KS-SAM: Kolmogorov-Arnold-driven Shape Alignment for Single-Domain Generalization in 3D Medical Image Segmentation

Rengmin Wu[1], Jiaqiang Li[1], Junqin Luo[1], Yuxing Wei[1], Liang Dan[2], Zhou Hui[3], Guangming Li[3], Xinhua Wei[2], Tianfu Wang[1], and Baiying Lei[1]✉ ⋆

[1] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University Medical School, Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, 518055, China leiby@szu.edu.cn

[2] Department of Radiology, Guangzhou First People's Hospital, Guangzhou, Guangdong,China

[3] Department of Radiology, Qingyuan People's Hospital, Qingyuan, Guangdong,China

Abstract. Single-Domain Generalization (SDG) in 3D medical image segmentation is inherently unstable: naive fine-tuning entangles domain-specific textures with anatomical semantics, leading to overfitting and catastrophic forgetting. The domain shifts caused by variations in imaging modalities and protocols make robust adaptation even more challenging. We argue that effective cross-domain adaptation requires constraining the model to rely on domain-invariant 3D geometry, rather than domain-specific textures.To address these challenges, we present Kolmogorov-Arnold Shape-aware Segment Anything Model (KS-SAM), a dual-engine framework that enforces geometry-driven adaptation for the Segment Anything Model (SAM). A domain-expert 3D network performs Uncertainty-Aware Feature Fusion (UAFF), adding noise to the input CT image and performing K forward passes to refine the SDF priors. These geometry-aware priors are injected into SAM via the Kolmogorov-Arnold-driven Prior-Semantic Synergistic Distillation (KPSSD) module, enabling bidirectional geometric-semantic anchoring without leaking appearance statistics. In parallel, Kolmogorov-Arnold-driven Low-Rank Adaptation (KAN-LoRA) introduces spline-based localized activations to model nonlinear modality shifts while mitigating catastrophic forgetting. Across multi-center and public benchmarks, KS-SAM significantly outperforms state-of-the-art methods on unseen domains, demonstrating that geometry-anchored adaptation is critical for reliable 3D medical SDG.

Keywords: 3D Medical Image Segmentation · Single-Domain Generalization · Segment Anything Model · Kolmogorov-Arnold Networks

⋆ Corresponding author.

1 Introduction

3D medical image segmentation is a critical task in medical imaging [1, 2], with broad applications in tumor detection, organ segmentation, and lesion identification. However, in the setting of Single-Domain Generalization (SDG), models are typically trained on data from a single source domain and then tested on a target domain. This task is especially challenging in 3D medical image segmentation due to significant domain shifts caused by differences in imaging modalities, scanner vendors, and institutional practices [3, 4]. These discrepancies are particularly pronounced in 3D volumetric data, where geometric and radiometric variations further exacerbate the problem, leading to poor model performance when deployed on new target domains.

While many methods attempt to address the SDG problem through transfer learning or data augmentation, these approaches are often limited in their ability to handle domain-specific variations. For instance, the Segment Anything Model (SAM) has shown remarkable zero-shot generalization in natural image segmentation tasks, inspiring its application in medical image segmentation [5–7]. However, despite SAM's success in natural images, it struggles in 3D medical imaging, primarily because SAM relies heavily on domain-specific texture and intensity statistics, leading to source domain overfitting and poor adaptation to modality shifts. When transferring the model from one source domain to a different target domain, SAM fails to effectively handle the domain shift, resulting in significant performance degradation.

Existing adaptation strategies, such as Linear Parameter-Efficient Fine-Tuning (LoRA) and Knowledge Distillation (KD), alleviate some overfitting issues but face limitations. LoRA assumes task-specific knowledge lies in a low-rank subspace, which fails to capture the nonlinear shifts in 3D medical imaging. KD, on the other hand, struggles with injecting anatomical priors without collapsing the model's open-world semantics, risking texture-statistics entanglement and propagating domain bias into global representations[8].

To address these issues, we argue that robust SDG requires explicitly anchoring adaptation to domain-invariant 3D anatomical geometry, rather than relying on domain-specific textures. Motivated by this insight, we propose Kolmogorov-Arnold Shape-aware Segment Anything Model (KS-SAM), a dual-engine interactive framework tailored for 3D medical SDG. Through the Uncertainty-Aware Feature Fusion (UAFF) mechanism, KS-SAM extracts consensus-aware Signed Distance Fields (SDF), ensuring structural consistency across domains. Meanwhile, the Kolmogorov-Arnold-driven Prior-Semantic Synergistic Distillation (KPSSD) module performs bidirectional geometric-semantic alignment and nonlinear distillation, effectively injecting structural priors into SAM while blocking texture leakage [9]. Additionally, the Kolmogorov-Arnold-driven Low-Rank Adaptation (KAN-LoRA) module employs localized spline activations instead of traditional linear projections to model nonlinear modality shifts and effectively mitigate catastrophic forgetting [10, 11].

