ContourDiff: Unpaired Image Translation with Contour-Guided Diffusion Models

Yuwen Chen[1] ⋆, Nicholas Konz[1], Hanxue Gu[1], Haoyu Dong[1], Yaqian Chen[1], Lin Li[4], Jisoo Lee[2], and Maciej A. Mazurowski[1,2,3,4]

[1] Department of Electrical and Computer Engineering, Duke University, NC, USA
[2] Department of Radiology, Duke University, NC, USA
[3] Department of Computer Science, Duke University, NC, USA
[4] Department of Biostatistics & Bioinformatics, Duke University, NC, USA

Abstract. Accurately translating medical images across different modalities (*e.g.*, CT to MRI) has numerous downstream clinical and machine learning applications. While several methods have been proposed to achieve this, they often prioritize perceptual quality with respect to output domain features over preserving anatomical fidelity. However, maintaining anatomy during translation is essential for many tasks, *e.g.*, when leveraging masks from the input domain to develop a segmentation model with images translated to the output domain. To address these challenges, we propose ContourDiff, a novel framework that leverages domain-invariant anatomical contour representations of images. These representations are simple to extract from images, yet form precise spatial constraints on their anatomical content. We introduce a diffusion model that converts contour representations of images from arbitrary input domains into images in the output domain of interest. By applying the contour as a constraint at every diffusion sampling step, we ensure the preservation of anatomical content. We evaluate our method by training a segmentation model on images translated from CT to MRI with their original CT masks and testing its performance on real MRIs. Our method outperforms other unpaired image translation methods by a significant margin, furthermore without the need to access any input domain information during training.

Keywords: Unpaired image-to-image translation · CT · MRI.

1 Introduction

Unpaired image-to-image (I2I) translation—the task of translating images from some input domain to an output domain with only unpaired data for training—offers extensive applications in medical image analysis [1]. A significant use case is facilitating segmentation across different imaging modalities (such as CT and MRI) [4], for anatomical locations such as the brain [11], abdomen [9], and pelvis [15]. This is especially useful due to the extensive time and labor investment

⋆ Corresponding author: yuwen.chen@duke.edu

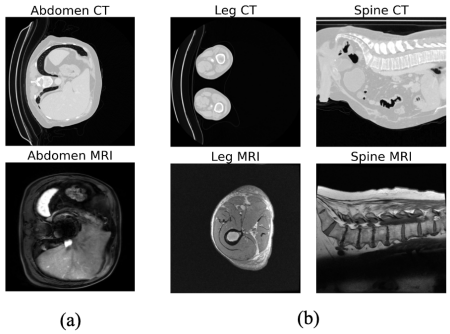

Fig. 1: **Structural biases between CT and MRI modalities in certain anatomical regions:** minor for the abdominal region from axial view (a), but severe for the leg from axial view and spinal regions from sagittal view (b).

required to annotate images in each individual modality, as annotations from one modality can be directly applied to images translated to another. However, doing so requires that anatomy is kept consistent through translation.

Ensuring anatomical consistency in unpaired I2I translation is challenging, particularly when the input and output domains exhibit substantial *structural biases*. An example of this is the drastic visual difference between CT and MRI for leg and spinal regions as captured in standard exams (Figure 1), where typically CT images display two legs while MRI scans only show one, and CT images capture entire abdominal body while MRI focuses on lumbar area, respectively. Due to the translation model learning these structural biases, the absence of such consistency can result in misalignment between the translated images and their corresponding segmentation masks, potentially leading to unreliable segmentation models trained on the translated images. Indeed, popular translation methods like CycleGAN [20] may yield undesirable outcomes when such substantial misalignment exists between modalities [13].

Inspired by previous works in spatially-conditioned diffusion models [10,19], we propose a translation diffusion model, "**ContourDiff**", that uses domain-invariant anatomical contour representations of images to guide the translation process, which enforces precise anatomical consistency even between modalities with severe structural biases. This model also has the added benefit that it is "source-free": it only requires a set of unlabeled output domain images for training. It so can potentially translate images from arbitrary unseen input domains at inference. To the best of our knowledge, this is the first unpaired image translation model that does not require input domain information for training.

We evaluate our method on CT to MRI translation for sagittal-view lumbar spine and axial-view hip-and-thigh body regions, which both possess severe structural biases (Figure 1). In addition to utilizing standard unpaired image generation quality metrics like Fréchet inception distance (FID), we evaluate the anatomical consistency of our translation model by training a segmentation

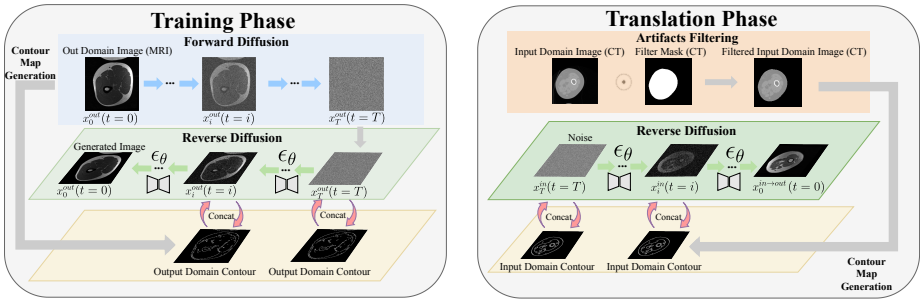

Fig. 2: Diagram of ContourDiff pipeline.

model on CT images translated to MRI given their original masks, and evaluating it for real MRI segmentation, as we find that FID does not properly capture anatomical fidelity. Our results show that ContourDiff outperforms existing *unpaired* I2I methods by at least 0.126 Dice coefficient over all test datasets, despite the fact that it requires no input domain information for training, unlike the competing methods.

2 Method

Problem definition: In unpaired image translation, only unpaired datasets of input and output domain examples are available for training. Our method is even more general in that it accomplishes *source-free* image translation, where only an unlabeled dataset of N_{out} output domain examples x_n^{out} $(n = 1, \ldots, N_{\text{out}})$ are available to train on. The goal is then to use the trained model at inference to translate unseen input domain data x_n^{in} to the output domain. In our case, we aim to translate CT images to the MRI domain, for usage with MRI-trained segmentation models. To do so, we propose a novel diffusion model-based image translation framework based on domain-invariant anatomical contour representations of images.

