This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/ICMCR64890.2025.10962807,
2026 11th International Conference on Electronic Technology and Information Science (ICETIS)

LGCF-Net: Local-Global Competitive Fusion for Ultra-Lightweight Medical Image Segmentation

1st Tao Lei*
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
*leitao@sust.edu.cn

2nd Bo Feng
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
m13892659163@163.com

3rd Shaoqing Liu
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
231611029@sust.edu.cn

4rd Yan Lu
College of Bioresources Chemical and Materials Engineering
Shaanxi University of Science and Technology
Xi'an, China
luyan@sust.edu.cn

5th Yong Wan
Department of Geriatric Surgery
The First Affiliated Hospital of Xi'an Jiaotong University
Xi'an, China
docwanyong@xjtu.edu.cn

6th Hongying Meng
Department of Electronic and Electrical Engineering
Brunel University of London
London, U.K.
hongying.meng@brunel.ac.uk

***Abstract*—Although deep neural networks possess powerful feature representation capabilities, they typically demand massive parameter scales and heavy computational burdens, while current lightweight alternatives that reduce model sizes through operations like pruning or depthwise separable convolutions often struggle to maintain the original representation capacity. To address these challenges, we propose a ultra-lightweight medical image segmentation network based on local-global competitive fusion, termed LGCF-Net. Firstly, to simultaneously enhance the perception of localized boundary textures and capture long-range contextual dependencies under limited computational budgets, we design a parallel dual-branch architecture at each resolution layer, integrating an Asymmetric Depthwise Separable Convolution (ADSC) module and a Mamba-based Visual State Space (VSS) module. Secondly, to mitigate feature redundancy and representation conflicts arising from this dual-branch integration, we introduce a Branch Competitive Gating (BCG) module to achieve adaptive feature selection via branch-wise competitive weight allocation. Finally, to suppress cross-layer noise propagation during feature transmission, we develop a Decoder-Guided Attention Gating (DGAG) mechanism within the skip connections to filter and refine shallow encoder features using higher-level decoder semantic priors. Extensive experiments on four public datasets demonstrate that LGCF-Net consistently outperforms state-of-the-art methods. Notably, our network compresses the 34.52M baseline parameters by nearly 96% to just 1.53M. On the ACDC dataset, it achieves a 91.56% Dice score with only 2.82G FLOPs, establishing a effective trade-off between segmentation accuracy and computational efficiency.**

***Index Terms*—Medical image segmentation, Local-global competitive fusion, Mamba state space model, Branch competitive gating, Decoder-guided attention gating**

I. Introduction

Medical image segmentation is foundational to computer-aided diagnosis and clinical decision-making. Unlike natural images, medical scans typically suffer from low contrast, indistinct boundaries, severe noise, and substantial morphological variations. This requires models to simultaneously capture fine-grained local details and maintain global structural consistency [1], [2]. Furthermore, deploying these models in resource-constrained clinical edge environments demands a compact parameter scale and low computational overhead. Balancing segmentation accuracy with lightweight design thus remains a critical, unresolved challenge [3].

Although CNNs excel at capturing fine-grained local details, their inherent locality limits long-range modeling despite multi-scale feature fusion. Conversely, global modules like Transformers capture long-range context but suffer from quadratic complexity, while linear-complexity alternatives like Mamba struggle with localized precision. To exploit their complementary strengths, hybrid local-global frameworks have emerged, broadly categorized into serial and parallel architectures [4], [5]. While serial designs inherently delay feature interaction, parallel structures concurrently extract dual-branch representations within the same layer. However, conventional parallel methods heavily rely on static fusion mechanisms—such as element-wise summation or channel concatenation—which fail to dynamically calibrate branch contributions [6], thereby inducing feature redundancy and undermining parallel synergy.

Copyright 'For the purpose of open access, the author has applied a 'Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.'

