

RKANet: Reconstruction-Consistent KAN–Mamba Framework for Medical Image Segmentation

Wei Wang¹, Junxiao Lv², Xin Wang¹

¹School of Physics and Electronic Science, Changsha University of Science and Technology, Changsha, China

²School of Computer Science and Technology, Changsha University of Science and Technology, Changsha, China

Abstract—Medical image segmentation remains challenging due to complex anatomy, ambiguous boundaries, and elongated structures, where structural inconsistency is frequently observed during feature reconstruction. In many encoder–decoder architectures, local upsampling operations are insufficient to preserve geometric continuity and long-range dependencies, resulting in fragmented predictions and boundary artifacts. To address this issue, a reconstruction-consistent segmentation framework, termed RKANet, is presented, in which decoding is explicitly strengthened as a structured reconstruction process. A cooperative modeling module is designed to jointly enforce local geometric refinement and global dependency propagation. In addition, a Boundary-Guided Skip Branch is incorporated to enhance skip feature fusion with boundary-aware shallow features. Extensive experiments on five public datasets demonstrate that RKANet achieves consistent performance improvements over representative methods, yielding a 3.00% Dice gain on BUSI and stable improvements of 0.25% – 0.55% on other benchmarks, with an average Dice score of 88.22%.

Index Terms—Medical image segmentation, encoder–decoder networks, state space models, Kolmogorov–Arnold networks, boundary-aware decoding

I. INTRODUCTION

Medical image segmentation is a fundamental task in computer-aided diagnosis and image-guided clinical applications. Beyond pixel-wise classification, accurate segmentation requires the reconstruction of anatomically consistent structures, which remains challenging in the presence of elongated targets, blurred boundaries, and complex anatomical topologies. Despite notable progress in recent years, structural degradation is still frequently observed during decoding, leading to fragmented predictions and boundary distortions.

Encoder–decoder architectures represented by U-Net [1] have become the dominant paradigm for medical image segmentation. Numerous variants have been proposed to enhance feature representation through improved encoders and skip connections, such as UNet++ [2], UNet 3+ [3], and MultiResUNet [4]. However, the decoding stage in most existing approaches is still largely simplified as a sequence of local upsampling and convolution operations, as exemplified by SegNet [5], resulting in a strong-encoder–weak-decoder asymmetry. Consequently, high-level semantic information cannot always be reliably reconstructed during spatial resolution recovery, especially in complex anatomical scenarios.

To alleviate the limitations of local decoding, global context modeling has been introduced into encoder–decoder

frameworks. To address this issue, Transformer-based methods have been incorporated into encoder–decoder architectures to enhance global context modeling. Representative approaches, including TransUNet [6], DS-TransUNet [7], and Swin-UNet [8], have demonstrated improved capability in capturing long-range dependencies, while their quadratic computational complexity has been found to limit scalability at high-resolution decoding stages.

As an efficient alternative, structured state space models have been explored for long-range modeling with linear computational complexity. Mamba-based segmentation networks, such as U-Mamba [9], BEFNet [10], and Swin-UMamba [11], have been shown to effectively preserve global structural dependencies. In parallel, Kolmogorov–Arnold Networks (KAN) [12] have been explored for strong nonlinear approximation and geometric modeling capability, as evidenced by U-KAN [13] and KANSeg [14]. Despite these advances, existing methods have typically focused on either global dependency modeling or geometric representation in isolation, while effective strategies for their cooperative integration within encoder–decoder architectures have remained insufficiently explored. The source code is available at <https://github.com/xiao-170/RKANet>.

Motivated by these observations, the main contributions of this work are summarized as follows:

- A reconstruction-consistent segmentation network, RKANet, is proposed to mitigate structural degradation during decoding.
- A cooperative modeling module is introduced to jointly enforce local geometric consistency and global topological dependencies.
- A Boundary-Guided Skip Branch (BGSB) is designed to suppress background propagation in skip connections using boundary-aware features.
- Extensive experiments on five multimodal datasets are conducted to validate the effectiveness of the proposed framework.

