MambaLiteUNet: Cross-Gated Adaptive Feature Fusion for Robust Skin Lesion Segmentation

Md Maklachur Rahman[1], Soon Ki Jung[2], Tracy Hammond[1]

[1]Texas A&M University [2]Kyungpook National University

CVPR JUNE 3-7, 2026 DENVER COLORADO

Paper ID# 42423

Paper

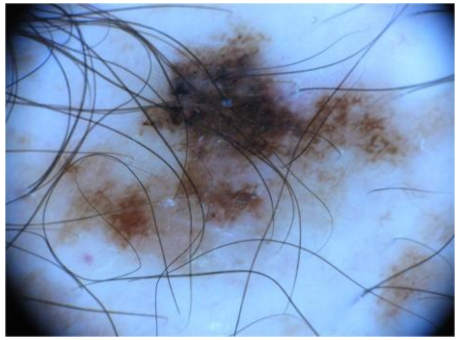
Code

Problem & Motivation: Why It Matters?

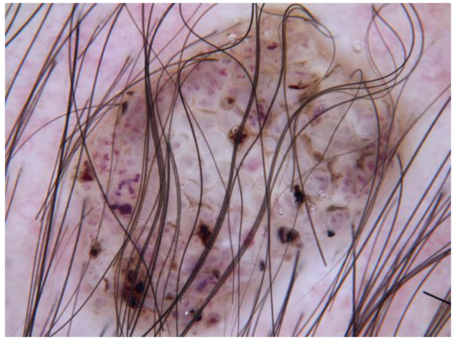
Clinical Need: Accurate lesion boundary analysis is essential for early melanoma detection, as melanoma is one of the deadliest forms of skin cancer.

Visual challenge: Color/shape variation, hair, artifacts, and low contrast often obscure lesion regions.

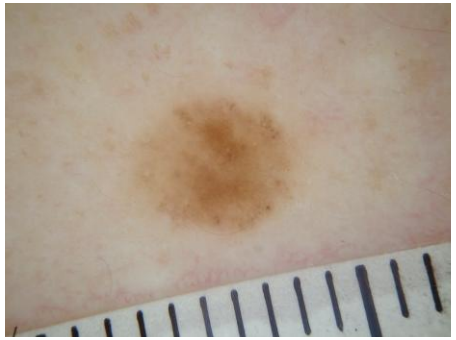
Efficiency gap: Heavy models are accurate but costly, while lightweight models can lose fine boundary details.



(a) Variations in lesion color/shape



(b) Hair occlusion



(c) Image acquisition artifacts

(d) Low contrast between lesion and skin

Core goal: Preserve lesion detail, capture global context, and filter background noise within a compact model.