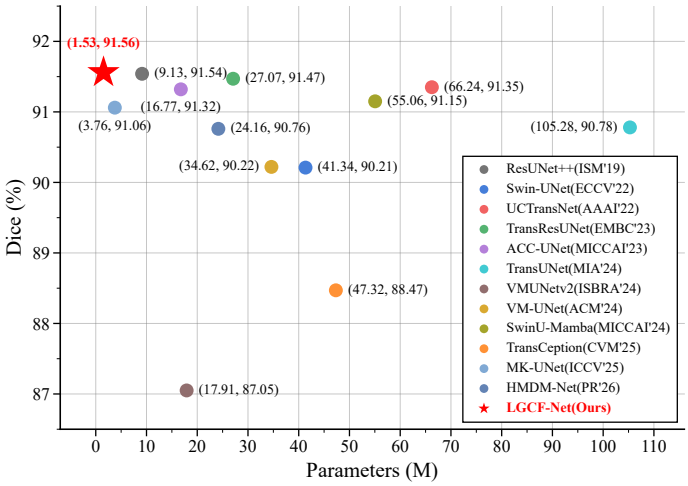

Figure 1. Visualization of comparative experimental results on the ACDC dataset.

To overcome these limitations, we present LGCF-Net, an ultra-lightweight network that parallelly models local details and global dependencies at each resolution layer. It utilizes an Asymmetric Depthwise Separable Convolution (ADSC) module for efficient local feature extraction and a Mamba-based Visual State Space (VSS) [7] module for linear-complexity global modeling. A Branch Competitive Gating (BCG) module then enforces channel-wise competition for adaptive feature selection, while a Decoder-Guided Attention Gating (DGAG) module suppresses cross-layer noise along skip connections. As plotted in the Figure 1 scatter plot, LGCF-Net (red star) strikes a superior accuracy-efficiency trade-off over mainstream methods (distinct colors). Our main contributions are summarized as follows:

- **Ultra-Lightweight Parallel Architecture:** To mitigate the prohibitive parameter scales of existing hybrid CNN-Mamba networks, we propose an ultra-lightweight parallel dual-branch architecture. By organically integrating an Asymmetric Depthwise Separable Convolution (ADSC) and a Mamba-based Visual State Space (VSS) module, it expands the receptive field and maintains robust representation capacity while compressing the parameter footprint by an order of magnitude.
- **Dynamic Local-Global Feature Fusion:** To resolve feature redundancy and conflicts in conventional static fusion (e.g., summation or concatenation), we design a Branch Competitive Gating (BCG) module. It applies a branch-wise Softmax constraint to dynamically calibrate channel weights for adaptive local-global aggregation.
- **Decoder-Guided Feature Refinement:** To suppress background noise inherent in standard skip connections, we introduce a Decoder-Guided Attention Gating (DGAG) mechanism for optimized cross-layer feature transmission. By leveraging high-level decoder semantic priors to conditionally modulate representations, it effectively filters background distractors and preserves sharp boundary details.
- **Comprehensive Experimental Validation:** Extensive evaluations across four public datasets (ISIC2018, GLAS, MoNuSeg, and ACDC) demonstrate that LGCF-Net consistently outperforms state-of-the-art methods, establishing a favorable trade-off between accuracy and efficiency.

II. Method

A. Overview

Figure 2 illustrates LGCF-Net, a symmetrical encoder-decoder network embedding parallel local-global tracks at each layer to process textures and context concurrently. The encoder utilizes Dual Path Competition Fusion (DPCF) modules to dynamically aggregate an asymmetric CNN branch and a Mamba branch via soft-constrained Branch Competitive Gating (BCG). Symmetrically, the decoder restores spatial resolution while deploying Decoder-Guided Attention Gating (DGAG) on skip connections to suppress background clutter.

B. Dual Path Competition Fusion

To overcome CNN locality, Transformer complexity, and static fusion redundancy, we propose the Dual Path Competition Fusion (DPCF) module for efficient local-global modeling. DPCF parallelizes an Asymmetric Depthwise Separable Convolution (ADSC) module to capture direction-sensitive boundaries cost-effectively, and a Visual State Space (VSS) module for linear-complexity long-range interactions. To eliminate cross-branch interference and optimize convergence, a Branch Competitive Gating (BCG) module applies a soft branch-wise competitive constraint, enabling dynamic, channel-wise feature selection between the dual representations.

