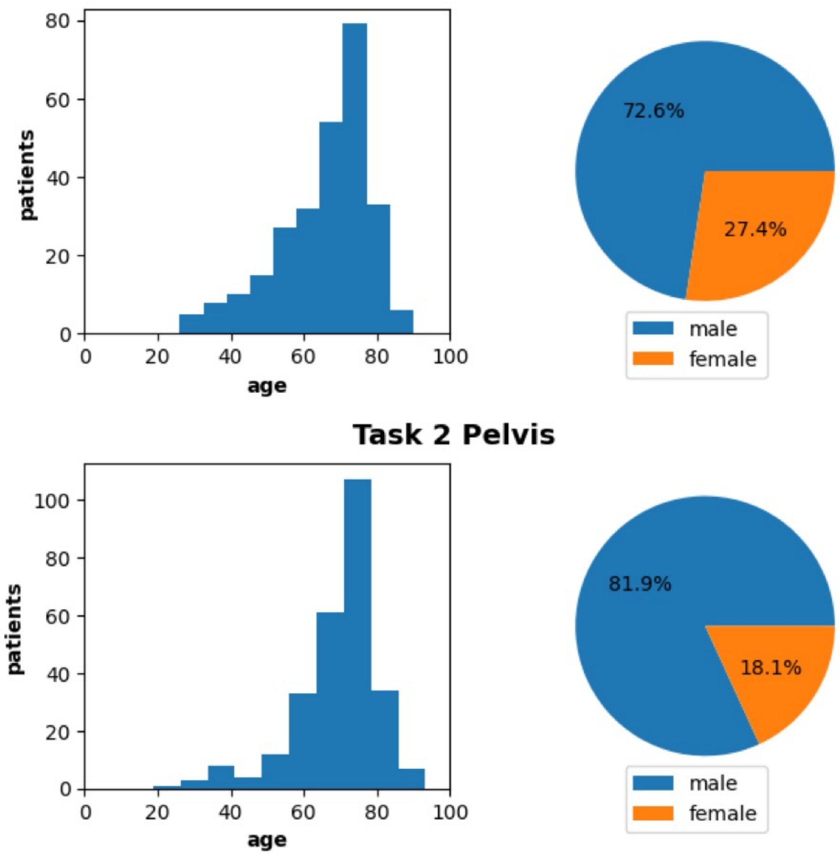
Gatidis S, Hepp T, Früh M, La Fougère C, Nikolaou K, Pfannenberg C, Schölkopf B, Küstner T, Cyran C, Rubin D. A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. Sci Data. 2022 Oct 4;9(1):601. doi: 10.1038/s41597-022-01718-3. PMID: 36195599; PMCID: PMC9532417.

Datasets and Pre-processing

SynthRAD2023 Dataset

SynthRAD2023 dataset, encompassing CT, CBCT, and T1-weighted MRI images of 540 brain and pelvic radiotherapy patients. There are **paired CT, MR images**, but corresponding **semantic maps are not included**



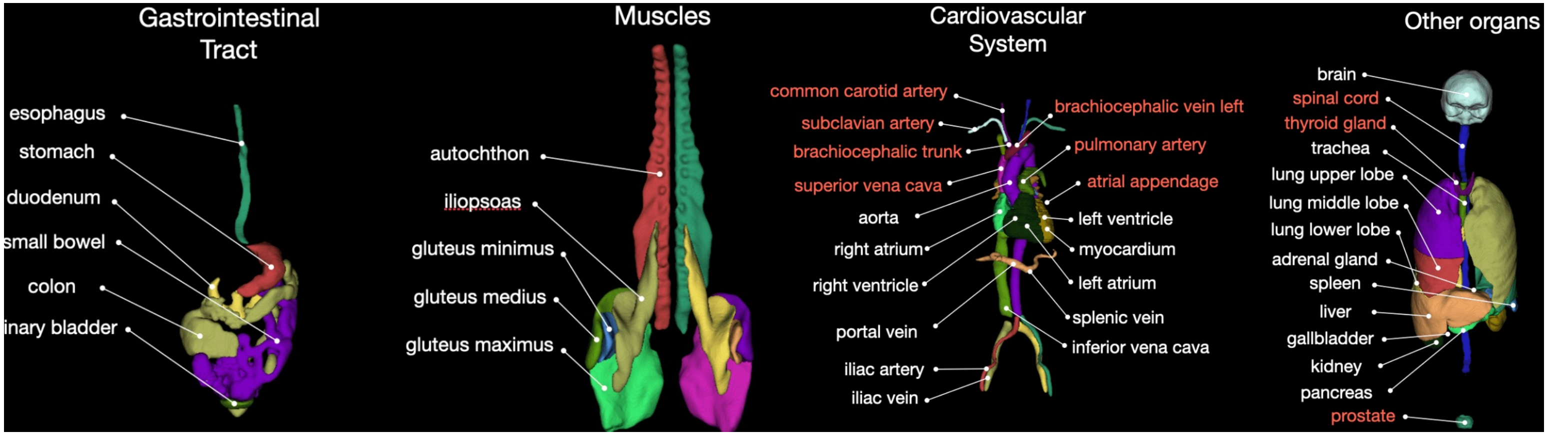


1. A. Thummerer, E. van der Bijl, A. Galapon, J. J. C. Verhoeff, J. A. Langendijk, S. Both, C. N. A. T. van den Berg and M. Maspero, "Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy," Medical Physics, vol. 50, no. 7, p.4664–4674, Jun. 2023

Datasets and Pre-processing

Implement TotalSegmentator to segment CT images

TotalSegmentator can robustly segment 117 anatomical structures in CT images. We use it to **generate CT semantic maps** for SynthRAD2023 Dataset, and selected, merged some classes and ended up with a total of **31 classes**.



Segmented CT classes of SynthRAD2023 [1] dataset

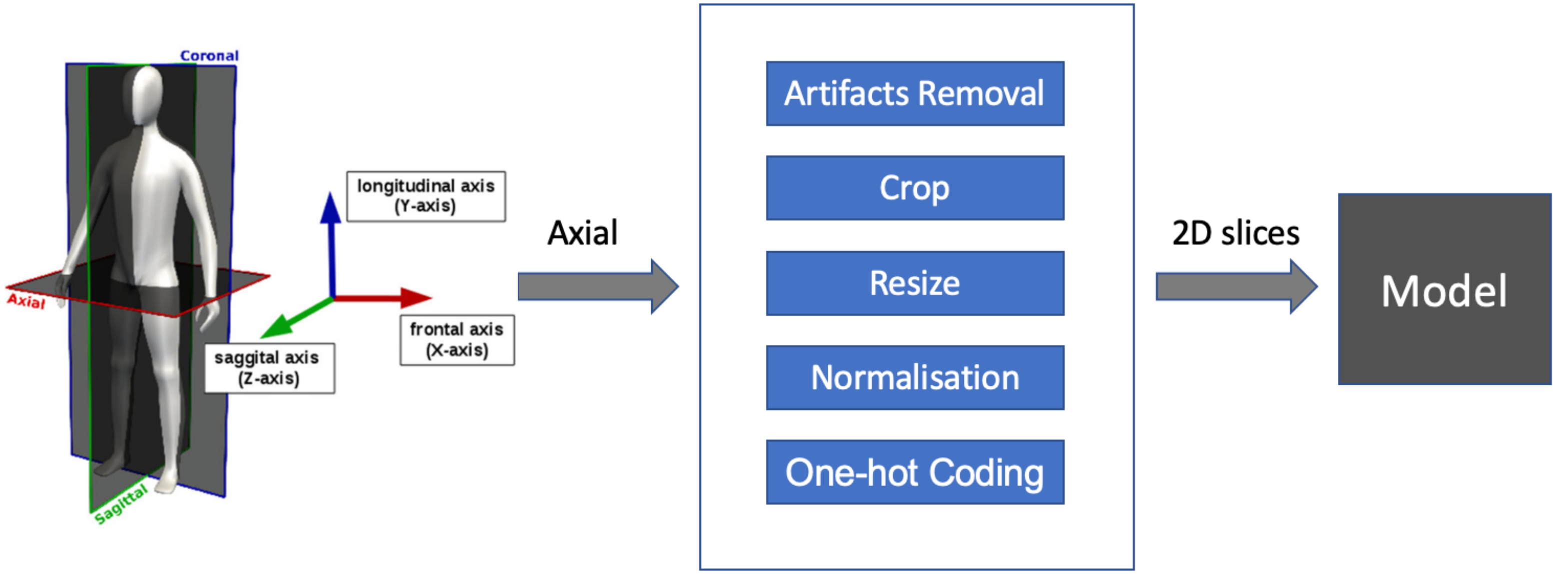
ID	Class	ID	Class	ID	Class	ID	Class
0	Background	8	Lung	16	Colon	24	Autochthon
1	Kidney	9	Vertebrae	17	Ribs	25	Iliopsoas
2	Vessels	10	Esophagus	18	Humerus	26	Urinary Bladder
3	Gallbladder	11	Trachea	19	Scapula	27	Skin
4	Liver	12	Heart	20	Clavicula	28	Spleen
5	Stomach	13	Pulmonary Artery	21	Femur	29	Fat
6	Pancreas	14	Small Bowel	22	Hips	30	Skeletal Muscle
7	Adrenal	15	Duodenum	23	Sacrum		

J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach and M. Segeroth, "Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images," Radiology: Artificial Intelligence, vol. 5, no. 5, Sep. 2023.

Datasets and Pre-processing

Preprocessing

Our model is based on 2D image, so the 3D .nifti images are cut into 2D slices, which are saves as PNG format later.



- Min-max normalisation

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \cdot (new\_max - new\_min) + new\_min$$

- One-hot coding

$$E_{i,x,y} = \begin{cases} 1, & \text{if } m_{x,y} = i \\ 0, & \text{otherwise} \end{cases}$$

Content

- Motivation
- Method
- Datasets and Pre-processing
- Experiments and Results
  - CT Label-to-CT Scan
  - CT Label-to-MR Scan
  - Failure Cases
  - Ablation on 3D noise
- Conclusion

Experiments and Results

Several Models were evaluated:

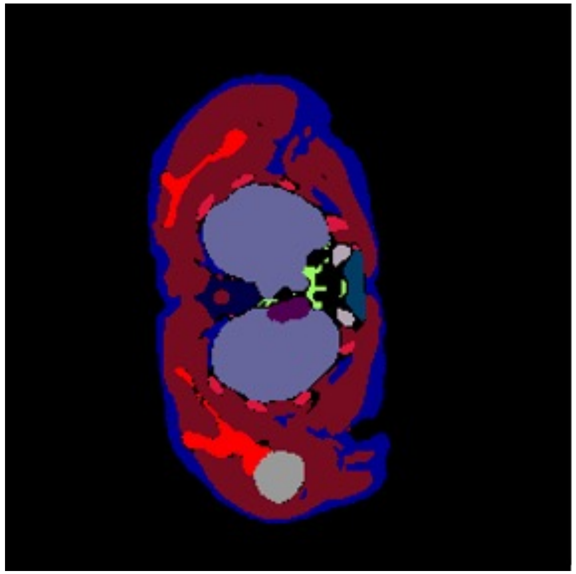
1. **Unpaired**
   - OASIS Generator
     - Without Mask Loss
     - With Mask Loss
   - Wavelet Generator
     - Without Mask Loss
     - With Mask Loss
2. **Paired**
   - OASIS Generator
   - Wavelet Generator

Results

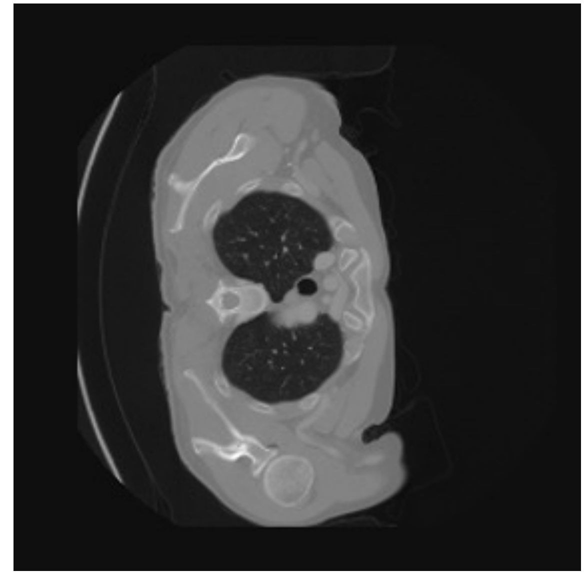
1- CT-Label-to-CT-Image based on AutoPET

