Supplementary Material for MedGSSR: Generalizable Medical Image Super-Resolution 3D Reconstruction via Hierarchical Feed-forward Gaussian Splatting

Chengkai Wang[1*], Luoyu Hong[2*], Yiting Zhao[2], Jiamin Wang[3], Xiang Feng[3], Feiwei Qin[2], Zhenzhong Kuang[2], Xuefei Yin[4]
Ali Bashashati[1✉], and Yanming Zhu[4✉]

[1] University of British Columbia, Vancouver, Canada
[2] Hangzhou Dianzi University, Hangzhou, China
[3] ShanghaiTech University, Shanghai, China
[4] Griffith University, Gold Coast, Australia
william2ai.github.io/medgssr

1 Additional Visual Comparisons for Intra-Domain 3D Super-Resolution

To complement the quantitative results in the main paper, we provide additional visual comparisons for intra-domain 3D super-resolution on both MRI and CT data. For each example, we show three anatomical planes, namely axial, coronal, and sagittal views, together with enlarged local regions to better highlight subtle structural differences. Results are presented at 2×, 3×, and 4× super-resolution scales.

Figure 1 shows the intra-dataset MRI super-resolution results on the Medical Segmentation Decathlon (MSD) dataset [1]. As the upsampling factor increases, the interpolation baselines become increasingly blurry and fail to preserve clear anatomical boundaries. Although CuNeRF and ArSSR recover part of the global contrast, they still exhibit noticeable over-smoothing and lose fine cortical folding patterns, especially at 4×. In contrast, MedGSSR restores sharper structures and more faithful local details across all three anatomical planes, with clearer recovery of thin folds and tissue boundaries in the zoomed regions. These visual results are consistent with the quantitative gains reported in the main paper and further demonstrate the advantage of Gaussian-based continuous volumetric modeling for MRI super-resolution.

Figure 2 presents the corresponding intra-dataset CT super-resolution results on the MELA dataset [5]. Compared with MRI, CT images contain both strong macroscopic structures and subtle high-frequency details, which makes accurate super-resolution particularly challenging. As shown in the figure, Trilinear and Cubic interpolation produce blurred boundaries and degraded local contrast, while CuNeRF and ArSSR still struggle to faithfully recover fine structures in challenging regions. By comparison, MedGSSR better preserves

