MAST-Pro: Dynamic Mixture-of-Experts for Adaptive Segmentation of Pan-Tumors with Knowledge-Driven Prompts

Runqi Meng[1,2,†], Sifan Song[2,†], Pengfei Jin[2], Lin Teng[1], Yulin Wang[1], Yiqun Sun[1], Ling Chen[2], Yujin Oh[2], Xiang Li[2], Quanzheng Li[2], Ning Guo[2(✉)], Dinggang Shen[1,3,4(✉)]

[1]School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2]Center of Advanced Medical Computing and Analysis, Massachusetts General Hospital and Harvard Medical School, Somerville, 02145, USA
[3]United Imaging Intelligence Co., Ltd., Shanghai 200230, China
[4]Shanghai Clinical Research and Trial Center, Shanghai 201210, China
Dinggang.Shen@gmail.com, Guo.Ning@mgh.harvard.edu

Abstract. Accurate tumor segmentation is crucial for cancer diagnosis and treatment. While foundation models have advanced general-purpose segmentation, existing methods still struggle with: (1) limited incorporation of medical priors, (2) imbalance between generic and tumor-specific features, and (3) high computational costs for clinical adaptation. To address these challenges, we propose MAST-Pro (**M**ixture-of-experts for **A**daptive **S**egmentation of pan-**T**umors with knowledge-driven **Pro**mpts), a novel framework that integrates dynamic Mixture-of-Experts (D-MoE) and knowledge-driven prompts for pan-tumor segmentation. Specifically, text and anatomical prompts provide domain-specific priors, guiding tumor representation learning, while D-MoE dynamically selects experts to balance generic and tumor-specific feature learning, improving segmentation accuracy across diverse tumor types. To enhance efficiency, we employ Parameter-Efficient Fine-Tuning (PEFT), optimizing MAST-Pro with significantly reduced computational overhead. Experiments on multi-anatomical tumor datasets demonstrate that MAST-Pro outperforms State-of-The-Art approaches, achieving up to a 5.20% improvement in average DSC while reducing trainable parameters by 91.04%, without compromising accuracy.

Keywords: Pan-tumor segmentation · Foundation model · Mixture-of-Expert.

1 Introduction

Cancer remains a leading cause of mortality worldwide, with incidence and death rates continuing to rise [1]. Since cancer originates from diverse tumor types,

† Equal contribution

early and accurate tumor segmentation is crucial for improving patient outcomes. However, existing methods [2,3,4,13] are often task-specific, failing to capture shared tumor characteristics and limiting their scalability in large-scale clinical applications. Therefore, developing a unified pan-tumor segmentation model is essential to enhance diagnostic efficiency and facilitate cross-tumor knowledge transfer. However, there are several challenges in the pan-tumor segmentation task, which includes two points: 1) the inherent heterogeneity of tumors across anatomical regions, *i.e.*, exhibiting remarkable diversity of tumors in shape, texture, and intensity, which hinders the adaptability; and 2) the pervasive imbalance in medical datasets, *i.e.*, imbalance distribution of medical datasets for robust feature learning (particularly for rare tumor types), which makes accurate tumor segmentation a daunting task.

Recently, inspired by foundation models such as the Segment Anything Model (SAM) [5] and Contrastive Language-Image Pre-training (CLIP) [6], prompt-driven approaches [7,8,9,10,14] have shown promising performance in medical image segmentation [11]. These methods can be broadly categorized into vision-prompt-driven and text-prompt-driven models. On the one hand, vision-prompt-driven models [7,8] leverage visual cues (*e.g.*, points, bounding boxes) to guide segmentation tasks. While effective, these models heavily depend on manual annotations and fail to incorporate anatomical and radiological priors, which are essential for addressing the high heterogeneity of tumors across different anatomical sites. On the other hand, text-prompt-driven methods [9,10,18] align image and text features within a shared latent space to enhance segmentation across diverse targets. However, their reliance on predefined text templates limits their ability to capture the extensive variability in tumor morphology and radiological presentation, making them less effective in handling domain shifts across anatomical regions. This issue is further compounded by dataset imbalance, resulting in inadequate feature learning and suboptimal segmentation performance, particularly for underrepresented tumor types.