a) Local-Global Parallel Feature Extraction

Standard lightweight designs like depthwise separable convolutions reduce computational overhead but often compromise directional details, leading to inaccurate boundary predictions near complex tissue contours [8]. To address this, we propose the Asymmetric Depthwise Separable Convolution (ADSC) module to reinforce directional feature capture via parallel asymmetric paths. Specifically, given an input feature map $X \in \mathbb{R}^{H\times W\times C}$, the depthwise phase splits the feature extraction into three parallel tracks with 3×3, 1×3, and 3×1 kernels. While the standard square kernel captures isotropic local patterns, the 1×3 and 3×1 kernels explicitly enhance horizontal and vertical boundary perception, respectively. These parallel outputs are fused via element-wise summation, followed by a pointwise convolution for cross-channel interaction, formulated as:

$$F_{local} = \text{PW-Conv}\Big(\text{DW-Conv}_{3\times 3}(X) + \text{DW-Conv}_{1\times 3}(X) + \text{DW-Conv}_{3\times 1}(X)\Big) \tag{1}$$

where $\text{DW-Conv}_{K_h\times K_w}$ denotes depthwise convolution with a $K_h \times K_w$ kernel, and PW-Conv represents the 1×1 pointwise convolution. ADSC substantially improves local detail representation with minimal parameter inflation.

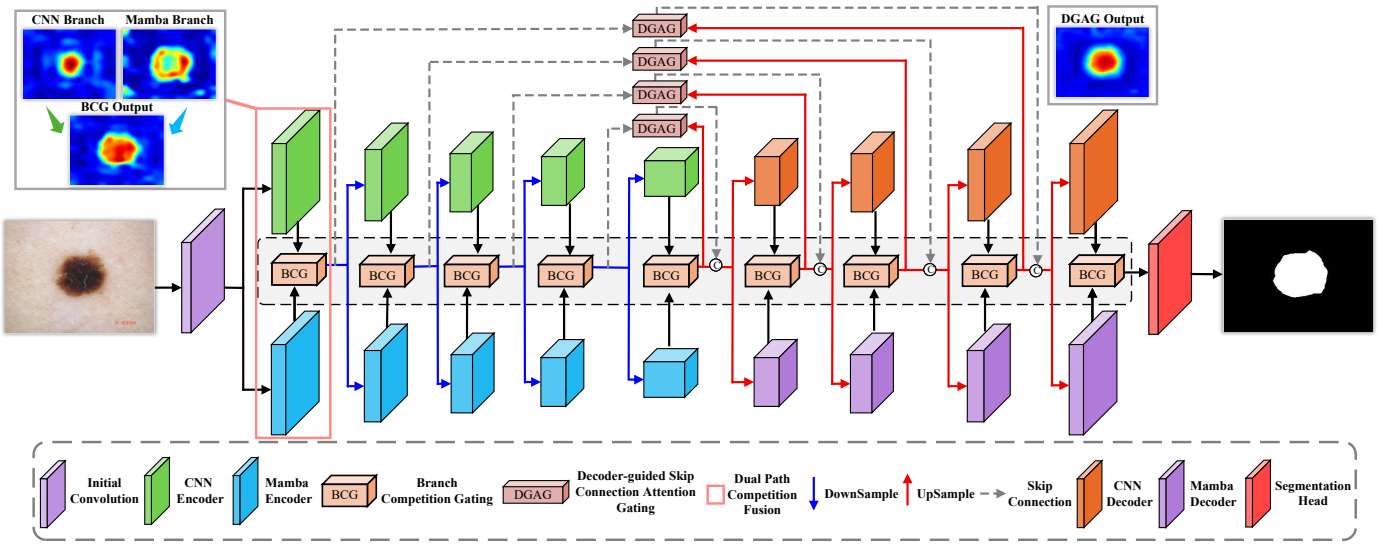

Figure 2. Overview of the proposed LGCF-Net framework.

