Breaking Grid Constraints: Dynamic Graph Reconstruction Network for Multi-organ Segmentation

Supplementary Material

In this supplementary material, we provide additional experimental details and results that were omitted from the main paper:

- **Appendix A**: Additional related works;
- **Appendix B.1**: Additional quantitative results;
- **Appendix B.2**: Additional qualitative analysis;
- **Appendix B.3**: Additional mis-segmentation analysis;
- **Appendix B.4**: Additional organ reconstruction analysis.

A. Additional Related Works

In this supplementary section, this paper systematically organizes the related works and conducts a comprehensive analysis of their strengths and limitations in multi-organ segmentation tasks. The involved methodologies include: CNN-based medical image segmentation approaches, Transformer-based medical image segmentation approaches, and graph neural network-based vision methods.

A.1. CNN-based Segmentation Methods

As the most widely adopted deep learning architecture for segmentation tasks, CNN-based methods play a pivotal role in multi-organ segmentation. The pioneering CNN approach is the Fully Convolutional Network (FCN) [11], which markedly improves the performance of deep learning methods in medical image segmentation through superior organ texture modeling capabilities and end-to-end training paradigms. Subsequently, Ronneberger et al. introduce U-Net [15], which establishes encoder-decoder connectivity through skip connections and remains the benchmark for most medical image segmentation tasks today. The success of U-Net catalyzes the development of numerous U-Net variants, including Attention U-Net [13] that suppresses background noise via attention mechanisms, Inception U-Net [18] for multi-scale feature fusion, U-Net++ [19] with dense skip connections enhancing feature interactivity, and V-Net [12] employing 3D convolution for direct volumetric segmentation. In recent years, CNN-based segmentation approaches have increasingly focused on lightweight architectures and irrelevant information suppression. Tiny U-Net [3] proposed by Chen et al. effectively balance performance and computational efficiency through cascaded multi-scale receptive fields. Ruan et al. [16] achieve multi-scale anatomical modeling via attention mechanisms integrated with group aggregation modules, maintaining performance while reducing model complexity. Notably, Zhu et al. [20] enhance segmentation accuracy by optimizing balanced supervision mechanisms between the encoder and decoder components to eliminate redundant representations.

A.2. Transformer-based Segmentation Methods

The remarkable performance of the Transformer in computer vision has attracted extensive research interest [5]. By enabling global attention through the self-attention mechanism, this paradigm effectively addresses the receptive field limitations inherent in CNN-based segmentation approaches. Consequently, a series of Transformer-based segmentation methods have emerged. Valanarasu et al. explore the feasibility of self-attention mechanisms in medical image segmentation, proposing MedT [17]. Inspired by Swin Transformer [10], Cao et al. develop SwinUNet [2] for multi-organ segmentation, achieving outstanding performance. However, Transformer-based methods excel at capturing global contextual relationships but require substantial training data, which restricts their applicability to centralized and fine-grained medical image segmentation tasks. To address these limitations, researchers have integrated CNN and Transformer methodologies to preserve local details while capturing global anatomical relationships. Representative hybrid approaches include TransUNet [4], UTNet [6], Daeformer [1], and EMCAD [14], which have demonstrated superior performance in medical image segmentation tasks.

A.3. Graph-based Vision Methods

Graphs demonstrate exceptional structural representation capabilities, as they can not only model diverse geometric morphologies but also represent inter-structural relationships through node connectivity. Compared to CNN and Transformer paradigms that operate on regular grid structures, graph representations break free from fixed grid constraints, enabling flexible modeling of irregular anatomies. Recently, graph-based vision architectures have garnered significant attention. Han et al. propose ViG [7], which transforms images into graph representations and employs message passing between nodes for feature learning. Building upon ViG, Han et al. further introduce ViHG [8] by incorporating hypergraph theory, where hyperedge representations replace conventional adjacency matrices, significantly reducing graph construction complexity. Notably, there exists an intrinsic connection between Graphs and Transformers. Joshi demonstrates that Transformers essentially constitute fully connected variants of graph neural

Table 1. The comprehensive performance comparison (in recall, sensitivity and 95HD (↓)) of MoDGR with SOTA segmentation architectures on CHAOS-T2, ACDC, BTCV, Cervix and CMR datasets. ♣ CNN-based, ♠ Transformer-based, ♦ Hybrid architectures, ♥ Graph-based.