To deal with the imbalance distribution of medical datasets, recent works [10,14,15,12] introduced a query-disentangling and self-prompting model to disentangle queries into organ-level and tumor-specific prompts. While this approach represents a step forward, they often overlook shared morphological patterns across tumor types that enhance feature learning for rare tumors [16], such as edge irregularities and contrast variations. Furthermore, striking a balance between generic and tumor-specific representations still remains challenging, as models struggle to simultaneously retain generic features across anatomical sites while preserving unique tumor characteristics. Beyond accuracy, computational efficiency is another major bottleneck—many approaches either train from scratch on small datasets, failing to leverage large-scale medical imaging data, or rely on full-model fine-tuning [17], incurring high computational costs and over-fitting risks. Therefore, a scalable and adaptive model is urgently needed to achieve robust and efficient pan-tumor segmentation, addressing both tumor heterogeneity and dataset imbalance while maintaining computational efficiency.

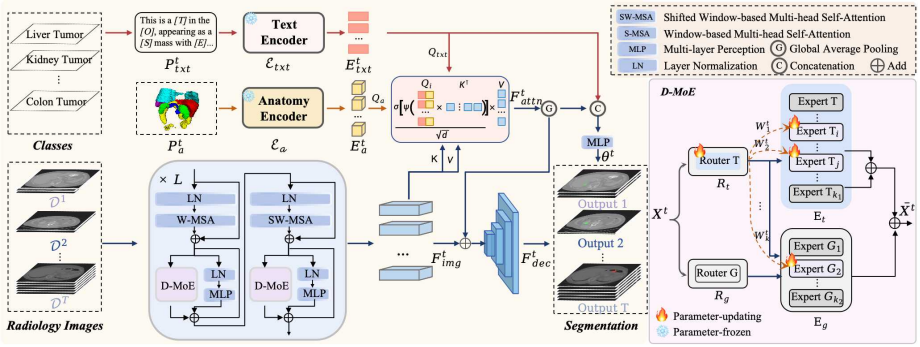

Fig. 1. An overview of the proposed MAST-Pro model for pan-tumor segmentation, with the text and anatomy prompts served as specific priors to enhance tumor representation learning. Multi-anatomical radiology images are processed through D-MoE-enhanced Swin UNETR, where task-dependent routers dynamically select experts to balance generic and tumor-specific feature learning.

To overcome the aforementioned limitations, we propose MAST-Pro (**M**ixture-of-experts for **A**daptive **S**egmentation of pan-**T**umors with knowledge-driven **Pro**mpts), a novel framework that integrates Dynamic Mixture-of-Experts (D-MoE) and knowledge-driven prompts for robust pan-tumor segmentation across diverse anatomical sites. Specifically, to enhance cross-tumor generalization, text and anatomical embeddings derived from higher-order medical knowledge are incorporated as domain-specific priors, guiding the segmentation process. To simultaneously capture generic tumor characteristics and preserve tumor-specific variations, we introduce a dynamic expert selection mechanism, which adaptively allocates computational resources to improve segmentation performance across heterogeneous datasets. Furthermore, we employ Parameter-Efficient Fine-Tuning (PEFT) for multi-anatomical tumor segmentation, significantly reducing computational overhead while enabling efficient adaptation to new tumor types and anatomical regions. Extensive experiments on assembly of eight public datasets demonstrate that MAST-Pro achieves performance comparable to State-of-The-Art (SOTA) methods.

2 Method

In this paper, we propose a novel universal model, called MAST-Pro, for pan-tumor segmentation. As illustrated in Fig. 1, our approach leverages text (P^t_{txt}) and anatomical (P^t_{a}) prompts as domain-specific priors, incorporating structured medical knowledge to improve segmentation performance (Sect. 2.1). Multi-anatomic radiology images are processed through D-MoE-enhanced Swin UNETR, where task-dependent routers dynamically select a mixture of generic and tumor-specific experts to optimize feature learning (Sect. 2.2). The extracted

prompts are fused with image features, contributing to both segmentation refinement and mask proposal generation. Notably, to improve training efficiency, we first pretrain a backbone on large-scale medical datasets, followed by fine-tuning using PEFT strategy, *i.e.*, up-dating only a small subset of experts within D-MoE rather than the entire model (Sect. 2.3).