Image source: International Skin Imaging Collaboration (**ISIC**) datasets

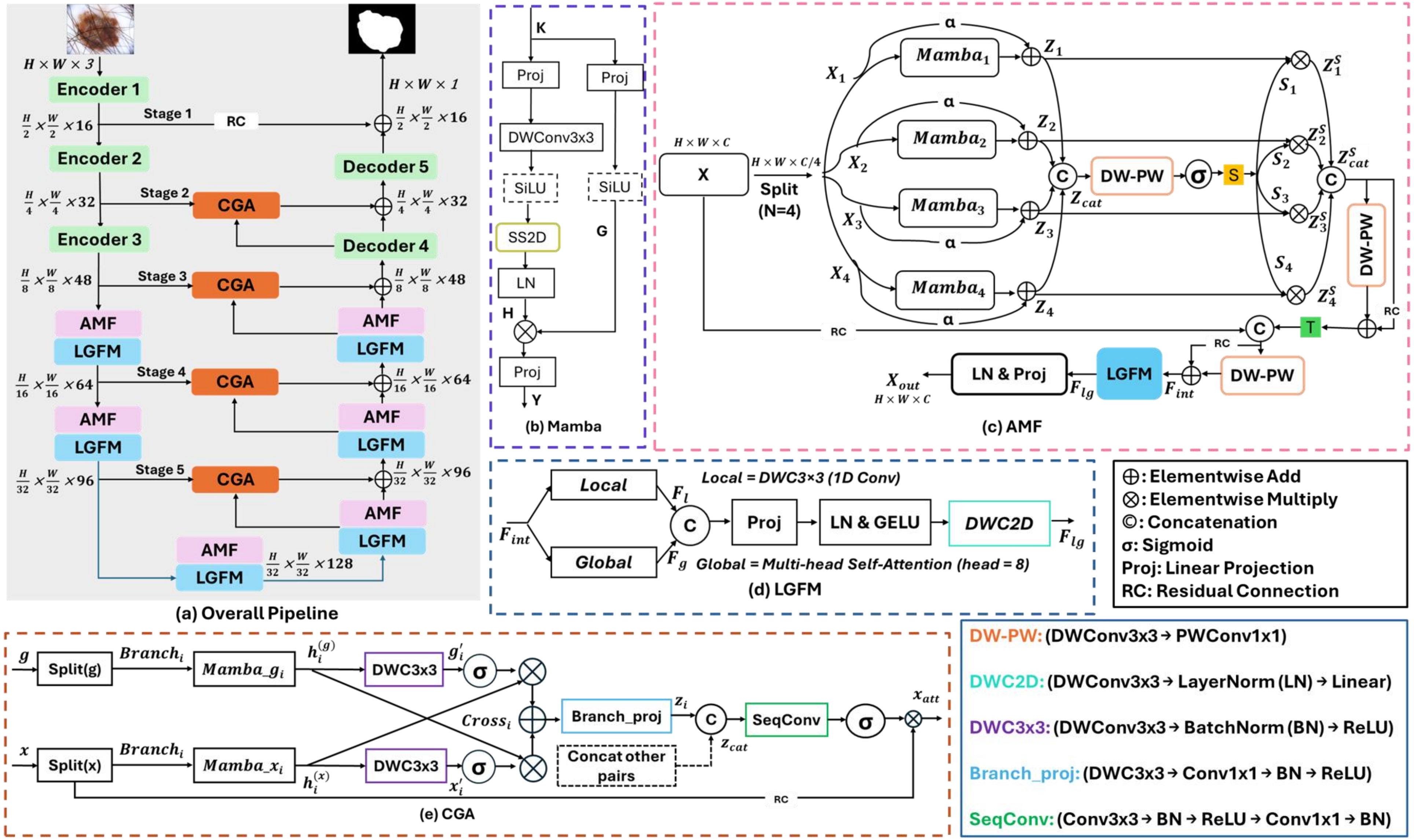
Proposed Solution: What We Introduce?

1. **Lightweight Design:**
   - Compact Mamba-based framework for robust skin lesion segmentation.

2. **Three Core Modules:**
   - **Adaptive Multi-Branch Mamba Feature Fusion (AMF):** Adaptively fuses informative Mamba branches to capture diverse lesion patterns.
   - **Local-Global Feature Mixing (LGFM):** Combines local texture with global context for stronger lesion representation.
   - **Cross-Gated Attention (CGA):** Filters skip connections to suppress background noise and sharpen lesion boundaries.

3. **Strong Generalization and Accuracy–Efficiency Trade-off:**
   - Outperforms SOTA models on ISIC2017, ISIC2018, HAM10000, and PH2 datasets.
   - Achieves 93.09% avg. Dice and 87.12% avg. IoU with only 0.494M Params and 0.326 GFLOPs.
   - Achieves robustness/generalization across benchmarks.

Method Overview: How MambaLiteUNet Works?