Qualitative Results - Shape inconsistency

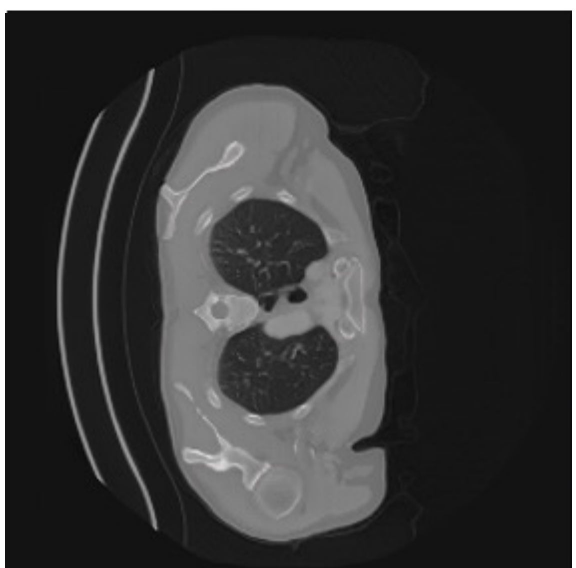
Label map



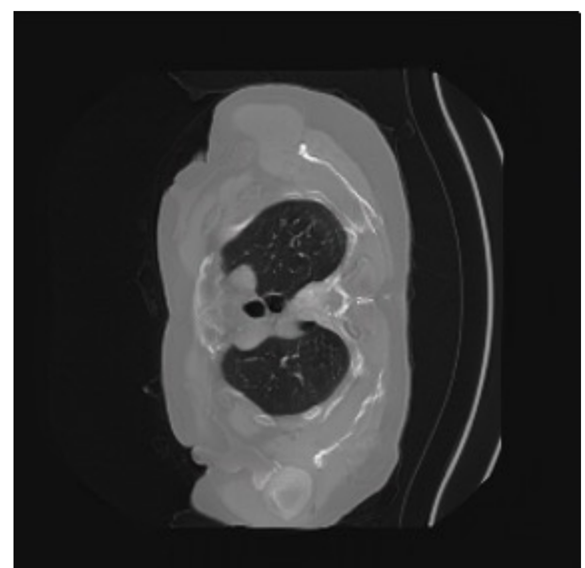
Ground truth



Generated 1

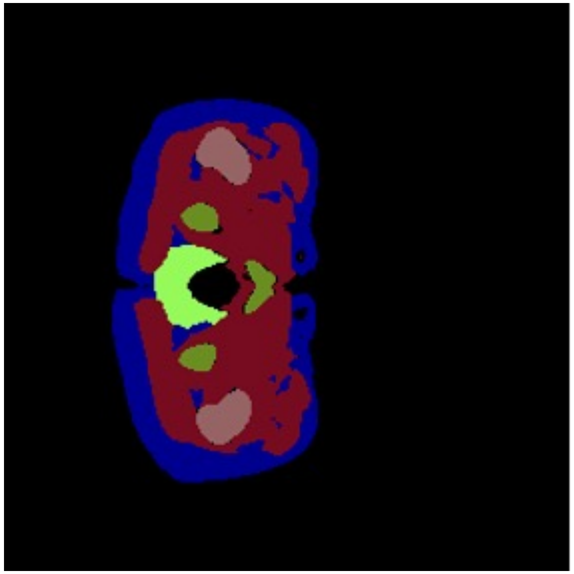


Generated 2

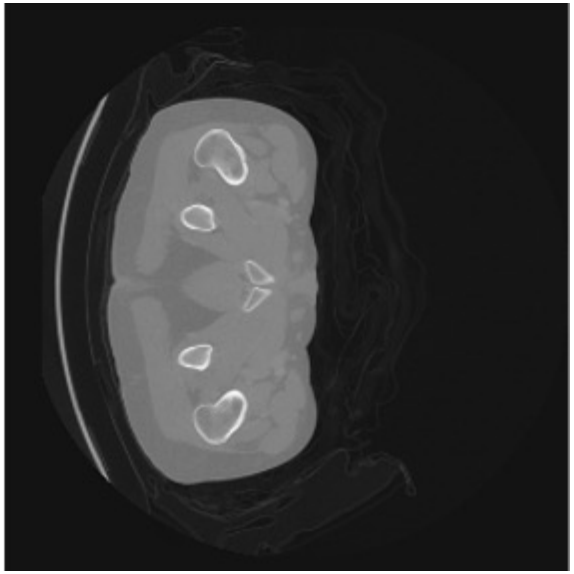


Qualitative Results - Shape inconsistency

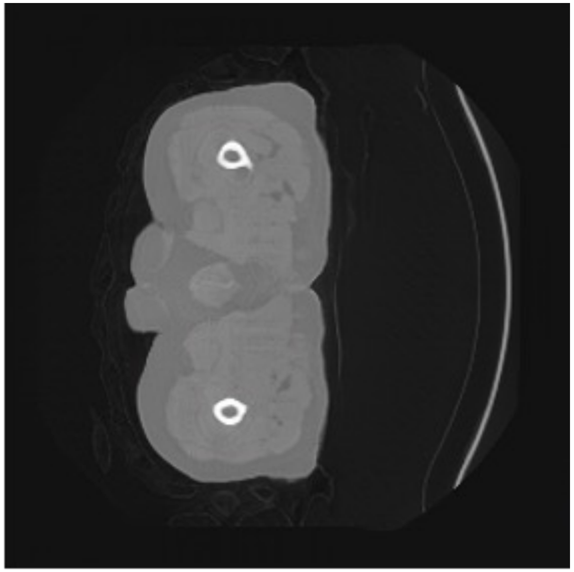
Label map



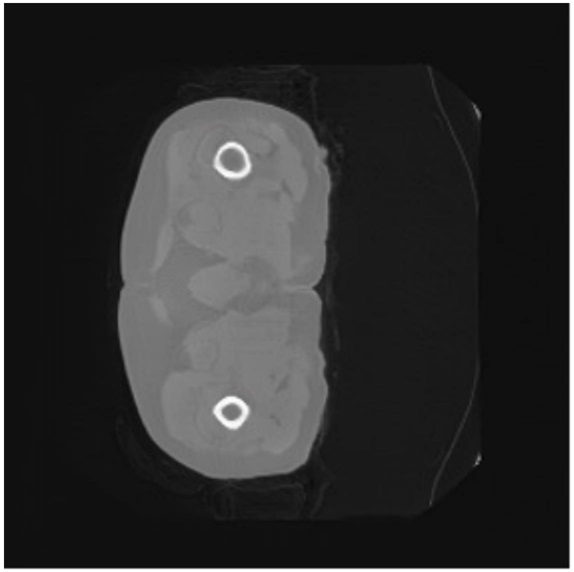
Ground truth



Generated 1

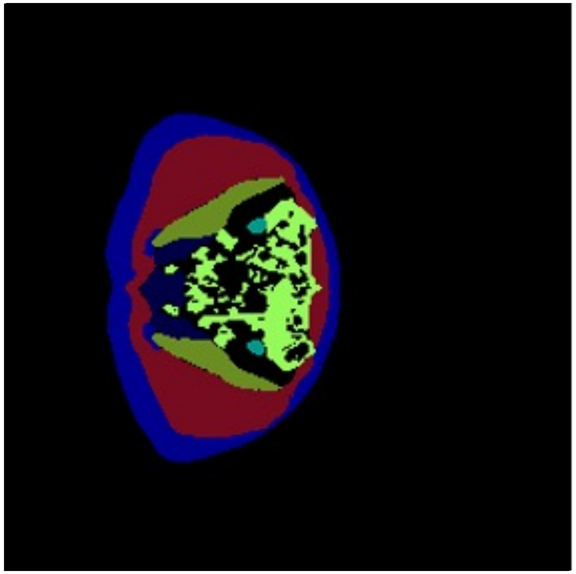


Generated 2

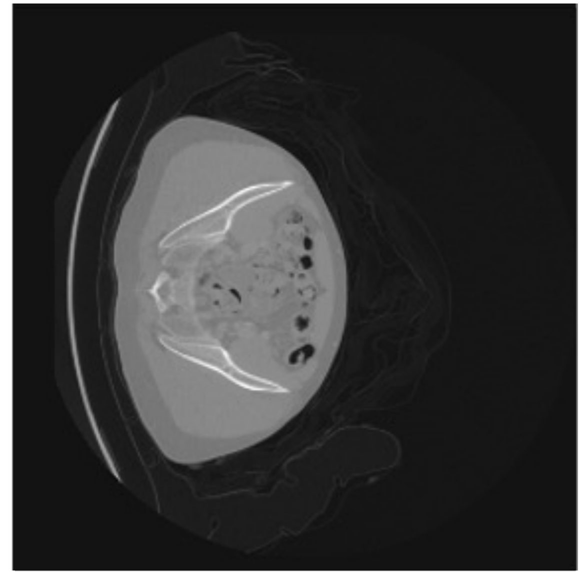


Qualitative Results - Shape inconsistency

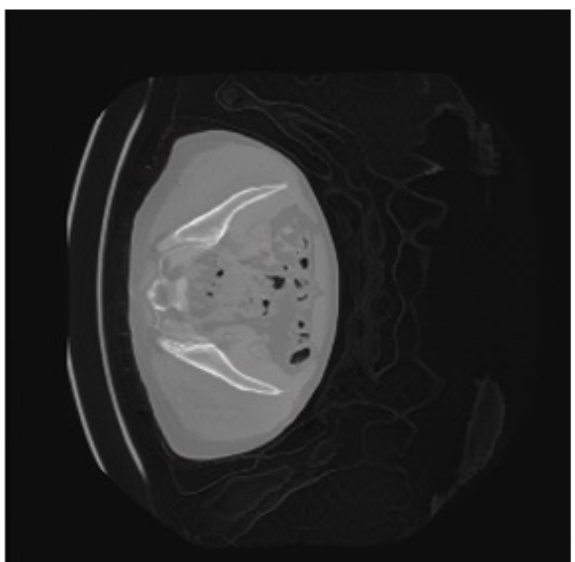
Label map



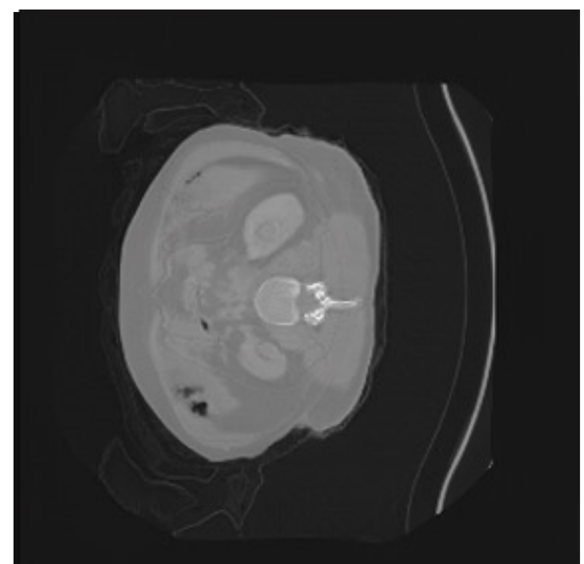
Ground truth



Generated 1



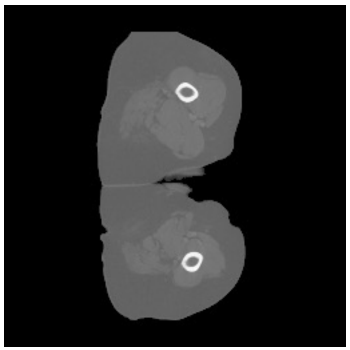
Generated 2



Model

Shape consistency loss (Mask loss)

Fake



Input

Traditional image
processing methods

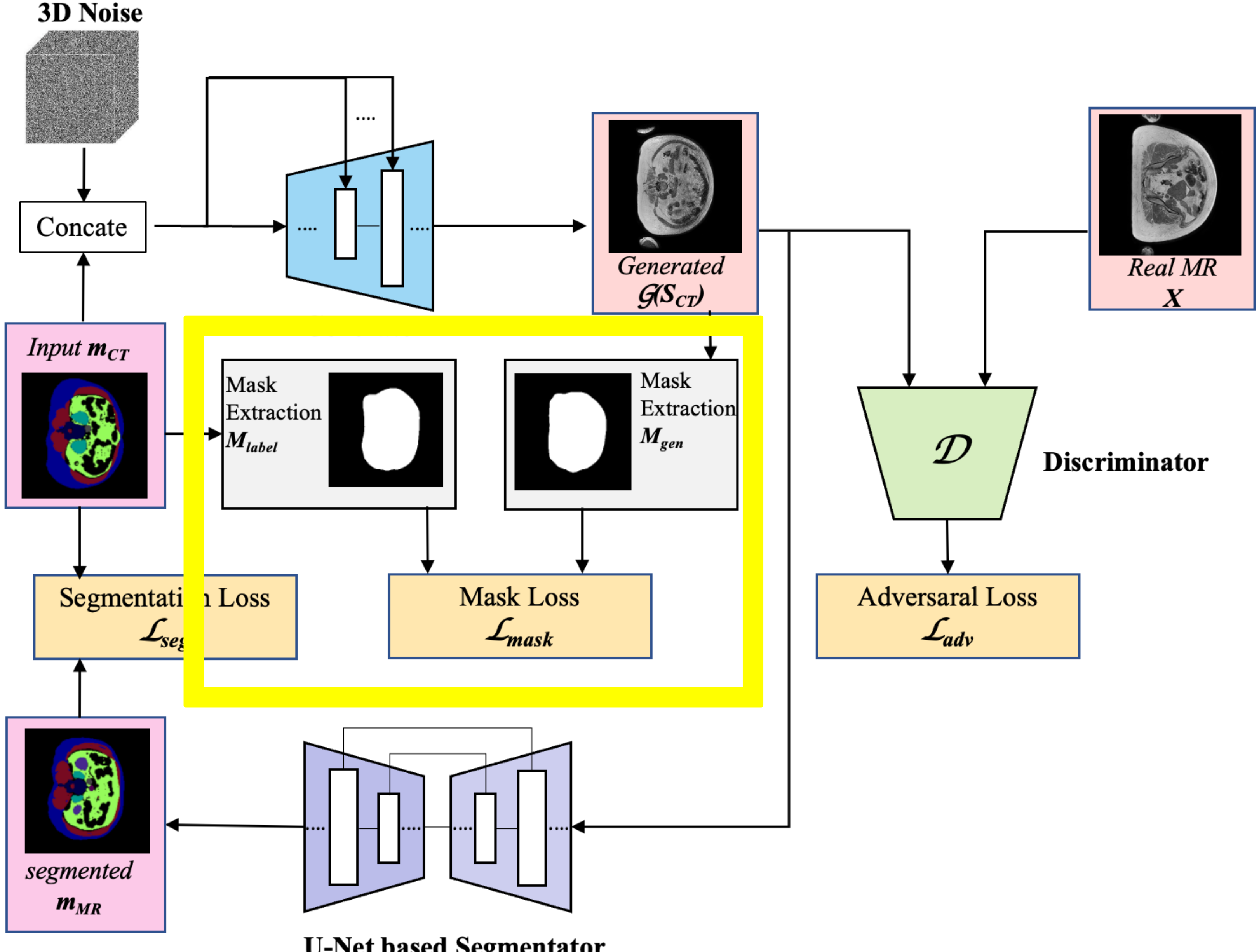
Binarization
Normalization
Binary opening
Connected component analysis

Mask from label
Use it as pseudo-groundtruth

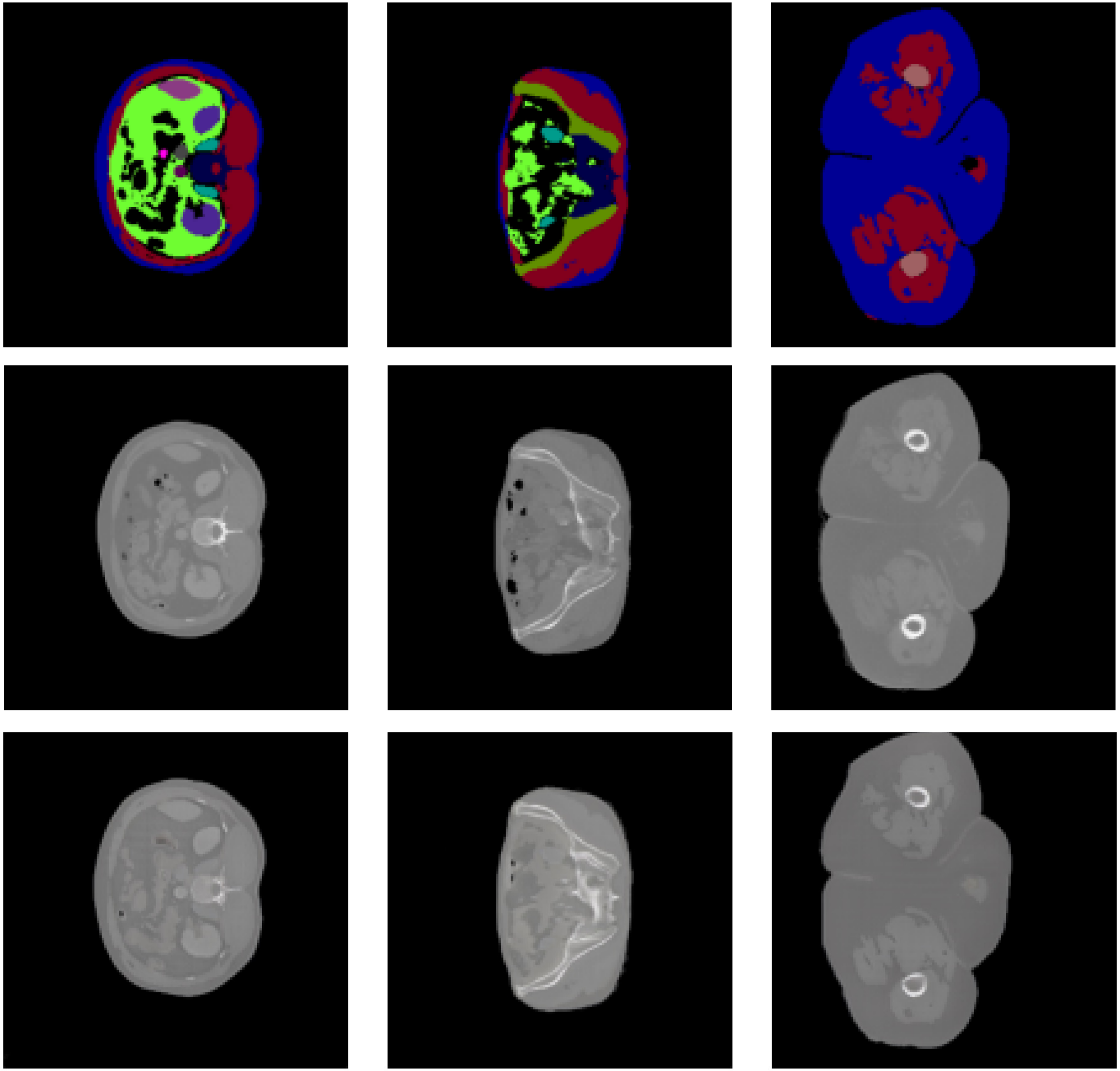
L2-Loss

Model

Med-USIS Framework



Qualitative Results with Mask Loss



Label

Real

Fake

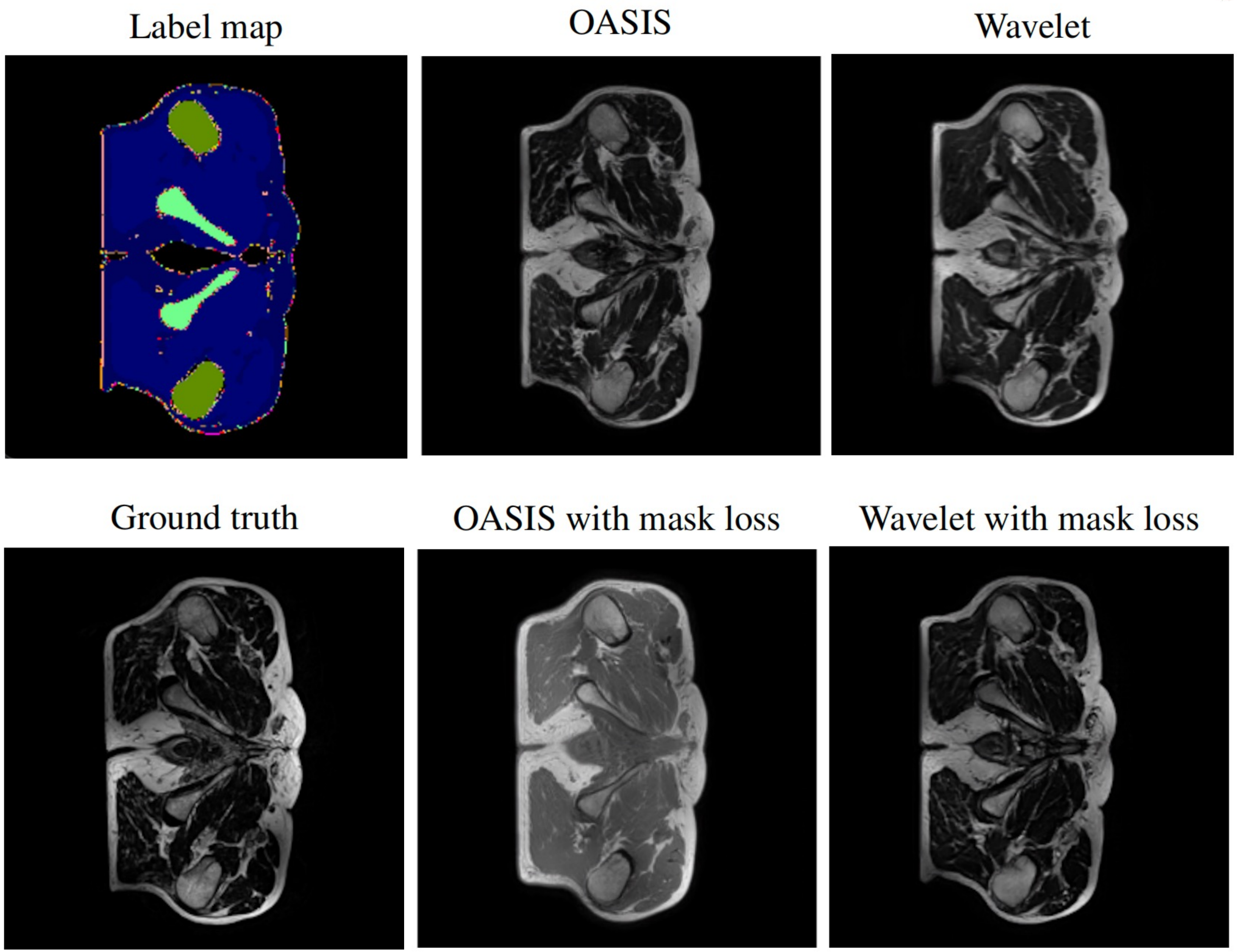
Results on AutoPET - Metrics

Learning Paradigm	Model	FID ↓	LPIPS (VGG) ↓	SSIM ↑	PSNR ↑	RMSE ↓
Paired	Wavelet	7.29	0.06	0.9995	**24.89**	0.06
	Wavelet+ Mask loss	10.68	**0.05**	**0.9995**	23.27	**0.06**
	OASIS	15.83	0.29	0.9714	15.39	0.923
	OASIS+ Mask loss	**5.67**	0.22	0.9713	7.04	0.45