2.1 Prompt Embedding

Text Prompt Embedding Each tumor type exhibits distinct characteristics, necessitating a structured text representation enriched with medical domain knowledge. To achieve this, we leverage a large language model (LLM) to generate concise yet informative text descriptions for each tumor type, following a standardized template: P_{txt} = "This is a [C] in the [O], appearing as a [S] mass with [E] borders on [M].", where [C], [O], [S], [E], and [M] represent placeholders for tumor-specific attributes, including tumor type, anatomical location, shape descriptor, edge characteristics, and imaging modality, respectively. These text prompts $\{P_{txt}^t\}$ are then processed using a pre-trained text encoder [6], denoted as $\mathcal{E}_{txt}$, to extract meaningful feature representations E_{txt}^t:

$$E_{txt}^t = \mathcal{E}_{txt}(P_{txt}^t). \tag{1}$$

Anatomical Prompt Embedding Incorporating anatomical priors enhances the model's ability to recognize tumor characteristics across different anatomy by providing structured spatial information. Given the strong segmentation performance of existing foundation models in organ segmentation, we leverage organ masks generated by TotalSegmentator [17] as anatomical prompts. The anatomical prompt embedding E_a is obtained by encoding $\{P_a^t\}$ using a pre-trained anatomical encoder [19] denoted as $\mathcal{E}_a$:

$$E_a^t = \mathcal{E}_a(P_a^t). \tag{2}$$

2.2 Pan-tumor Adaptive Mixture-of-Experts

Integrated into the Swin UNETR block, D-MoE enhances pan-tumor segmentation by dynamically selecting experts to balance generic and tumor-specific features. To be specific, The general router R^g captures common features by selecting experts only from the generic pool E^g and is used for datasets with mixed tumor types (*e.g.*, AbdomenCT-1K). In contrast, the task-specific router R^t is used for tumor-specific datasets, and it selects top-k experts from both E^t and E^g, enabling a flexible combination of specific and general knowledge.

Given a segmentation task $\mathcal{T}^t$, the router adaptively selects the top-k experts based on extracted feature representations:

$$\bar{X}^t = \sum_{i=1}^{k} R_i^t(X^t) \cdot \left(E_i^t(X^t); E_i^g(X^t)\right), \tag{3}$$

where X^t represents the input feature representation before expert adaptation, and $\bar{X}^t$ is the refined feature representation for task t after processing through D-MoE. The selection weight $R_i^t(X^t)$ determines the contribution of each expert. $E_i^t(X^t)$ and $E_i^g(X^t)$ denote the tumor-specific and generic experts, respectively.

The expert selection is determined through a gated mechanism, where the router assigns selection weights via:

$$R(X) = \text{Softmax}(\text{KeepTop-}k(X^\top W, k)), \tag{4}$$

where W is a learnable projection matrix, and KeepTop-k retains the highest-weighted expert activations:

$$\text{KeepTop-}k(v,k)_i = \begin{cases} v_i, & \text{if } v_i \text{ is in top } k \text{ elements of } v, \\ -\infty, & \text{otherwise.} \end{cases} \tag{5}$$

In our implementation, we set $k_1 = k_2 = 4$ for R^t and R^g, respectively.

To integrate domain-specific priors into visual representations, the extracted prompts act as queries $\mathcal{Q}_j$, while the image features serve as keys $\mathcal{K}$ and values $\mathcal{V}$ within the cross attention mechanism, formulated as:

$$\mathcal{F}_{attn} = \text{Softmax}\left(\frac{[\mathcal{Q}_{txt}; \mathcal{Q}_a]\mathcal{K}^\top}{\sqrt{d}}\right)\mathcal{V}, \tag{6}$$

where d is a scaling factor for stability. The attention-refined features ($\mathcal{F}_{attn}$) are fused with image features ($\mathcal{F}_{img}$) and decoded by Swin UNETR:

$$\mathcal{F}_{dec} = \mathcal{D}(\mathcal{F}_{attn} + \mathcal{F}_{img}). \tag{7}$$