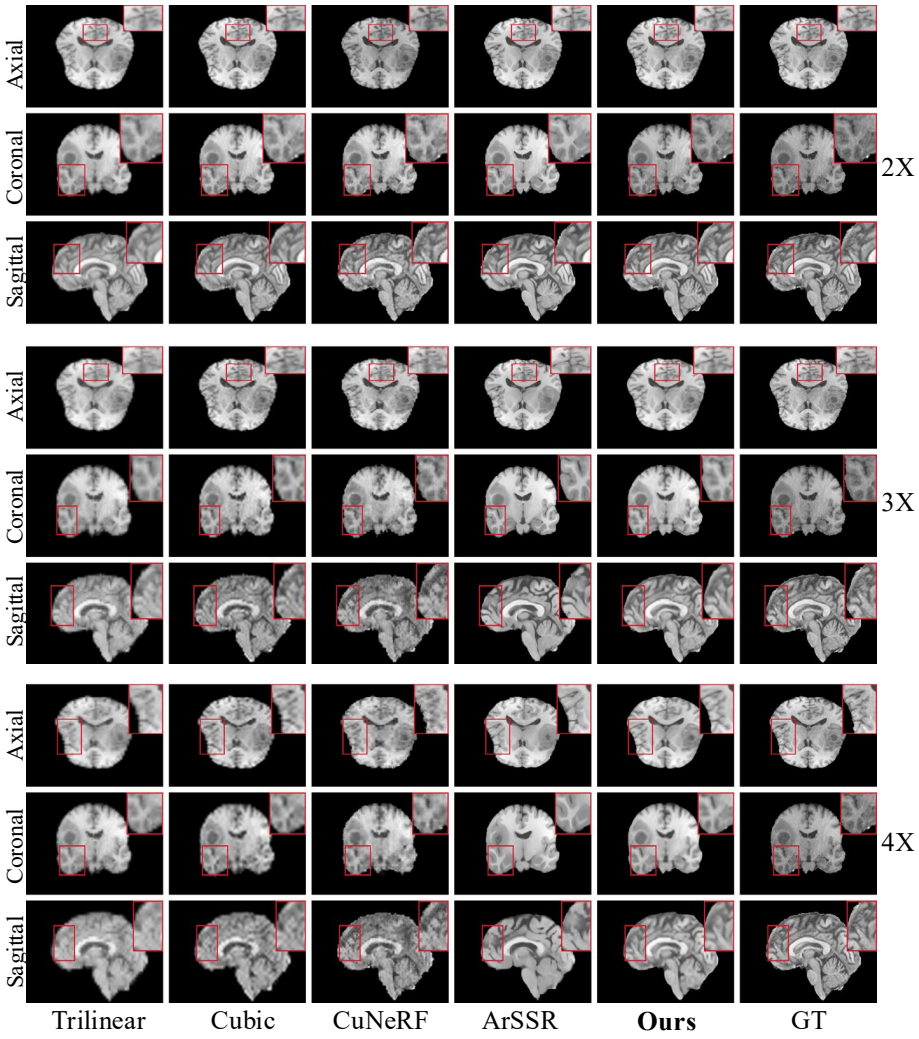

Fig. 1: Additional visual comparisons for intra-domain 3D super-resolution on the MSD (MRI) dataset. Results are shown across axial, coronal, and sagittal views at 2×, 3×, and 4× upsampling scales, with zoomed-in regions for detailed inspection. From left to right: Trilinear, Cubic, CuNeRF, ArSSR, Ours, and GT. MedGSSR better preserves cortical folds, tissue boundaries, and local anatomical details than competing methods.