Methods	CHAOS-T2			ACDC			BTCV			Synapse-C			CMR		
	Sen.	Pre.	95HD	Sen.	Pre.	95HD	Sen.	Pre.	95HD	Sen.	Pre.	95HD	Sen.	Pre.	95HD
♣ EGE (MICCAI 23')	.7673	.8271	78.5321	.9091	.9482	22.7375	.7030	.8136	74.2004	.5218	.6680	53.6636	.8845	.8952	25.4532
♣ SelfRegUNet (MICCAI 24')	.7420	.8287	68.7118	.9056	.9181	20.2689	.7034	.7304	83.3533	.4619	.5775	58.4830	.8966	.8774	22.7412
♣ TinyUNet (MICCAI 24')	.7482	.8059	54.9545	.8937	.9239	24.9172	.6875	.7713	89.6362	.5021	.6090	54.9693	.8535	.8823	25.7416
♠ MedT (MICCAI 21')	.5955	.7232	83.0137	.9056	.9424	22.5387	.6227	.6895	94.5674	.4719	.5986	58.9834	.7986	.8272	30.3032
♠ SwinUNet (ECCV 22')	.7018	.7617	83.9388	.8185	.8548	33.7631	.6543	.7438	97.1356	.4333	.6065	56.4255	.7709	.7790	38.5128
♦ UTNet (MICCAI 21')	.7592	.8498	61.4286	.9063	.9581	19.3013	.7339	.8445	84.9672	.5477	.6112	49.5132	.9116	.8967	21.6566
♦ Daeformer (MICCAI 23')	.7658	.8290	61.2261	.8929	.9352	27.6894	.6607	.7718	64.7749	.4335	.6353	57.1806	.8230	.8526	30.2255
♦ EMCAD (CVPR 24')	.7761	.8728	92.0817	.8965	.9541	28.8562	.6627	.7658	67.9781	.5299	.6766	51.9387	.8615	.8857	23.3042
♥ **DGRNet(ours)**	**.8471**	**.8792**	**39.2901**	**.9155**	**.9667**	**17.9127**	**.8551**	**.8795**	**34.5149**	**.6594**	**.7364**	**23.8489**	**.9157**	**.9191**	**14.9538**

networks [9]. Given the inherent irregularity of anatomical structures in multi-organ segmentation tasks and the critical importance of inter-organ relationships for precise anatomical modeling, this paper proposes DGRNet to leverage graph representations for capturing diverse anatomical structures. This approach effectively overcomes the fixed-grid limitations of CNN and Transformer-based methods, thereby enhancing multi-organ segmentation performance.

B. Additional Experiment Results

This section provides additional comparative analyses to validate the superior performance of DGRNet further. The supplementary evaluation comprises quantitative analysis, qualitative analysis, mis-segmentation analysis and organ-reconstruction analysis, which holistically reinforce the experimental findings in the main text.

B.1. Additional Quantitative Analysis

In supplementary quantitative analysis, we further validate the effectiveness of DGRNet through Sensitivity (Sen.), Precision (Pre.), and 95% Hausdorff Distance (95HD). Notably, Sen. and Pre. reflect the discriminative capability for organ classes, while 95HD quantifies boundary alignment between predictions and GT. As shown in Table 1, CNN-based methods generally outperform Transformer-based counterparts, attributed to their local receptive fields better suited for organ region characteristics. Hybrid networks combining CNN and Transformer components achieve improved segmentation performance through synergistic global-local spatial dependency modeling of anatomical structures. Nevertheless, DGRNet surpasses all compared methods, with graph-based organ modeling demonstrating exceptional adaptability to irregular morphology, evidenced by its lowest 95HD values across all datasets. DGRNet can explicitly enforce boundary constraints through category-specific priors, which simultaneously improve organ class discriminability. Consequently, DGRNet achieves optimal Sen. and Pre. on all five datasets, outperforming state-of-the-art methods.

B.2. Additional Qualitative Analysis

In supplementary qualitative analyses, we provide more comprehensive visual comparisons to demonstrate the superior segmentation capability of DGRNet. As illustrated in Figure 1, magnified views of critical anatomical regions are provided with corresponding GT annotations for reference. Our method exhibits remarkable adaptability to inter-organ morphological variations while maintaining precise delineation across organs of diverse scales. Notably, DGRNet achieves anatomically consistent segmentation even in regions with ambiguous tissue boundaries (shown in Row 2 of the BTCV dataset), where information aggregation methods typically produce fragmented predictions. Extended visualization results further validate the robustness of DGRNet in handling complex multi-organ scenarios, particularly outperforming existing approaches in preserving topological correctness for small-scale anatomical structures. This visual evidence aligns with our quantitative findings, confirming the effectiveness of the proposed dynamic graph reconstruction paradigm.