Quantitative Results: Accuracy-Efficiency Balance

Cls.	Model	Complexity		ISIC2017			ISIC2018			HAM10000			PH2			Avg.		
		Params↓	GFLOPs↓	IoU↑	DSC↑	HD95↓	IoU↑	DSC↑	HD95↓	IoU↑	DSC↑	HD95↓	IoU↑	DSC↑	HD95↓	IoU↑	DSC↑	HD95↓
C	U-Net [26]	7.773	13.758	79.55	88.61	16.48	74.64	85.48	19.67	83.07	90.75	15.35	80.33	89.09	18.40	79.40	88.48	17.48
C	SCR-Net [35]	0.801	1.567	78.57	88.00	17.06	79.27	88.44	15.82	85.86	92.39	13.90	74.10	85.13	22.80	79.45	88.49	17.39
T	TransFuse [42]	26.270	11.530	80.17	89.00	15.04	78.75	88.11	16.77	84.59	91.65	14.76	83.54	91.03	16.10	81.76	89.95	15.67
T	UTNetV2 [11]	12.800	15.500	78.35	87.86	17.22	77.46	87.30	17.23	75.82	86.25	18.50	75.21	85.85	22.00	76.71	86.81	18.74
T	ASwin U-Net [1]	46.910	14.181	78.37	87.87	15.84	74.62	85.46	19.79	81.96	90.09	16.17	81.41	89.75	18.10	79.09	88.29	17.48
C	C^2SDG [15]	22.001	7.972	80.73	89.34	14.30	80.00	88.88	15.21	81.34	89.71	16.00	82.23	90.25	17.30	81.08	89.55	15.70
C	UNeXt-S [33]	0.302	0.103	80.91	89.45	14.30	80.29	89.07	15.03	84.69	91.71	14.20	82.51	90.41	16.85	82.10	90.16	15.10
C	MALUNet [28]	0.175	0.083	80.37	89.11	14.66	81.03	89.52	14.72	86.18	92.58	13.70	85.99	92.47	15.10	83.39	90.92	14.54
C	EGE-UNet [29]	0.053	0.072	83.08	90.76	12.49	79.82	88.78	15.40	87.78	93.48	12.97	86.97	93.03	14.90	84.41	91.51	13.94
M	VM-UNet [27]	27.427	4.112	82.55	90.44	14.43	80.96	89.48	14.31	86.68	92.87	13.40	84.73	91.73	15.90	83.73	91.13	14.51
M	VM-UNet2 [41]	22.771	4.400	82.38	90.34	14.06	80.65	89.29	14.77	87.79	93.50	12.84	85.30	92.07	15.30	84.03	91.30	14.24
M	LightM-UNet [17]	0.403	0.391	81.49	89.80	13.80	80.16	88.99	15.10	88.56	93.93	12.52	85.38	92.11	15.40	83.90	91.21	14.21
C	LB-UNet [39]	**0.038**	0.098	82.40	90.35	12.05	81.22	89.64	14.61	89.33	94.36	9.72	87.12	93.12	14.70	85.02	91.87	12.77
M	ULVM-UNet [36]	0.049	**0.060**	83.05	90.74	12.93	80.64	89.29	15.06	88.78	94.06	12.23	87.10	93.10	12.40	84.89	91.80	13.16
M	**Ours**	0.494	0.326	**85.55**	**92.21**	**10.73**	**83.60**	**91.07**	**12.94**	**90.77**	**95.16**	**8.65**	**88.54**	**93.92**	**9.88**	**87.11**	**93.09**	**10.55**

Overall Performance

Avg. Performance:

- IoU = 87.12%, DSC = 93.09%, HD95 = 10.55 pixels

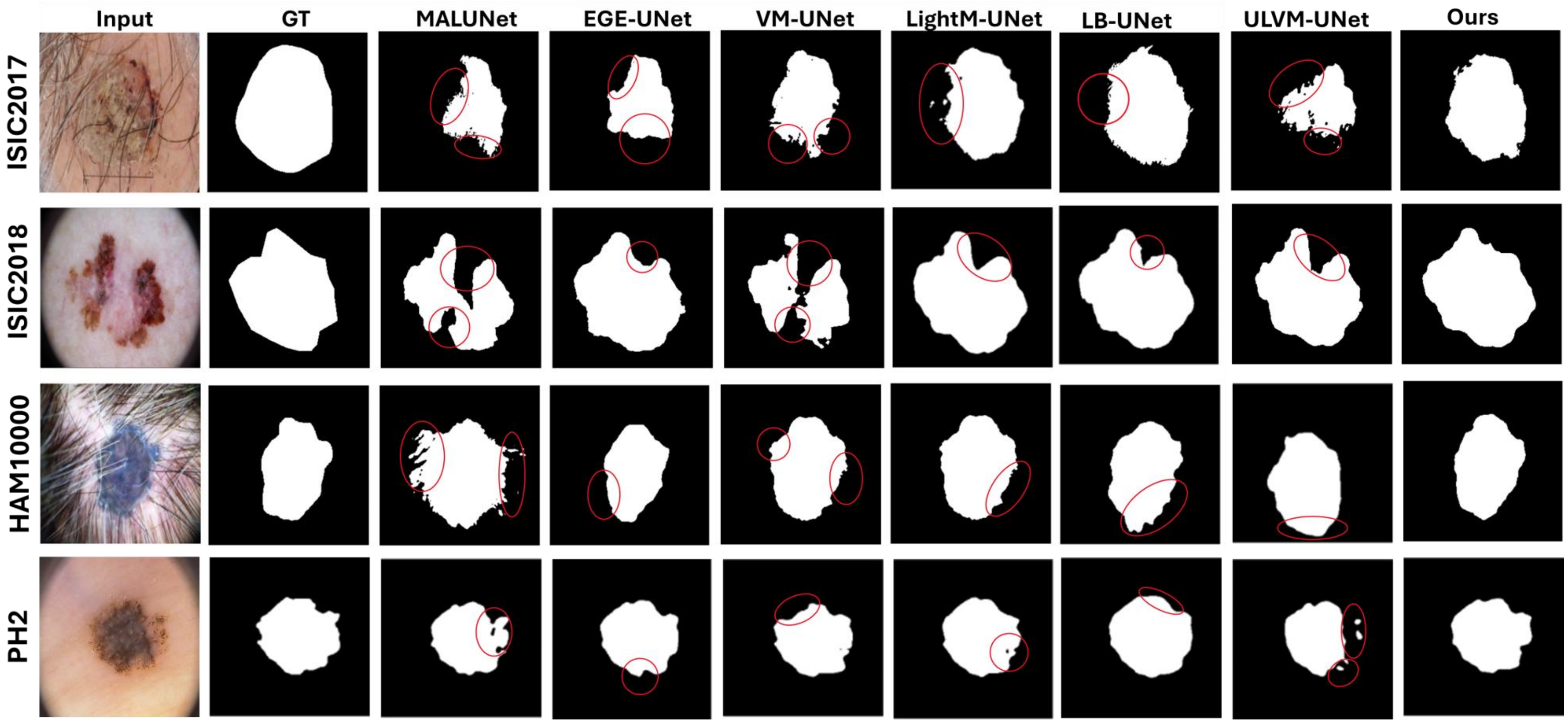
Model Complexity

- 0.494M params & 0.326 GFLOPs

- (↑) indicates higher is better, while (↓) indicates lower is better.
- **Bold** values indicate the best results.
- Our results are averaged over five independent runs.

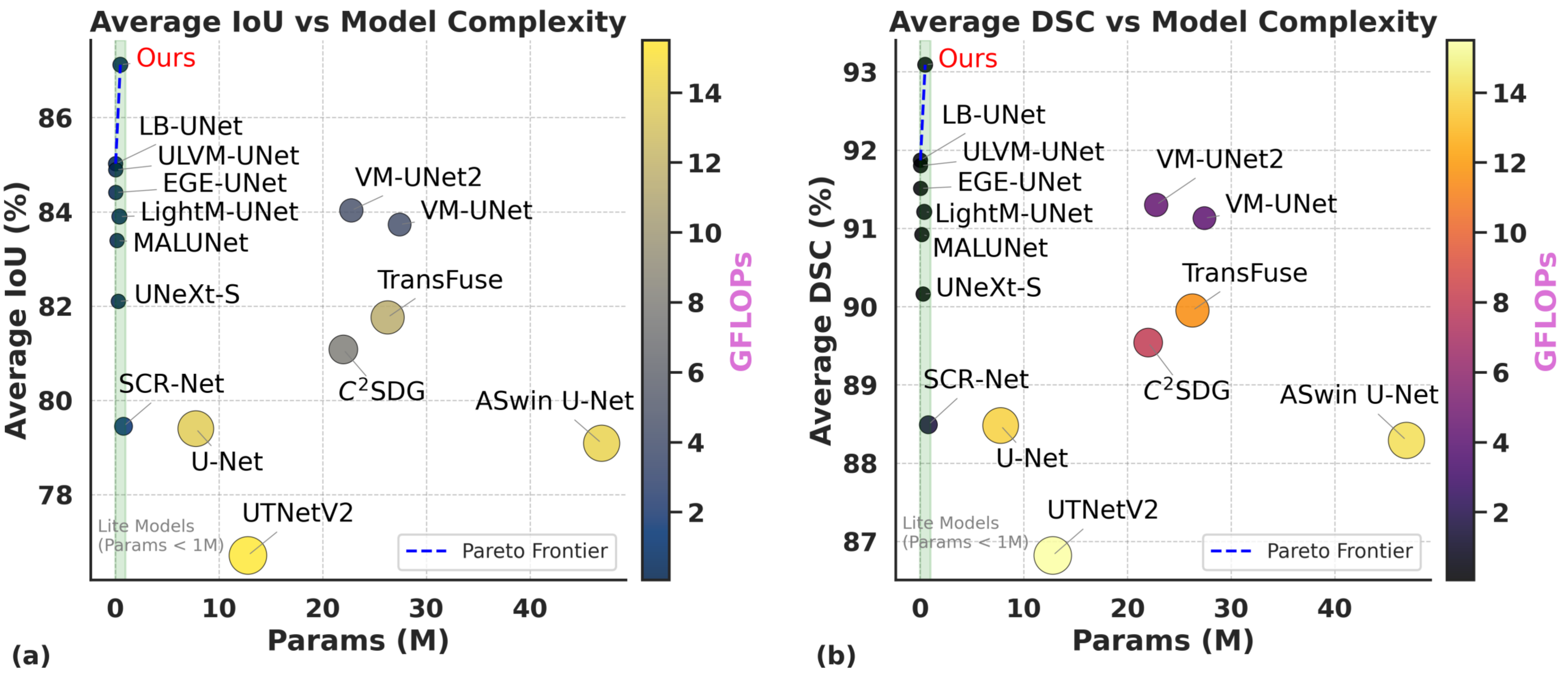
C: CNN, **T:** Transformer, **M**: Mamba