In summary, our main contributions are threefold:

– We propose KS-SAM, a geometry-anchored dual-engine framework that mitigates source-domain overfitting in 3D medical SDG by explicitly disentangling structural priors from appearance features.
– We introduce an uncertainty-filtered volumetric UAFF mechanism to extract consensus-aware 3D Signed Distance Field (SDF) priors, and develop the KPSSD module to enable bidirectional geometric-semantic distillation without propagating texture biases.
– We incorporate KAN-LoRA for nonlinear out-of-distribution modulation, replacing linear low-rank adaptation with localized spline-based activations to preserve foundation model generalization under complex volumetric modality shifts.

2 Methodology

2.1 Overall Architecture

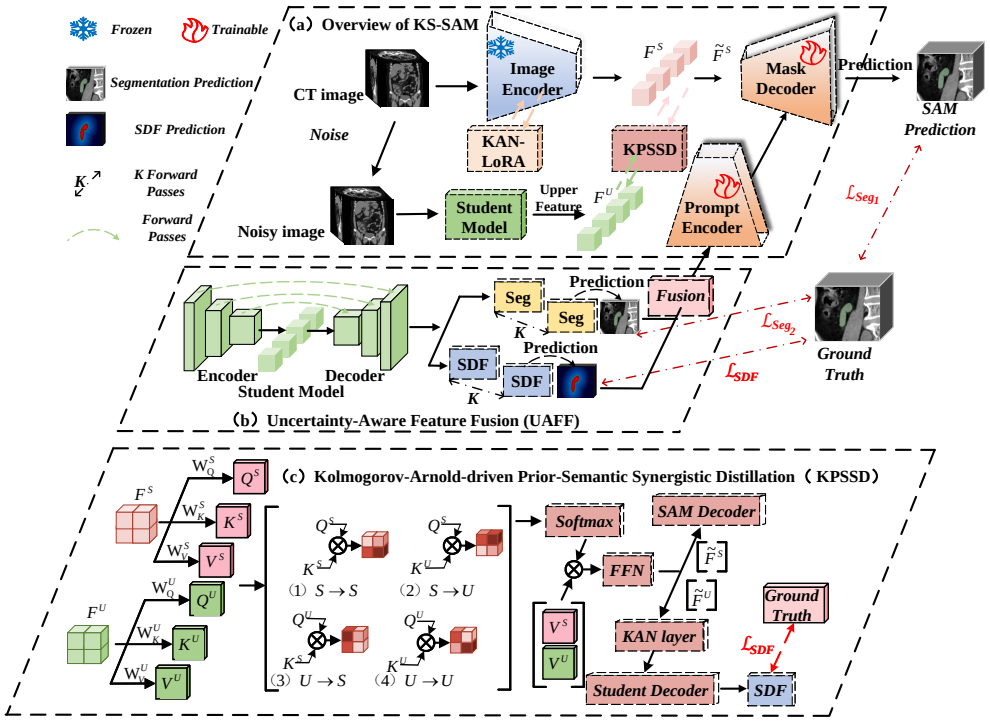

Fig. 1. (a) Overview of the proposed KS-SAM framework; (b) extraction of domain-invariant 3D SDF priors via UAFF; (c) the inner workings of KPSSD featuring four-way synergistic attention and KAN-based non-linear distillation.

As shown in Fig. 1, the KS-SAM framework processes the input CT image through two parallel streams. The CT image is fed into the SAM decoder with

KAN-LoRA for nonlinear feature modulation, producing one set of features. Simultaneously, the noisy CT image is passed through the student network to generate another set of features. These features are fused for SDF distillation via the KPSSD module. Additionally, the student network performs UAFF to refine the SDF priors through multiple forward passes and uncertainty fusion. The merged outputs are then passed to the Prompt Encoder, and the final distilled features are input into the SAM mask decoder to generate the segmentation mask.

