

Supplementary Materials: Devil is in Details: Locality-Aware 3D Abdominal CT Volume Generation for Self-Supervised Organ Segmentation

Anonymous Authors

1 SYNTHETIC VOLUMES FOR SELF-SUPERVISED ORGAN SEGMENTATION ON TOTALSEGMENTATOR DATASET

To assess the efficacy of synthetic volumes for self-supervised learning (SSL) tasks, we conduct experiments on TotalSegmentator dataset [4]. Specifically, we utilize 614 synthetic data samples generated from each method, including Medical Diffusion [1] and Lad, as the training set for the self-supervised segmentation model, SSL-ALPNet [3]. Similar to our experiments on AbdomenCT-1K dataset [2], we perform five training runs of the segmentation model and average the test results to evaluate its performance for each training set. The segmentation tests are conducted on models trained on different training sets from TotalSegmentator dataset, and the dice scores obtained are presented in Figure 1.

1.1 Synthetic Data for Training.

In line with our main paper’s experiments on AbdomenCT-1K dataset, we utilize synthetic data from TotalSegmentator dataset as substitutes for real data in training the segmentation model. This allows us to evaluate the genuine impact of synthetic data on downstream feature learning tasks, all without incorporating any real data in the process. As illustrated in Figure 1a, the segmentation model trained on synthetic volumes from Lad outperforms the one trained on synthetic volumes from Medical Diffusion in terms of mean dice scores for four abdominal organs. This mirrors the phenomenon observed in Figure 6 of our main paper, where higher dice scores are achieved for small organs such as the pancreas and spleen. Specifically, our method’s emphasis on granularity significantly enhances the delineation of anatomical structures’ details, resulting in an improvement of 1.6% and 3.6% over Medical Diffusion for the pancreas and spleen, respectively. Despite prioritizing local features, our method achieves comparable performance to Medical Diffusion on larger organs such as the liver and kidney.

1.2 Synthetic Data for Augmentation.

In line with our main paper’s experiments on the AbdomenCT-1K dataset, we integrate synthetic volumes, each approximately 20% the size of the training set, into the self-supervised segmentation model as a form of data augmentation. Figure 1b demonstrates that our method achieves the highest mean dice scores, indicating its effectiveness in facilitating the learning of visual representations by the SSL model. Notably, our method exhibits particular strength in addressing the challenges associated with small organs, effectively bridging the performance gap observed when compared to Medical Diffusion.

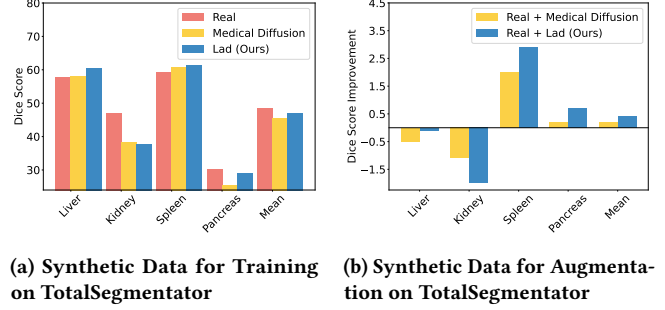


Figure 1: Abdominal Organ Segmentation Performance Comparison on Different Trainsets of TotalSegmentator Dataset. (a) Dice Score on Synthetic Data. The segmentation model trained on synthetic volumes from Lad outperforms that trained on synthetic volumes from Medical Diffusion in mean dice scores for four abdominal organs. Higher dice scores are achieved for small organs, resulting in a 1.6% and 3.6% improvement over Medical Diffusion for the pancreas and spleen, respectively. Despite prioritizing local features, our method achieves comparable performance to Medical Diffusion on larger organs such as the liver and kidney. (b) Dice Score Improvement on Augmented Trainset. The mean dice score of four abdominal organs improves when using synthetic volumes from Lad as an augmentation method, especially on small organs such as pancreas and spleen, effectively bridging the performance gap observed compared to Medical Diffusion.

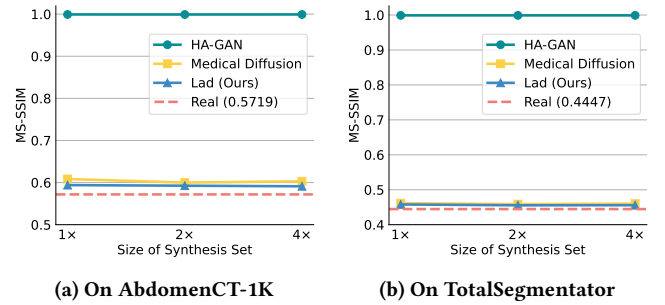


Figure 2: MS-SSIM Scores of Synthesis Sets of Varying Sizes Across Different Methods on Abdomenct-1K dataset and TotalSegmentator dataset. “1x” denotes a synthesis set size identical to that of the trainset, while “2x” and “4x” indicate set sizes two and four times larger than the trainset respectively. Lad showcases diversity comparable to real data across a broad spectrum of synthesized volumes, evidence by lower MS-SSIM scores.

2 GENERATING MASSIVE VOLUMES

Lad exhibits a significant advantage in its ability to generate extensive quantities of high-fidelity abdominal CT volumes using common augmentation techniques applied to the original mask set, while maintaining diversity. The augmented mask set provides a substantial source of conditions for generating diverse volumes. As demonstrated by the multi-scale structural similarity metric (MS-SSIM) scores of synthesis sets of different sizes from various methods on the AbdomenCT-1K dataset in 2, our method achieves lower MS-SSIM scores compared to others on both datasets. This illustrates Lad’s proficiency in generating a wide array of volumetric data with high diversity, regardless of whether the synthesis set size matches that of the training set or is quadruple its size.

REFERENCES

- [1] Firas Khader, Gustav Müller-Franzes, Soroosh Tayebi Arasteh, Tianyu Han, Christoph Haarbuerger, Maximilian Schulze-Hagen, Philipp Schad, Sandy Engelhardt, Bettina Baeßler, Sebastian Foersch, et al. 2023. Denoising diffusion probabilistic models for 3D medical image generation. *Scientific Reports* 13, 1 (2023), 7303.
- [2] Jun Ma, Yao Zhang, Song Gu, Cheng Zhu, Cheng Ge, Yichi Zhang, Xingle An, Congcong Wang, Qiyuan Wang, Xin Liu, et al. 2021. Abdomenct-1k: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 10 (2021), 6695–6714.
- [3] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. 2022. Self-supervised learning for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging* 41, 7 (2022), 1837–1848.
- [4] Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. 2023. Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence* 5, 5 (2023).