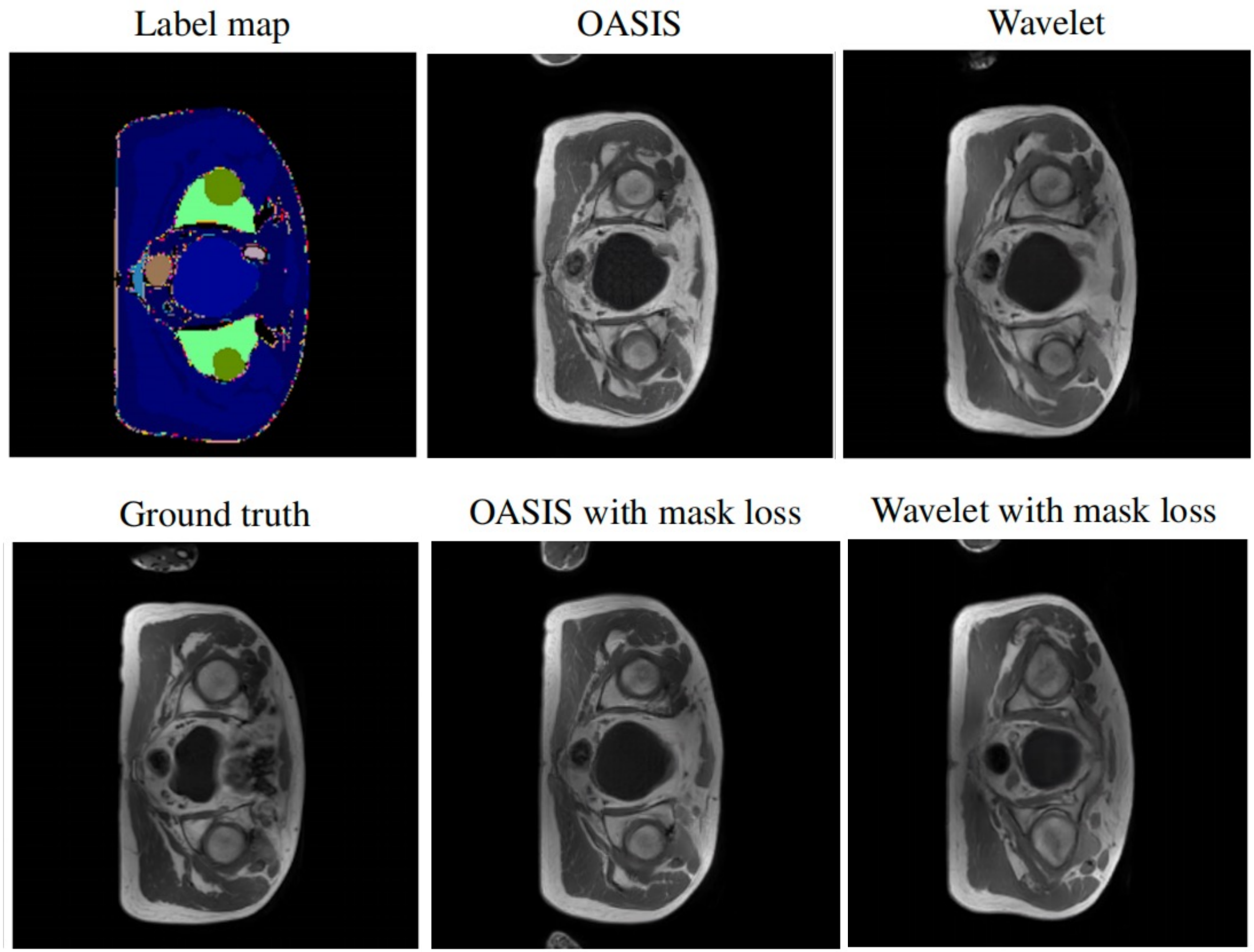
IoU of the The four most important Classes is about 0.75

2- CT-Label-to-MR-Image based on SynthRAD2023

Qualitative Results- Unsupervised



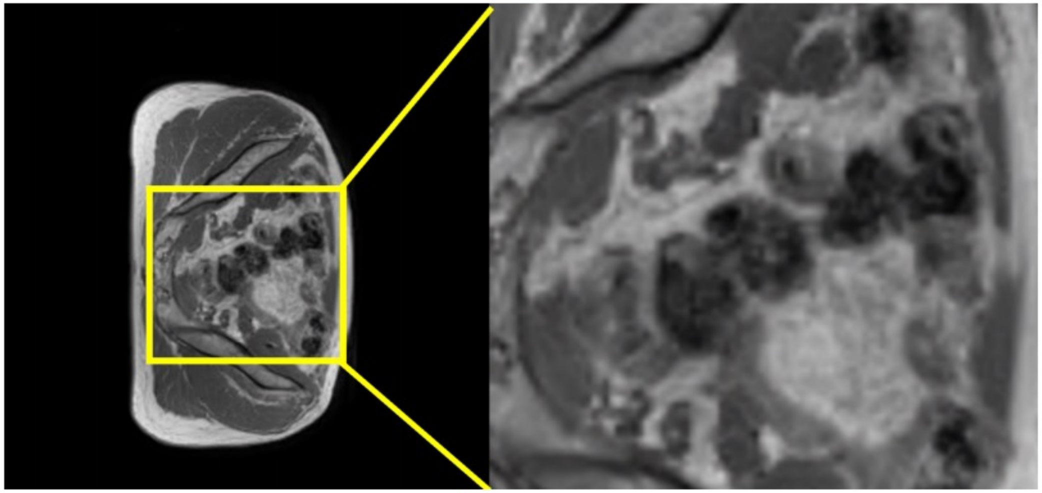
Qualitative Results- Unsupervised



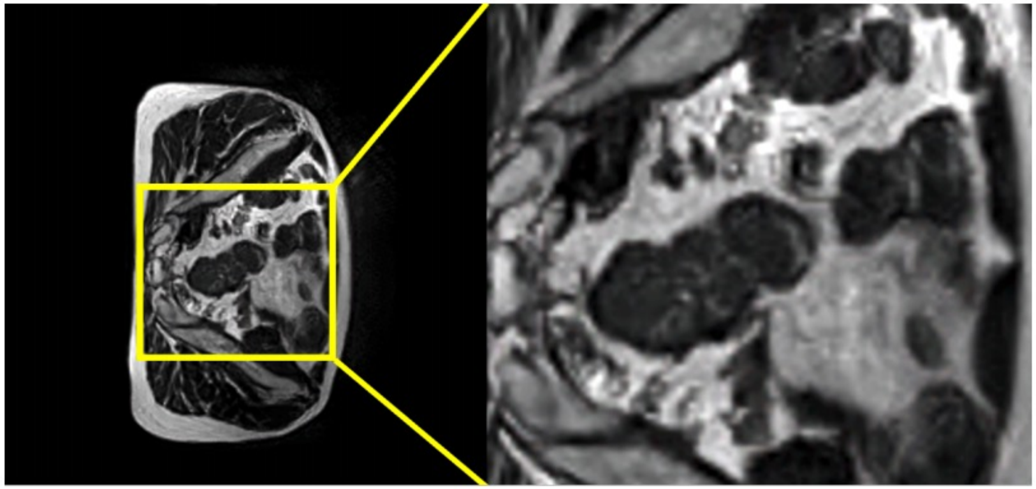
Qualitative Results- Unsupervised

Qualitative comparation of generators

OASIS generator



Wavelet generator

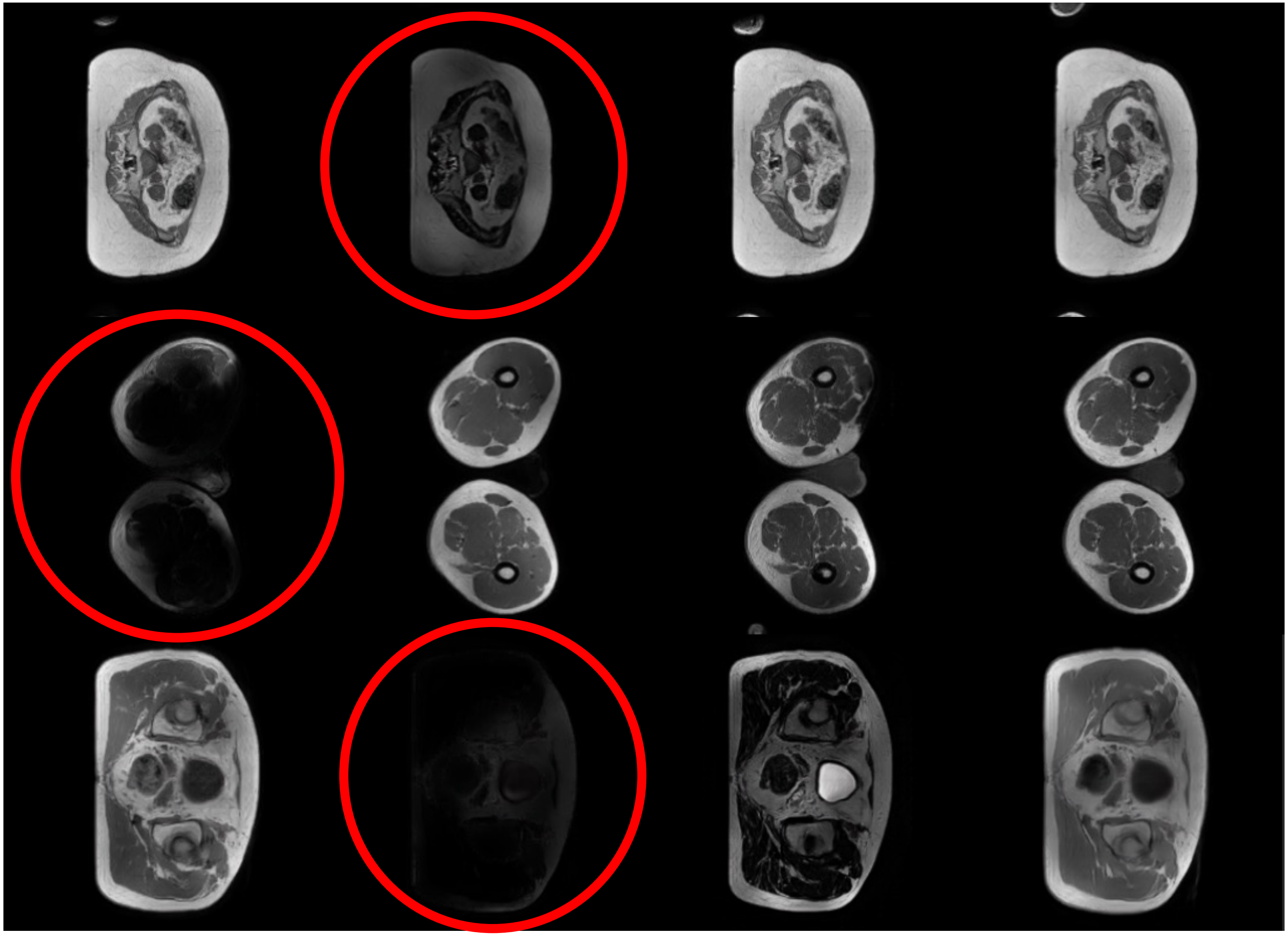


Results on SynthRAD2023 - Metrics

Learning Paradigm	Model	FID ↓	LPIPS (VGG) ↓	SSIM ↑	PSNR ↑	RMSE ↓
Unpaired	Wavelet	**51.92**	0.15	0.9983	18.69	0.12
	Wavelet + mask loss	54.53	**0.15**	**0.9984**	**18.77**	**0.12**
	OASIS	60.93	0.15	0.9983	18.54	0.12
	OASIS + mask loss	59.97	0.16	0.9980	18.09	0.12

Failure Cases

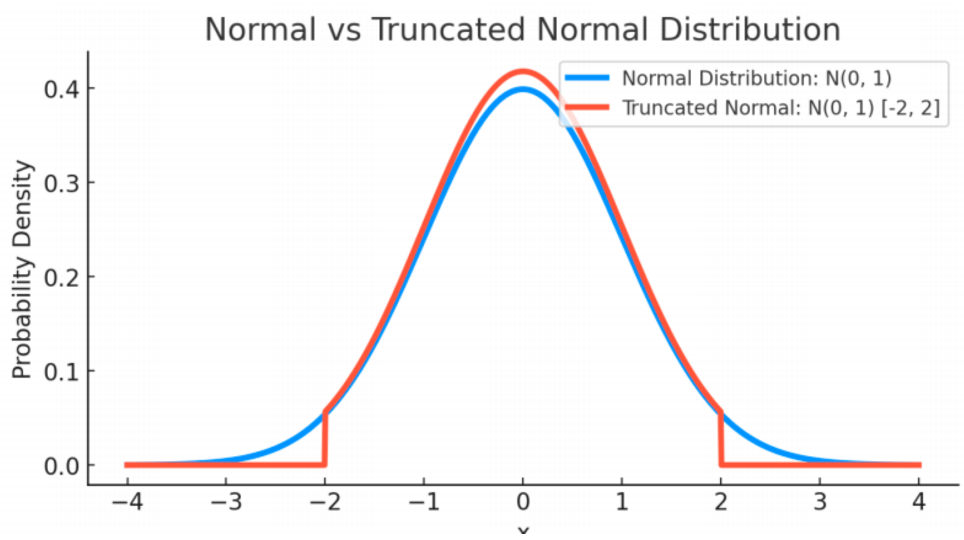
Unstable grey value of different random noise inputs.

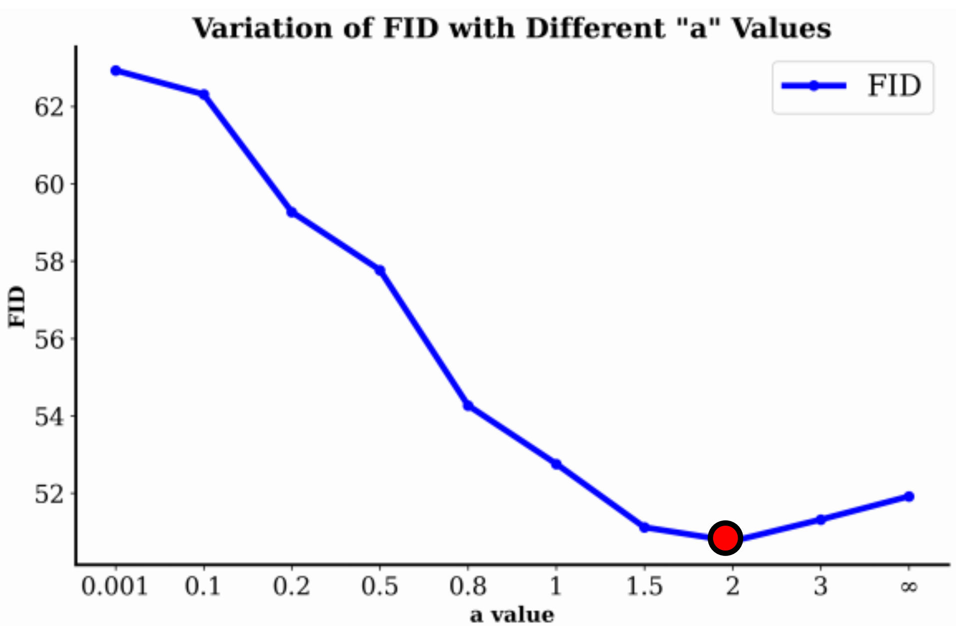


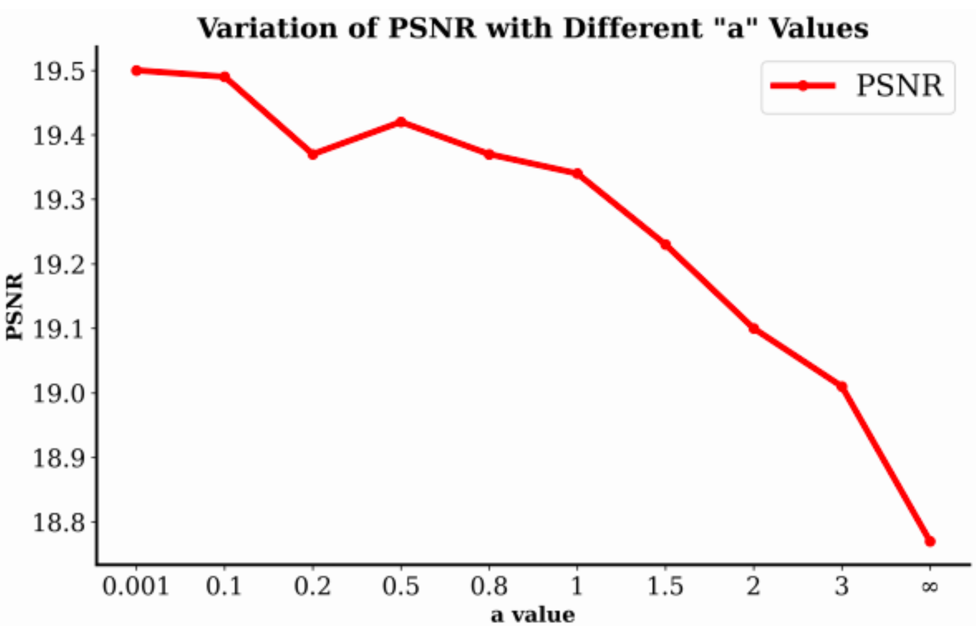
Ablation on 3D noise

Truncated normal distribution as noise input during inference

- The noise input remains within a certain range [-a, a]
- Controlled variability and enhances model stability without the risk of extreme values







Ablation on 3D noise

One to one mapping without 3D noise input during training

$$\mathcal{G} : (\mathbf{z}, \mathbf{m}) \rightarrow \hat{\mathbf{x}} \quad \Longrightarrow \quad \mathcal{G} : (\mathbf{m}) \rightarrow \hat{\mathbf{x}}$$

3D noise input	FID	LPIPs	SSIM	RMSE	PSNR
w	51.92	0.15	0.9984	0.12	18.77
w/o	60.35	0.26	0.9951	0.19	14.19

- Due to the guidance of semantic map, the variation because of random noise are only showed in Image brightness and small details
- It will reduce the diversity of generated images and not good for privacy preservation.