2.1 Adding contour guidance to diffusion models

Denoising diffusion probabilistic models [7], or just *diffusion models*, are generative models that learn to reverse a gradual process of adding noise to an image over many time steps $t = 0, \ldots, T$. New images can be generated by starting with a (Gaussian) noise sample x_T and iteratively applying the model to obtain x_{t-1} from x_t for $t = T, \ldots, 0$ until an image x_0 is recovered. In practice, the neural network itself $\epsilon_\theta(x_t, t)$ is an image-to-image architecture (*e.g.*, a UNet [14]) that is trained to predict the noise ϵ added to an image x_0 at various timesteps t, via the simple loss $L = \mathbb{E}_{x_0,t,\epsilon}\left[||\epsilon - \epsilon_\theta(x_t, t)||^2\right]$ [12], where θ is the model parameters.

Algorithm 1: Contour-guided diffusion model training.

Input: Output domain training distribution $p(x_0^{\text{out}})$.
repeat
  $x_0^{\text{out}} \sim p(x_0^{\text{out}})$
  $c^{\text{out}} = \texttt{Canny}(x_0^{\text{out}})$
  $\epsilon \sim \mathcal{N}(0, I_n)$
  $t \sim \text{Uniform}(\{1, \dots, T\})$
  $x_t^{\text{out}} = \sqrt{\bar{\alpha}_t} x_0^{\text{out}} + \sqrt{1-\bar{\alpha}_t}\epsilon$
  Update θ with $\nabla_\theta \left\| \epsilon - \epsilon_\theta(x_t^{\text{out}}, t | c^{\text{out}}) \right\|^2$
until *converged*;

Algorithm 2: Contour-guided image translation.

Input: Input domain image x_0^{in}.
Output: Translated image $x_0^{\text{in}\to\text{out}}$
$c^{\text{in}} = \texttt{Canny}(x_0^{\text{in}})$
$x_T^{\text{out}} \sim \mathcal{N}(0, I_n)$
for $t = T, \dots, 1$ **do**
  $\epsilon \sim \mathcal{N}(0, I_n)$ if $t > 1$, else $\epsilon = 0$
  $x_{t-1}^{\text{out}} = \frac{1}{\sqrt{\alpha_t}} \left(x_t^{\text{out}} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t^{\text{out}}, t | c^{\text{in}}) \right) + \sigma_t \epsilon$
end
return $x_0^{\text{in}\to\text{out}}$

However, for standard diffusion models, how to constrain the semantics/anatomy of generated images is unclear. To address this, we propose to utilize *contour* representations of images to provide guidance in generating the image. While training the model, we use the Canny edge detection filter [2] to extract the contour representation c of each training image x_0, and concatenate it with the network input at every denoising step, a practice similar to [10,19]. This modifies the network to become $\epsilon_\theta(x_t, t|c)$ and the diffusion training objective to become

$$L = \mathbb{E}_{(x_0,c),t,\epsilon} \left[||\epsilon - \epsilon_\theta(x_t, t|c)||^2 \right], \tag{1}$$

where (x_0, c) is a training set image and its accompanying contour.

2.2 Contour-guided image translation

One important feature of contours is that they can be viewed as domain-invariant yet anatomy-preserving representations of images. This allows for a contour-guided diffusion model trained in some output domain to serve as a source-free image translation method, as follows.

First, we train a contour-guided diffusion model on output domain images with accompanying computed contours $(x_n^{\text{out}}, c_n^{\text{out}})$, shown in Algorithm 1. Next, to translate some *input domain* image x^{in} to the output domain, we extract its contour c^{in} after removing irrelevant backgrounds using F_{filter}, and use the output domain-trained model ϵ_θ conditioned on c^{in} to generate the image $x^{\text{in}\to\text{out}}$. Therefore, $x^{\text{in}\to\text{out}}$ maintains the anatomical content of x^{in}, while possessing the visual domain characteristics of the output domain. Our translation algorithm is shown in Algorithm 2, where $\alpha_t = 1 - \beta_t$ with the variance of the additive pre-scheduled noise β_t, and $\overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

Filtering Out Image Artifacts. We also apply additional pre-processing to all network input images x to filter out non-anatomical features/artifacts (*e.g.*, the motorized table in CT), by applying a binary mask M_{filter} as $x \leftarrow M_{filter} \odot$

x. M_{filter} is defined by sequentially computing the follow Scikit-Image [17] functions on x: `threshold_multiotsu`, `binary_erosion`, `remove_small_objects`, and `remove_small_holes`.

Enforcing Translation Consistency for Adjacent Slices. Additionally, we propose to enforce the consistency of translating adjacent input domain image slices taken from 3D images (*e.g.*, CT) to a output domain as follows. Firstly, we translate the first slice image x_1^{in} in the 3D volume to its output domain version $x_1^{\text{in}\to\text{out}}$. We repeat the generation until the mean of one generated image is less than a specified threshold m_{thresh}. Then, we translate the successive slices x_i^{in} ($i = 2, \ldots, N_{\text{slices}}$) by generating different candidate translations $x_i^{\text{in}\to\text{out}}$ (using different sampled noise inputs to the model) until one is within an L_2 distance of δ of the previous slice translation $x_{i-1}^{\text{in}\to\text{out}}$, and use that for the final $x_i^{\text{in}\to\text{out}}$. We use $m_{thresh} = 110$, $\delta = 50$ for lumbar spine and $m_{thresh} = 100$, $\delta = 40$ for hip/thigh anatomical views, respectively. During each iteration, if multiple candidates satisfy the L_2 distance criterion, we choose the one with the smallest δ. We generate 4 candidates per iteration, allowing up to 5 attempts; should none of the candidates meet the specified requirements after this, we select the one with the smallest δ.