B.3. Additional Mis-segmentation Analysis

In the supplementary mis-segmentation analysis, we provide additional visual evidence demonstrating the superior capability of DGRNet to classify segmented regions into target organ categories correctly. As shown in Figure 2, state-of-the-art methods exhibit mis-segmentation conditions in anatomically interleaved regions, where correctly segmented areas are classified to incorrect organ classes. Notably, Transformer-based methods show fewer mis-segmentations than CNN-based approaches, attributed to their global receptive fields better capturing inter-organ spatial dependencies for class discrimination. Nevertheless, DGRNet performs better than the comparative baseline, a superiority enabled by its category-aware guidance mechanism. This mechanism injects category-specific priors during organ reconstruction, simultaneously encoding anatomical features and their semantic identities to mitigate feature ambiguity. Consequently, DGRNet maintains precise class-aware representations even in high-complexity regions.

B.4. Additional Organ Reconstruction Analysis

In supplementary organ reconstruction analyses, we present visual exemplars from five datasets to validate the organ modeling capabilities of DGRNet. As demonstrated in Figure 3, CNN-based methods exhibit finer-grained organ morphology modeling compared to Transformer-based approaches. This advantage stems from two factors: (1) convolutional kernels, being significantly smaller than Transformer patch blocks, enable finer anatomical detail preservation; (2) the overlapping nature of convolutional operations contrasts with Transformers' disjoint patch processing. However, both paradigms rely on information aggregation approaches that fundamentally conflict with irregular organ geometries, inducing cross-organ interference that manifests as blurred morphological and boundary representations. In contrast, DGRNet leverages the inherent flexibility of graph topology to align with anatomical irregularities while incorporating category-specific priors to reinforce boundary delineation. This dual mechanism enables precise, anatomy-aware modeling that robustly adapts to both morphological variations and spatial dependencies across multi-organ configurations. Visualizations of deep feature maps reveal that the dynamic graph reconstruction of DGRNet maintains a sharper focus on target organs and aligns the referenced GT.

References

[1] Reza Azad, René Arimond, Ehsan Khodapanah Aghdam, Amirhossein Kazerouni, and Dorit Merhof. Dae-former: Dual attention-guided efficient transformer for medical image segmentation. In *International Workshop on Predictive Intelligence in Medicine*, pages 83–95. Springer, 2023. 1

[2] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 1

[3] Junren Chen, Rui Chen, Wei Wang, Junlong Cheng, Lei Zhang, and Liangyin Chen. Tinyu-net: Lighter yet better u-net with cascaded multi-receptive fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 626–635. Springer, 2024. 1

[4] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 97:103280, 2024. 1

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1

[6] Yunhe Gao, Mu Zhou, and Dimitris N Metaxas. Utnet: a hybrid transformer architecture for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 61–71. Springer, 2021. 1

[7] Kai Han, Yunhe Wang, Jianyuan Guo, Yehui Tang, and Enhua Wu. Vision gnn: An image is worth graph of nodes. *Advances in Neural Information Processing Systems*, 35:8291–8303, 2022. 1

[8] Yan Han, Peihao Wang, Souvik Kundu, Ying Ding, and Zhangyang Wang. Vision hgnn: An image is more than a graph of nodes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19878–19888, 2023. 1

[9] Chaitanya K Joshi. Transformers are graph neural networks. *arXiv preprint arXiv:2506.22084*, 2025. 2

[10] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021. 1

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015. 1

[12] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016. 1

[13] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 1

[14] Md Mostafijur Rahman, Mustafa Munir, and Radu Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024. 1

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer, 2015. 1

[16] Jiacheng Ruan, Mingye Xie, Jingsheng Gao, Ting Liu, and Yuzhuo Fu. Ege-unet: an efficient group enhanced unet for

skin lesion segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 481–490. Springer, 2023. 1

[17] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, part I 24*, pages 36–46. Springer, 2021. 1

[18] Ziang Zhang, Chengdong Wu, Sonya Coleman, and Dermot Kerr. Dense-inception u-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, 192: 105395, 2020. 1

[19] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2019. 1

[20] Wenhui Zhu, Xiwen Chen, Peijie Qiu, Mohammad Farazi, Aristeidis Sotiras, Abolfazl Razi, and Yalin Wang. Selfreg-unet: Self-regularized unet for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 601–611. Springer, 2024. 1

Figure 1. Additional visual comparison with SOTA segmentation networks on CHAOS-T2, ACDC, BTCV, Cervix, and CMR datasets. Red box is the zoomed-in GT. Green box is the zoomed-in predicted mask.