2.2 Domain-Invariant 3D Prior via the UAFF Module

To decouple anatomical structures from domain-sensitive textures, we employ a 3D UNet with volumetric UAFF. We implement a progressive augmentation schedule where the number of stochastic passes K and noise intensity σ (Gaussian $\epsilon \sim \mathcal{N}(0, \sigma^2)$) scale dynamically with training epochs E. This curriculum strategy ensures initial convergence on stable structures before addressing complex domain shifts.

To distill consensus from K inferences, an **Uncertainty-Filtered Fusion** module is introduced. For segmentation maps $\hat{Y}_k$ and 3D SDFs $\hat{Z}_k$, we compute voxel-wise categorical entropy U_{seg} and epistemic variance U_{sdf}:

$$U_{seg} = -\sum_c \bar{Y}_c \log \bar{Y}_c, \quad U_{sdf} = \frac{1}{K}\sum_{k=1}^{K}(\hat{Z}_k - \bar{Z})^2 \tag{1}$$

where $\bar{Y}$ and $\bar{Z}$ are ensemble means. These maps are transformed into adaptive confidence weights to generate the filtered 3D shape prior F_{UNet}, prioritizing structural consensus over domain-specific artifacts.

2.3 Pseudo-Inverse Distillation via the KPSSD Module

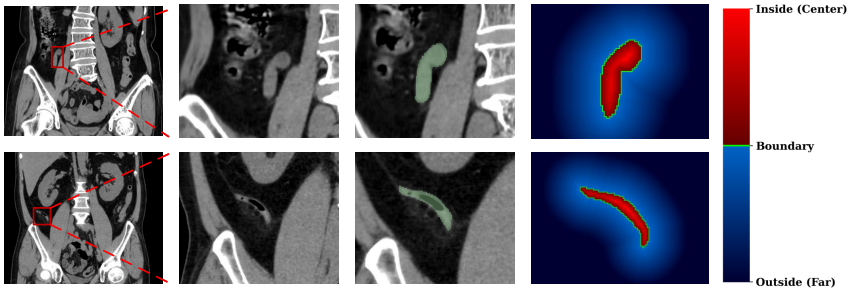

Fig. 2. Visualization of the shape-aware SDF generated by the KPSSD module.

The proposed KPSSD module, detailed in Fig.1 (c), serves as the core mechanism for geometric knowledge transfer.

Prior-Semantic Bi-Attention (PSBA). To synergize anatomical priors F^S and volumetric semantic features F^U, we design the PSBA module (Fig. 1 (c)). Unlike standard self-attention, PSBA employs a quadruple scheme to model both intra-domain spatial context and inter-domain correlations. Let d_s be the scaling factor; the bi-attention operation is formulated via a unified matrix transformation:

$$\begin{bmatrix} \widetilde{F}^S \\ \widetilde{F}^U \end{bmatrix} = \mathrm{FFN}\left(\sigma\left(\frac{1}{\sqrt{d_s}}\begin{bmatrix} Q^S \\ Q^U \end{bmatrix}\begin{bmatrix} K^S & K^U \end{bmatrix}^\top\right)\begin{bmatrix} V^S \\ V^U \end{bmatrix}\right) \tag{2}$$

where $\sigma(\cdot)$ denotes Softmax. This framework facilitates bidirectional knowledge flow: $S \to U$ injects anatomical constraints into semantic features, while $U \to S$ refines prior localization using high-level context.

Non-linear Shape Distillation. The enhanced features $[\widetilde{F}^S; \widetilde{F}^U]$ are processed by a **KAN layer** for non-linear SDF reconstruction. By mapping features into a continuous distance field Φ_{pred}, the model captures fine-grained boundary details often lost in mask-based supervision. The distillation is governed by:

$$\mathcal{L}_{SDF} = ||\Phi_{pred} - \Phi_{GT}||^2 \tag{3}$$

where Φ_{GT} is the ground-truth SDF. This ensures high-frequency geometric fidelity even in low-contrast regions.