3 Experiments and Results

3.1 Experiment Setup

Dataset collection. In this paper we study one of the most common translation scenarios: CT to MRI. For the MRI datasets where we train the contour-guided diffusion model, we collect a private dataset with T-1 weighted lumbar spine (L) and hip & thigh (H&T) body regions. 40 sagittal lumbar MRI volumes (670 2D slices), and 10 axial MRI volumes from thigh and hip (404 2D slices) are selected. Correspondingly, we obtain 54 sagittal (2,333 2D slices) and 29 axial (4,937 2D slices) CT volumes from the TotalSegmentator project [18] in L and H&T, respectively. This experiment setting mimics an important real-world application of utilizing external label information to accelerate the annotation process of an internal dataset. To train the downstream task models for bone segmentation, we further randomly split the two CT sets by patients (43:11 for L and 23:6 for H&T) for training and validation. We evaluate the test performance of the segmentation models with held-out annotated MRI sets (10 L volumes including 158 2D slices, 12 H&T volumes including 426 2D slices). In addition, to study the generalization ability of our method, we test the lumbar segmentation model on 40 volumes (731 2D slices) from the SPIDER lumbar spine (L-SPIDER) public dataset [5] [5].

Evaluation Metrics. For evaluation metrics, we quantitatively evaluate MRI segmentation model performance on translated images using Dice coefficient

[5] We crop the slices to exclude sacrum as it is not annotated

(Dice) and average symmetric surface distance (ASSD). As there are no paired images, we also calculate the FID [6] between the translated image and real output domain image distributions for reference, although we find that it does not capture global anatomical realism.

Implementation Details. For the image translation model, we adopt the UNet architecture [14] for the denoising model ϵ_θ with a two-channel input (grayscale image and its contour). The training settings for the diffusion model follow the same as that in [10]. We use the DDIM algorithm [16] for sampling, with 50 steps. For the segmentation models, we use the convolution-based UNet [14] and transformer-based SwinUNet [3]. All images are resized to 256×256 and normalized to $[0, 255]$. For the training of other baselines, we mostly follow the default settings from each official GitHub. We train up to 200 epochs and set the $\lambda_{idt} = 0.5$ to include identity loss if the methods are provided. For CycleGAN and MaskGAN, we train the segmentation model with a cosine learning rate scheduler up to 100 epochs with the initial learning rate of 1×10^{-3}.

3.2 Image Translation Performance

Comparision with Other Methods. We compare our framework to other translation/adaptation methods, including CycleGAN [20], SynSeg-Net [9], CyCADA [8] and MaskGAN [13], via the performance of output domain-trained downstream task segmentation models on translated images. CycleGAN and SynSeg-Net translate the images solely at the image level, while CyCADA also aligns the latent feature output from the downstream task model encoder. MaskGAN incorporates the automatically extracted coarse masks to preserve the structures better when translating input images to the output domain. For CyCADA, we utilized the same segmentation architecture as the other models without the skip connection to enable feature-level alignment. In terms of other methods, we trained and evaluated the performance of the segmentation model based on generated images from image translators across multiple training epochs and reported the best ones[6]. We note that these methods are all unpaired (trained on input and output domain images), while our *source-free* method only requires output domain images for training; this is because there are no other source-free image-level adaptation (translation) methods to compare to, as ours is the first of its kind.

The segmentation model results are shown in Table 1. "w/o Adap." is a baseline referring to the model trained on the CTs without any adaptation and tested on the MRIs directly. For the three test sets, our method outperforms previous image adaptation methods by a significant margin: *e.g.*, the output domain model Dice is higher by at least 0.199, 0.126 and 0.196 for L, L-SPIDER and H&T, respectively for UNet.

[6] For SynSeg-Net and CyCADA, we evaluate the segmentation model every 20 epochs. For CycleGAN and MaskGAN, as we need to train the segmentation model separately, we evaluate at 20, 60, 100, 150, and 200 epochs.

Method	M_{seg}	L		L-SPIDER		H & T	
		Dice ↑	ASSD ↓	Dice ↑	ASSD ↓	Dice ↑	ASSD ↓
w/o Adap.	UNet	0.287	6.515	0.236	8.275	0.004	45.730
CycleGAN	UNet	0.484	2.479	0.507	3.629	0.535	9.140
SynSeg-Net	UNet	0.316	3.014	0.364	3.207	0.370	4.708
CyCADA	UNet	0.331	5.942	0.364	4.389	0.349	11.247
MaskGAN	UNet	0.428	3.192	0.458	3.729	0.289	16.228
Ours	UNet	**0.683**	**1.432**	**0.633**	**2.066**	**0.731**	**3.139**
w/o Adap.	SwinUNet	0.171	7.386	0.187	8.327	0.003	48.624
CycleGAN	SwinUNet	0.362	3.505	0.412	3.701	0.464	9.791
SynSeg-Net	SwinUNet	0.288	3.527	0.291	5.502	0.059	12.869
CyCADA	SwinUNet	0.319	3.691	0.260	4.726	0.155	13.004
MaskGAN	SwinUNet	0.322	4.692	0.385	5.355	0.292	17.591
Ours	SwinUNet	**0.654**	**1.434**	**0.534**	**2.353**	**0.659**	**5.780**

Table 1: Comparison of our model to other image translation methods in terms of segmentation model (M_{seg}) performance on held-out output domain images. (L: Lumbar dataset, L-SPIDER: Public SPIDER Lumbar dataset, H & T: Hip & Thigh dataset)

We provide example image translations in Figure 3. These datasets form a challenging task due to (1) the noticeable shift in image features between the input and output domains and (2) the high anatomical variability between different scans. Moreover, we see that adversarially-trained models (*e.g.*, CycleGAN) have trouble with the consistent *structural bias* between the input and output domains, *i.e.*, when one domain is absent of certain features seen in the other. As shown in Figure 1 and Appendix D.1,D.2,D.3 and D.4, this is particularly evident in our H&T dataset, where MRIs are dominant by a single leg, and CTs often contain two legs. Such a bias may lead the adversarial mechanism to over-emphasize these features and, therefore, tend to translate CTs of two legs into MRIs depicting only one leg. For the lumbar spine from the sagittal view, MRIs often start from the lowest thoracic spine and end at the sacrum. On the other hand, CTs often include the upper leg and sometimes the abdominal body. Our model explicitly enforces anatomical consistency through translation despite these domain feature differences through its contour guidance, generating MRIs that strictly follow input CT images, resulting in better mask alignment and better segmentation model performance.