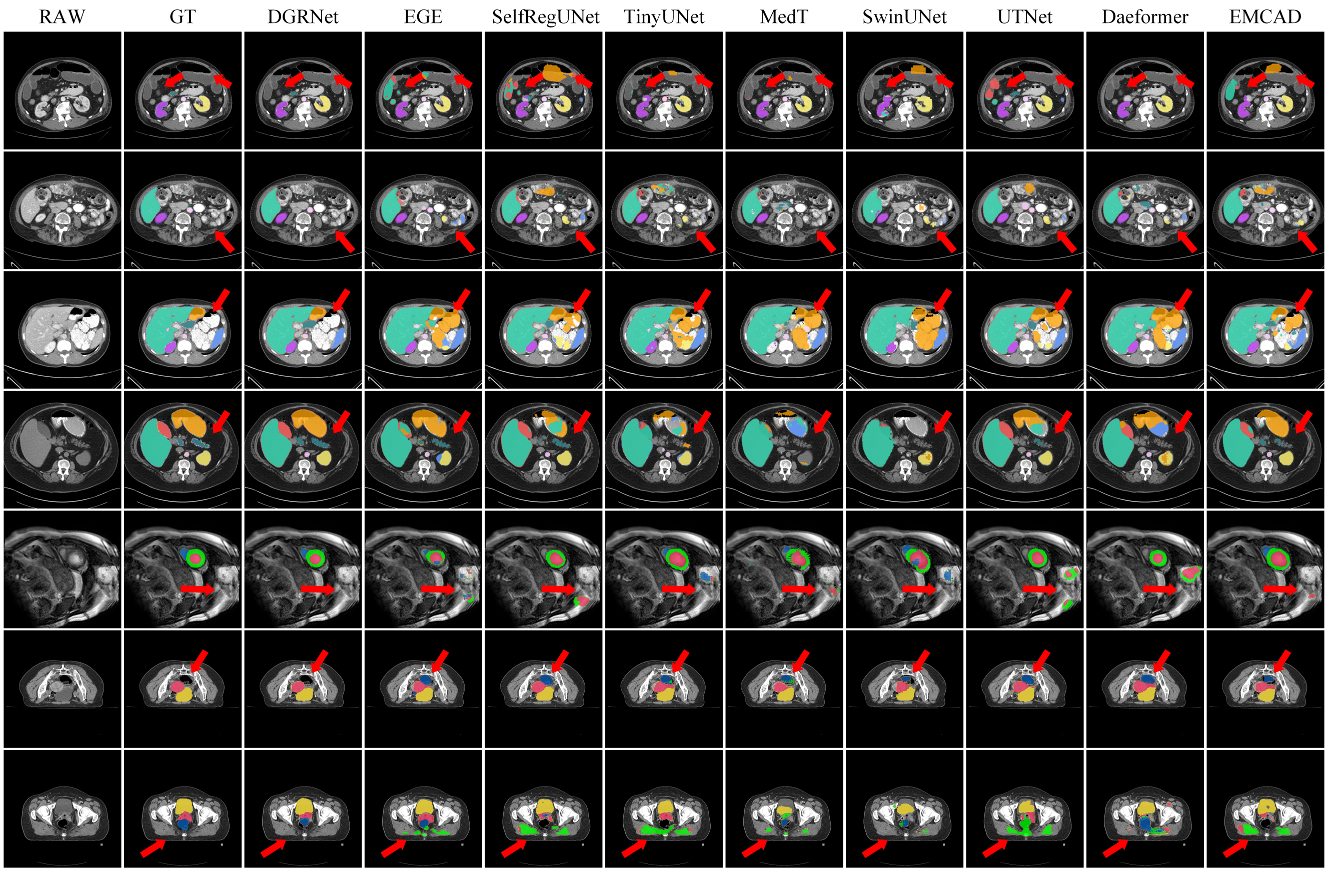

Figure 2. Additional visual comparison of the mis-segmentation between DGRNet and SOTA segmentation networks. ➚ points out the key regions.