To complement these local features, a Mamba-based Visual State Space (VSS) module parallelly extracts global representations $F_{global} \in \mathbb{R}^{H\times W\times C}$, capturing long-range spatial dependencies with linear complexity.

b) Branch Competitive Gating

To overcome the redundancy inherent in static fusion strategies, we introduce the Branch Competitive Gating (BCG) module. Given local features $F_{local} \in \mathbb{R}^{H\times W\times C}$ and global features $F_{global} \in \mathbb{R}^{H\times W\times C}$, we first align their distributions via 1×1 convolutions and Batch Normalization to obtain F'_{local} and F'_{global}. To capture cross-path correlations, these aligned features are concatenated and spatially aggregated via Global Average Pooling (GAP). The resulting channel descriptor is mapped by a 1D convolution and reshaped into a branch-channel score matrix $M_s \in \mathbb{R}^{2\times C}$. Crucially, a Softmax function is applied to M_s along the branch dimension to enforce a soft competitive constraint, generating decoupled weight vectors $W_{local}, W_{global} \in \mathbb{R}^{1\times C}$:

$$[W_{local}, W_{global}] = \text{Softmax}_{branch}(M_s) \tag{2}$$

This constraint ensures cross-path complementarity. Finally, these weights are spatially broadcast and element-wise multiplied ($\odot$) to compute the fused output $F_{out} \in \mathbb{R}^{H\times W\times C}$:

$$F_{out} = W_{local} \odot F'_{local} + W_{global} \odot F'_{global} \tag{3}$$

This dynamic selection effectively minimizes feature redundancy while optimizing dual-path synergy.

C. Decoder-Guided Attention Gating

Standard skip connections transfer encoder features directly to the decoder to recover spatial details, but shallow encoder streams invariably introduce background noise and semantic inconsistencies that degrade boundary delineation [9]. To resolve this, we present the Decoder-Guided Attention Gating (DGAG) module, which leverages high-level decoder semantics to conditionally filter skip-connection clutter via sequential channel and spatial attention tracks.

DGAG operates on the encoder skip feature $F_{skip} \in \mathbb{R}^{H\times W\times C}$ and the adjacent upsampled decoder feature $F_{dec} \in \mathbb{R}^{H\times W\times C}$. To align their distinct distributions, both streams are processed via independent 1×1 convolutions and Batch Normalization (BN), then aggregated via element-wise summation to generate a joint representation $F_{joint} \in \mathbb{R}^{H\times W\times C}$:

$$F_{joint} = \text{BN}\Big(\text{Conv}_{1\times1}(F_{dec})\Big) + \text{BN}\Big(\text{Conv}_{1\times1}(F_{skip})\Big) \tag{4}$$

Based on F_{joint}, the module executes dual-dimensional feature filtering. First, to model cross-channel dependencies, F_{joint} is spatially aggregated via parallel global average pooling (GAP) and global max pooling (GMP). The descriptors are mapped through a shared multi-layer perceptron (MLP) to compute the channel attention weights $A_c \in \mathbb{R}^{1\times1\times C}$:

$$A_c = \sigma\left(\text{MLP}\Big(\text{GAP}(F_{joint})\Big) + \text{MLP}\Big(\text{GMP}(F_{joint})\Big)\right) \tag{5}$$

where σ denotes the Sigmoid function. The joint feature map is then modulated along channels as $F'_{joint} = A_c \otimes F_{joint}$ ($\otimes$ signifies channel-wise multiplication).

Second, to isolate precise boundaries, F'_{joint} is pooled along the channel axis using average and maximum operators, concatenated, and mapped through a 7×7 convolution to generate the spatial attention map $A_s \in \mathbb{R}^{H\times W\times1}$. This map gating controls the original skip feature to produce the final output $F_{out} \in \mathbb{R}^{H\times W\times C}$:

$$A_s = \sigma\left(\text{Conv}_{7\times7}\Big(\big[\text{AvgPool}(F'_{joint});\ \text{MaxPool}(F'_{joint})\big]\Big)\right),$$
$$F_{out} = A_s \odot F_{skip} \tag{6}$$