2.4 Out-of-Distribution Modulation via KAN-LoRA

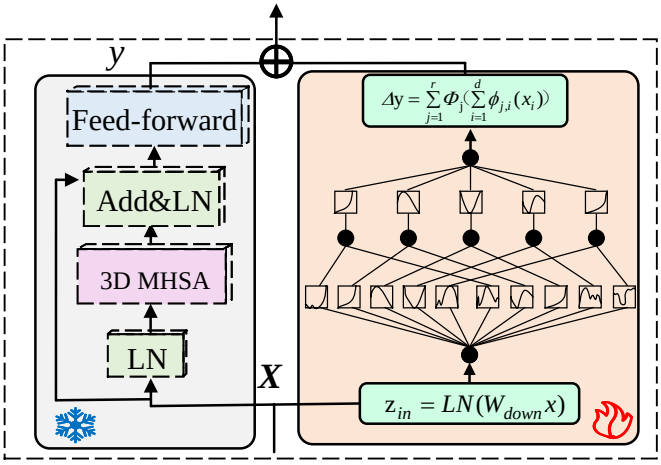

Fig. 3. Illustration of KAN-LoRA

To bridge the modality gap between natural images and 3D medical volumes, we introduce **KAN-LoRA** (Fig. 3). Unlike standard LoRA's linear projections,

KAN-LoRA utilizes a parallel bottleneck with a KAN[11] branch to capture nonlinear volumetric features while keeping the 3D Transformer blocks frozen. For an input feature X, the adaptation is formulated as:

$$y_{out} = f_{frozen}(X) + \sum_{j=1}^{r} \Phi_j \left(\sum_{i=1}^{d} \phi_{j,i}(LN(W_{down}X)) \right) \tag{4}$$

where $\phi_{j,i}$ are learnable B-spline activation functions. By leveraging the local support property of splines, KAN-LoRA models complex intensity distributions in CT/MRI scans while mathematically mitigating catastrophic forgetting, thereby preserving the foundation model's open-world generalizability.

3 Experiments and Results

3.1 Datasets and Evaluation Metrics

The dataset used for evaluating SDG consists of two parts: the Multi-center Appendix Dataset and the Public Kidney Benchmark. The Multi-center Appendix Dataset includes data from four centers, with Guangzhou First People's Hospital (Source A) having 976 cases, Qingyuan People's Hospital (OOD B) having 152 cases, Huizhou Central People's Hospital (OOD C) with 27 cases, and Zhongda Hospital Southeast University (OOD D) with 84 cases, totaling 1,239 cases. The Public Kidney Benchmark includes three datasets: WORD[12, 13] (Source E) with 100 cases, BTCV[14] (OOD F) with 30 cases, and KiTS23 [1] (OOD G) with 489 cases, totaling 619 cases. Domains B–D and F–G in these datasets serve as unseen out-of-domain (OOD) targets for SDG evaluation

We employ two standard volumetric metrics: the Dice Similarity Coefficient (DSC, %) to measure spatial overlap, and the 95th percentile Hausdorff Distance (HD95, mm) to assess 3D boundary quality and topological accuracy.

3.2 Implementation and Optimization

The framework is implemented in PyTorch, with all 3D CT volumes resampled to $1.0 \times 1.0 \times 1.0$ mm^3 and normalized within $[-150, 250]$ HU.

Optimization Objective. KS-SAM is trained end-to-end by minimizing a multi-task joint objective:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{prior} + \lambda_2 \mathcal{L}_{SDF} + \lambda_3 \mathcal{L}_{SAM} \tag{5}$$

where $\mathcal{L}_{prior}$ and $\mathcal{L}_{SAM}$ combine Dice and Cross-Entropy losses for the UNet-prior and SAM-decoder, respectively. $\mathcal{L}_{SDF}$ denotes the MSE loss supervising the geometric continuity of the Signed Distance Field. We empirically set $\lambda_1 = \lambda_2 = \lambda_3 = 1.0$.

Training Protocol. We employ the AdamW optimizer with a learning rate of 1×10^{-4} and weight decay of 1×10^{-5}. To strictly adhere to the SDG protocol, no target-domain exposure or weight updates are performed during inference.

Table 1. Quantitative comparison of segmentation performance on the multi-center appendix datasets. Best results are in **bold**.

Models	Dice (%) ↑				HD95 (mm) ↓			
	B	C	D	Avg	B	C	D	Avg
nnUNet[3]	59.72±3.09	66.26±6.27	57.76±3.64	61.25	31.56±0.85	13.91±5.23	16.75±9.12	20.75
nnFormer[15]	55.25±5.29	60.62±9.33	55.77±9.21	57.21	24.41±5.05	17.39±2.21	21.09±10.06	20.96
UNETR++[16]	58.34±8.38	64.34±10.03	54.00±10.06	58.89	18.76±7.39	8.42±6.08	28.45±8.33	18.54
VNET[5]	56.35±10.00	56.53±8.72	60.42±8.49	57.77	19.53±3.95	26.95±9.35	10.57±8.66	19.02
MedSAM3D[6]	61.52±15.65	62.38±9.97	61.58±8.14	61.83	32.04±10.06	21.47±4.31	19.73±10.01	24.41
Ours	**66.51±2.33**	**69.26±4.46**	**63.92±3.35**	**66.56**	**7.86±0.72**	**7.96±1.63**	**9.10±1.05**	**8.31**

Table 2. Quantitative comparison of segmentation performance on the multi-center Kidney datasets. The best results are highlighted in **bold**.