Additionally, it appears that the FID score does not reliably reflect the anatomical consistency between the translated and real output domain images (see Appendix C). Based on the Table 1 and the Figure 3, translated images from ContourDiff appear to best follow anatomical fidelity compared to other models, both quantitatively and qualitatively, despite achieving the highest score in FID.

Ablation Study. (1) We verify the effectiveness of introducing contours to each denoising step during training by conditionally training on empty map (*i.e.*, all

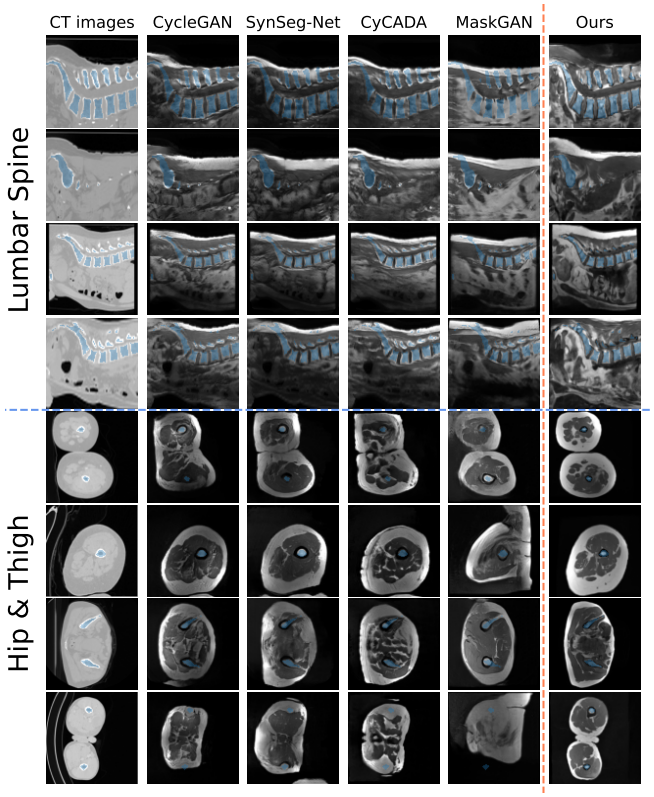

Fig. 3: Generated MRIs given CTs from Lumbar and Hip & Thigh areas from different translation models. The masks (in blue) from the original CTs are added to all the generated images to visualize the alignment.

zeros) and adding the CTs contours during the translation steps. Figure 4 showed that the denoised model ϵ_θ trained without contours hardly followed the introduced CTs contours. Furthermore, the UNet trained on these unconditionally generated MRIs experienced a dramatic performance drop (see Appendix A). (2) We generate the images directly (*i.e.*, by single candidate) without enforcing translation consistency for adjacent slices. The qualitative result shows a reduced quality of the generated images by using a single candidate (see Appendix B).

4 Conclusion

In this paper, we introduce a novel framework (ContourDiff) to preserve the anatomical fidelity in unpaired image translation. Our method constrains the generated images in the output domain to align with the anatomical contour of images from the input domain. Both quantitative and qualitative results on medical datasets show that ContourDiff significantly outperforms multiple existing image translation methods in maintaining anatomical structures.

References

1. Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B.: Medgan: Medical image translation using gans. Computerized medical imaging and graphics **79**, 101684 (2020)
2. Canny, J.: A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence (6), 679–698 (1986)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
4. Chen, J., Chen, S., Wee, L., Dekker, A., Bermejo, I.: Deep learning based unpaired image-to-image translation applications for medical physics: a systematic review. Physics in Medicine & Biology (2023)
5. van der Graaf, J.W., van Hooff, M.L., Buckens, C.F.M., Rutten, M., van Susante, J.L.C., Kroeze, R.J., de Kleuver, M., van Ginneken, B., Lessmann, N.: Lumbar spine segmentation in mr images: a dataset and a public benchmark (2023)
6. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
8. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A.A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation (2017)
9. Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T.K., Savona, M.R., Abramson, R.G., Landman, B.A.: Synseg-net: Synthetic segmentation without target modality ground truth. IEEE Transactions on Medical Imaging **38**(4), 1016–1025 (Apr 2019). https://doi.org/10.1109/tmi.2018.2876633, http://dx.doi.org/10.1109/TMI.2018.2876633
10. Konz, N., Chen, Y., Dong, H., Mazurowski, M.A.: Anatomically-controllable medical image generation with segmentation-guided diffusion models. arXiv preprint arXiv:2402.05210 (2024)
11. Li, W., Li, Y., Qin, W., Liang, X., Xu, J., Xiong, J., Xie, Y.: Magnetic resonance image (mri) synthesis from brain computed tomography (ct) images based on deep learning methods for magnetic resonance (mr)-guided radiotherapy. Quantitative imaging in medicine and surgery **10**(6), 1223 (2020)
12. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
13. Phan, V.M.H., Liao, Z., Verjans, J.W., To, M.S.: Structure-preserving synthesis: Maskgan for unpaired mr-ct translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 56–65. Springer (2023)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
15. Rossi, M., Cerveri, P.: Comparison of supervised and unsupervised approaches for the generation of synthetic ct from cone-beam ct. Diagnostics **11**(8), 1435 (2021)
16. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=St1giarCHLP

17. van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T.: scikit-image: image processing in python. PeerJ **2**, e453 (Jun 2014). https://doi.org/10.7717/peerj.453, http://dx.doi.org/10.7717/peerj.453
18. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5) (Sep 2023). https://doi.org/10.1148/ryai.230024, http://dx.doi.org/10.1148/ryai.230024
19. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023)
20. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)

A Performance of Segmentation Model Trained on Generated MRIs from Unconditional Diffusion Model

		L		L-SPIDER		H & T	
Method	M_{seg}	Dice ↑	ASSD ↓	Dice ↑	ASSD ↓	Dice ↑	ASSD ↓
Uncon-Diff	UNet	0.354	5.360	0.197	7.251	0.281	19.895
Ours	UNet	**0.683**	**1.432**	**0.633**	**2.066**	**0.731**	**3.139**

Table 2: Comparison of diffusion model with/without contour-guided training in terms of downstream task segmentation model (M_{seg}) performance.