II. RELATED WORK

A. Encoder–Decoder Segmentation Networks

Convolutional encoder–decoder architectures represented by U-Net [1] have been widely adopted for medical image segmentation due to their effectiveness in multi-scale feature extraction. Numerous variants, including UNet++ [2], UNet 3+ [3], KiU-Net [15], and MultiResUNet [4], have

been developed to enhance feature representation via improved encoders and skip connections. However, in most existing approaches, the decoding stage has been simplified as successive upsampling and local convolution operations, as observed in U-Net and SegNet [5], which may lead to the loss of structural details and boundary accuracy during spatial resolution recovery.

To alleviate the limitations of local decoding, global context modeling has been introduced into encoder-decoder frameworks. Transformer-based models, such as Vision Transformer [16] and Swin-UNet [8], have been employed to capture long-range dependencies, but the quadratic computational complexity of self-attention has limited their scalability, particularly for high-resolution medical images.

B. State Space Models and Kolmogorov–Arnold Networks

Structured state space models [17], particularly Mamba [18], have been explored as an efficient alternative to Transformers for long-range dependency modeling with linear computational complexity. Mamba-based segmentation networks, including U-Mamba [9], SegMamba [19], Semi-Mamba-UNet [23], and Swin-UMamba [11], have been shown to effectively capture global structural dependencies. In parallel, Kolmogorov–Arnold Networks (KAN) have been introduced for high-order nonlinear geometric modeling via learnable spline functions, and their effectiveness has been demonstrated in segmentation tasks by U-KAN [13] and KANSeg [14]. Nevertheless, global dependency modeling and geometric representation have typically been addressed in isolation, while effective strategies for their cooperative integration within encoder-decoder architectures remain insufficiently explored.

C. Boundary-Aware Skip Connections

Skip connections are widely adopted to recover spatial details during decoding by reintroducing high-resolution encoder features. However, they may also propagate background noise and boundary artifacts, particularly in low-contrast and complex scenarios. To alleviate this issue, attention- and boundary-aware mechanisms have been introduced to emphasize informative features or incorporate boundary-related cues, as exemplified by Attention U-Net [21], ET-Net [22], and BASNet [25].

More recent studies have explored structure-aware designs to enhance shallow feature utilization. For example, GBA-Net [26] employs multi-scale Gaussian boundary attention to emphasize structural information. Despite these advances, skip feature propagation is typically left unconstrained during decoding, and explicit regulation using structural constraints remains insufficiently explored.

III. METHODOLOGY

A. Overall Architecture

Structural inconsistency is frequently observed during decoding, as local upsampling operations fail to preserve geometric continuity and long-range dependencies. To address this issue, decoding is reformulated as a structured reconstruction

process to achieve reconstruction consistency, where local geometric refinement and global dependency modeling are jointly considered, while skip feature propagation is regulated to suppress background interference.

As shown in Fig. 1, RKANet adopts a U-Net-like encoder-decoder architecture reinterpreted as a reconstruction pipeline. A convolutional encoder extracts multi-scale local features, and T-KMamba blocks are integrated to enable co-operative local-to-global modeling. In addition, a Boundary-Guided Skip Branch (BGSB) refines skip connections using boundary-aware features, suppressing irrelevant information and enhancing structural coherence. Through this coordinated design, the proposed framework promotes structurally consistent reconstruction during decoding.

B. Convolutional Encoder Module

Reliable low-level cues are required to support geometric refinement during reconstruction. Therefore, a convolutional encoder is employed to extract stable edge- and texture-level features at multiple scales. Cascaded convolutional stages are adopted. At stage ℓ , the input feature is processed by two 3×3 convolutional layers with batch normalization and ReLU activation, followed by a 2×2 max-pooling operation for spatial downsampling:

$$\mathbf{E}_\ell = \text{Pool}(\mathcal{F}(\mathbf{E}_{\ell-1})), \quad (1)$$

where $\mathcal{F}(\cdot)$ denotes the composite mapping of the double convolution and nonlinear activation, and $\ell = 1, \dots, L$. This hierarchical encoding produces multi-scale feature maps $\{\mathbf{E}_\ell\}$ for subsequent decoding and structural modeling.