To further enhance contextual understanding, $\mathcal{F}_{attn}$ is first subjected to global average pooling (GAP) and then concatenated with text-based prompts. The combined feature vector is passed through a multi-layer perceptron (MLP) to generate an initial mask proposal θ^t [9], where θ^t comprises weights ($W \in \mathbb{R}^{T\times c\times d\times h\times w}$) and biases ($b \in \mathbb{R}^T$), with T denoting the number of tumor types and c the number of latent channels. This initial proposal is supervised using cross-entropy loss with tumor category labels to ensure category-aware representation learning. Subsequently, the learned θ^t is used to generate category-specific mask proposals for each tumor type, which guide the final segmentation output, while a Dice loss is applied to further refine prediction accuracy and enforce boundary alignment.

$$\mathcal{F}_{final} = \mathcal{F}_{dec}\theta^t, \quad \text{where} \quad \theta^t = \text{MLP}(\text{GAP}(\mathcal{F}_{attn}); \mathcal{E}_{txt}). \tag{8}$$

2.3 Parameter-Efficient Fine-Tuning (PEFT) Strategy

In D-MoE, task-dependent routers selectively fine-tune low-rank experts to preserve generic tumor characteristics while refining tumor-specific features. Specifically, we employ k_2 experts to encode generic tumor features and k_1 experts to capture domain-specific variations. The routers dynamically select the top k experts from both groups, enabling adaptive feature learning.

3 Experiment

3.1 Dataset

Following prior work [9], we pretrained our model's backbone on a diverse collection of large-scale medical imaging datasets, including BTCV [20], CT-ORG [21], Pancreas-CT [22], CHAOS [23], 3D-IRCADb [24], WORD [26], and AMOS [27], along with tumor-specific datasets such as AbdomenCT-1K [25], LiTS [28], KiTS [29], and CT images from the MSD dataset [30]. Importantly, only the training partitions of these datasets were used during pretraining to prevent data leakage. Building on this foundation, we curated over 2,000 tumor cases from eight tumor-specific datasets to train and evaluate a pan-tumor segmentation model that generalizes across datasets without cohort-specific fine-tuning. The combined cohort was randomly partitioned into training and testing subsets using an 8:2 ratio.

To ensure consistency, all CT scans were reoriented, resampled to $1.5 \times 1.5 \times 1.5\ mm^3$ isotropic spacing, and cropped to focus on tumor-relevant regions. During training, we extracted $96 \times 96 \times 96$ voxel patches, ensuring balanced tumor and background sampling. Data augmentation included random 90-degree rotations and intensity shifting to enhance robustness and generalization.

3.2 Implementation Details

All experiments were conducted using PyTorch on four NVIDIA Tesla H100 GPUs (80GB RAM). The backbone was pretrained for 1,000 epochs, followed by PEFT applied to D-MoE. Both pretraining and fine-tuning were performed with identical hyperparameters, utilizing the AdamW optimizer with a base learning rate of 5×10^{-5} and a batch size of 4. Multi-GPU training was performed using Distributed Data Parallel (DDP) to ensure efficient scalability.

3.3 Comparison with State-of-The-Art Methods

To evaluate the effectiveness of our method, we conducted comparative experiments against SOTA universal medical image segmentation approaches, which are categorized into baseline methods, vision-prompt-based methods, and automatic text-prompt-based methods. Specifically, we selected nnU-Net (3D full resolution) [32] and Swin UNETR [19] as baselines, Med-SAM3D [7], MA-SAM [31], and SegVol [8] as vision-prompt-based methods, and the Universal Model [9] and ZePT [10] as automatic text-prompt-based models. All baselines followed their original training protocols for fair comparison.

Quantitative Comparison Table 1 presents the segmentation performance across eight datasets, demonstrating that MAST-Pro achieves the highest mean DSC of 68.71%, and outperforms both vision-prompt-based (*e.g.*, Med-SAM3D, SegVol) and text-prompt-based (*e.g.*, ZePT, Universal Model) approaches. Compared to the strongest baseline (ZePT), our model improves the average DSC by 5.2%, highlighting the effectiveness of D-MoE and knowledge-driven prompts

Table 1. DSC (%) results for multimodal segmentation methods across all datasets. The best results are bolded. Abbreviations: "M-Li" – "MSD-Liver", "M-Lu" – "MSD-Lung", "M-Pa" – "MSD-Pancreas", "M-HT" – "MSD-HepaticVessel Tumor", "M-Co" – "MSD-Colon", and "Abd" – "AbdomenCT-1K".