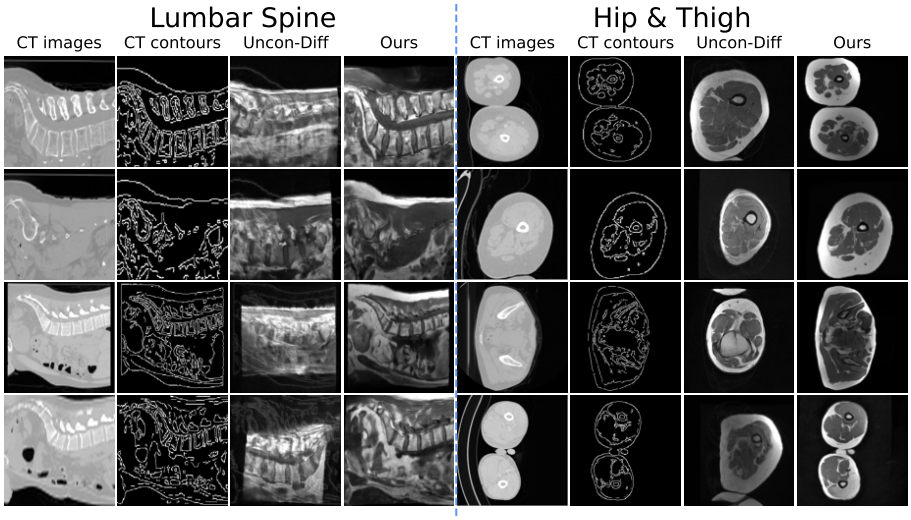

Fig. 4: Ablation study 1: Comparison of the images generated by unconditional diffusion model and that with contour-guided diffusion model

B Comparison between generated images from single candidate and multiple candidates

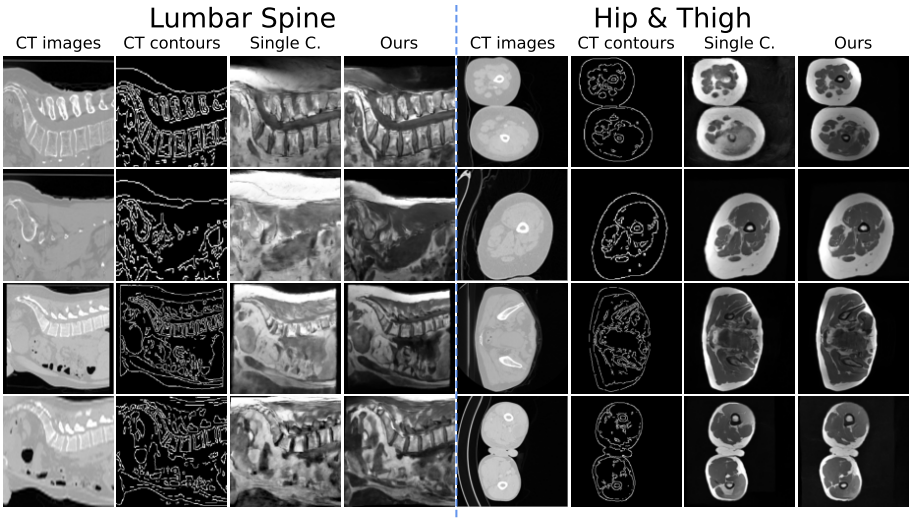

Fig. 5: Ablation study 2: Comparison of the images generated by single candidate (Single C.) without selection and generation using multiple candidates

C FID

Body Part	CycleGAN	SynSeg-Net	CyCADA	MaskGAN	Ours
Lumbar (L)	5.959	9.461	6.139	4.884	21.980
Hip & Thigh (H & T)	17.276	24.241	17.429	24.444	26.430

Table 3: FID between generated MRIs from each translation pipeline and real MRIs.

D CT and MRI Examples

D.1 Hip & Thigh CT

We present the initial, quarter, half, three-quarter, and final slices from several Hip & Thigh CT volumes utilized in this project.