C. T-KMamba Cooperative Modeling Module

When feature modeling relies primarily on local operations, fine-grained geometric continuity and long-range dependencies are difficult to preserve simultaneously. To address this limitation, a cooperative modeling strategy is employed, in which local nonlinear geometric details are first enhanced and global dependencies are subsequently propagated. A serial KAN→Mamba design with post-normalization residual blocks is adopted to ensure stable and efficient feature modeling, where ordered local-to-global modeling plays a critical role in maintaining reconstruction consistency.

Given an input feature map $\mathbf{X} \in \mathbb{R}^{B \times C \times H \times W}$, it is projected into a sequence representation:

$$\mathbf{S} = \mathcal{P}(\mathbf{X}), \quad \mathbf{S} \in \mathbb{R}^{B \times L \times C}, \quad (2)$$

where $\mathcal{P}(\cdot)$ denotes the flattening and linear embedding operation, and $L = H \times W$. The sequence is first refined by KAN-based nonlinear modeling,

$$\mathbf{S}_1 = \mathbf{S} + \text{KAN}(\mathbf{S}), \quad (3)$$

and then processed by Mamba to propagate long-range dependencies,

$$\mathbf{S}_2 = \mathbf{S}_1 + \text{Mamba}(\mathbf{S}_1). \quad (4)$$

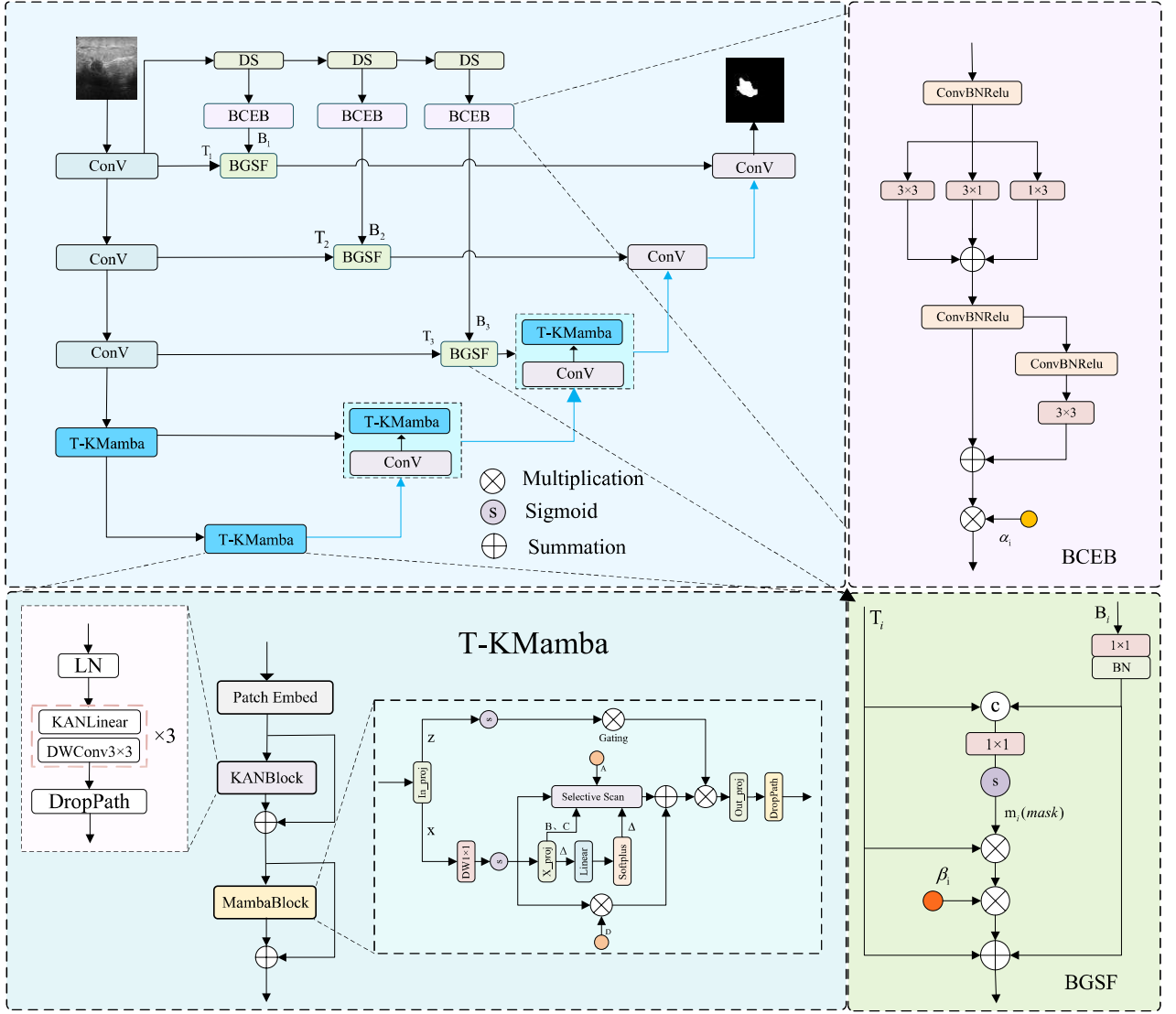


Fig. 1. Overall architecture of the proposed RKANet.

The sequence features are reshaped back to the spatial domain:

$$\mathbf{X}' = \mathcal{R}(\mathbf{S}_2), \quad (5)$$

where $\mathcal{R}(\cdot)$ denotes the reshape operation.

D. Boundary-Guided Skip Branch (BGSB)

The proposed Boundary-Guided Skip Branch (BGSB) consists of a Boundary Compensation Enhancement Block (BCEB) for boundary-aware feature extraction and a Boundary-Guided Skip Fusion (BGSF) module for gated residual fusion, enabling refined skip representations with reduced noise propagation.

Given a shallow encoder feature $\mathbf{F}_0$, multi-scale boundary-aware features are generated via hierarchical downsampling