Method	M-Li	M-Lu	M-Pa	M-HT	M-Co	LiTS	KiTS	Abd	Mean
nnU-Net [32]	60.22	68.54	52.75	69.50	45.07	57.15	65.18	62.85	60.16
Swin UNETR [19]	63.24	66.70	53.24	66.23	42.55	66.79	65.23	64.82	61.10
Med-SAM3D [7]	47.81	24.28	40.26	57.89	48.21	22.32	67.15	-	43.98
MA-SAM [31]	69.16	51.70	31.22	63.57	39.98	57.22	**75.91**	-	55.53
SegVol [8]	69.07	65.53	54.35	68.75	**48.23**	62.18	57.74	-	60.84
Universal Model [9]	65.92	67.11	54.72	66.31	42.82	76.07	62.86	66.53	62.79
ZePT [10]	69.58	69.07	53.39	70.65	43.18	79.66	57.83	64.76	63.51
MAST-Pro (Ours)	**72.96**	**72.10**	**59.34**	**74.76**	46.79	**82.12**	72.99	**68.65**	**68.71**

in enhancing generalization. Our model achieves top performance on six out of eight datasets, with notable improvements in M-Pa (+4.26%), M-HT (+4.11%), and LiTS (+2.46%), demonstrating its robustness across diverse tumor types. Particularly in Liver tumor segmentation (LiTS), MAST-Pro surpasses Med-SAM3D by 59.8% and SegVol by 19.94%, showcasing its ability to capture tumor-specific features autonomously. Furthermore, the M-Lu dataset, one of the smallest cohorts in terms of sample size, serves as a representative "rare tumor" case. On this dataset, MAST-Pro achieves a DSC of 72.10%, outperforming previous SOTA methods by a large margin (+19.3 vs. Med-SAM3D, +20.6 vs. MA-SAM, +5.2 vs. ZePT), thereby demonstrating its robustness in data-scarce scenarios.

Qualitative Comparison Fig. 2 presents qualitative segmentation results, demonstrating the superiority of our method in capturing tumor boundaries and preserving structural details. To be specific, our method demonstrates superior segmentation accuracy, capturing finer tumor details with fewer false positives and better boundary adherence compared to the competing methods. Particularly in small or complex tumors (first and second rows), our results closely matches the ground truth, whereas other models either under-segment (ZePT, Universal Model) or over-segment (MA-SAM, SegVol) tumors, leading to inaccuracy. Additionally, in large and irregular tumors (third and fourth rows), our model produces more accurate contours and reduces misclassification errors, highlighting its robustness in handling diverse tumor morphology.

3.4 Ablation Study

Table 2 shows the ablation study, evaluating the contribution of each component.

Effect of Knowledge-Driven Prompts Removing all prompts leads to a significant drop in performance (61.10% mean DSC), underscoring their importance. Anatomical prompts alone improves segmentation (+2.27% mean DSC), particularly in M-Li and M-Pa, by incorporating structural priors. Text prompts

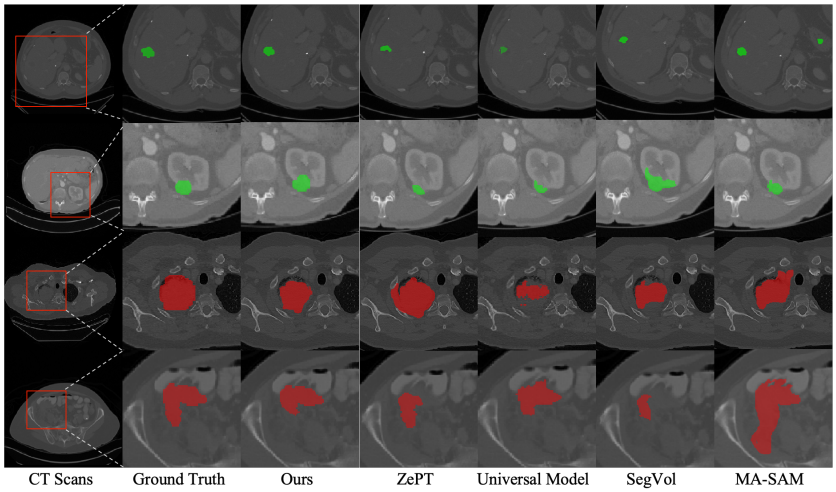

Fig. 2. Qualitative comparisons of the proposed MAST-Pro model (Ours) with other prompt-driven methods for multi-tumor segmentation. The first column shows the original CT scans, and the second column presents the ground-truth segmentations. The segmentation results in rows one to four are liver tumors, kidney tumors, lung tumors, and colon tumors, respectively.