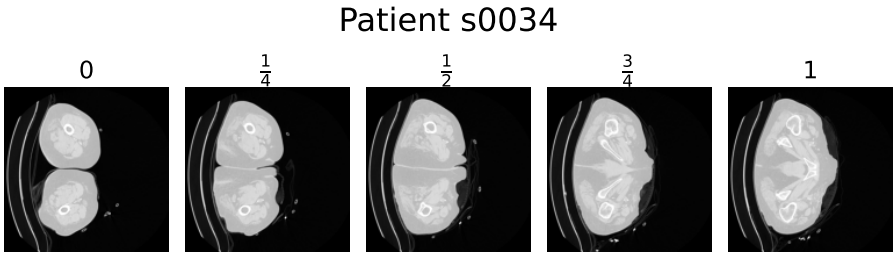

Fig. 6: Example slices for Patient $s0034$ in Hip & Thigh CT Dataset

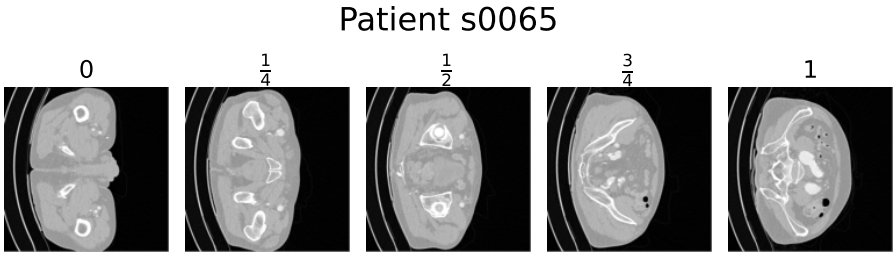

Fig. 7: Example slices for Patient $s0065$ in Hip & Thigh CT Dataset

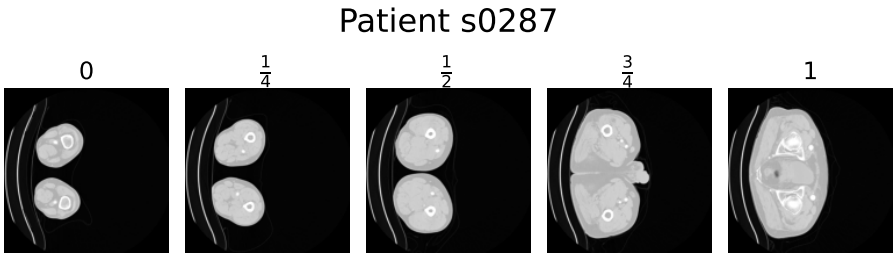

Fig. 8: Example slices for Patient $s0287$ in Hip & Thigh CT Dataset

D.2 Hip & Thigh MRI

We present the initial, quarter, half, three-quarter, and final slices from all Hip & Thigh MRI volumes utilized in this project.