and BCEB. A learnable scale parameter α_i , constrained by a softplus function, controls the modulation strength:

$$\mathbf{B}_i = \text{softplus}(\alpha_i) \cdot \text{BCEB}(\mathcal{D}_i(\mathbf{F}_0)). \quad (6)$$

At skip level i , the boundary-aware feature is aligned to the channel dimension of the skip feature and used to enhance the encoder feature T_i through a gated residual mechanism:

$$M_i = \sigma(\text{Conv}_{1 \times 1}([T_i, A_i(B_i)])), \quad (7)$$

$$T'_i = T_i + \text{softplus}(\beta_i) \cdot (M_i \odot A_i(B_i)), \quad (8)$$

where $\sigma(\cdot)$ denotes the sigmoid function, $A_i(\cdot)$ denotes channel alignment implemented by a 1×1 convolution followed by normalization, and $[\cdot, \cdot]$ denotes channel-wise concatenation. The gating map M_i adaptively modulates $A_i(B_i)$ and injects it into T_i in a residual manner, suppressing irrelevant responses while enhancing structure-relevant information.

TABLE I
QUANTITATIVE COMPARISON WITH REPRESENTATIVE SEGMENTATION METHODS ON FIVE MEDICAL IMAGE SEGMENTATION DATASETS.

Method	BUSI		CVC-ClinicDB		GlaS		COVID-19		ISIC-17	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	F1	IoU	F1
U-Net(2015)	58.10	73.49	83.45	90.98	81.87	90.03	55.76	71.60	76.81	86.88
Att-UNet(2018)	58.45	73.78	84.52	91.46	82.55	90.44	68.80	81.52	78.13	87.72
Roll-UNet(2024)	60.27	74.89	85.97	92.40	85.69	92.28	69.94	79.55	84.70	91.49
U-Mamba(2024)	62.99	74.42	84.79	91.63	87.01	93.02	71.08	78.70	86.23	91.37
U-KAN(2025)	62.59	76.77	85.20	91.88	86.25	92.55	67.41	79.47	85.86	92.34
KANSeg(2025)	64.08	74.26	83.66	89.46	88.29	93.52	61.98	69.86	83.36	89.76
RKANet(ours)	66.65	79.77	86.90	92.76	88.36	93.77	71.44	82.07	86.59	92.75

Through this gated residual fusion mechanism, boundary-aware features effectively suppress background noise propagation and enhance structure-relevant representations, thereby improving structural continuity and supporting reconstruction consistency during decoding.