Content

- Motivation
- Model
- Datasets and pre-processing
- Experiments and Results
  - CT Label-to-CT Scan
  - CT Label-to-MR Scan
  - Failure Cases
  - Ablation on 3D noise
- Conclusion

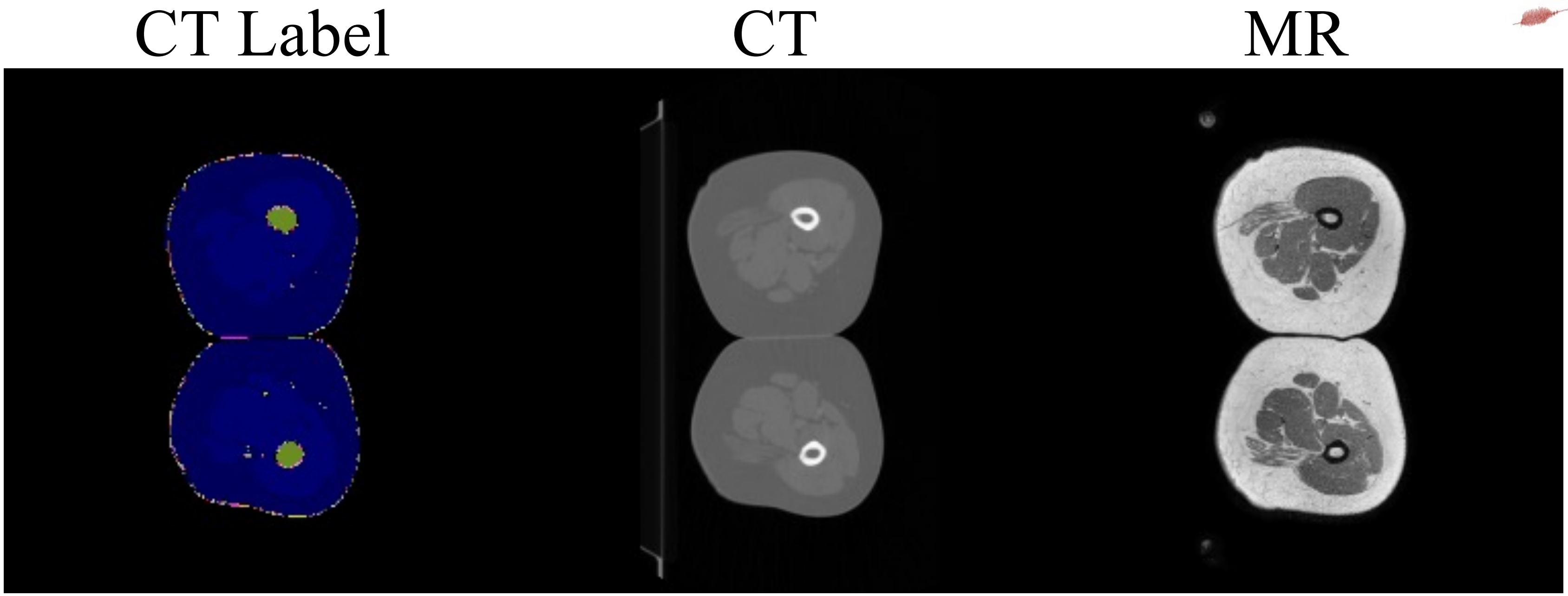
Conclusion

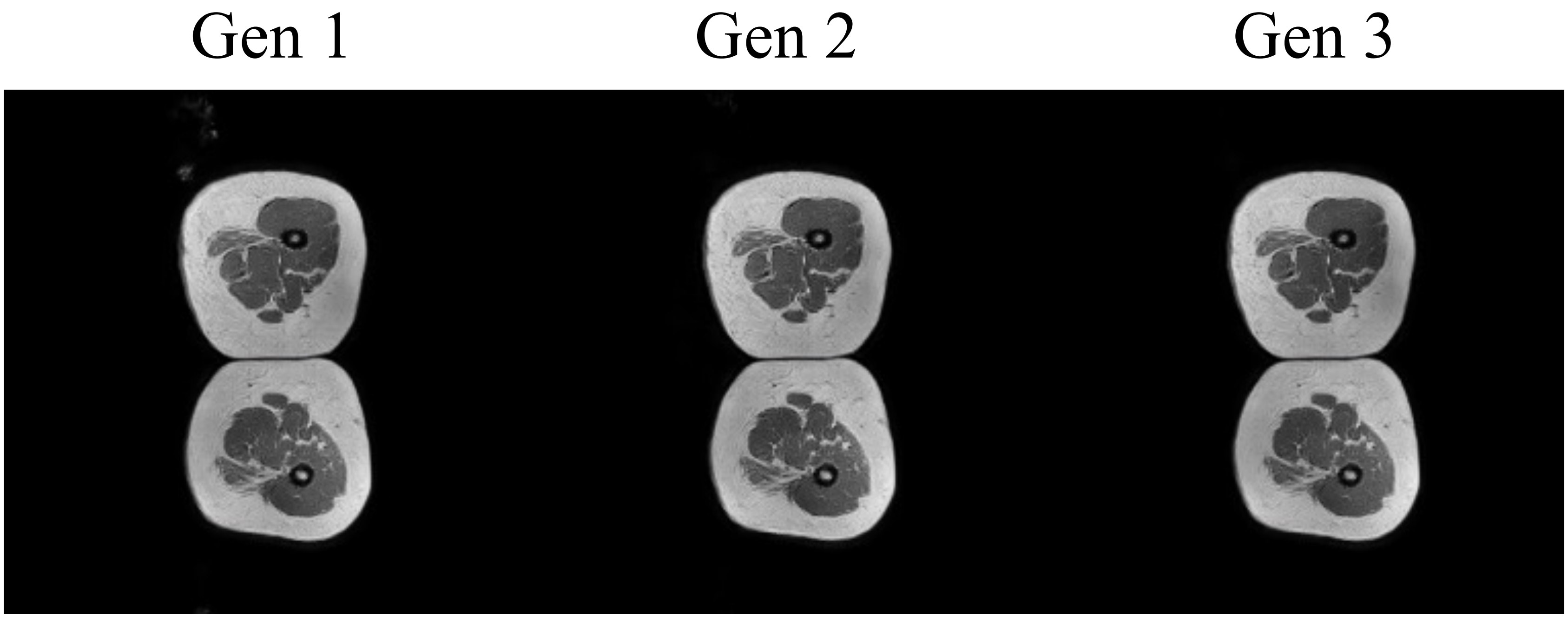
- We propose a framework Med-USIS for unsupervised semantic image synthesis for medical imaging
- Wavelet approach has better performance in generating high frequency details and aligning the shapes
- Mask Loss can effectively penalize the shape inconsistency of the generated images
- Future Work
  - Transformer Architecture for better localization
  - Implement Diffusion models
  - Extension to 3D volume generation
  - Experiments: Using generated images to pre-train a segmentation network

Thanks for your attention!