Qualitative Results: Cleaner Lesion Boundaries



- Qualitative comparison with SOTA lightweight models across datasets.
- Red boundaries indicate segmentation errors in other methods.
- Our MambaLiteUNet produces better boundaries and reduces false positives and false negatives.

Complexity-Performance Trade-off



- In the ranking plot, MambaLiteUNet consistently appears in the **top-left corner**, the optimal region of high accuracy and low model complexity.

Ablation Study: What Matters Most?

- **Effect of different core modules:**

Components				ISIC2017		ISIC2018	
Mamba	AMF	LGFM	CGA	IoU↑	DSC↑	IoU↑	DSC↑
–	✓	–	–	83.80	91.18	81.96	90.09
–	–	✓	–	82.82	90.60	81.20	89.62
–	–	–	✓	83.70	91.13	82.17	90.21
–	✓	✓	–	83.15	90.80	82.02	90.12
–	✓	–	✓	84.03	91.32	82.48	90.40
–	–	✓	✓	83.94	91.27	82.50	90.41
–	✓	✓	✓	84.26	91.46	82.66	90.51
✓	–	–	–	82.45	90.38	80.59	89.25
✓	✓	–	–	84.35	91.51	82.57	90.45
✓	–	✓	–	84.88	91.82	82.25	90.26
✓	–	–	✓	84.68	91.71	82.61	90.48
✓	✓	✓	–	85.23	92.03	82.90	90.65
✓	✓	–	✓	85.22	92.02	83.28	90.87
✓	–	✓	✓	85.21	92.02	83.07	90.75
✓	✓	✓	✓	**85.55**	**92.21**	**83.60**	**91.07**

- **Effect of channel configurations and input sizes:**

Setting	Channels / Input	Complexity		ISIC2018	
		Params↓	GFLOPs↓	IoU↑	DSC↑
Effect of Channel Configuration					
C1	{8, 16, 24, 32, 48, 64}	**0.131**	**0.092**	82.10	90.17
C2	{8, 16, 32, 48, 64, 96}	0.242	0.118	82.88	90.64
C3	{16, 32, 48, 64, 96, 128}	0.494	0.326	**83.60**	**91.07**
C4	{8, 16, 32, 64, 128, 256}	0.878	0.168	81.31	89.69
C5	{16, 32, 64, 128, 256, 512}	3.439	0.622	82.52	90.42
Effect of Input Size using C3					
224×224	C3	0.494	**0.250**	83.08	90.76
256×256	C3	0.494	0.326	**83.60**	**91.07**
288×288	C3	0.494	0.413	81.93	90.07
320×320	C3	0.494	0.510	82.42	90.36
512×512	C3	0.494	1.305	83.33	90.91

Note: Extensive ablation studies across core modules, channel settings, input sizes, branch numbers, and loss functions show that the **best performance** is achieved when **Mamba** is **combined** with our **three core modules**, using **four branches**, the **C3** channel setting, a **256×256** input size, and the **joint loss** formulation.