IV. EXPERIMENTS

A. Datasets

Experiments are conducted on five publicly available medical image segmentation datasets, including BUSI [27], CVC-ClinicDB [28], GlaS [29], COVID-19 [30], and ISIC-17 [31]. The BUSI, CVC-ClinicDB, GlaS, COVID-19, and ISIC-17 datasets contain 647, 612, 165, 894, and 2000 images, respectively. These datasets cover diverse imaging modalities and anatomical structures, and involve challenging scenarios such as low-contrast regions, elongated targets, and ambiguous boundaries. For each dataset, images are independently divided into training and testing subsets with a ratio of 8:2, and the two subsets are mutually exclusive.

B. Implementation Details

RKANet is implemented in PyTorch and trained on an NVIDIA A800 GPU. Images from the GlaS dataset are resized to 512×512 , while all other datasets are resized to 256×256 . The network is trained for 400 epochs with a batch size of 8. The Adam optimizer with an initial learning rate of 1×10^{-4} and a cosine-annealing schedule is adopted. Standard data augmentation, including random rotation, flipping, resizing, and normalization, is applied. A composite BCEDice loss is used for optimization, and all methods are trained under identical settings for fair comparison.

C. Results

1) Quantitative Results Quantitative evaluation is performed using IoU and Dice scores, which jointly reflect region overlap accuracy and boundary consistency. RKANet is compared with six representative segmentation networks, including U-Net [1], Attention U-Net [21], Rolling-UNet [24], U-Mamba [9], U-KAN [13], and KANSeg [14], under identical experimental settings. The quantitative results on five datasets are summarized in Table I. Overall, RKANet demonstrates competitive and often superior performance across all evaluated datasets.

In particular, on the BUSI dataset, which is characterized by severe boundary ambiguity and elongated lesion structures,

TABLE II
OVERALL COMPARISON WITH STATE-OF-THE-ART SEGMENTATION MODELS IN TERMS OF SEGMENTATION ACCURACY AND COMPUTATIONAL EFFICIENCY.

Method	Average Seg.		Efficiency	
	IoU	Dice	Params (M)	FLOPs (G)
U-Net(2015)	71.19	82.59	34.53	65.52
Att-UNet(2018)	74.49	84.98	34.88	66.63
Roll-UNet(2024)	77.31	86.12	1.78	4.20
U-Mamba(2024)	78.42	85.83	64.01	1147.65
U-KAN(2025)	77.46	86.02	6.36	1.75
KANSeg(2025)	76.27	83.37	6.70	15.64
RKANet(ours)	79.99	88.22	8.05	6.26

RKANet outperforms the second-best method by 3.00% in terms of Dice score. On CVC-ClinicDB, GlaS, COVID-19, and ISIC-17, consistent performance gains are observed, with Dice improvements ranging from 0.25% to 0.55%, indicating stable segmentation performance across diverse anatomical targets and imaging modalities.

2) Qualitative Visualization Results Qualitative visualizations are provided to further analyze segmentation quality beyond numerical metrics, particularly in terms of boundary delineation and structural continuity. As shown in Fig. 2, for challenging scenarios with irregular contours and low contrast, such as BUSI and ISIC-17, RKANet preserves fine-grained geometric details while reducing boundary over-smoothing. On the GlaS dataset, adjacent glands with ambiguous boundaries are more clearly separated. For COVID-19 samples with scattered infection regions, improved topological continuity is observed, indicating effective long-range dependency modeling during decoding. In CVC-ClinicDB, the reduction of background interference caused by reflections and uneven illumination leads to sharper polyp boundaries.