Models	Dice (%) ↑			HD95 (mm) ↓		
	F	G	Avg	F	G	Avg
nnUNet[3]	79.12±9.41	80.12±6.55	80.62	11.66±8.60	12.09±10.03	11.88
nnFormer[15]	79.48±8.24	82.44±9.92	82.96	12.11±8.39	17.39±8.94	14.75
UNETR++[16]	79.26±6.90	80.56±6.11	82.41	20.93±12.80	14.91±12.86	17.92
VNET[5]	77.72±6.91	77.45±6.78	80.59	29.02±15.07	24.58±6.53	26.80
MedSAM3D[6]	80.11±7.44	83.89±8.66	84.00	15.67±9.94	13.82±12.5	14.75
Ours	**87.70±7.12**	**89.88±2.93**	**89.29**	**8.17±10.34**	**8.40±9.35**	**8.29**

3.3 Comparison with State-of-the-Art Methods

KS-SAM is compared against five SOTA 3D segmentors: nnUNet, nnFormer, UNETR++, VNET, and MedSAM3D. As shown in Table 1, KS-SAM achieves a mean Dice of 66.56%, outperforming MedSAM3D by 4.73%. Notably, our method exhibits significantly lower variance (±2.33%) and reduces the average HD95 from 24.41mm to 8.31mm. This demonstrates improved boundary recovery under domain shifts On the WORD-to-Kidney task (Table 2), KS-SAM maintains a robust Dice of 89.29%. While baselines struggle with institutional intensity variations, our uncertainty-filtered SDF prior ensures structural consistency, mitigating over-reliance on domain-specific textures. Qualitative results (Fig. 4) further confirm superior topological correctness in low-contrast regions.

3.4 Ablation Study

We ablate key components on Domain F (Table 3) using MedSAM3D as the baseline. Replacing linear LoRA with KAN-LoRA boosts Dice to 81.56% (+1.45%), validating non-linear B-spline activations. The KPSSD module yields the largest gain (+3.55%), highlighting the value of SDF-guided constraints, while UAFF further stabilizes predictions. Our full model achieves a peak Dice of 87.70%, confirming that decoupling structural priors from appearance is crucial for robust SDG.

Table 3. Ablation study on Domain F (BTCV). We evaluate KAN-LoRA, KPSSD, and UAFF against the baseline (MedSAM3D + LoRA).

Models	LoRA	KAN-LoRA	KPSSD	UAFF	DSC (%)
MedSAM3D	✓				80.11 ± 7.44
		✓			81.56 ± 15.66
			✓		83.66 ± 10.23
				✓	81.11 ± 5.20
		✓	✓		86.66 ± 10.78
		✓		✓	85.11 ± 7.36
Ours		✓	✓	✓	87.70 ± 7.12

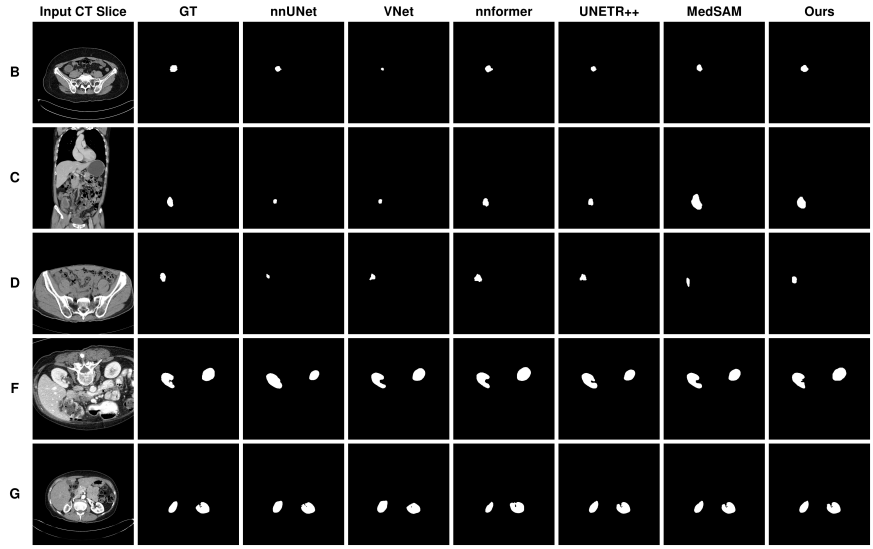

Fig. 4. Qualitative comparison of 3D segmentation results on unseen target domains.