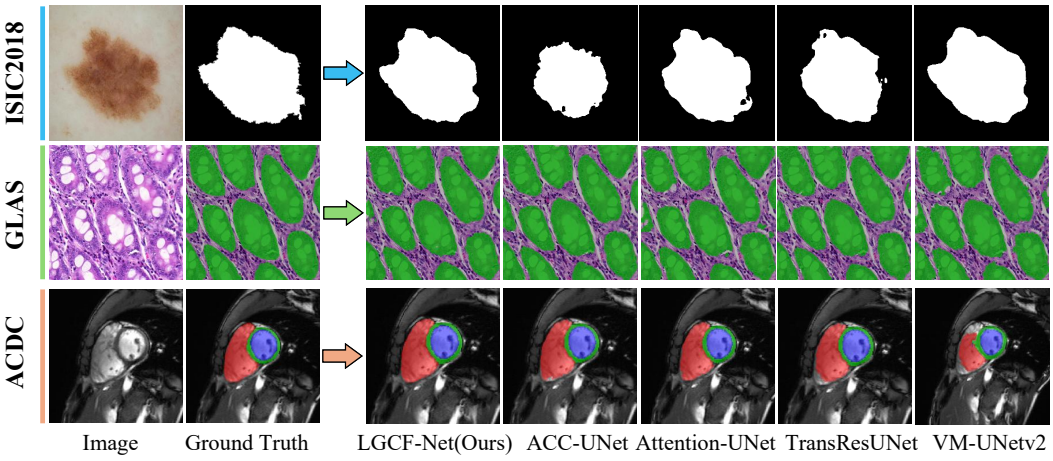

Figure 3. Visual comparison of our method with existing methods.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT METHODS ON THE GLAS, ACDC, ISIC2018 AND MONUSEG DATASETS.

Method	Architecture	GLAS		ACDC		ISIC2018		MoNuSeg		Param.(M)	FLOPs(G)
		Dice	IoU	Dice	HD95	Dice	IoU	Dice	IoU		
UNet	CNN	85.45	74.78	90.68	1.51	85.54	78.47	76.45	62.86	34.52	65.46
ResUNet++	CNN	86.46	77.26	91.54	1.12	85.57	81.35	79.44	66.25	9.13	35.45
Attention-UNet	CNN	88.80	80.69	90.91	1.52	88.41	81.49	76.67	63.74	34.88	66.63
MK-UNet	CNN	89.32	81.73	91.06	1.33	89.25	81.74	80.25	67.05	3.76	8.19
Swin-UNet	Transformer	89.58	82.06	90.21	1.32	88.87	81.67	77.85	64.00	41.34	8.68
TransResUNet	Transformer	89.86	82.01	91.47	1.24	89.29	81.86	80.31	67.41	27.07	24.07
UCTransNet	Transformer	90.08	82.16	91.35	1.51	89.03	82.02	79.08	65.50	66.24	32.98
TransCeption	CNN+Transformer	87.70	78.27	88.47	1.78	87.24	80.02	80.08	67.05	47.32	14.96
TransUNet	CNN+Transformer	88.40	80.40	90.78	1.58	84.99	77.00	78.53	65.05	105.28	24.68
ACC-UNet	CNN+Transformer	90.19	82.37	91.32	1.54	88.18	81.11	79.49	66.27	16.77	45.33
VMUNetv2	Mamba	86.08	74.86	87.05	1.40	89.73	81.37	81.31	68.56	17.91	35.28
VM-UNet	Mamba	89.30	81.68	90.22	1.59	89.71	81.35	79.52	66.31	34.62	5.79
HMDM-Net	Mamba	89.29	81.03	90.76	1.46	89.95	81.78	80.69	68.17	24.16	23.30
SwinU-Mamba	Mamba	89.51	81.24	91.15	1.80	90.27	82.71	81.10	68.25	55.06	33.64
LGCF-Net(Ours)	**CNN+Mamba**	**90.37**	**82.63**	**91.56**	**1.12**	**90.95**	**83.72**	**82.00**	**69.53**	**1.53**	**2.82**

where $[\cdot;\cdot]$ denotes channel-wise concatenation and $\odot$ represents element-wise spatial multiplication.

This decoder-guided dual-attention constraint effectively suppresses cross-layer noise propagation, safeguarding geometric fidelity for boundary reconstruction.