- **Effect of branch number:**

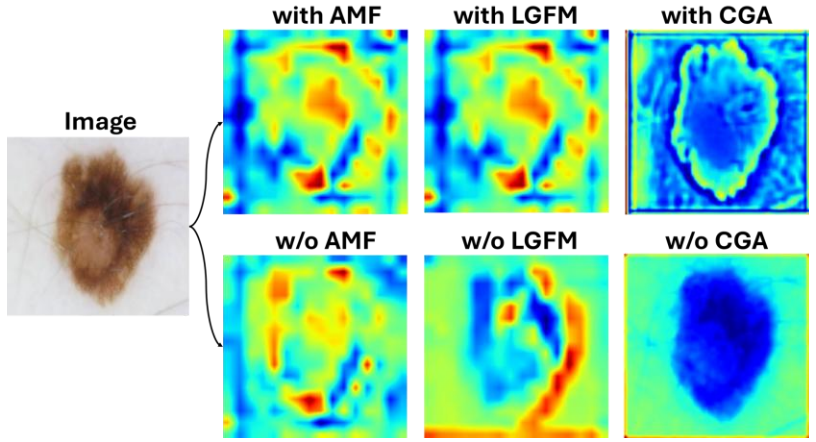
Branch	Params (M)↓	GFLOPs↓	IoU↑	DSC↑	AC↑	SP↑	SE↑
1	**0.381**	**0.270**	81.74	89.95	95.60	97.29	**89.62**
2	0.418	0.289	82.50	90.41	95.82	97.57	**89.62**
4	0.494	0.326	**83.60**	**91.07**	**96.13**	**97.95**	89.69
8	0.646	0.402	82.38	90.34	95.79	97.55	89.57
16	0.949	0.553	81.24	89.65	95.53	97.64	88.04

- **Effect of different losses:**

Loss		ISIC2017		ISIC2018		HAM10000	
BCE	Dice	IoU↑	DSC↑	IoU↑	DSC↑	IoU↑	DSC↑
✓	–	84.88	91.82	82.51	90.42	90.23	94.87
–	✓	85.15	91.98	81.12	89.58	90.55	95.04
✓	✓	**85.55**	**92.21**	**83.60**	**91.07**	**90.77**	**95.16**

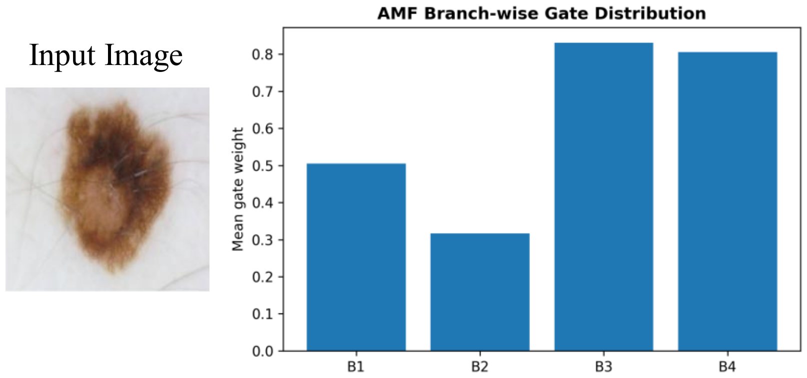
Feature Visualizations: What the Model Learns?

- **Module-wise visualization:**



- The top row shows feature maps when each module is enabled, while the bottom row shows maps without the module.