3) Computational Efficiency In addition to segmentation accuracy, computational efficiency is evaluated to assess the practical applicability of the proposed method. As reported in Table II, RKANet achieves a competitive balance between segmentation accuracy and computational cost. Compared with lightweight models such as Roll-UNet and U-KAN, RKANet introduces moderately higher complexity due to additional structural modeling components. Nevertheless, this increase enables RKANet to achieve the highest average Dice score of 88.22% among the compared methods.

4) Response Heatmaps To provide further insight into the model's focus during decoding, response heatmaps are

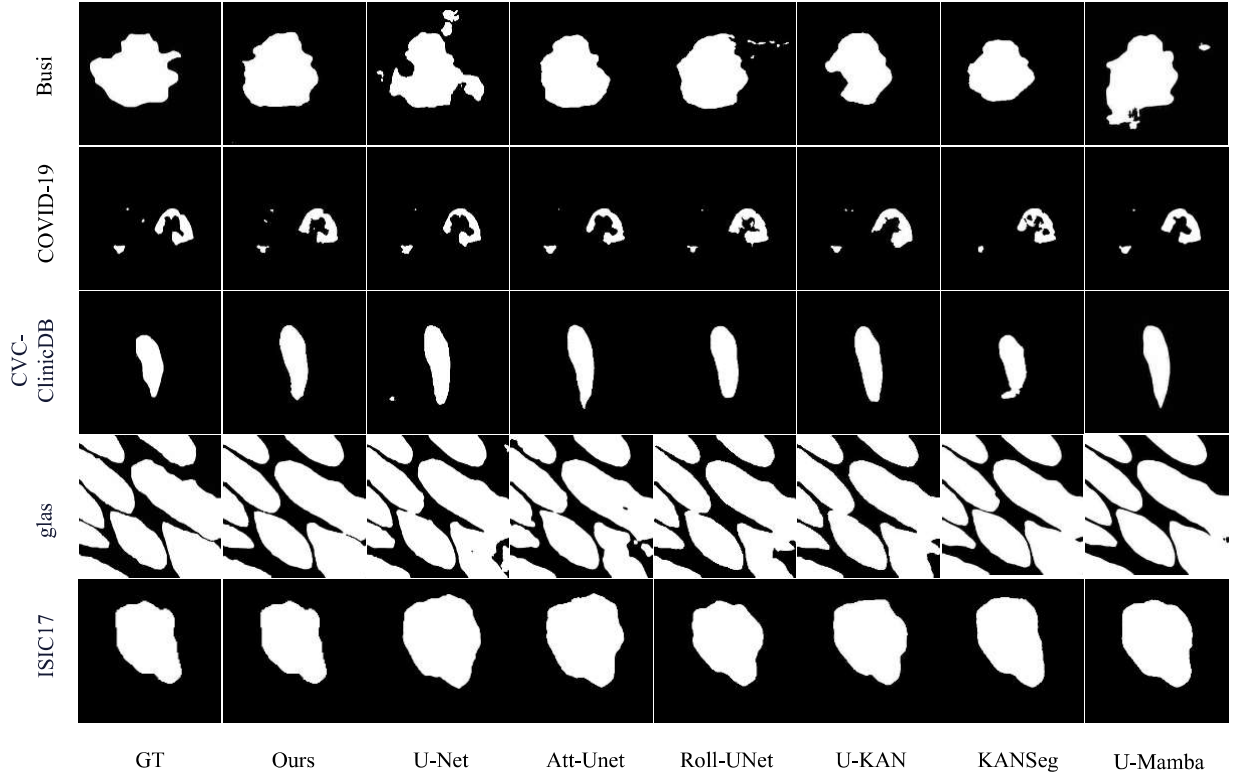


Fig. 2. Qualitative comparison of segmentation results produced by different methods on multiple medical image segmentation datasets.

TABLE III
ABLATION STUDY ON THE CVC-CLINICDB DATASET.

Method	IoU	Dice
U-Net	83.45	90.98
U-Net + KAN	85.20	91.88
U-Net + KAN + Mamba	86.14	92.39
U-Net + KAN + Mamba + BGSB	86.90	92.97

visualized in Fig. 3. For BUSI and ISIC-17, concentrated activations are aligned with irregular lesion boundaries, with minimal response in surrounding background regions. In the GlaS dataset, distinct activations corresponding to individual glands are observed. Similar focused responses are obtained for infection and polyp regions in COVID-19 and CVC-ClinicDB, respectively, indicating consistent feature attention across.