III. EXPERIMENTS AND RESULTS

a) Datasets

We validate LGCF-Net across four public datasets spanning diverse modalities and clinical targets. For skin lesion segmentation, ISIC-2018 comprises 2,594 training and 1,000 testing dermoscopy images. For gland segmentation, GlaS contains 165 highly variable H&E-stained colorectal histology images, split into 85 training and 80 testing cases [10]. For multi-organ nuclei segmentation, MoNuSeg features 30 training and 14 validation H&E images cropped and resized into non-overlapping 256×256 patches. Finally, for multi-class cardiac segmentation, ACDC includes MRI scans from 100 patients partitioned into 70 training, 10 validation, and 20 testing cases.

b) Implementation Details

LGCF-Net is implemented in PyTorch 1.10.1 on an NVIDIA RTX 3090 GPU. Training utilizes the Adam optimizer with a batch size of 8 and an initial learning rate of 1×10^{-3} for 30,000 iterations. The network minimizes a joint cross-entropy and Dice loss under a uniform data augmentation strategy to ensure unbiased comparisons.

c) Comparison with State-of-the-Art Methods

Our evaluation covers baselines across four architectures—CNNs, Transformers, hybrids, and Mamba.

As summarized in Table I, LGCF-Net achieves highly competitive quantitative results across four clinical validation cohorts, with the best and second-best outcomes highlighted in bold black and green, respectively. These metrics demonstrate an optimal trade-off between segmentation accuracy and parameter economy. Specifically, on the GlaS dataset, it achieves a Dice score of 90.37% and an IoU of 82.63%, outperforming the hybrid ACC-UNet while reducing the parameter footprint from 16.77M to 1.53M. Similarly, on the morphologically diverse ISIC2018 cohort, our framework yields a 90.95% Dice score and 83.72% IoU, outpacing SwinU-Mamba by 0.68%

and 1.01% respectively, while requiring merely 2.7% of its parameter scale. LGCF-Net also delivers leading performance on the ACDC (91.56% Dice) and MoNuSeg (69.53% Iou) datasets. Collectively, these evaluations prove that our ultra-lightweight architecture (1.53M parameters and 2.82G FLOPs) consistently delivers state-of-the-art precision across diverse imaging modalities.

Figure 3 compares three validation cohorts, showing LGCF-Net achieves the highest structural similarity to ground truth. On ISIC 2018 (white), it captures irregular morphology while suppressing under-segmentation and false positives. On GLAS (green), it accurately delineates multi-scale glands, decouples adjacent structures, and reduces misses. On ACDC (red: LV, green: MYO, blue: RV), it reconstructs anatomical interfaces with sharper boundaries and geometric completeness, confirming robust multi-class parsing.

d) Ablation Studies

Ablation studies on the ACDC dataset systematically validate the architectural efficacy and necessity of each component within LGCF-Net (Tables II and III), where "Number" identifies distinct method variants. Component-wise evaluation (Table II) reveals that while naive channel reduction (Lightweight UNet) compromises representation capacity (90.02% Dice), progressively embedding the Dual Path Competition Fusion (DPCF) and Decoder-Guided Attention Gating (DGAG) modules drives continuous performance gains. The finalized LGCF-Net achieves a peak Dice score of 91.56% and an HD95 of 1.12 mm, while drastically compressing baseline UNet parameters and FLOPs by over 95% (down to 1.53M and 2.82G). Furthermore, inner-structural ablation within the DPCF module (Table III, where "w/o" and "w" designate theq exclusion and inclusion of functional modules) demonstrates that replacing Branch Competitive Gating (BCG) with conventional concatenation degrades the Dice score to 91.22%, verifying the value of dynamic channel-wise competition. Either VSS or ADSC ablation lowers accuracy, validating their complementary necessity.

TABLE II
ABLATION STUDY OF DIFFERENT MODULES IN LGCF-NET.

Number	Configuration	Dice	HD95	Param.(M)	FLOPs(G)
N1	UNet	90.68	1.51	34.52	65.46
N2	Lightweight UNet	90.02	1.86	3.53	7.34
N3	N2+DPCF	91.32	1.24	**1.48**	**2.79**
N4	**N3+DGAG(Ours)**	**91.56**	**1.12**	1.53	2.82

TABLE III
EFFECTIVENESS VALIDATION OF THE DPCF MODULE DESIGN.