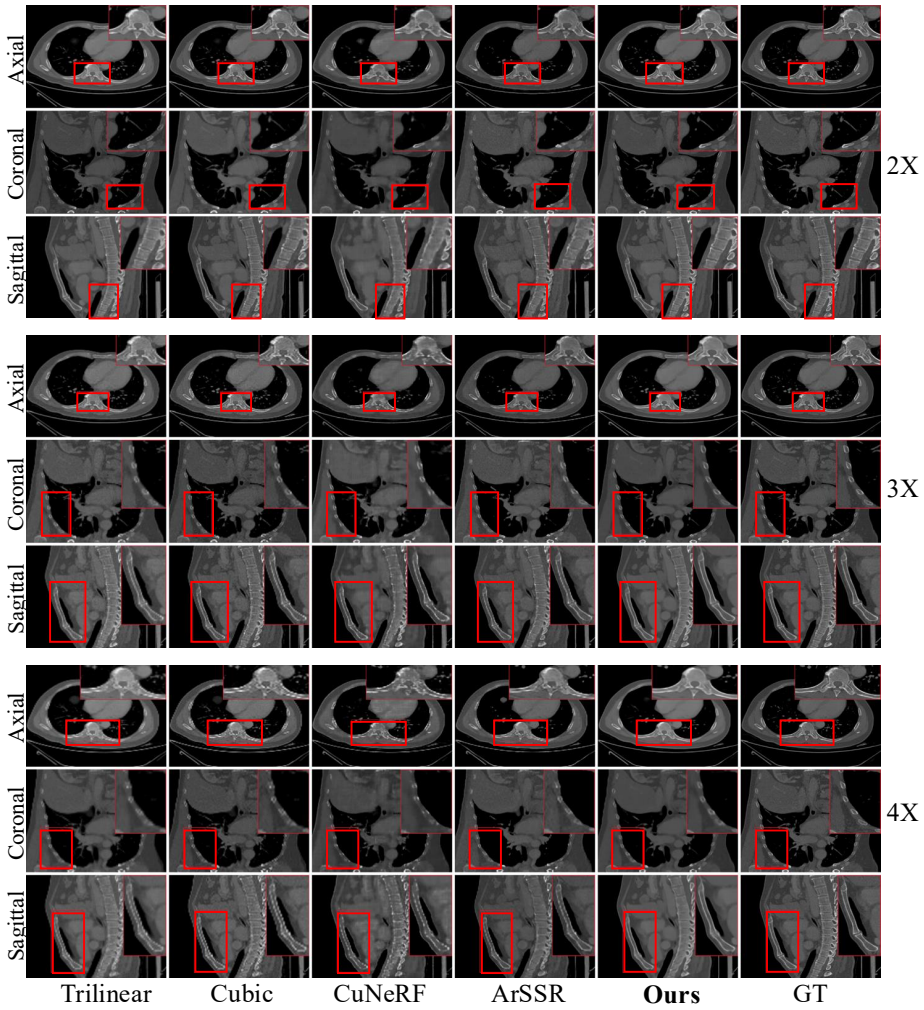

Fig. 2: Additional visual comparisons for intra-domain 3D super-resolution on the MELA (CT) dataset. Results are shown across axial, coronal, and sagittal views at 2×, 3×, and 4× upsampling scales, with zoomed-in regions for detailed inspection. From left to right: Trilinear, Cubic, CuNeRF, ArSSR, Ours, and GT. Compared with the baselines, MedGSSR produces clearer boundaries and more faithful recovery of fine anatomical structures across all super-resolution.

anatomical boundaries and restores clearer local details across different scales and planes, while introducing fewer artifacts. The improvements are especially visible in the magnified regions, where our method produces reconstructions that are consistently closer to the ground truth.

Overall, these visual comparisons further support the main paper by showing that the proposed Gaussian-based super-resolution framework consistently preserves both global anatomical structure and local high-frequency detail under intra-dataset settings.

2 Additional Visual Comparisons for Cross-Domain 3D Super-Resolution Generalization

To complement the cross-domain quantitative results reported in the main paper, we provide additional visual comparisons on unseen MRI and CT datasets. These visualizations are intended to further illustrate the robustness of the proposed method under distribution shift.

Figure 3 presents additional cross-dataset MRI super-resolution results on the unseen Human Connectome Project (HCP) dataset [6]. All models are trained exclusively on the MSD dataset and directly evaluated on HCP at 2×, 3×, and 4× upsampling scales. As shown across the axial, coronal, and sagittal views, interpolation based baselines become increasingly blurry as the scale factor increases, while CuNeRF and ArSSR exhibit noticeable degradation in fine anatomical structures under dataset shift. In contrast, MedGSSR preserves clearer tissue boundaries, more coherent cortical patterns, and more faithful local details across all three planes. These visual observations are consistent with the cross-dataset quantitative results in the main paper and further support the strong generalization capability of the proposed Gaussian-based super-resolution framework.

Figure 4 shows the corresponding cross-dataset CT results on the unseen Ultra-High-Resolution CT (UHRCT) dataset [2], using models trained only on the MELA dataset. Following the evaluation protocol of this benchmark, we visualize the axial plane, which is also the most reliable plane for qualitative comparison under the native acquisition setting. In addition to the 4× case considered in the main paper, we further include more challenging 6× and 8× visual comparisons as an additional stress test of arbitrary-scale generalization. As the resolution gap increases, the competing methods show progressively stronger blurring and structural degradation, whereas MedGSSR still preserves clearer anatomical boundaries and more faithful local structures. These results further demonstrate that the proposed method remains robust under both dataset shift and large upsampling factors.