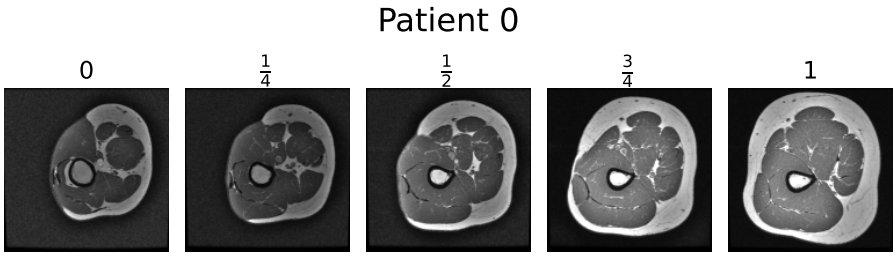

Fig. 9: Example slices for Patient 0 in Hip & Thigh MRI Dataset

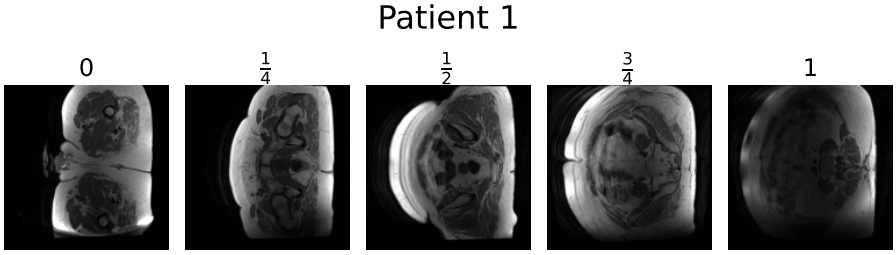

Fig. 10: Example slices for Patient 1 in Hip & Thigh MRI Dataset

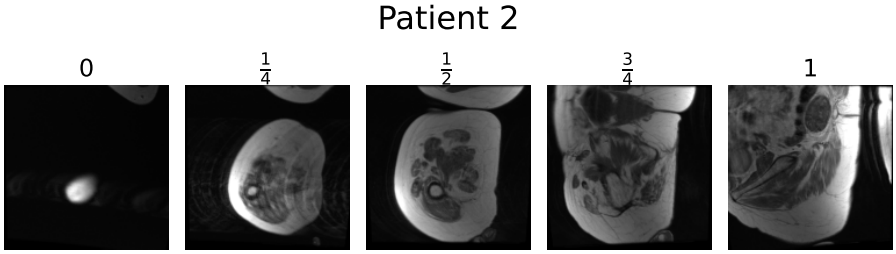

Fig. 11: Example slices for Patient 2 in Hip & Thigh MRI Dataset

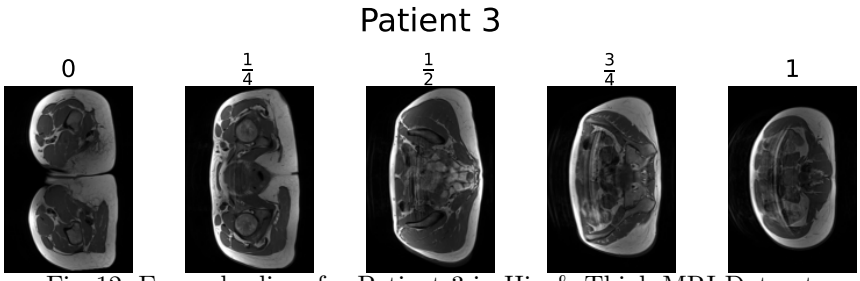

Fig. 12: Example slices for Patient 3 in Hip & Thigh MRI Dataset

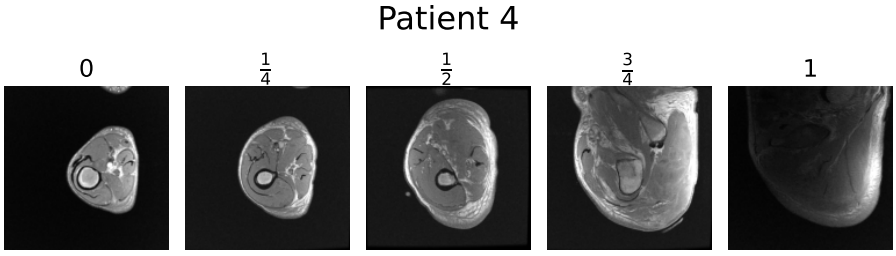

Fig. 13: Example slices for Patient 4 in Hip & Thigh MRI Dataset

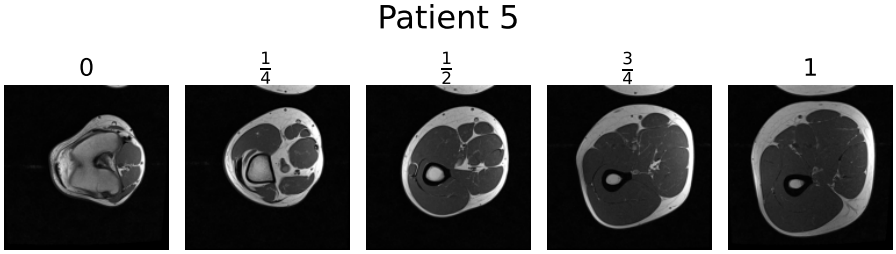

Fig. 14: Example slices for Patient 5 in Hip & Thigh MRI Dataset

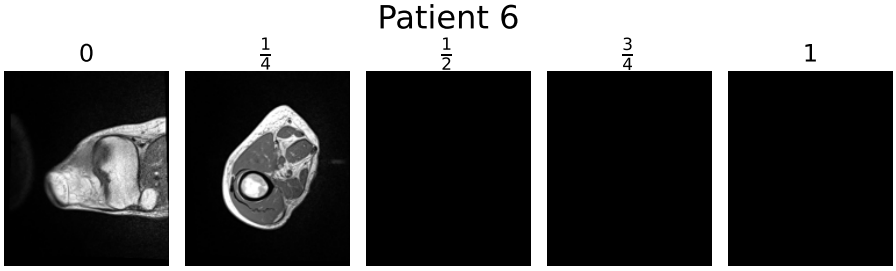

Fig. 15: Example slices for Patient 6 in Hip & Thigh MRI Dataset

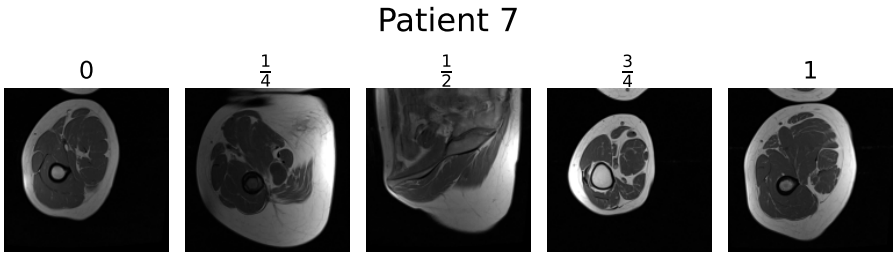

Fig. 16: Example slices for Patient 7 in Hip & Thigh MRI Dataset

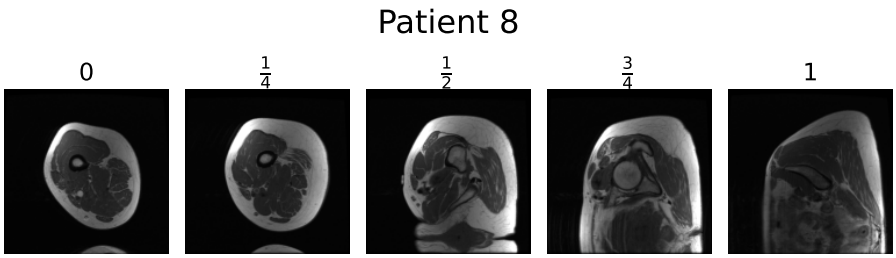

Fig. 17: Example slices for Patient 8 in Hip & Thigh MRI Dataset

D.3 Lumbar CT

We present the initial, quarter, half, three-quarter, and final slices from several lumbar spine CT volumes utilized in this project.

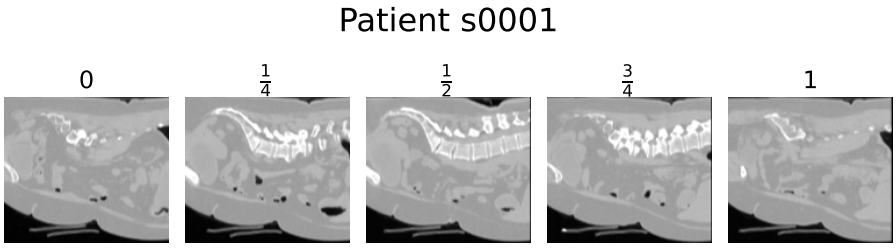

Fig. 18: Example slices for Patient $s0001$ in Lumbar CT Dataset

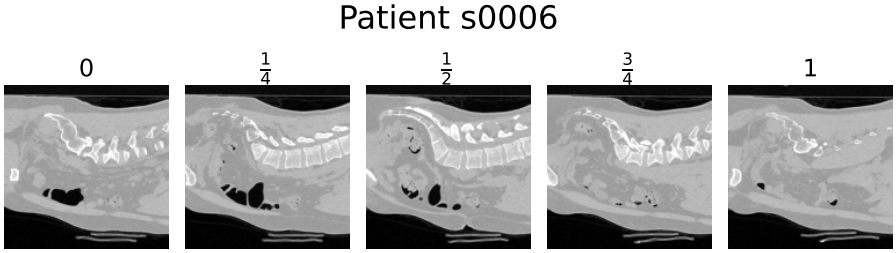

Fig. 19: Example slices for Patient $s0006$ in Lumbar CT Dataset

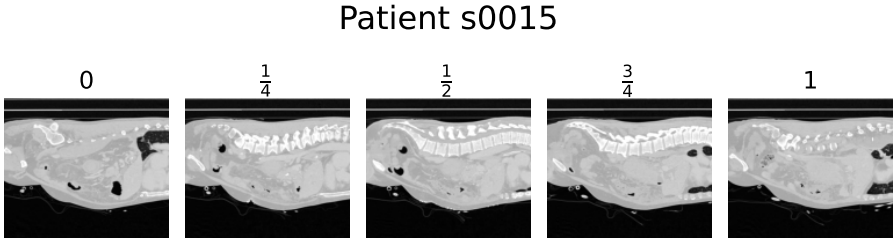

Fig. 20: Example slices for Patient $s0015$ in Lumbar CT Dataset

D.4 Lumbar MRI

We present the initial, quarter, half, three-quarter, and final slices from several lumbar spine MRI volumes utilized in this project.