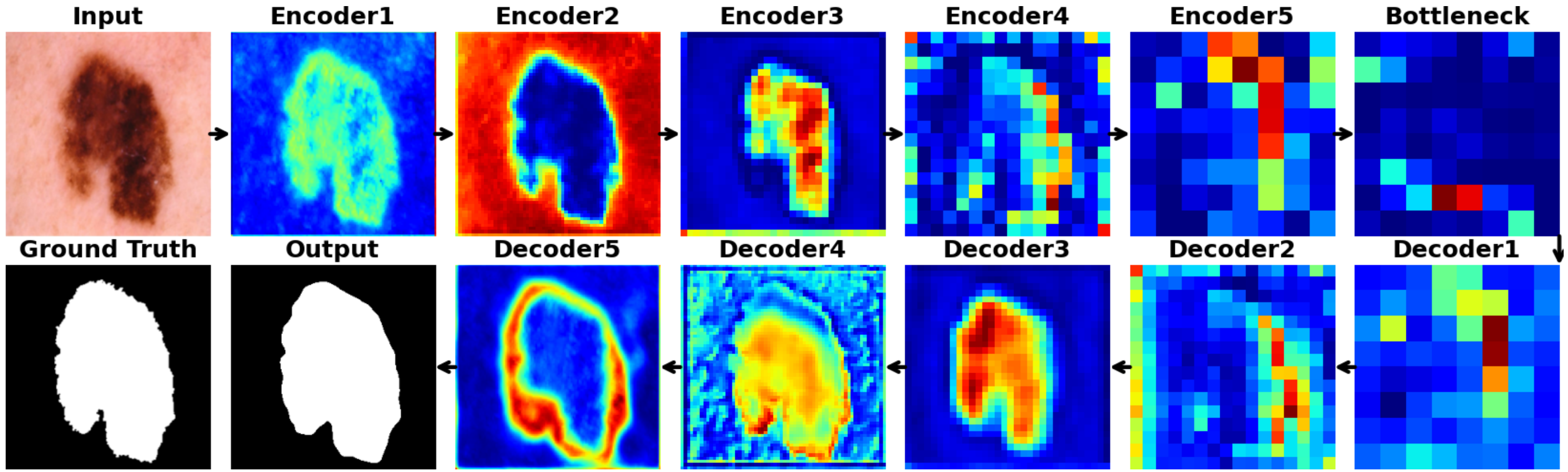
- **Branch-wise weight learning:**



- **Stage-wise visualization:**

Top row: Features become progressively abstract while highlighting lesion boundaries.

Bottom row: Decoding flows from Decoder1 to Decoder5, where early blocks recover coarse structure and later blocks sharpen lesion contours.



Generalization Ability & Robustness

- **Cross-dataset analysis:**

Model	Train on ISIC2018 → Test on PH2		
	IoU↑	DSC↑	HD95↓
U-Net [26]	77.02	87.02	22.95
SCR-Net [35]	78.93	88.23	19.54
ASwin U-Net [1]	75.01	85.72	21.76
TransFuse [42]	80.56	89.23	18.70
UTNetV2 [11]	79.94	88.85	18.82
C^2SDG [15]	79.83	88.79	21.53
UNeXt-S [33]	80.70	89.32	18.42
MALUNet [28]	79.87	88.81	19.62
EGE-UNet [29]	81.11	89.57	17.36
VM-UNet [27]	80.75	89.35	17.12
VM-UNet2 [41]	80.94	89.47	17.76
LightM-UNet [17]	81.10	89.56	16.63
LB-UNet [39]	81.17	89.61	17.38
ULVM-UNet [36]	81.35	89.72	17.07
Ours	**81.71**	**89.93**	**15.58**