Number	Configuration	Dice	HD95	Param.(M)	FLOPs(G)
N4	**LGCF-Net(Ours)**	**91.56**	**1.12**	1.53	2.82
N5	w/o BCG	91.22	1.52	1.50	2.80
N6	w/o VSS	90.93	1.28	**0.62**	**1.18**
N7	w/o ADSC	91.01	1.32	1.11	2.09

IV. CONCLUSION

We propose LGCF-Net, a lightweight medical image segmentation framework that balances architectural efficiency and boundary fidelity. Its Dual-Path Asymmetric Convolutional Fusion (DPCF) and Decoder-Guided Attention Gating (DGAG) modules mitigate representation trade-offs and cross-layer noise inherent in conventional designs. Extensive experiments on ACDC, ISIC 2018, GLAS, and MoNuSeg show that LGCF-Net consistently outperforms CNN, Transformer, and Mamba-based state-of-the-art methods. With only 1.53M parameters and 2.82G FLOPs, it achieves accurate delineation and boundary tracking, providing a practical solution for resource-constrained clinical assistance.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (Program No. 62271296, 62201334), in part by The Innovation Capability Support Plan Project in Shaanxi Province (2025RS-CXTD-012), in part by Scientific Research Program Funded by Shaanxi Provincial Education Department (25JP023, 23JP014, 23JP022), and in part by Young Science and Technology Innovation Leading Talents Program of Xi'an (25ZQRC00019).

REFERENCES

[1] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical image segmentation: A comprehensive review of deep learning-based methods," *Tomography*, vol. 11, no. 5, p. 52, 2025. doi:10.3390/tomography11050052.

[2] R. R. Kumar and R. Priyadarshi, "Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches," *Multimedia Tools and Applications*, vol. 84, no. 12, pp. 10817–10875, 2025. doi:10.1007/s11042-024-19313-6.

[3] Z. Zhu, H. Wang, G. Qi, Y. Li, N. Mazur, Y. Liu, H. Li, B. Cong, and L. Bai, "A survey on lightweight technology of neural networks for medical image segmentation," *Pattern Recognition*, p. 113870, 2026. doi:10.1016/j.patcog.2026.113870.

[4] T. Lei, R. Sun, X. Wang, Y. Wang, X. He, and A. Nandi, "Cit-net: convolutional neural networks hand in hand with vision transformers for medical image segmentation," *arXiv preprint arXiv:2306.03373*, 2023. doi:10.24963/ijcai.2023/113.

[5] S. Somvanshi, M. M. Islam, M. S. Mimi, S. B. B. Polock, G. Chhetri, and S. Das, "From s4 to mamba: A comprehensive survey on structured state space models," *arXiv preprint arXiv:2503.18970*, pp. 1–30, 2025. doi:10.48550/arXiv.2503.18970.

[6] P. Cao, Y. Zhao, T. Duan, L. Li, C. Xian, and S. Li, "A review of multimodal image feature fusion technology and application.," *Applied Sciences (2076-3417)*, vol. 16, no. 11, p. 5290, 2026. doi:10.3390/app16115290.

[7] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, J. Jiao, and Y. Liu, "Vmamba: Visual state space model," *Advances in neural information processing systems*, vol. 37, pp. 103031–103063, 2024. doi:10.52202/079017-3273.

[8] H. Lu, Y. Zhao, G. Yang, and S. Wu, "Dsgnet: A lightweight network integrating depthwise separable and ghost convolutions for real-time surface defect segmentation," *Expert Systems*, vol. 43, no. 2, p. e70187, 2026. doi:10.1111/exsy.70187.

[9] Y. Peng, D. Z. Chen, and M. Sonka, "U-net v2: Rethinking the skip connections of u-net for medical image segmentation," in *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2025. doi:10.1109/ISBI60581.2025.10980742.

[10] K. Bräutigam, A.-M. Baker, V. H. Koelzer, J. N. Kather, and T. A. Graham, "Integrating artificial intelligence (ai) into colorectal cancer reporting," *The Journal of Pathology*, vol. 268, no. 4, pp. 367–382, 2026. doi:10.1002/path.70029.