Appendix

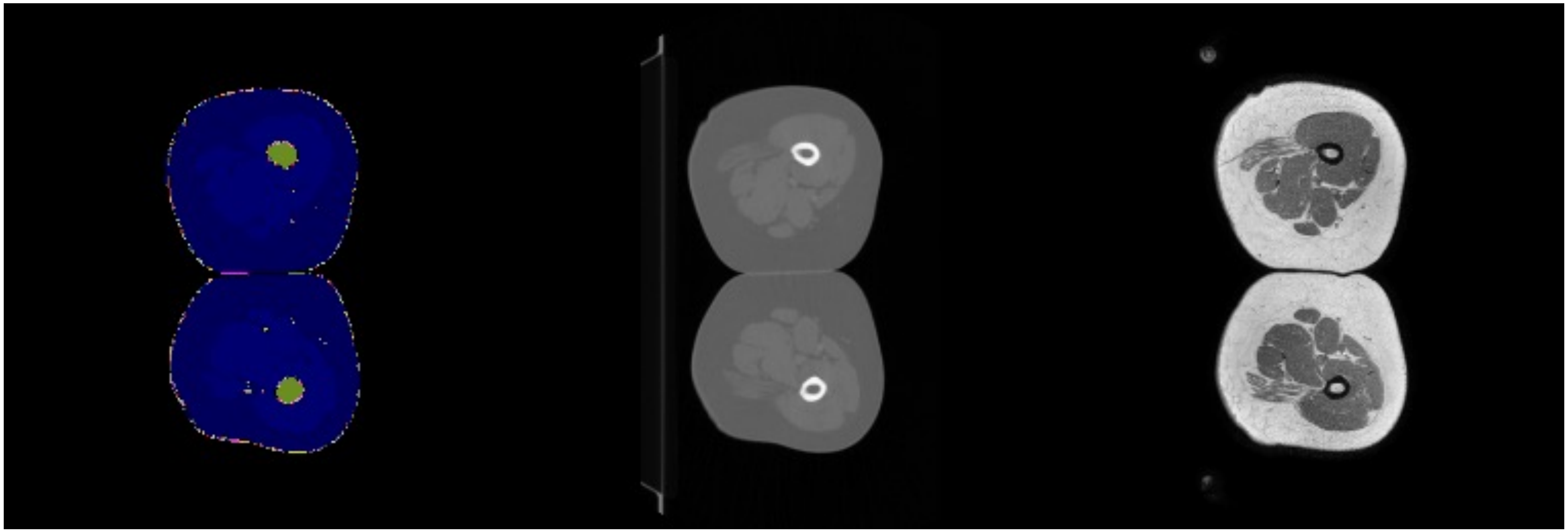
Qualitative Results- Unsupervised



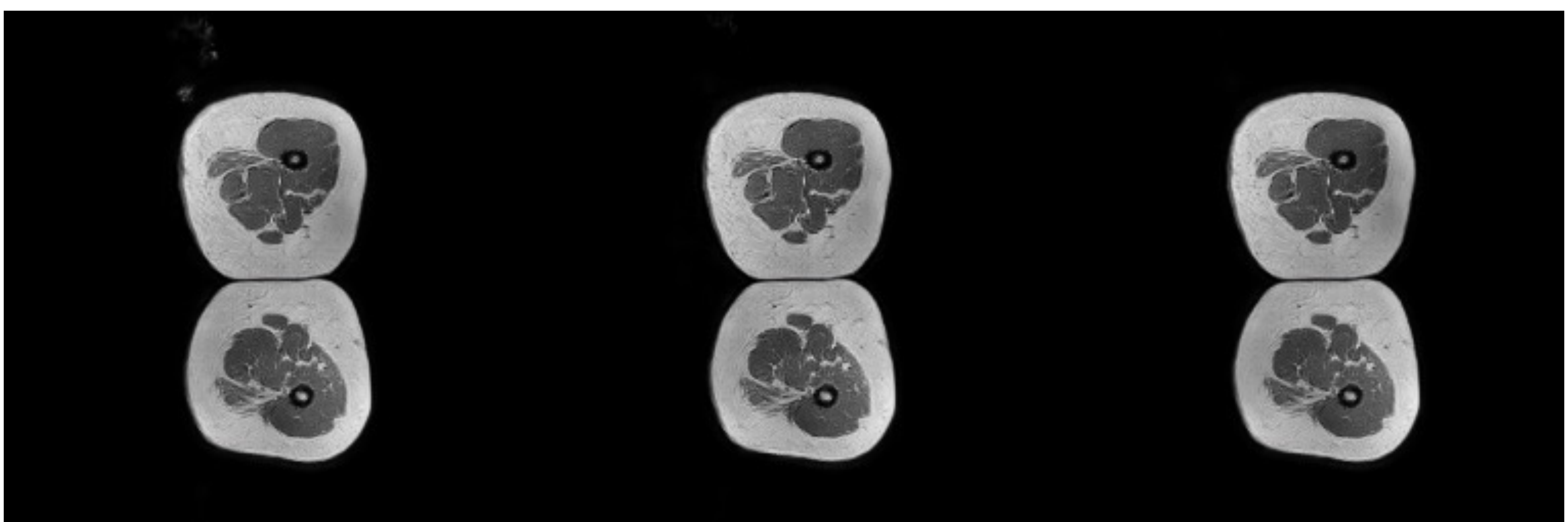


Qualitative Results- Unsupervised

CT Label CT MR

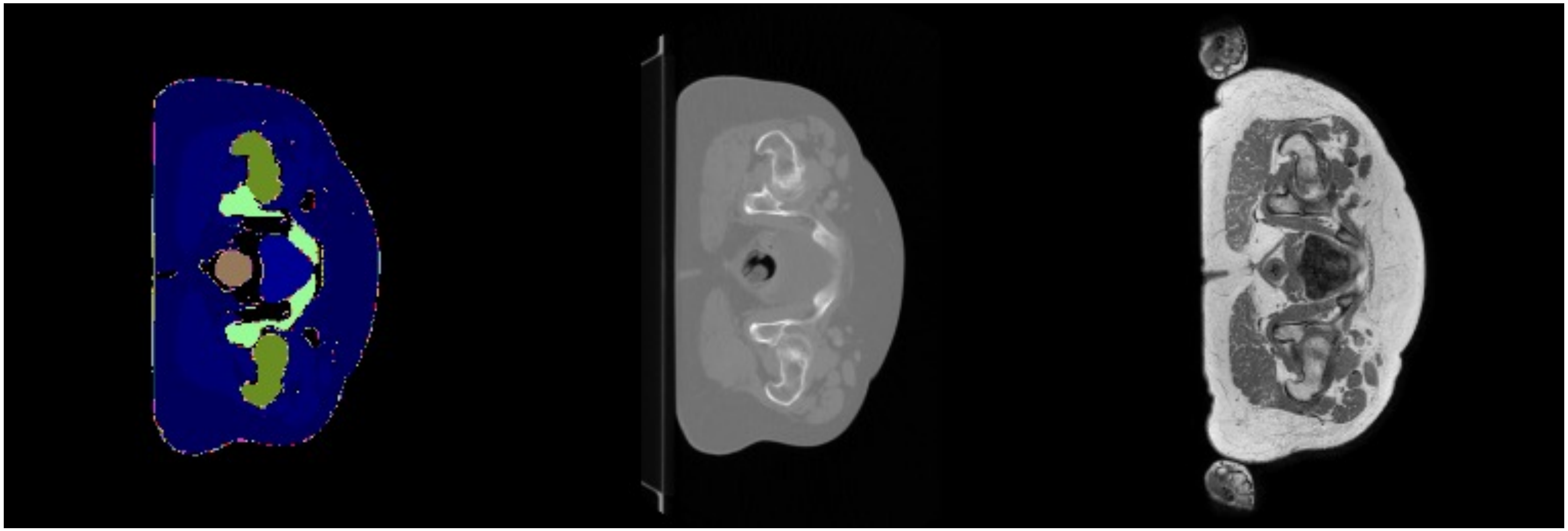


Gen 1 Gen 2 Gen 3



Qualitative Results- Unsupervised

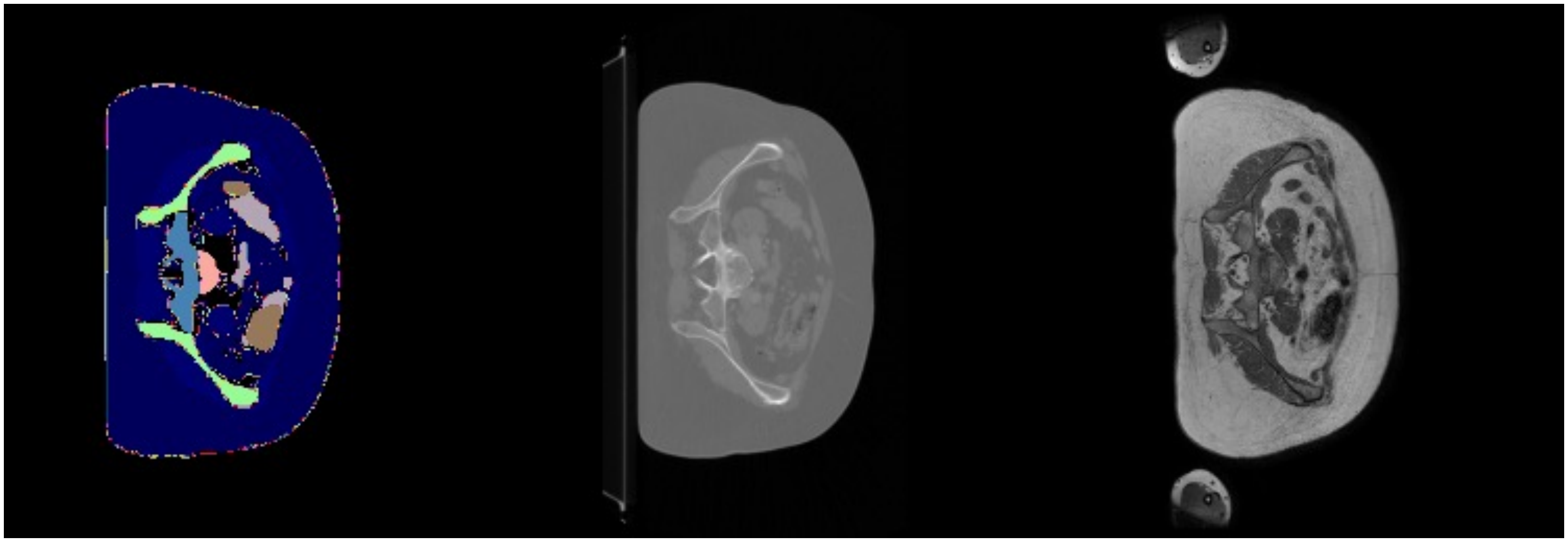
CT Label | CT | MR



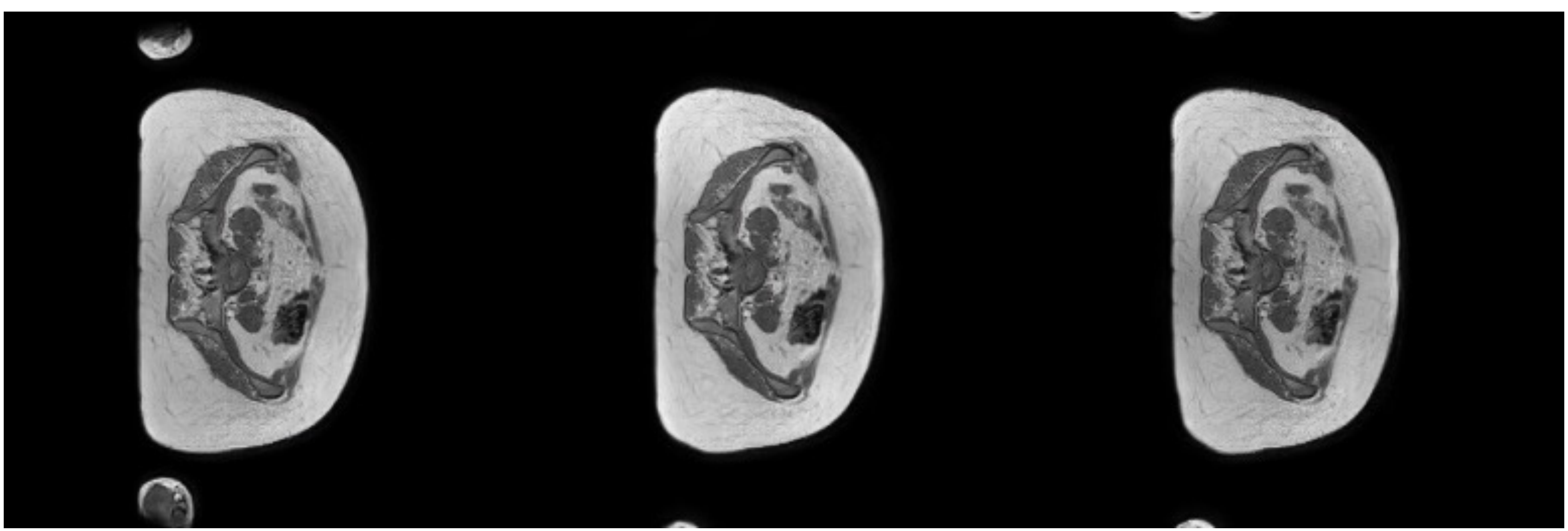
Gen 1 | Gen 2 | Gen 3

Qualitative Results- Unsupervised

CT Label CT MR

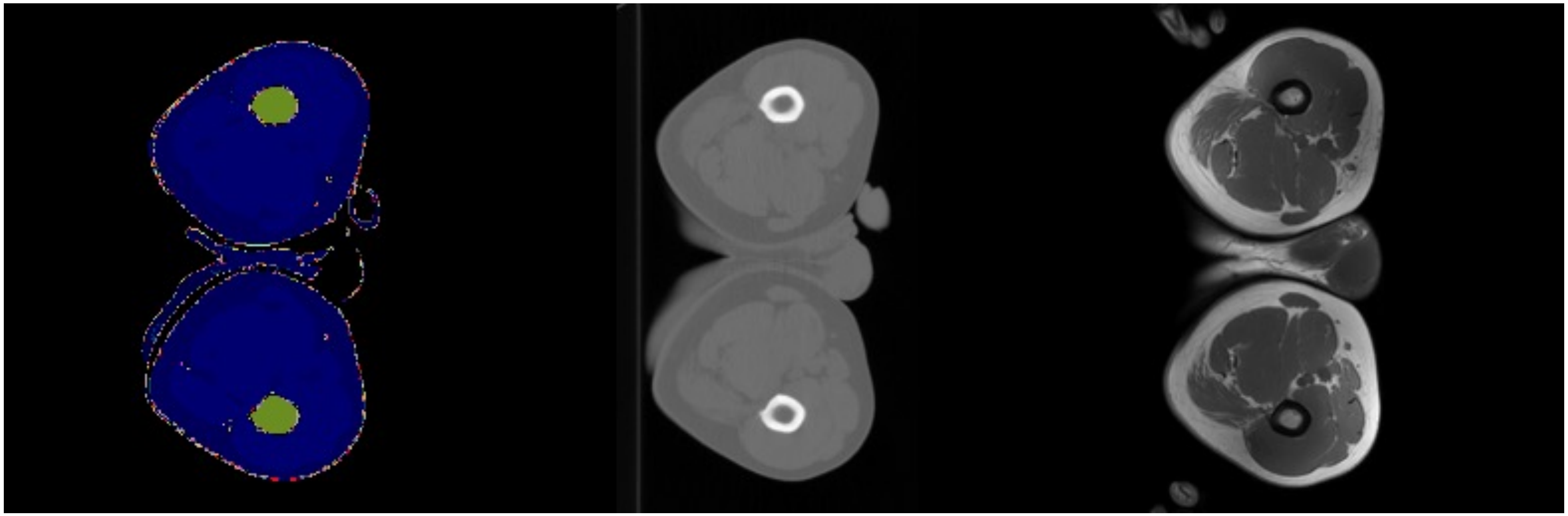


Gen 1 Gen 2 Gen 3

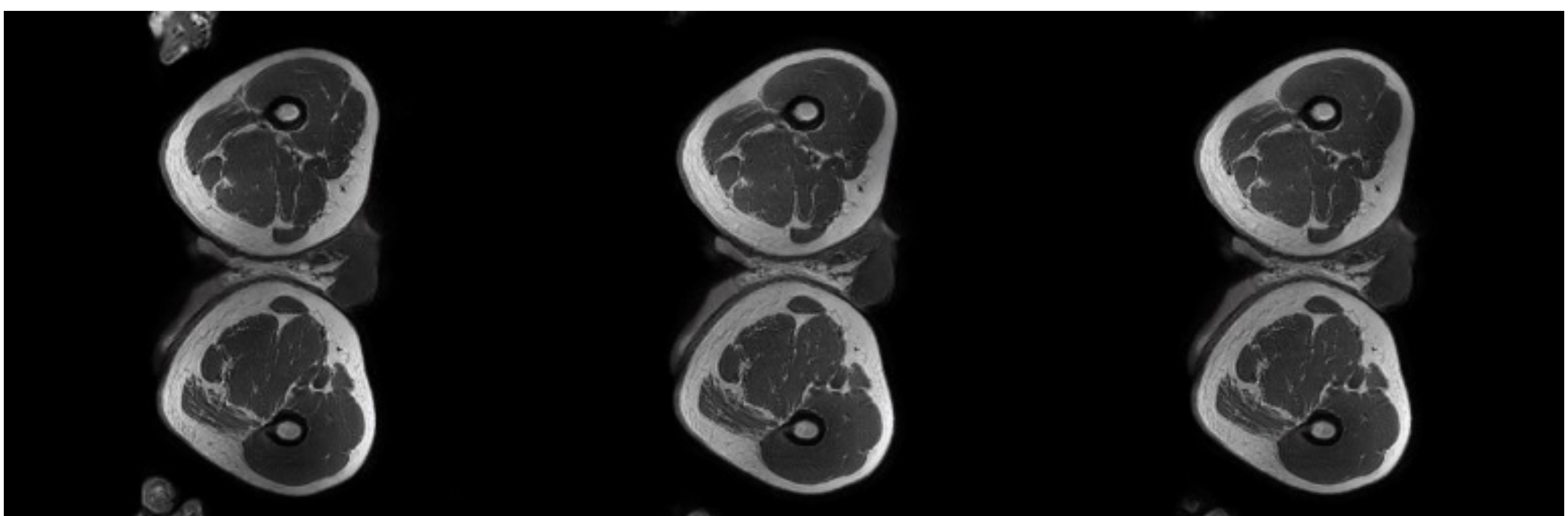