Table 2. Ablation study on the effectiveness of anatomical prompts (AP), text prompts (TP), and the D-MoE module. The results are reported in DSC (%).

AP	TP	D-MoE	M-Li	M-Lu	M-Pa	M-HT	M-Co	LiTS	KiTS	Abd	Mean
×	×	×	63.24	66.70	53.24	66.23	42.55	66.79	65.23	64.82	61.10
✓	×	×	63.45	70.46	56.82	68.92	44.82	72.72	68.44	62.16	63.47
×	✓	×	66.48	68.24	53.42	68.02	42.26	77.03	64.86	67.08	63.42
×	×	✓	67.45	69.21	56.20	66.24	44.25	72.16	63.42	65.13	63.01
✓	✓	✓	**72.96**	**72.10**	**59.34**	**74.76**	**46.79**	**82.12**	**72.99**	**68.65**	**68.71**

alone enhances performance in LiTS but struggles with fine-grained boundaries, highlighting its limitations in handling tumor variability.

Effect of D-MoE Introducing D-MoE alone improves the mean DSC to 63.01%, particularly benefiting LiTS (+5.37%) and M-Li (+4.21%), demonstrating its ability to balance generic and tumor-specific features. However, without domain priors, its effectiveness is constrained in highly variable datasets.

Computational Efficiency To assess computational efficiency, we compare trainable parameters and GPU memory usage during training. As shown in Table 3, our model requires only 21.04M parameters and the lowest GPU memory, demonstrating the effectiveness of PEFT in reducing computational overhead while maintaining high segmentation accuracy.

Table 3. Comparison of computational cost during training between our and other methods in terms of training parameters and GPU memory usage.

Method	Train Params (M) ↓	Memory Usage (MB) ↓
MA-SAM	363.68	74544.02
SegVol	449.08	19898.00
Universal Model	244.80	9710.55
ZePT	495.10	24705.40
MAST-Pro (Ours)	21.04	8961.30

4 Conclusion

In this paper, we propose MAST-Pro, a novel framework that integrates Dynamic Mixture-of-Experts (D-MoE) and knowledge-driven prompts for pan-tumor segmentation. Text and anatomical prompts provide domain-specific priors, while D-MoE dynamically balances generic and tumor-specific feature learning, improving segmentation across diverse tumor types. Additionally, Parameter-Efficient Fine-Tuning (PEFT) reduces computational overhead without compromising accuracy. Experiments on multi-anatomical tumor datasets show MAST-Pro outperforms other SOTA methods by 5.20% DSC while reducing trainable parameters by 91.04%, demonstrating its effectiveness in accurate, generalizable, and efficient tumor segmentation.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China (Grants U23A20295, 82441023, 62131015, and 82394432), the Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001), and the HPC Platform of ShanghaiTech University.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bray, F., et al. "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: A Cancer Journal for Clinicians, vol. 74, no. 3, 2024, pp. 229–263.
2. Meng, R., et al. "NaMa: Neighbor-aware multi-modal adaptive learning for prostate tumor segmentation on anisotropic MR images." Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 5, 2024.
3. Wang, H., et al. "Dual-reference source-free active domain adaptation for nasopharyngeal carcinoma tumor segmentation across multiple hospitals." IEEE Transactions on Medical Imaging, 2024.

