

Supplementary Materials: Multi-view X-ray Image Synthesis with Multiple Domain Disentanglement from CT Scans

Anonymous Authors

A OVERVIEW

In our supplementary material, we present additional experimental results in Appendix B. Specifically, we present visualizations on multi-view X-ray image synthesis in Appendix B.1. To further illustrate the effectiveness of our method, we provide additional qualitative comparisons in Appendix B.2. Finally, we present the limitations of current study and future work in Appendix C.

B ADDITIONAL EXPERIMENTS

B.1 Results on Multi-view X-ray Image Synthesis

In this section, we evaluate the performance of multi-view image synthesis by the proposed model. In detail, we randomly selected several volume datas from the CTSpine1K dataset and synthesize seven views with uniformly distributed camera poses for each volume. The multi-view synthesis results are presented in Figure 1. As shown in the Figure, our method can synthesize X-ray images from different view angles.

B.2 Qualitative Comparisons on X-ray Image Synthesis

We provide more qualitative results in this section. Figure 2 illustrates more qualitative comparisons of our method with other state-of-the-art 3D-aware generation methods, including π -GAN and EG3D. From left to right of the Figure, we show multi-view results from -90° to 90° . For each method, we provide X-ray synthesis results from volumes obtained from two patients.

C FAILURE CASES

Recent studies have shown that it is fundamentally impossible to fully disentangle features [1, 2]. Therefore, the use of unpaired X-ray image style information extracted by the style decoupling encoder would unavoidably introduce structural information from the X-ray image and impose an influence on the performance of our model. Despite adding supervised constraints during training, this approach still results in inaccuracies for faulty structures. For instance, as shown in the top row of Figure 3, the model fails to generate bone structures in the correct positions. Furthermore, our model lacks distance perception. In the training dataset, as CT Volumes occupy various physical positions in space, the resulting DRR images exhibit varying scales, which the model erroneously interprets as style cues, leading to inaccurately scaled generated outputs, illustrated in the bottom row of Figure 3. Exploring generative models with better structural constraints and awareness of distance would further improve the performance of our model.

REFERENCES

- [1] Daniella Horan, Eitan Richardson, and Yair Weiss. 2021. When is unsupervised disentanglement possible? *Advances in Neural Information Processing Systems* 34 (2021), 5150–5161.

- [2] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. 2019. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*. PMLR, 4114–4124.



Figure 1: Illustrated results of multi-view image synthesis. From top to bottom: results using different volumes as input. From left to right: results from -90° to 90° with an interval of 30° .