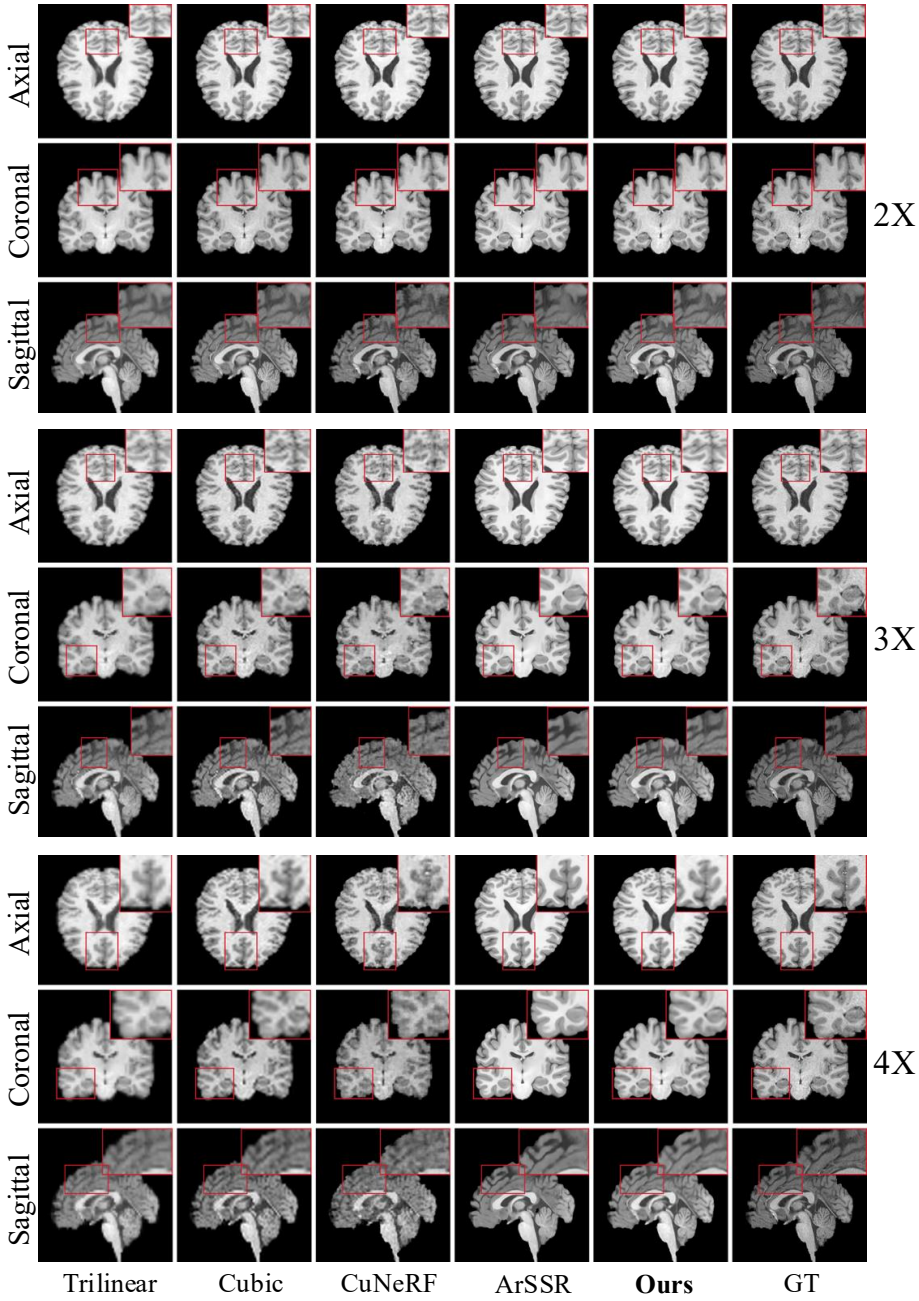

Fig. 3: Additional visual comparisons for cross-domain 3D super-resolution on the unseen HCP (MRI) dataset [6]. Models are trained exclusively on the MSD dataset and evaluated at 2×, 3×, and 4× upsampling scales. Results are shown across axial, coronal, and sagittal views, with zoomed-in regions for detailed inspection. Compared with competing methods, MedGSSR better preserves tissue boundaries, cortical structures, and local anatomical details under dataset shift.