Qualitative Results- Unsupervised

CT Label | CT | MR

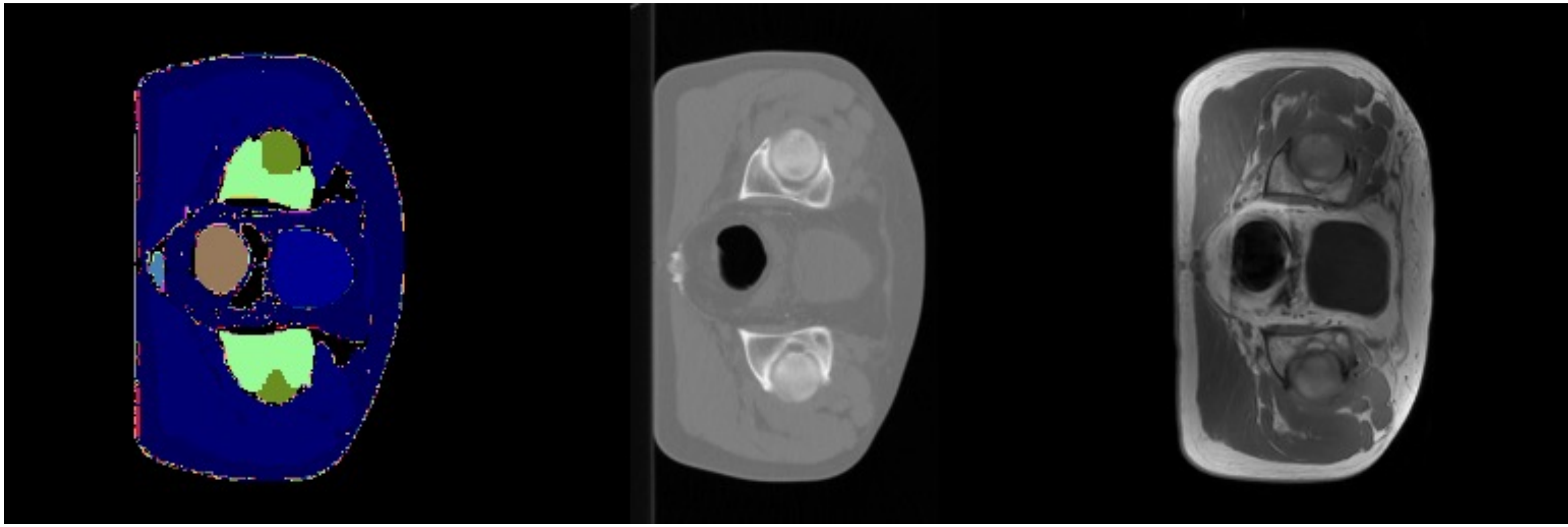


Gen 1 | Gen 2 | Gen 3

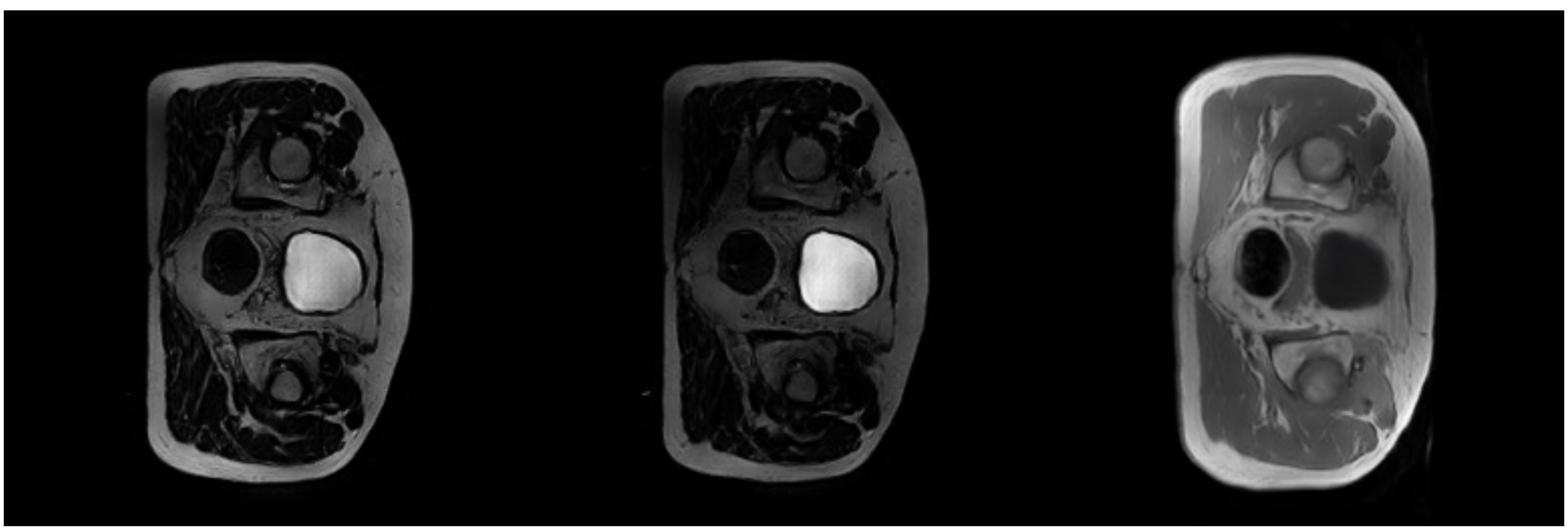


Qualitative Results- Unsupervised

CT Label CT MR

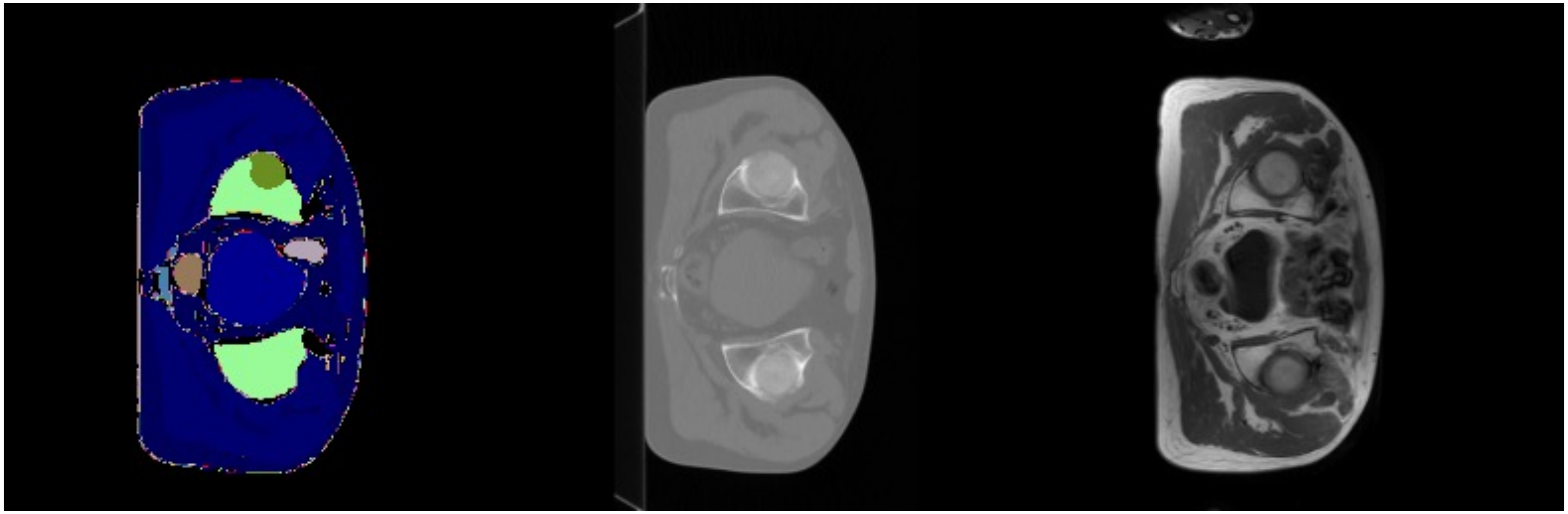


Gen 1 Gen 2 Gen 3

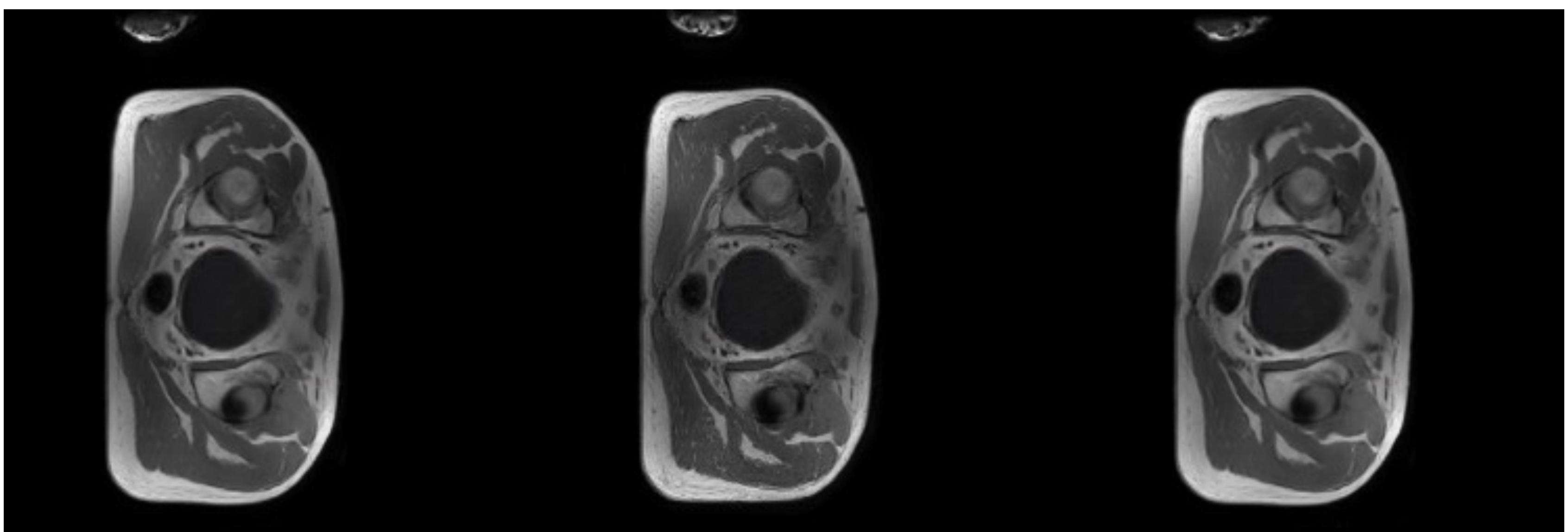


Qualitative Results- Unsupervised

CT Label | CT | MR

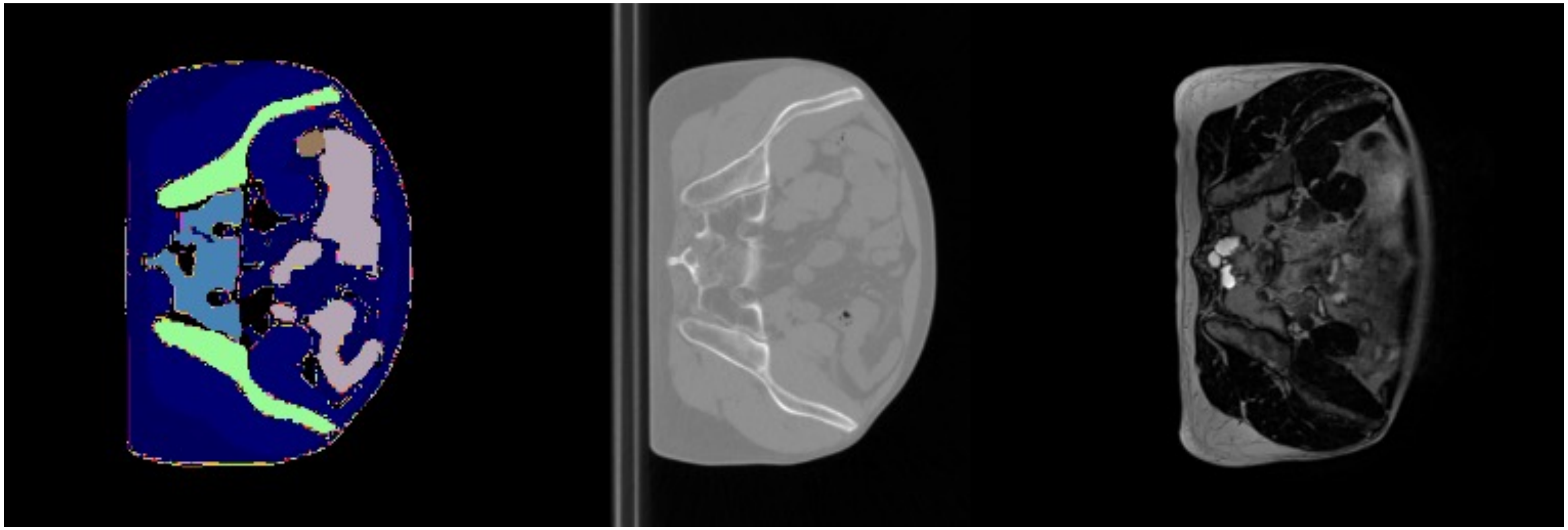


Gen 1 | Gen 2 | Gen 3

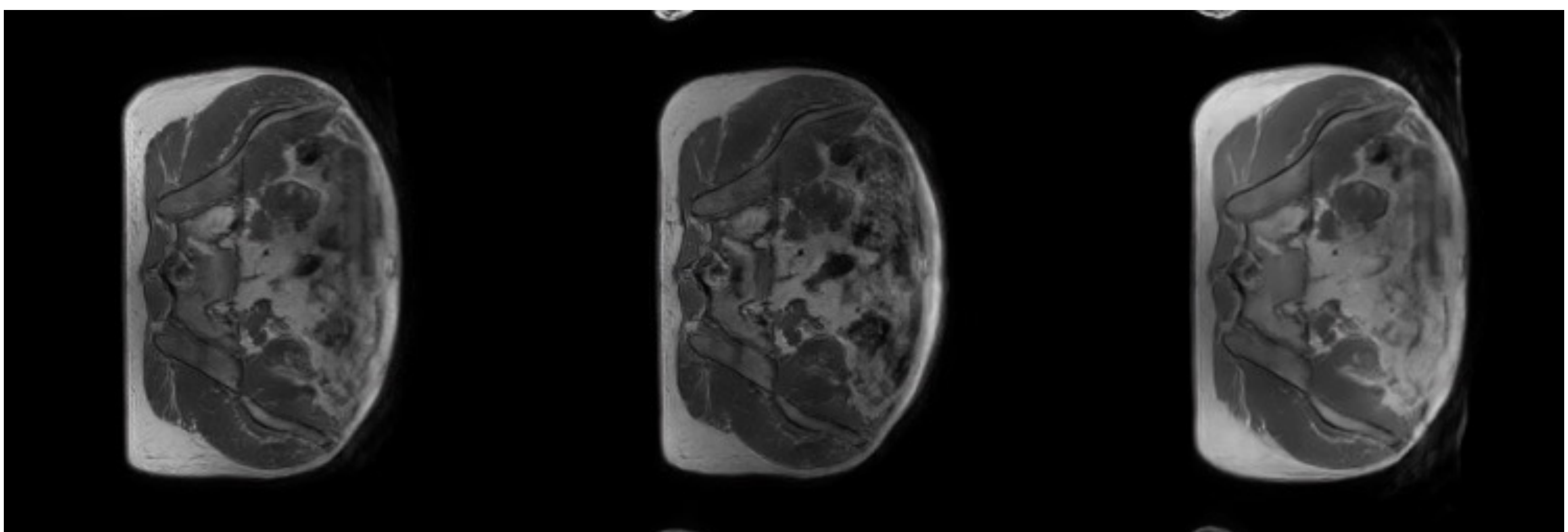


Qualitative Results- Unsupervised