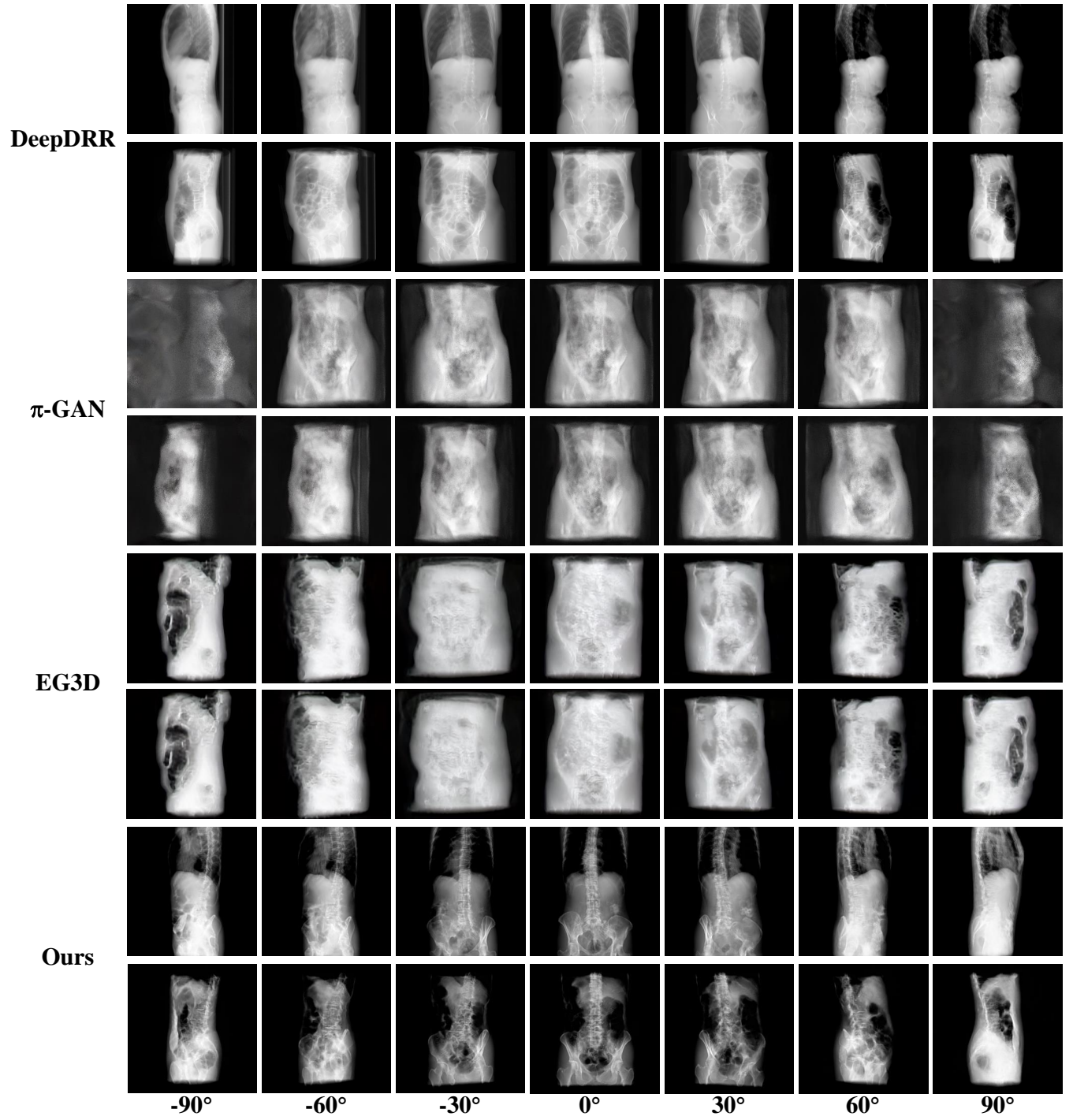


Figure 2: More comparison results between ours and state-of-the-art 3D aware generation methods, including π -GAN and EG3D. As shown in the figure, our model can effectively maintain the anatomical structures and propel the style of the results more similar to real X-ray images.

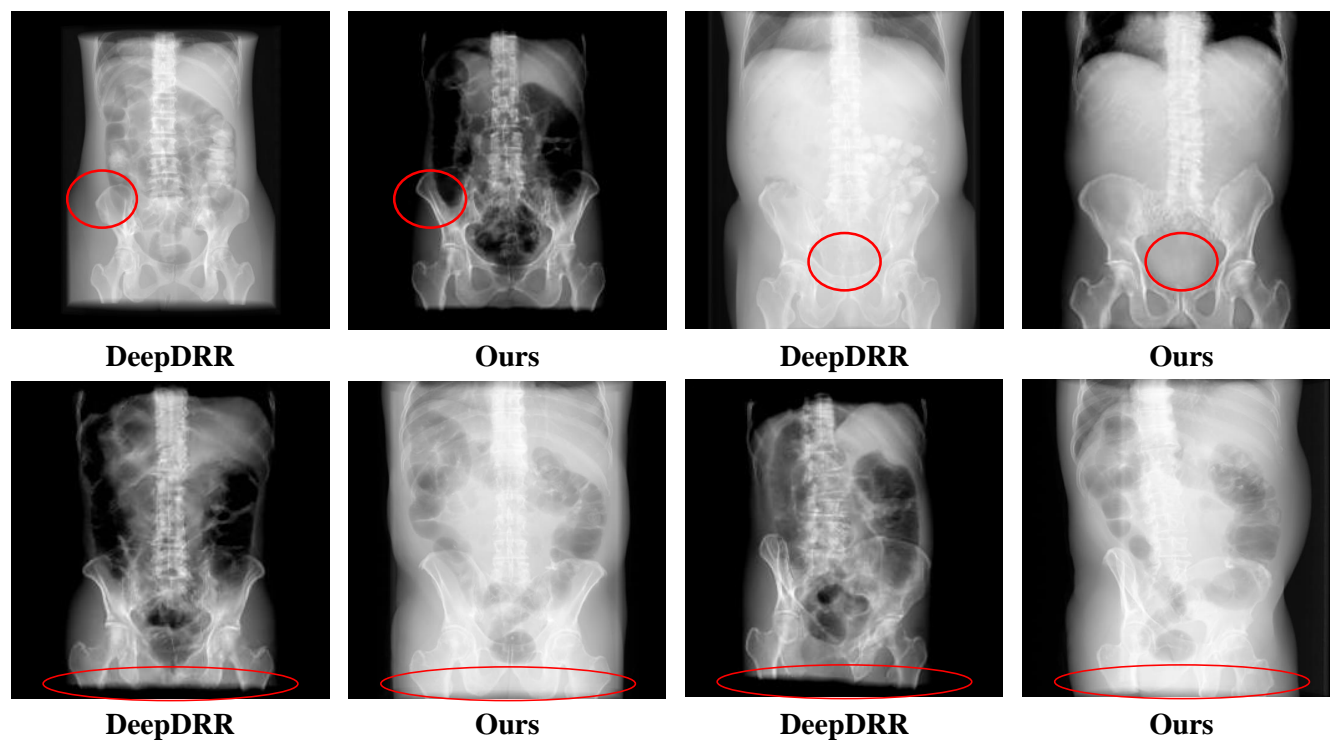


Figure 3: Illustration of failed cases. The top row shows wrongly generated structures. The bottom row shows wrongly scaled results. The model fails to generate bones in the correct position due to insufficient structural constraints and the absence of distance awareness. Exploring models with better capability for capturing structure information and distance information would further improve performance.