4 Conclusion

We present KS-SAM, a geometry-anchored dual-engine framework for 3D medical SDG. By leveraging uncertainty-filtered UAFF to extract consensus-aware SDF priors and injecting them into SAM via KPSSD, our approach enforces geometry-driven adaptation while suppressing texture-induced bias. Meanwhile, KAN-LoRA provides localized nonlinear modulation to bridge modality shifts and mitigate catastrophic forgetting.Extensive evaluations across multiple unseen domains demonstrate that geometry-constrained foundation model adaptation significantly improves robustness and boundary fidelity, highlighting a promising direction for reliable deployment of foundation models in real-world 3D medical segmentation.

Acknowledgments. This work was supported partly by Shenzhen Medical Research Funds (Nos.C2301005 and D250403004), National Natural Science Foundation of China (Nos. U22A2024 and 62271328) and Shenzhen Science and Technology Program (Grant Nos. JCYJ20241202124202004, and JCYJ20240813142111015).

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

References

1. Nicholas Heller et al. The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct. *arXiv preprint arXiv:2307.01984*, 2023.
2. Zifeng violinist Qiu, Peng Yang, Chunlun Xiao, et al. 3d multimodal fusion network with disease-induced joint learning for early alzheimer's disease diagnosis. *IEEE Trans. Med. Imaging*, 43:3161–3175, 2024.
3. Fabian Isensee, Paul F Jaeger, Simon AA Kohl, et al. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods*, 18:203–211, 2021.
4. Jungkyu Park, Jakub Chłędowski, Stanisław Jastrzębski, et al. An efficient deep neural network to classify large 3d images with small objects. *IEEE Trans. Med. Imaging*, 43:351–365, 2023.
5. Arnick Abdollahi, Biswajeet Pradhan, and Abdullah Alamri. Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *IEEE Access*, 8:179424–179436, 2020.
6. Haoyu Wang, Sizheng Guo, Jin Ye, et al. Sam-med3d: a vision foundation model for general-purpose segmentation on volumetric medical images. *IEEE Trans. Neural Netw. Learn. Syst.*, 2025.
7. Jia Wei, Xiaoqi Zhao, Jonghye Woo, et al. Mixture-of-shape-experts (mose): End-to-end shape dictionary framework to prompt sam for generalizable medical segmentation. In *Proc. CVPR*, pages 6448–6458, 2025.
8. Zhengshan Wang, Long Chen, Xuelin Xie, et al. Teacher-student instance-level adversarial augmentation for single domain generalized medical image segmentation. *IEEE Trans. Med. Imaging*, 2025.
9. Rongchang Zhao, Zhangyue Wu, Jian Zhang, et al. Radkam: Attention-driven kolmogorov-arnold model for automatic radiation-induced lymphopenia prediction by multimodal learning. In *Proc. MICCAI*, pages 513–522, 2025.
10. Mohammad Marufur Rahman, Guanchu Wang, Kaixiong Zhou, et al. Catastrophic forgetting in kolmogorov-arnold networks. *arXiv preprint arXiv:2511.12828*, 2025.
11. Ziming Liu, Yixuan Wang, Sachin Vaidya, et al. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2025.
12. Wenjun Liao, Xiangde Luo, Yuan He, et al. Comprehensive evaluation of a deep learning model for automatic organs-at-risk segmentation on heterogeneous computed tomography images. *Int. J. Radiat. Oncol. Biol. Phys.*, 117:994–1006, 2023.
13. Xiangde Luo, Jianghong Liao, Wenjun to Xiao, et al. WORD: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from ct image. *Med. Image Anal.*, 82:102642, 2022.
14. Bennett Landman, Zhoubing Xu, J Igelsias, et al. Miccai multi-atlas labeling beyond the cranial vault. In *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault*, page 12, 2015.

15. Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, et al. nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE Trans. Image Process.*, 32:4036–4045, 2023.
16. Abdelrahman Shaker, Muhammad Maaz, Hanoona Rasheed, et al. Unetr++: delving into efficient and accurate 3d medical image segmentation. *IEEE Trans. Med. Imaging*, 43:3377–3390, 2024.