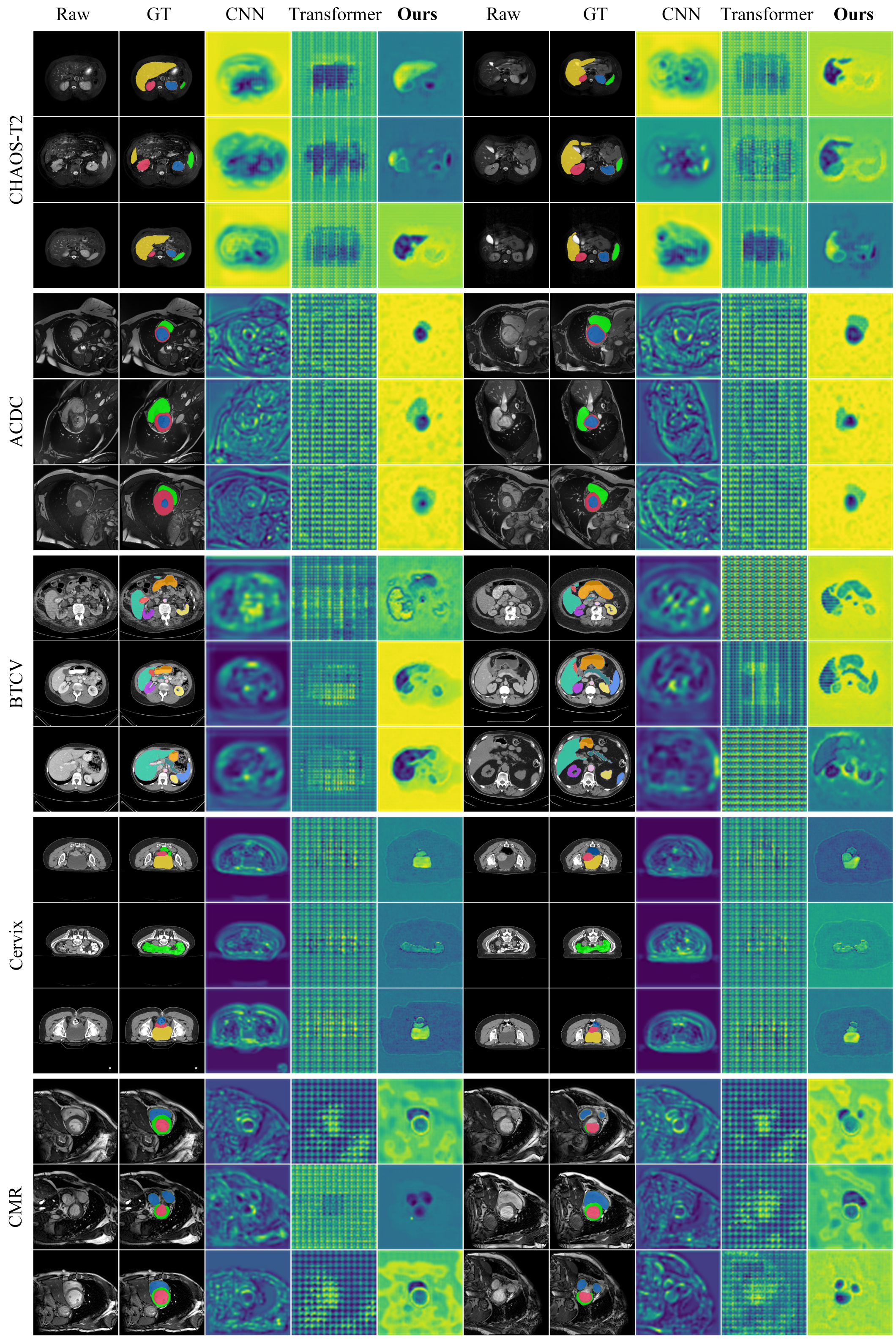

Figure 3. Additional visual comparison of the organ-reconstruction between DGRNet and SOTA segmentation networks on CHAOS-T2, ACDC, BTCV, Cervix, and CMR datasets.