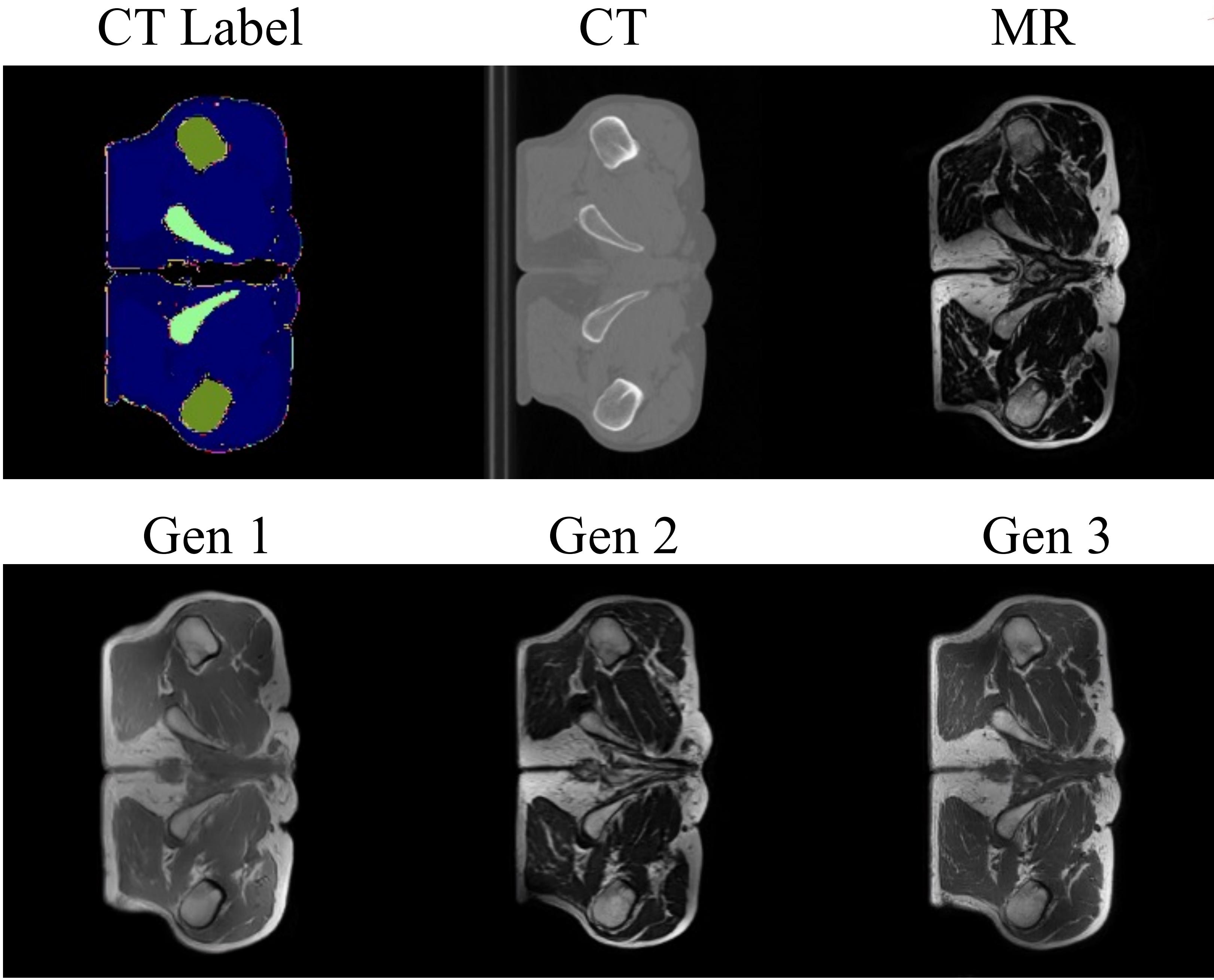
CT Label | CT | MR



Gen 1 | Gen 2 | Gen 3

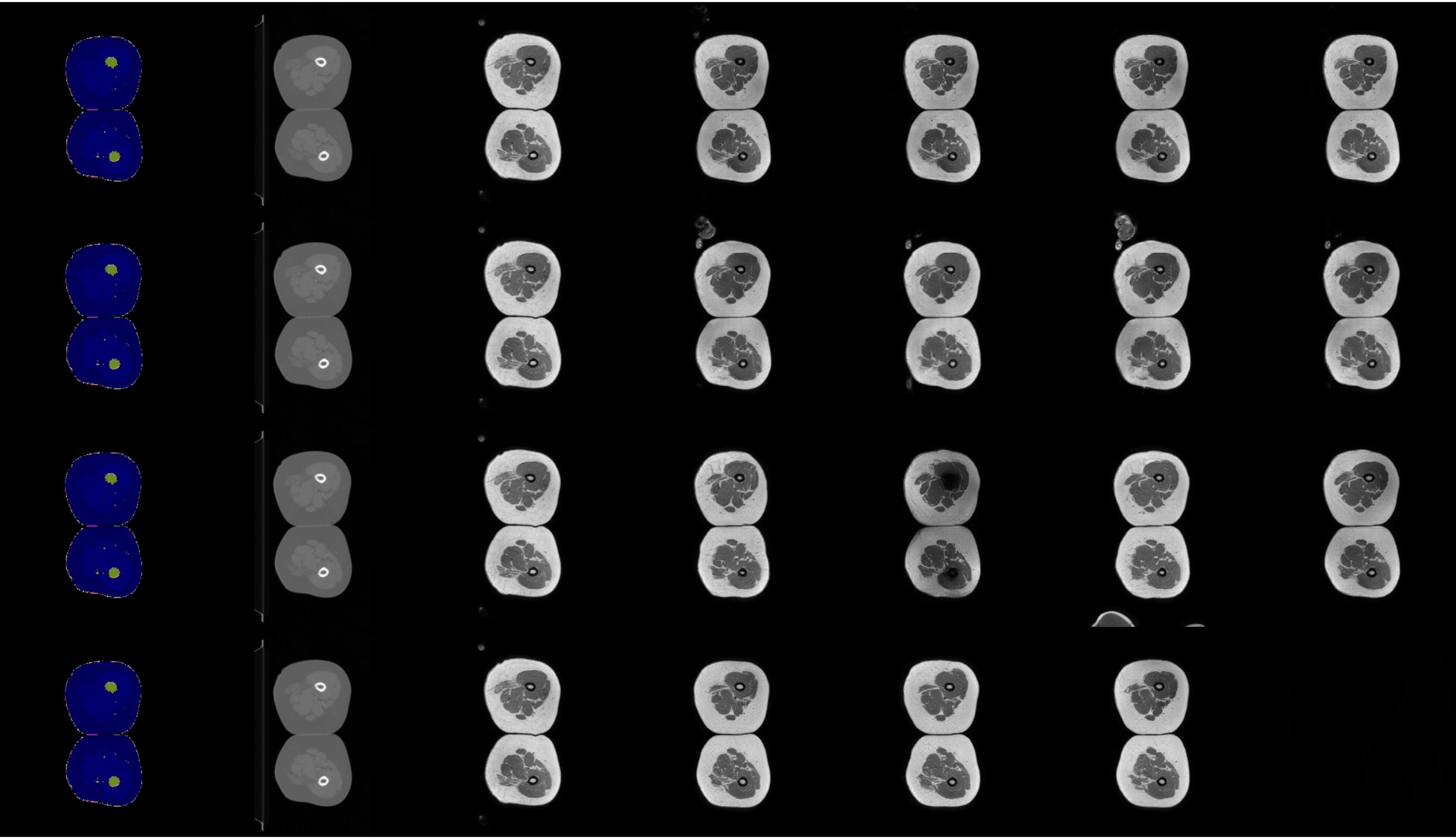


Qualitative Results- Unsupervised

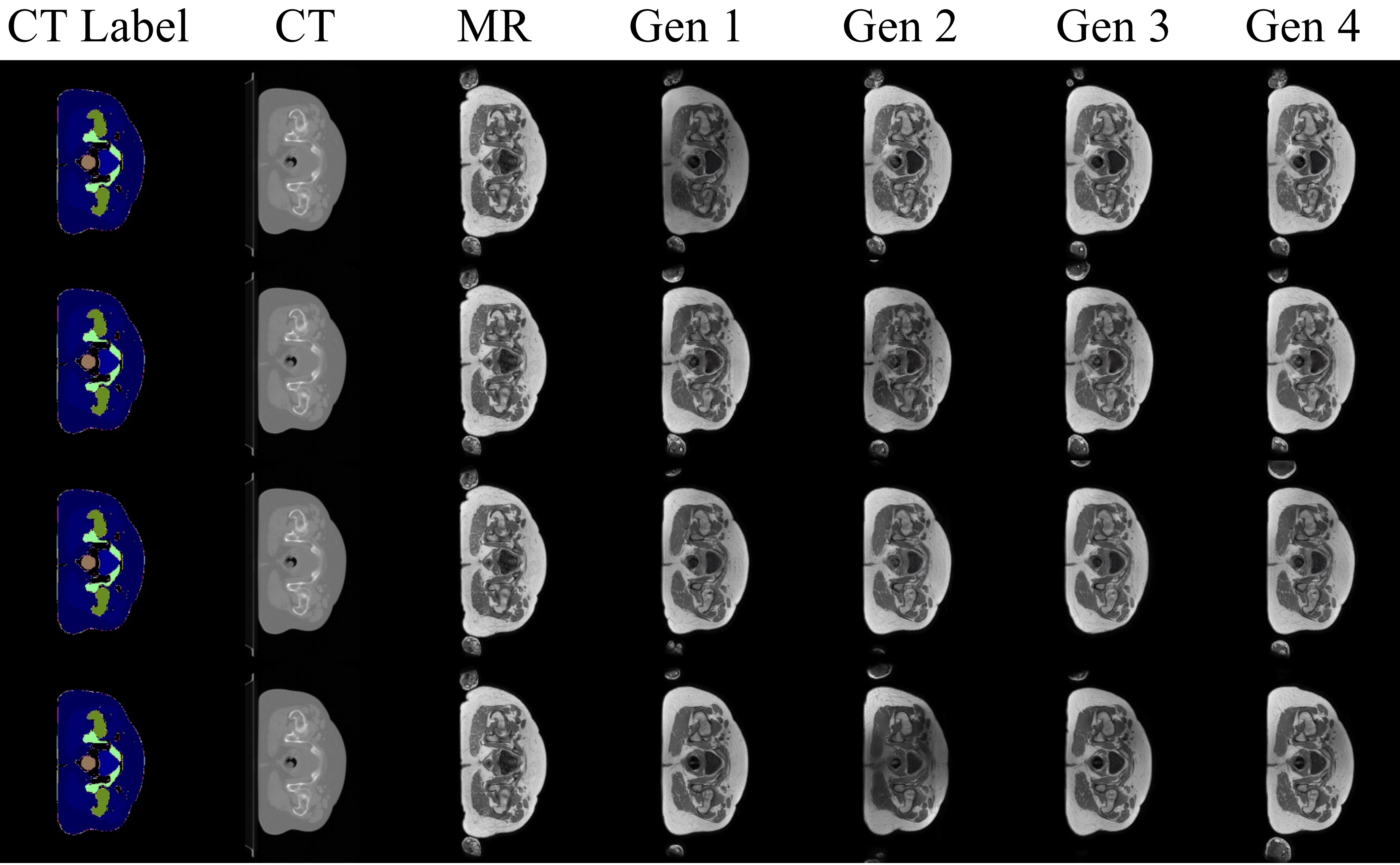


Qualitative Results- Unsupervised

CT Label | CT | MR | Gen 1 | Gen 2 | Gen 3 | Gen 4

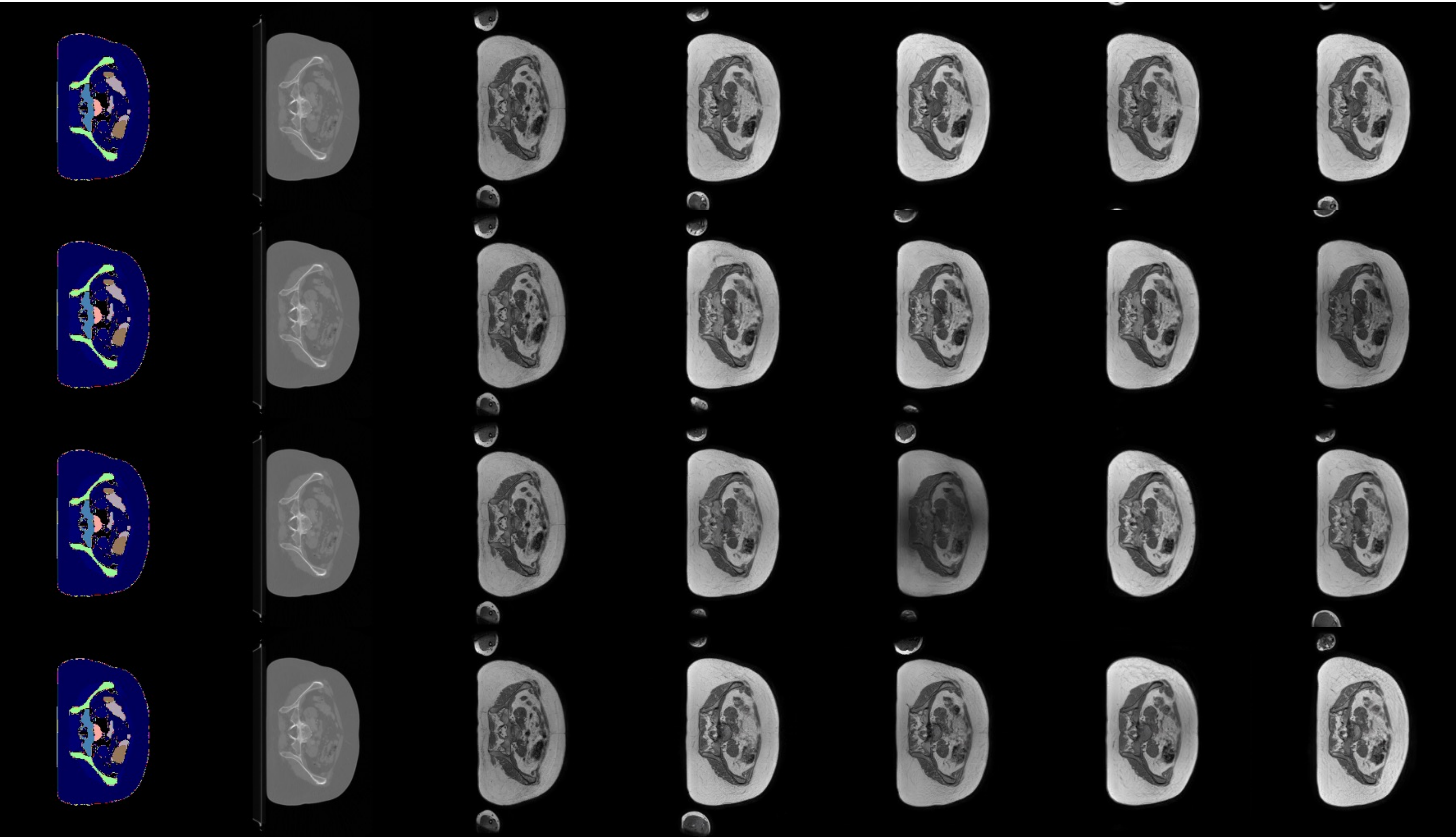


Qualitative Results- Unsupervised

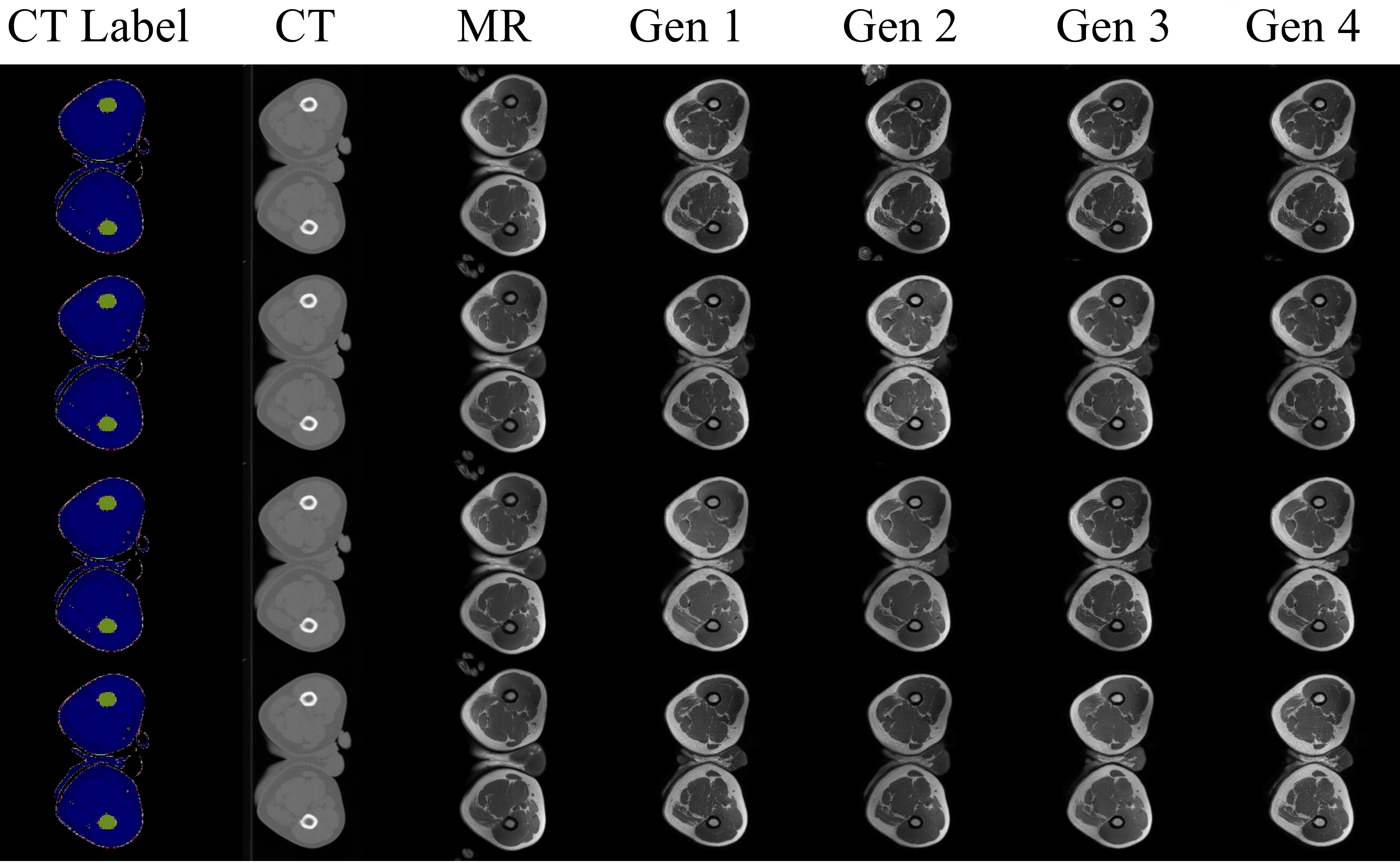


Qualitative Results- Unsupervised

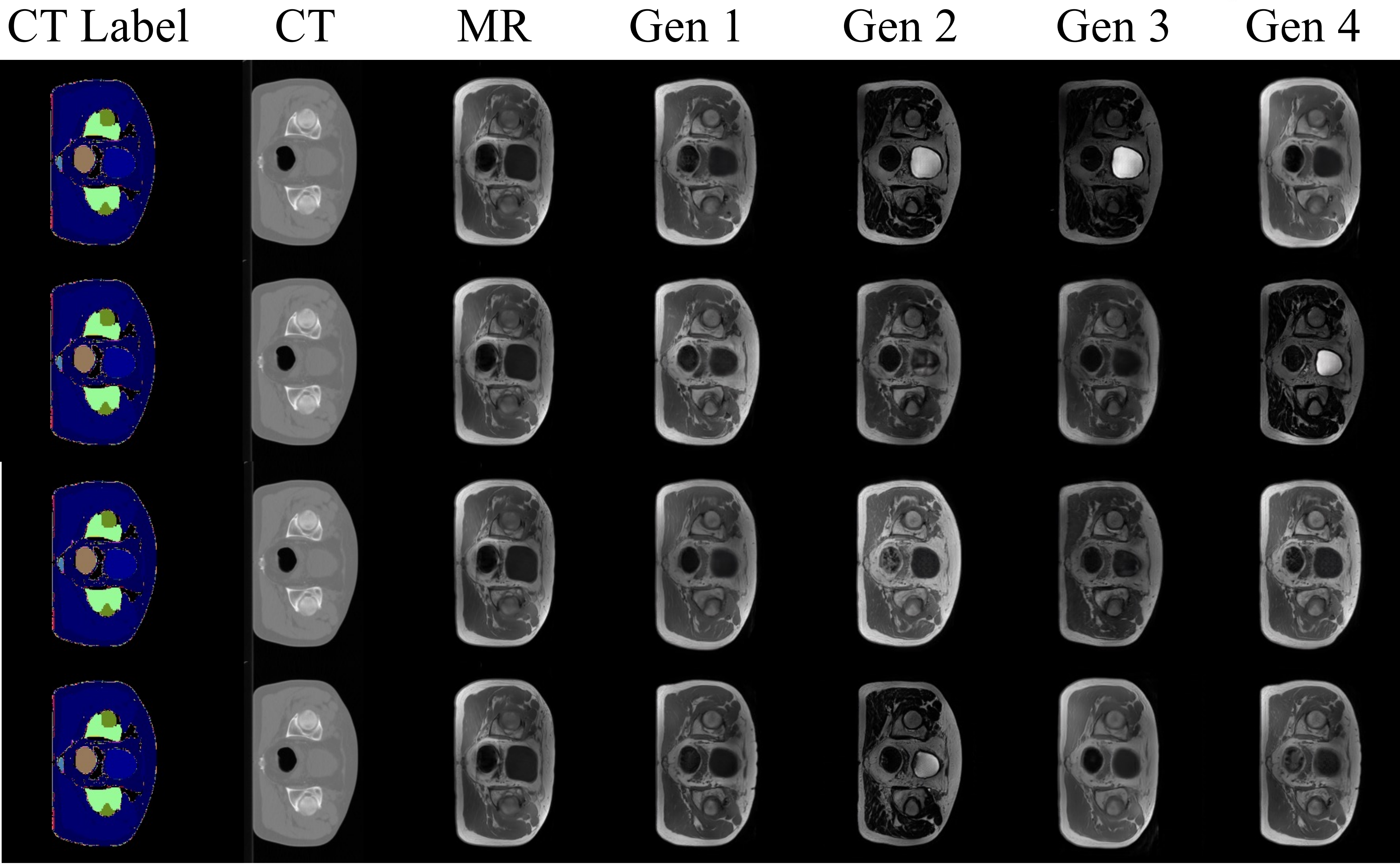
CT Label | CT | MR | Gen 1 | Gen 2 | Gen 3 | Gen 4