4. Liu, H., et al. "Multimodal brain tumor segmentation boosted by monomodal normal brain images." IEEE Transactions on Image Processing, vol. 33, 2024, pp. 1199–1210.
5. Kirillov, A., et al. "Segment anything." arXiv preprint arXiv:2304.02643, 2023.
6. Radford, A., et al. "Learning transferable visual models from natural language supervision." arXiv preprint arXiv:2103.00020, 2021.
7. Wang, H., et al. "SAM-Med3D." arXiv preprint arXiv:2310.15161, 2023.
8. Du, Y., et al. "SegVol: Universal and interactive volumetric medical image segmentation." Advances in Neural Information Processing Systems, vol. 37, 2025, pp. 110746–110783.
9. Liu, J., et al. "CLIP-driven universal model for organ segmentation and tumor detection." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 21152–21164.
10. Jiang, Y., et al. "ZePT: Zero-shot pan-tumor segmentation via query-disentangling and self-prompting." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 11386–11397.
11. Teng, L., et al. "Knowledge-Guided Prompt Learning for Lifespan Brain MR Image Segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention, 2024.
12. Yujin, O., et al. "Distribution-aware Fairness Learning in Medical Image Segmentation From A Control-Theoretic Perspective." arXiv preprint arXiv:2502.00619, 2025.
13. Runqi, M., et al. "A Neighbor-sensitive Multi-modal Flexible Learning Framework for Improved Prostate Tumor Segmentation in Anisotropic MR Images." IEEE Transactions on Biomedical Engineering, 2025.
14. Chen, J., et al. "CancerUniT: Towards a single unified model for effective detection, segmentation, and diagnosis of eight major cancers using a large collection of CT scans." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 21270–21281. DOI: 10.1109/ICCV51070.2023.01950.
15. Huang, Z., et al. "CAT: Coordinating anatomical-text prompts for multi-organ and tumor segmentation." Advances in Neural Information Processing Systems, vol. 37, 2025, pp. 3588–3610.
16. Shah, A., and Rojas, C. A. "Imaging modalities (MRI, CT, PET/CT), indications, differential diagnosis and imaging characteristics of cystic mediastinal masses: A review." Mediastinum, 2022.
17. Wasserthal, J., et al. "TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images." Radiology: Artificial Intelligence, 2023. DOI: 10.1148/ryai.230024.
18. Ziheng, Z., et al. "One model to rule them all: Towards universal segmentation for medical images with text prompts." arXiv preprint arXiv:2312.17183, 2023.
19. Tang, Y., et al. "Self-supervised pre-training of Swin Transformers for 3D medical image analysis." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 20730–20740.
20. Landman, B., et al. "MICCAI multi-atlas labeling beyond the cranial vault–workshop and challenge." Proceedings of MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge, vol. 5, 2015, pp. 12.
21. Rister, B., et al. "CT-ORG, a new dataset for multiple organ segmentation in computed tomography." Scientific Data, vol. 7, no. 1, 2020, pp. 381.
22. Roth, H. R., et al. "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation." Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 556–564.

23. Kavur, A. E., et al. "CHAOS challenge: Combined (CT-MR) healthy abdominal organ segmentation." Medical Image Analysis, vol. 69, 2021, pp. 101950.
24. Soler, L., et al. "3D image reconstruction for comparison of algorithm database: A patient-specific anatomical and medical image database." IRCAD, Strasbourg, France, Tech. Rep., 2010.
25. Ma, J., et al. "AbdomenCT-1K: Is abdominal organ segmentation a solved problem?" IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.
26. Luo, X., et al. "WORD: A large-scale dataset, benchmark, and clinically applicable study for abdominal organ segmentation from CT images." Medical Image Analysis, vol. 82, 2022, pp. 102642.
27. Ji, Y., et al. "AMOS: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation." Advances in Neural Information Processing Systems, vol. 35, 2022, pp. 36722–36732.
28. Bilic, P., et al. "The liver tumor segmentation benchmark (LiTS)." Medical Image Analysis, vol. 84, 2023, pp. 102680.
29. Heller, N., et al. "An international challenge to use artificial intelligence to define the State-of-The-Art in kidney and kidney tumor segmentation in CT imaging." Medical Image Analysis, 2020.
30. Antonelli, M., et al. "The medical segmentation decathlon." Nature Communications, vol. 13, no. 1, 2022, pp. 4128.
31. Chen, C., et al. "MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation." Medical Image Analysis, vol. 98, 2024, pp. 103310.
32. Isensee, F., et al. "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation." Nature Methods, vol. 18, no. 2, 2021, pp. 203–211.