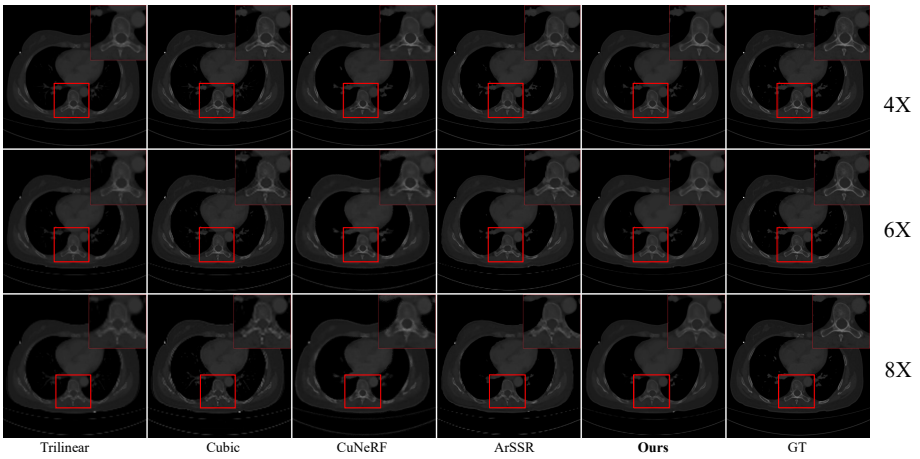

Fig. 4: Additional visual comparisons for cross-dataset 3D super-resolution on the unseen UHRCT (CT) dataset [2]. Models are trained on MELA and evaluated at 4×, 6×, and 8× upsampling scales. Following the benchmark protocol, results are shown on the axial plane, with zoomed-in regions for detailed inspection. MedGSSR remains more robust than competing methods as the resolution gap increases, preserving clearer anatomical boundaries and finer local structures.

3 Extended Baseline and Efficiency Analysis on Cross-Domain HCP MRI

To further validate the cross-dataset generalization and computational efficiency of MedGSSR, we provide an extended comparison on the unseen HCP MRI dataset under the 4× setting. In addition to the baselines reported in the main paper, we include two transformer-based 3D super-resolution methods, MTVNet [4] and SuperFormer [3]. We also report computational cost for all compared methods, including FLOPs, peak memory, parameter count, and inference time.

As shown in Table 1, MedGSSR achieves the best PSNR and SSIM, while requiring the fewest FLOPs and the shortest inference time. Although ArSSR obtains the lowest LPIPS in this setting, it shows lower PSNR/SSIM and higher computational cost than MedGSSR. The two transformer-based baselines, MTVNet and SuperFormer, achieve competitive reconstruction quality but require substantially higher FLOPs and inference time. These results further demonstrate that MedGSSR offers a favorable accuracy-efficiency trade-off under domain shift.

4 Additional Visualization of Downstream Segmentation on Super-Resolved MRI Volumes

To complement the downstream segmentation results reported in the main paper, we provide additional visual comparisons on the unseen HCP dataset using

Table 1: Extended quantitative and computational comparison under cross-domain 4× 3D SR on the HCP dataset. In addition to the baselines reported in the main paper, MTVNet and SuperFormer are included as transformer-based baselines. Computational cost is reported for all methods. Best and second-best results are bolded and underlined, respectively.

Methods	PSNR ↑	SSIM ↑	LPIPS ↓	FLOPs (G)	Peak Mem. (GB)	Params (M)	Inference Time (ms)
MTVNet	<u>34.16±2.86</u>	0.9246±0.0267	0.1368±0.0416	1359.50	1.75	22.61	1083.18
SuperFormer	33.87±3.02	0.9142±0.0273	0.1446±0.0435	1888.20	2.06	19.66	1592.91
CuNeRF	33.12±3.56	<u>0.9382±0.0365</u>	0.1587±0.0557	2061.58	**0.12**	**0.98**	259.76
ArSSR	31.87±3.37	0.9220±0.0324	**0.1141±0.0513**	<u>955.15</u>	8.15	<u>6.28</u>	<u>105.98</u>
MedGSSR(Ours)	**35.84±2.73**	**0.9533±0.0224**	<u>0.1253±0.0378</u>	**385.87**	<u>1.29</u>	18.13	**56.52**

the same segmentation setting as described in the main paper. In particular, while the main paper reports quantitative segmentation results and includes one representative visualization for the 4× case, here we present more comprehensive visual results across axial, coronal, and sagittal views to further illustrate the impact of super-resolution quality on the downstream segmentation performance.