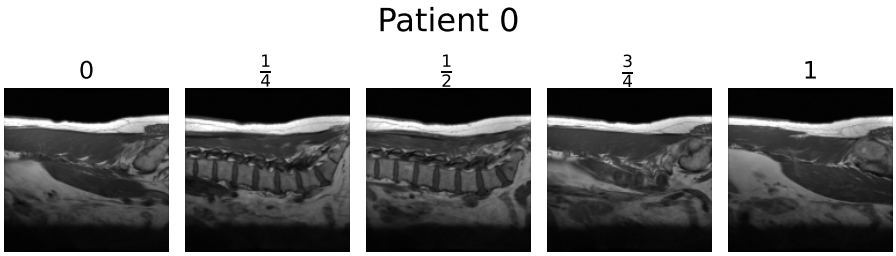

Fig. 21: Example slices for Patient 0 in Lumbar MRI Dataset

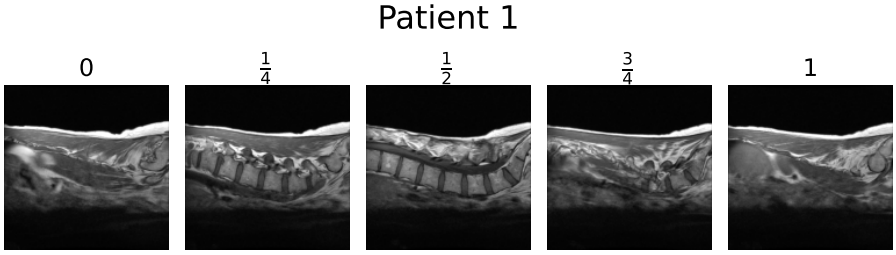

Fig. 22: Example slices for Patient 1 in Lumbar MRI Dataset

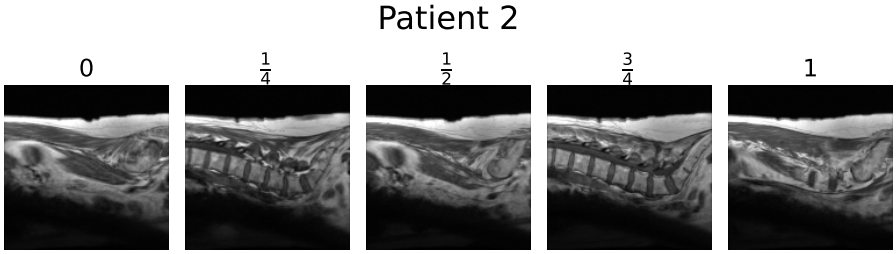

Fig. 23: Example slices for Patient 2 in Lumbar MRI Dataset

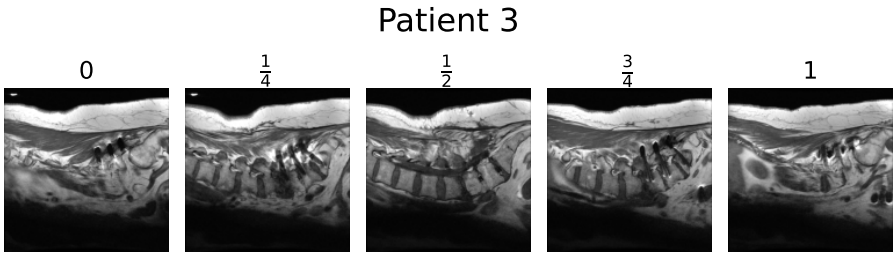

Fig. 24: Example slices for Patient 3 in Lumbar MRI Dataset

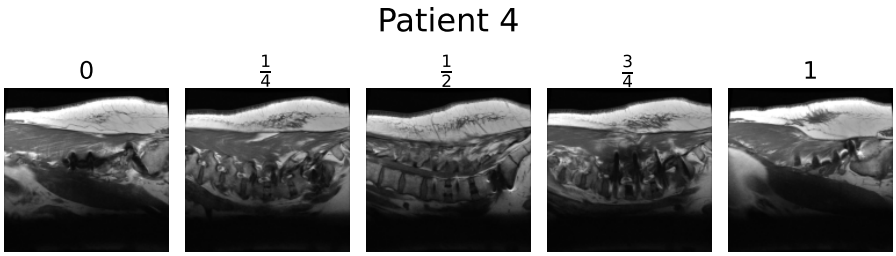

Fig. 25: Example slices for Patient 4 in Lumbar MRI Dataset

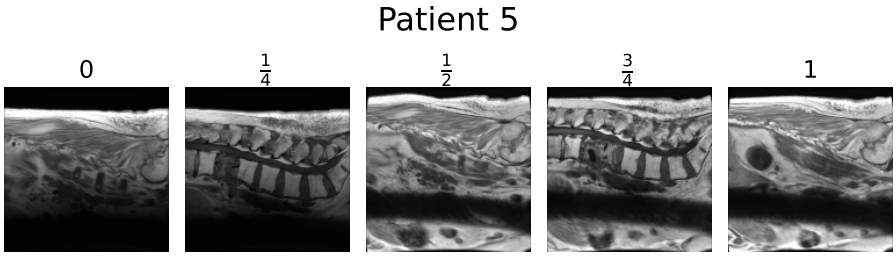

Fig. 26: Example slices for Patient 5 in Lumbar MRI Dataset

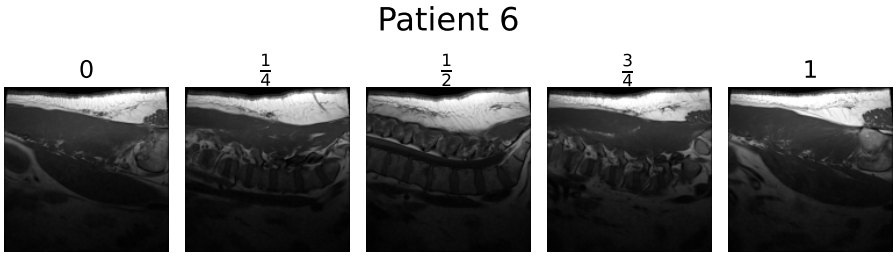

Fig. 27: Example slices for Patient 6 in Lumbar MRI Dataset