D. Ablation Study

An ablation study on the CVC-ClinicDB dataset is conducted to evaluate the contribution of each component. As summarized in Table III, starting from the U-Net baseline, successive performance improvements are observed with the introduction of KAN-based geometric modeling, Mamba-based global dependency modeling, and the Boundary-Guided Skip Branch, with the full model achieving the best results.

Table IV further investigates different fusion strategies within the T-KMamba module. Compared with parallel fusion,

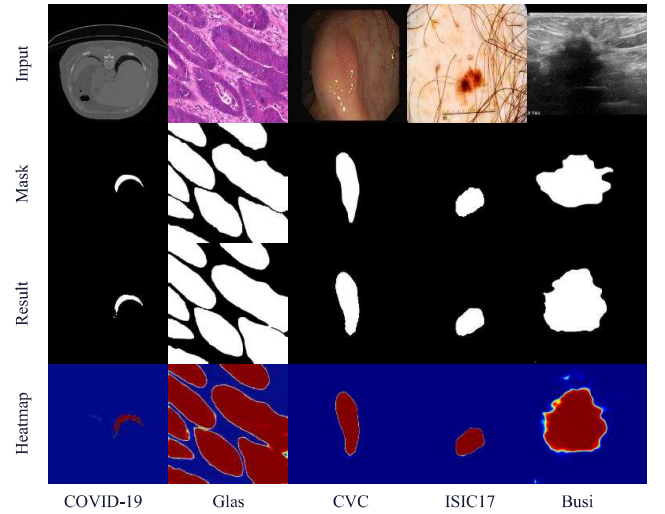


Fig. 3. Visualization of response heatmaps generated by RKANet on different datasets.

serial fusion yields better performance, and among the serial designs, the K→M configuration achieves the best results. This suggests that ordered local-to-global feature modeling facilitates more stable and coherent feature reconstruction during decoding.

TABLE IV
ABLATION STUDY OF DIFFERENT KAN-MAMBA FUSION STRATEGIES
WITHIN THE T-KMAMBA MODULE ON THE BUSI DATASET.

Model Variant	Fusion Strategy	IoU	Dice
Variant A	Parallel ($K \parallel M$)	63.82	77.72
Variant B	Serial ($M \rightarrow K$)	64.14	77.89
Ours	Serial ($K \rightarrow M$)	66.65	79.77

V. CONCLUSION

In this work, a reconstruction-consistent medical image segmentation framework, termed RKANet, has been presented to address structural degradation during the decoding stage. By reformulating decoding as a structured reconstruction pathway, local geometric refinement and global dependency modeling are jointly considered, highlighting the importance of strengthening the decoder beyond simple upsampling operations. A cooperative modeling strategy is employed to integrate nonlinear geometric representation with long-range dependency propagation in an ordered local-to-global manner, while skip feature fusion is regulated using boundary-aware features. This design improves reconstruction consistency, leading to more structurally coherent and reliable segmentation results. Experimental results on multiple public medical image segmentation datasets demonstrate that the proposed framework effectively enhances structural continuity and boundary fidelity across diverse anatomical scenarios. These findings suggest that cooperative and structure-aware decoding designs provide a meaningful direction for improving the robustness and reliability of encoder-decoder segmentation networks.

REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [2] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learning in Medical Image Analysis*, 2018, pp. 3–11.
- [3] H. Huang, L. Lin, R. Tong, *et al.*, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1055–1059.
- [4] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [6] J. Chen, Y. Lu, Q. Yu, *et al.*, "TransUNet: Transformers make strong encoders for medical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2021, pp. 196–206.
- [7] A. Lin, B. Chen, J. Xu, *et al.*, "DS-TransUNet: Dual Swin Transformer U-Net for medical image segmentation," *IEEE Trans. Instrumentation and Measurement*, 2022.
- [8] H. Cao, Y. Wang, J. Chen, *et al.*, "Swin-Unet: Unet-like pure transformer for medical image segmentation," in *Proc. European Conf. Computer Vision (ECCV)*, 2022, pp. 205–218.
- [9] J. Ma, F. Li, and B. Wang, "U-Mamba: Enhancing long-range dependency for biomedical image segmentation," *arXiv preprint arXiv:2401.04722*, 2024.
- [10] W. Wang, Y. Luo, and X. Wang, "BEFNet: A Hybrid CNN-Mamba Architecture for Accurate Skin Lesion Image Segmentation," in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, Lisbon, Portugal, 2024, pp. 3795–3798, doi: 10.1109/BIBM62325.2024.10822106.
- [11] J. Liu, H. Yang, H. Zhou, *et al.*, "Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining," *arXiv preprint arXiv:2402.03302*, 2024.
- [12] Z. Liu, Y. Wang, N. Vaidya, *et al.*, "Kolmogorov–Arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.
- [13] C. Li, X. Liu, W. Li, *et al.*, "U-KAN makes strong backbone for medical image segmentation and generation," in *Proc. AAAI Conf. Artificial Intelligence*, 2025, vol. 39, no. 5, pp. 4652–4660.
- [14] J. Zhu, Z. Tang, Z. Liang, *et al.*, "KANSeg: An efficient medical image segmentation model based on Kolmogorov–Arnold networks for multi-organ segmentation," *Digital Signal Processing*, vol. 147, p. 105472, 2025.
- [15] J. M. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, "KiU-Net: Overcomplete convolutional architectures for biomedical image segmentation," *IEEE Trans. Medical Imaging*, vol. 41, no. 4, pp. 965–976, 2022.
- [16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.
- [17] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," *arXiv preprint arXiv:2111.00396*, 2021.
- [18] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [19] Z. Xing, T. Ye, Y. Yang, *et al.*, "SegMamba: Long-range sequential modeling with Mamba for 3D medical image segmentation," in *Proc. MICCAI*, 2024, pp. 578–588.
- [20] W. Liao, Y. Zhu, X. Wang, *et al.*, "LightM-UNet: Mamba assists in lightweight UNet for medical image segmentation," *arXiv preprint arXiv:2403.05246*, 2024.
- [21] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, "Attention U-Net: Learning where to look for the pancreas," in *Proc. MICCAI*, 2018, pp. 39–46.
- [22] Z. Zhang, H. Fu, H. Dai, *et al.*, "ET-Net: A generic edge-attention guidance network for medical image segmentation," in *Proc. MICCAI*, 2019, pp. 442–450.
- [23] H. Liu, S. Wang, X. Chen, *et al.*, "Semi-Mamba-UNet: Pixel-level contrastive cross-supervised visual Mamba-based UNet for semi-supervised medical image segmentation," *arXiv preprint arXiv:2402.07245*, 2024.
- [24] Y. Liu, H. Zhu, M. Liu, *et al.*, "Rolling-UNet: Revitalizing MLP's ability to efficiently extract long-distance dependencies for medical image segmentation," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 38, no. 4, pp. 3819–3827, 2024.
- [25] X. Qin, Z. Zhang, C. Huang, *et al.*, "BASNet: Boundary-aware segmentation network for salient object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7479–7489.
- [26] W. Wang, L. Wang, X. Wang, *et al.*, "GBA-Net: A method for 3D brain tumor segmentation based on multi-scale Gaussian boundary attention," in *Proc. ICASSP*, 2025.
- [27] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104863, 2020.
- [28] J. Bernal, J. Sánchez, G. Fernández-Esparrach, *et al.*, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps," *Medical Image Analysis*, vol. 43, pp. 99–111, 2018.
- [29] J. M. J. Valanarasu, *et al.*, "Medical Transformer: Gated axial-attention for medical image segmentation," in *Proc. MICCAI*, 2021, pp. 36–46.
- [30] M. Jun, *et al.*, "COVID-19 CT lung and infection segmentation dataset," 2020.
- [31] N. C. F. Codella, D. Gutman, M. E. Celebi, *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at ISBI 2017," in *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI)*, 2018, pp. 168–172.