Figure 5 shows additional visual comparisons for downstream brain tissue segmentation based on 4× super-resolved MRI volumes. Consistent with the quantitative and visual results in the main paper, the segmentation masks obtained from MedGSSR are more faithful to the ground truth, with clearer structural boundaries and better preservation of fine anatomical regions. In contrast, the competing super-resolution methods introduce blurring, boundary leakage, or structural distortion, which propagate to the downstream segmentation results and lead to visibly less accurate tissue delineation.

Figure 6 further presents a more challenging 8× setting as an additional stress test beyond the main experiments. As the resolution gap becomes substantially larger, the quality of the super-resolved input becomes even more critical for reliable downstream analysis. Under this challenging setting, the baseline methods exhibit much stronger degradation in the predicted segmentation masks, including boundary confusion, region bleeding, and loss of structural consistency. By comparison, MedGSSR still preserves more coherent anatomical structures and produces segmentation results that remain substantially closer to the ground truth across all three views. These visual results further support that the proposed method preserves task-relevant structural information effectively, even under large upsampling factors and cross-dataset generalization.

5 Additional CT Ablation Study

To further examine whether the design observations on MRI generalize to another imaging modality, we conduct additional ablations on the MELA CT dataset under the intra-domain 4× 3D SR setting. As shown in Table 2, the CT results exhibit trends consistent with the MRI ablations: reducing model capacity or narrowing the Gaussian support consistently degrades reconstruction quality, while moderate data reduction only causes a limited performance drop. Specifically, the full model achieves the best performance across all metrics.

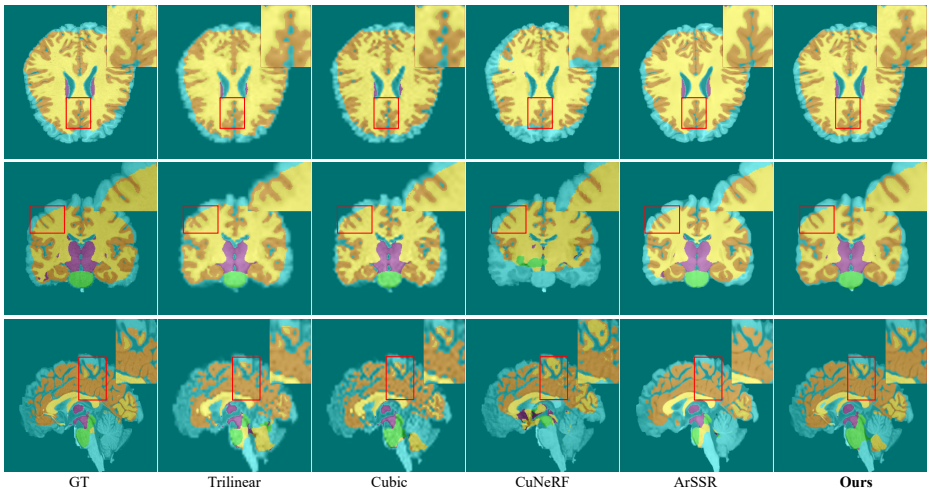

Fig. 5: Additional visual comparisons for downstream brain tissue segmentation under cross-domain on the unseen HCP dataset based on 4× super-resolved MRI volumes. Results are shown across axial, coronal, and sagittal views, with zoomed-in regions for detailed inspection. Compared with competing methods, MedGSSR yields segmentation masks that are more consistent with the ground truth, with clearer boundaries and more faithful anatomical structures.