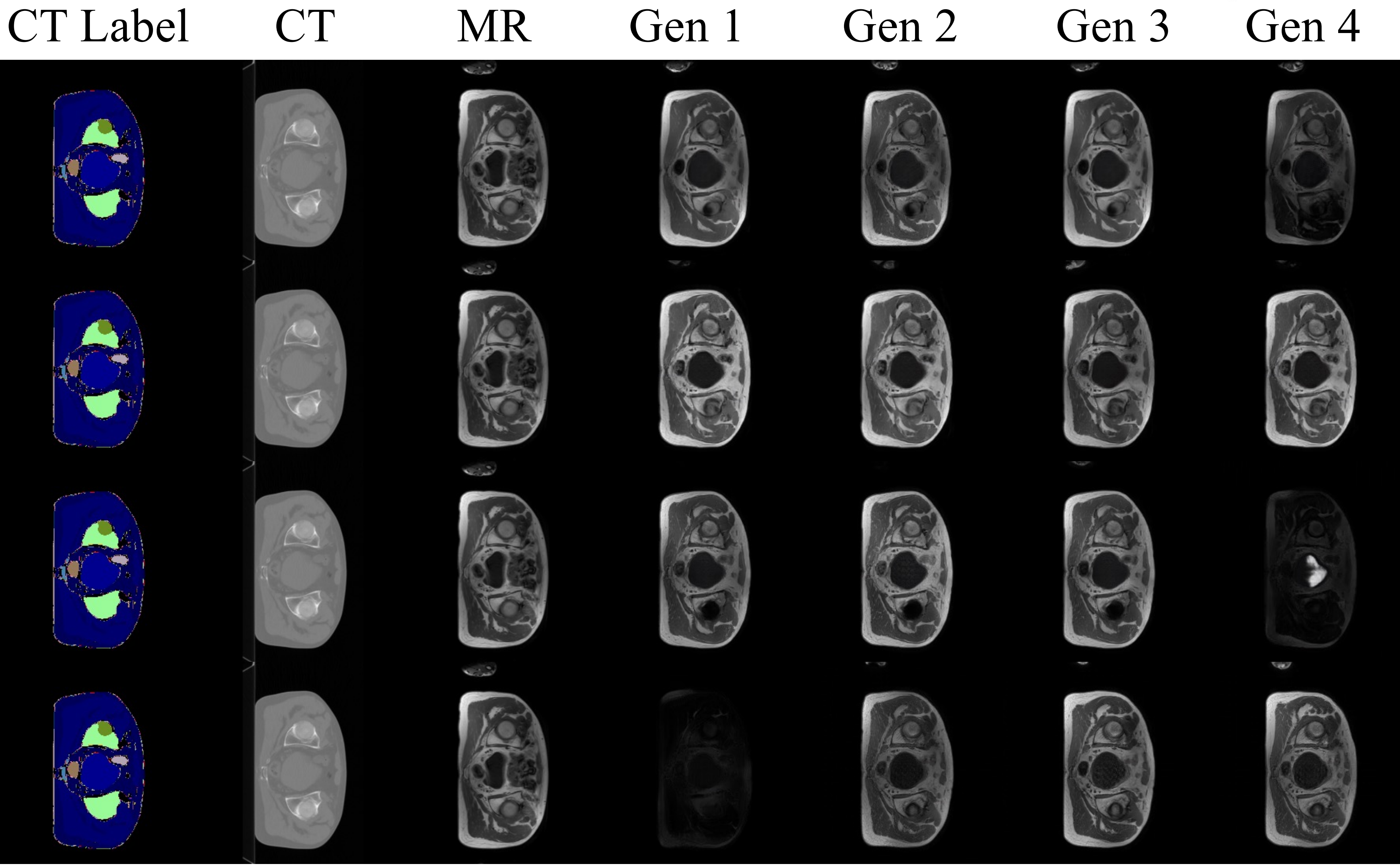
Qualitative Results- Unsupervised



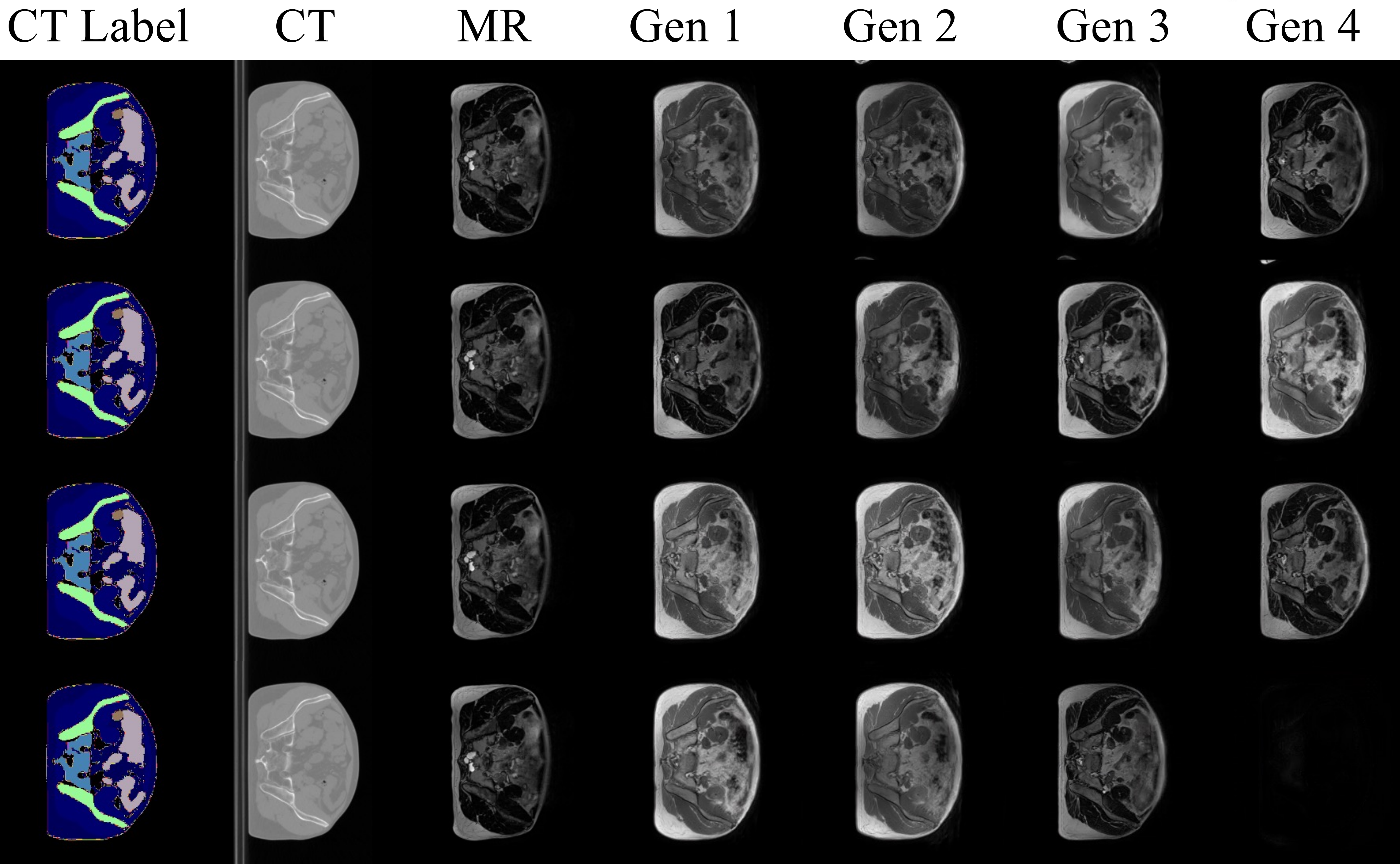
Qualitative Results- Unsupervised



Qualitative Results- Unsupervised

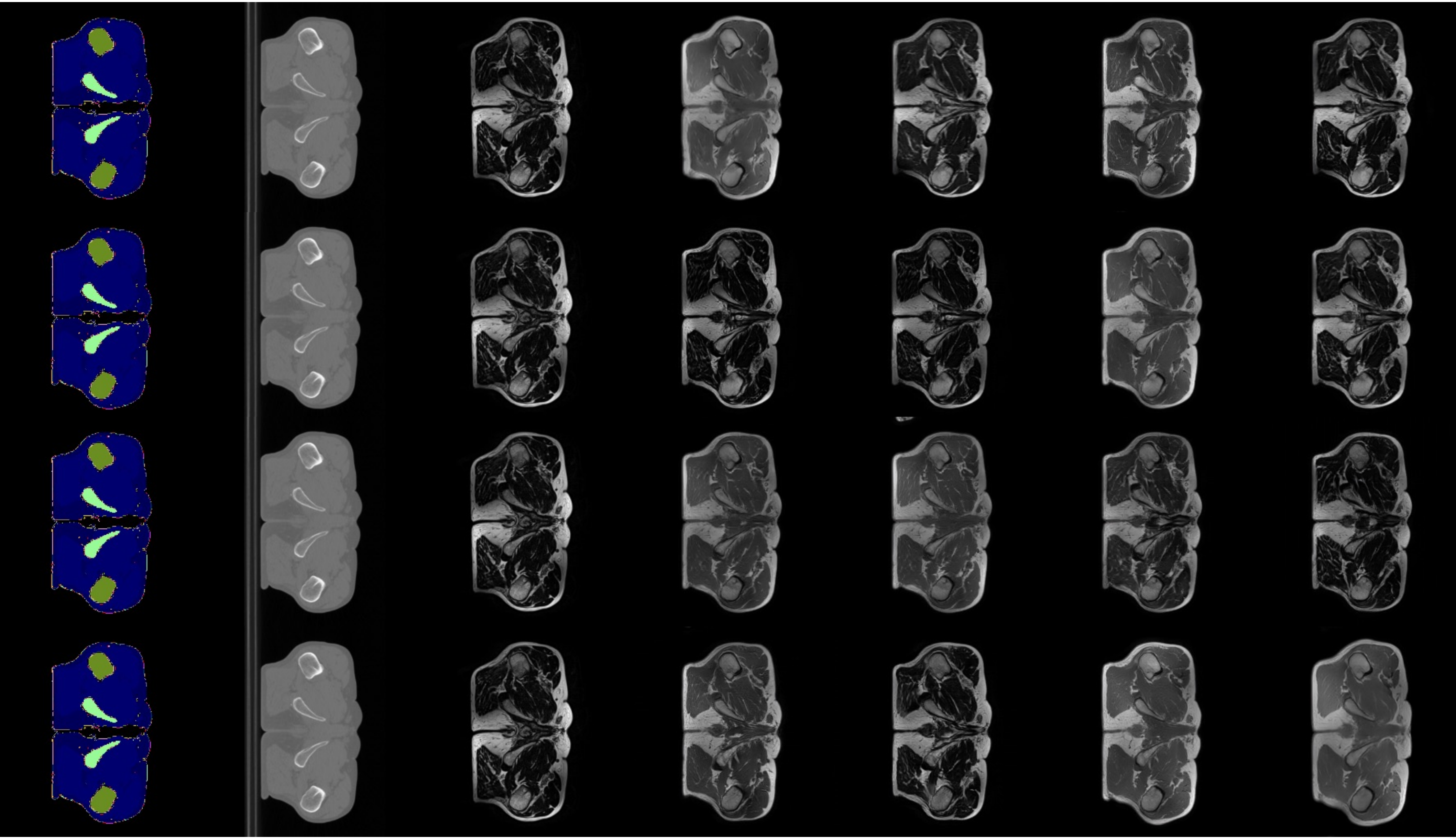


Qualitative Results- Unsupervised

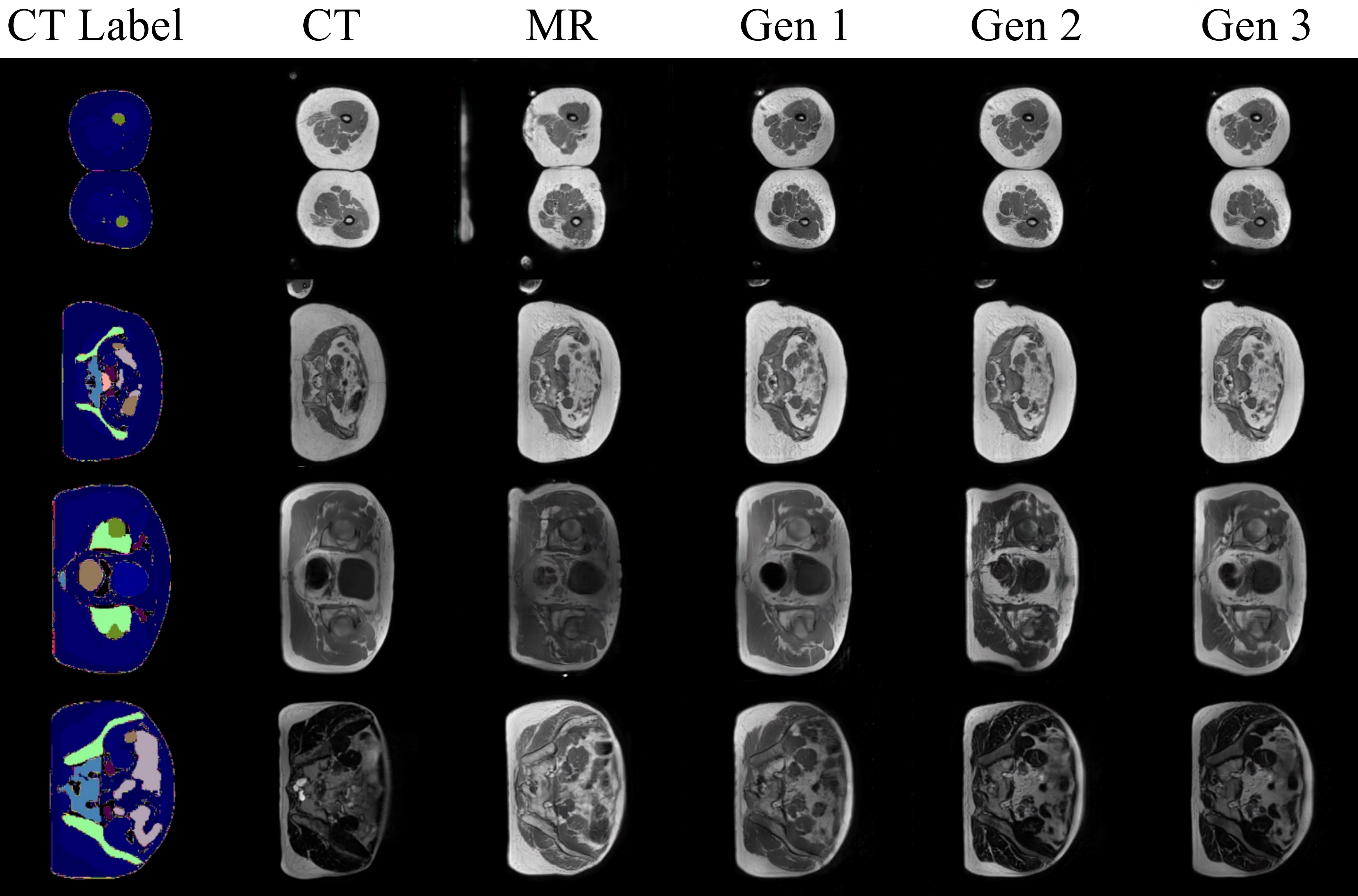


Qualitative Results- Unsupervised

CT Label	CT	MR	Gen 1	Gen 2	Gen 3	Gen 4

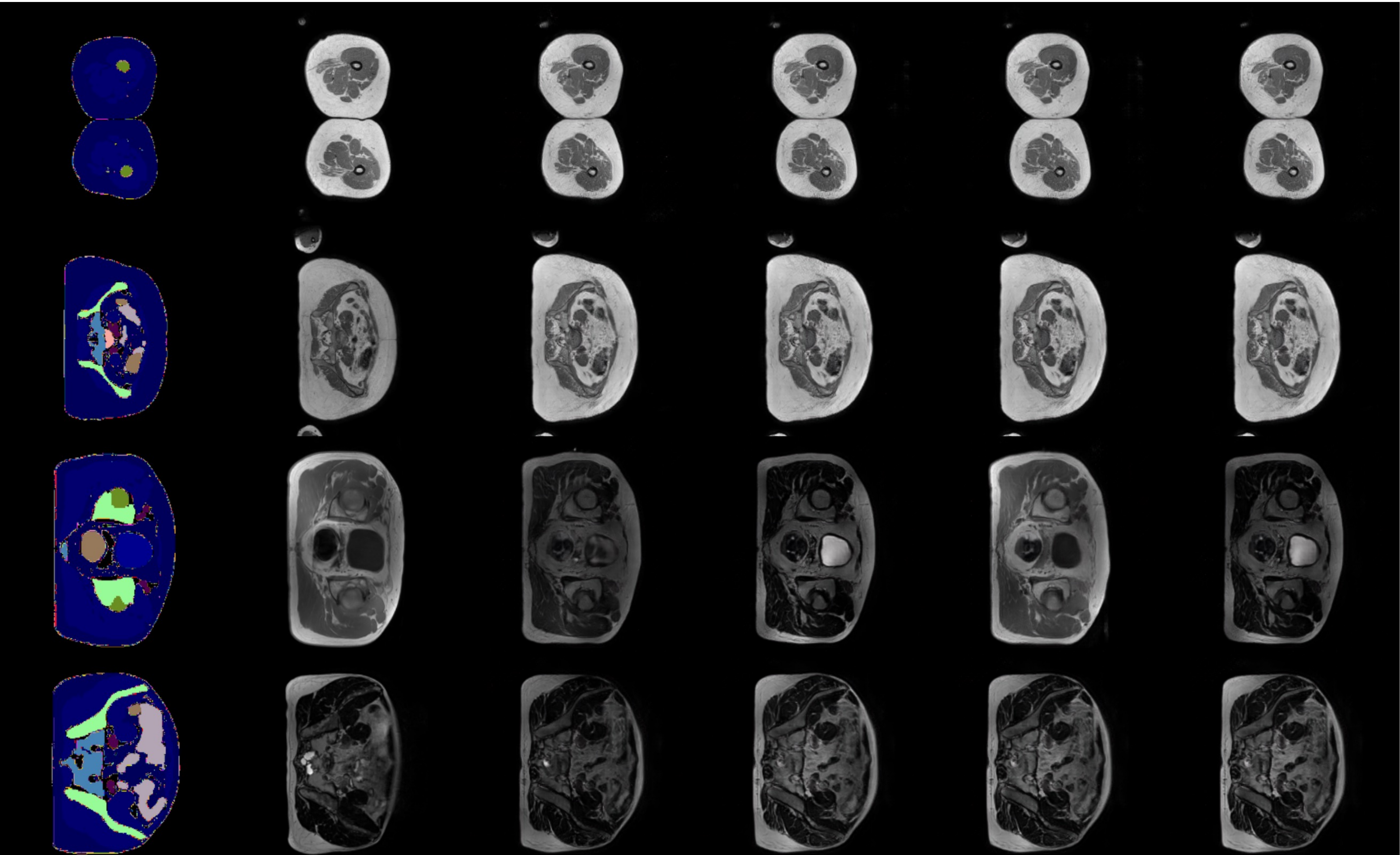


Qualitative Results- Supervised

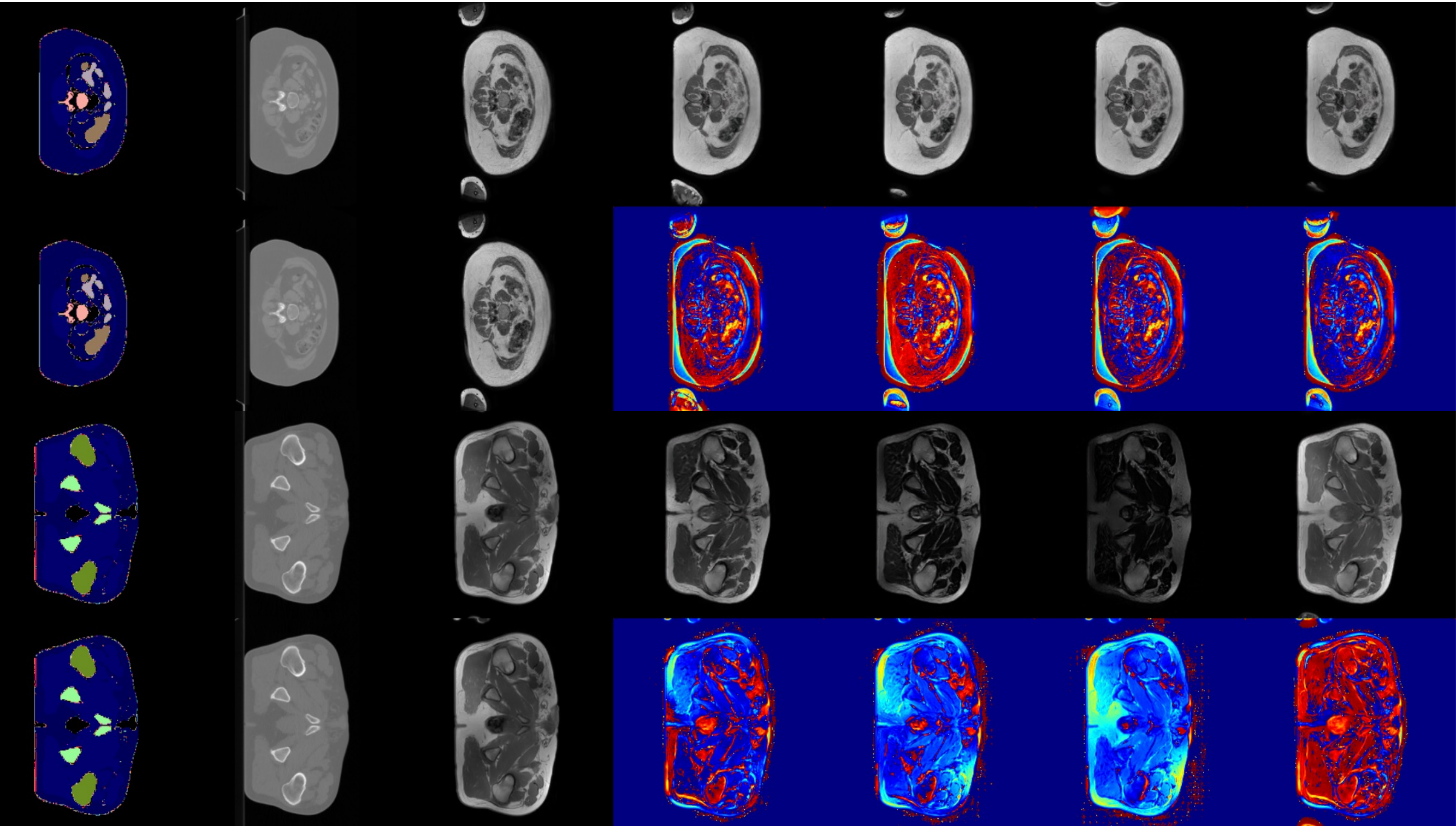


Qualitative Results- Supervised

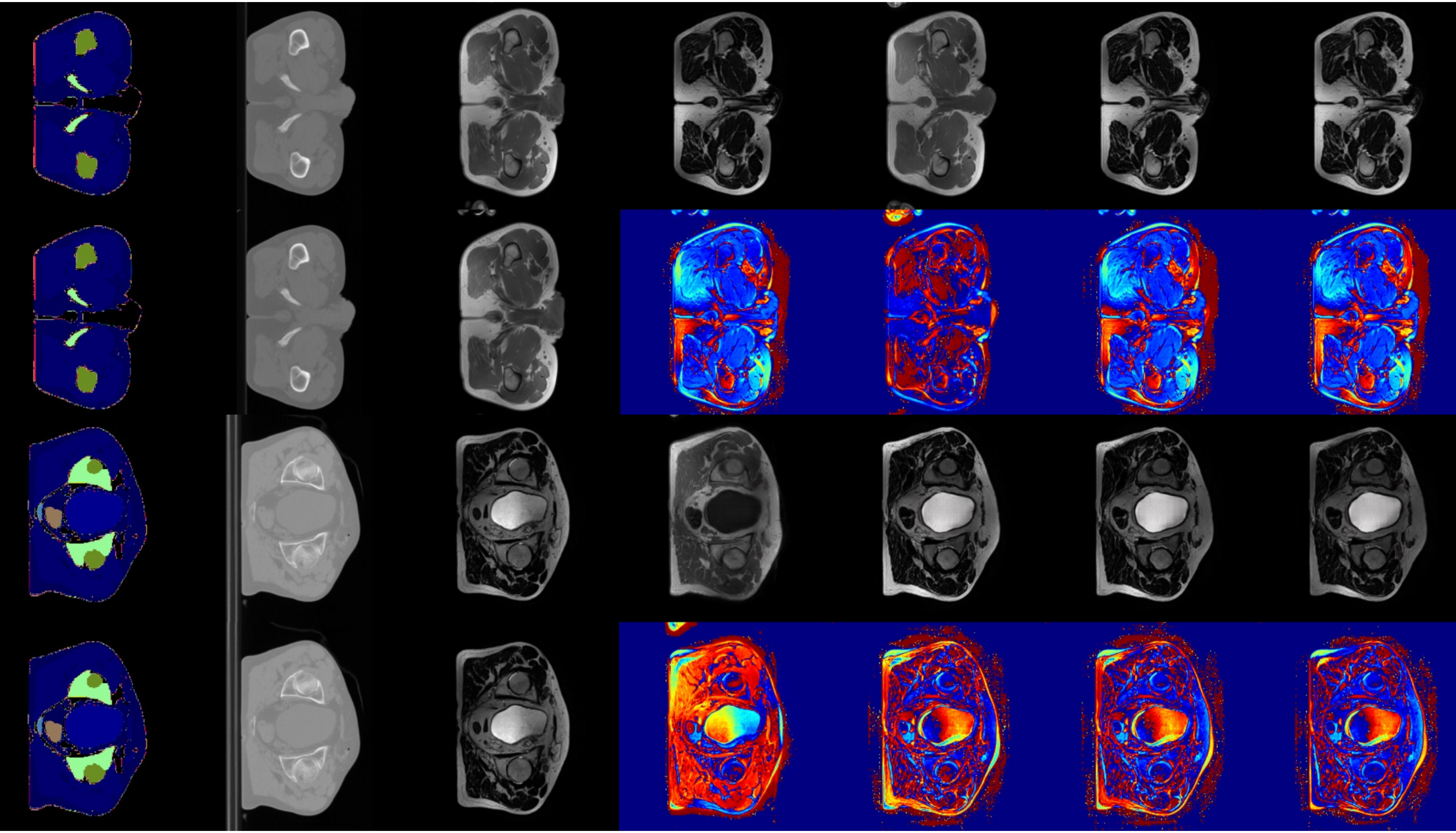
CT Label | CT | MR | Gen 1 | Gen 2 | Gen 3



MAE Visualization



MAE Visualization



MAE Visualization