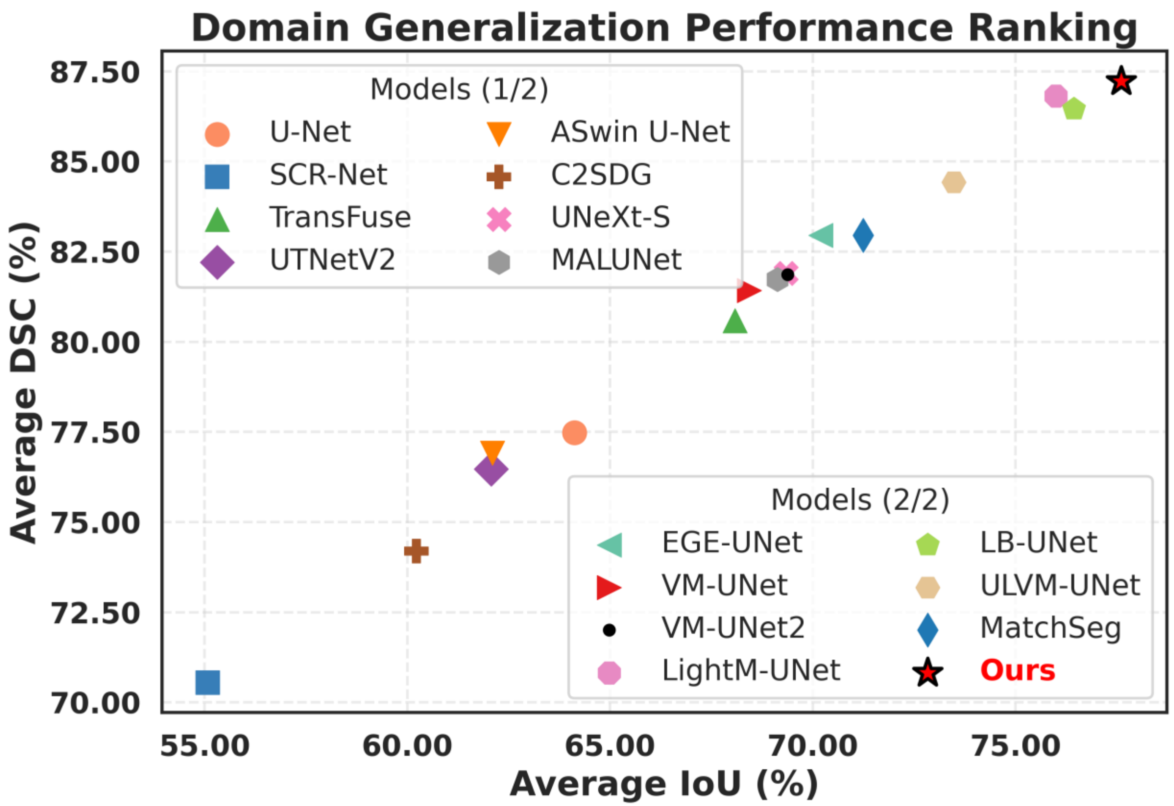
- **Non-dermoscopic datasets:**

Model	BUS (Ultrasound) [2]			GlaS (Histopathology) [31]		
	IoU↑	DSC↑	HD95↓	IoU↑	DSC↑	HD95↓
U-Net [26]	67.03	80.26	22.72	72.69	84.19	25.30
TransFuse [42]	70.16	82.46	18.46	73.49	84.72	25.49
UTNetV2 [11]	68.63	81.40	25.15	67.67	80.72	25.12
C^2SDG [15]	73.11	84.47	13.32	75.49	86.04	24.28
UNeXt-S [33]	72.11	83.80	15.88	74.41	85.33	25.17
MALUNet [28]	67.19	80.37	22.75	74.64	85.48	24.17
EGE-UNet [29]	65.81	79.38	19.29	71.25	83.21	25.06
VM-UNet [27]	72.02	83.74	15.29	72.64	84.15	24.94
LightM-UNet [17]	71.44	83.34	15.37	69.40	81.94	28.06
LB-UNet [39]	63.75	77.86	14.49	71.30	83.24	24.66
ULVM-UNet [36]	70.19	82.49	15.23	73.33	84.61	26.66
Ours	**77.68**	**87.44**	**11.55**	**78.63**	**88.04**	**21.62**

Cross-Dataset & Cross-Modality Generalization

Generalization Ability & Robustness

- **Domain generalization on HAM10000**



- Models are **trained only on NV (1 class from HAM10000)** and tested on **six unseen lesion types**: AKIEC, BCC, BKL, DF, MEL, and VASC. Ranking uses avg. IoU and DSC.

Domain Generalization on Unseen Lesion Types

Inference Time, Memory Usage & Data Efficiency

- **Inference time & memory usage:**

Model	Sec/Image ↓	Memory (MB) ↓
VM-UNet [27]	0.1718	582.5
VM-UNet2 [41]	0.1836	613.7
LightM-UNet [17]	0.0194	63.6
ULVM-UNet [36]	**0.0058**	**17.4**
Ours	0.0167	54.5

- Latency (Sec/Image) and peak GPU memory (MB) at 256×256 with batch size 1 (GPU - RTX 3090 Ti)

- **Data efficiency:**

Training Data Size	ISIC2017			ISIC2018		
	mIoU↑	DSC↑	HD95↓	mIoU↑	DSC↑	HD95↓
50%	83.30	90.89	13.24	81.49	89.80	14.99
70%	84.14	91.39	12.06	82.23	90.25	13.61
100%	**85.55**	**92.21**	**10.73**	**83.60**	**91.07**	**12.94**

- Results are obtained by randomly subsampling {50, 70, 100}% of the training split, while keeping the test set unchanged.

Key Takeaways

- SOTA segmentation with only 0.494M parameters and 0.326 GFLOPs.
- Robust generalization ability across datasets and modalities.
- Fast, lightweight design with strong deployment potential.
- Our code and pretrained models are publicly available.

Paper

Code

Thank you for your time.