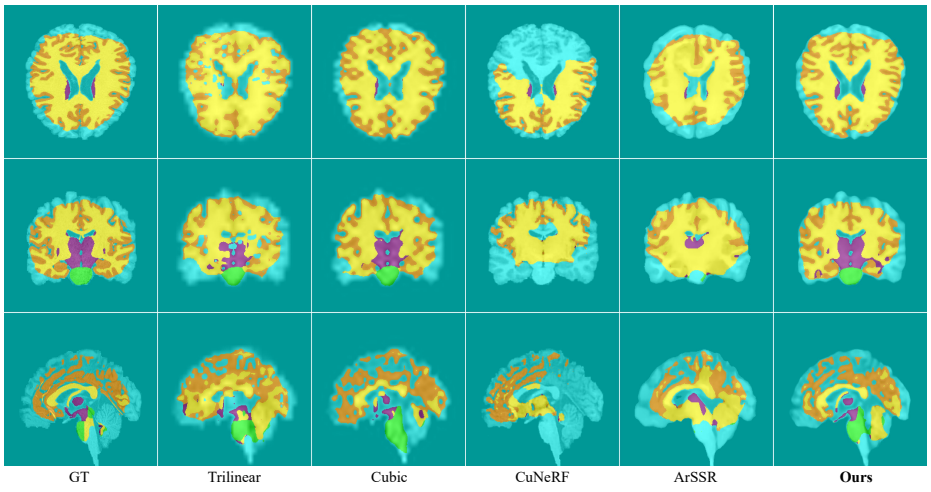

Fig. 6: Additional visual comparisons for downstream brain tissue segmentation under cross-domain on the unseen HCP dataset based on 8× super-resolved MRI volumes. Results are shown across axial, coronal, and sagittal views. This more challenging setting serves as an additional stress test beyond the main experiments. As the upsampling factor increases, the competing methods exhibit substantial degradation in structural consistency, whereas MedGSSR still preserves more faithful anatomical regions and produces segmentation masks that remain noticeably closer to the ground truth.

Table 2: Additional CT ablation on the MELA dataset under the intra-domain 4× 3D SR setting. The results show trends consistent with the MRI ablations and evaluate sub-voxel decomposition, Gaussian truncation radius, and reduced training data. Best results are highlighted in bold.

Variants	PSNR ↑	SSIM ↑	LPIPS ↓
Ours	**37.31**	**0.9362**	**0.1472**
Ours w/ $m = 1$	36.42	0.9234	0.1564
Ours w/ 1σ truncation	33.47	0.8971	0.1827
Ours w/ 75% training data	37.04	0.9338	0.1502

Reducing the sub-voxel count to $m = 1$ decreases PSNR by 0.89 dB and worsens LPIPS, confirming the importance of sub-voxel Gaussian decomposition for representing fine CT structures. The 1σ truncation variant leads to the largest degradation, indicating that sufficient Gaussian support is also critical for CT reconstruction. In addition, the model trained with 75% of the data remains close to the full-data model, demonstrating stable performance under reduced supervision.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
2. Chu, Y., Zhou, L., Luo, G., Qiu, Z., Gao, X.: Topology-preserving computed tomography super-resolution based on dual-stream diffusion model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 260–270. Springer (2023)
3. Forigua, C., Escobar, M., Arbelaez, P.: Superformer: Volumetric transformer architectures for mri super-resolution. In: International workshop on simulation and synthesis in medical imaging. pp. 132–141. Springer (2022)
4. Høeg, A.L., Bardenfleth, S.W., Kjer, H.M., Dyrby, T.B., Dahl, V.A., Dahl, A.: Mtvnet: Mapping using transformers for volumes–network for super-resolution with long-range interactions. arXiv preprint arXiv:2412.03379 (2024)
5. Song, S., Xu, R., Luo, Y., Du, B., Yang, J., Kuang, K., She, Y., Zhao, M.: Mela dataset: A benchmark for mediastinal lesion analysis (annotation v2.0) (May 2022). `https://doi.org/10.5281/zenodo.6520000`
6. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The wu-minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)