Master's Thesis Presentation for the 144th Dissertation Evaluation

A Study on Storage-Efficient Segmentation Network for On-Device Skincare Device with Application to Skin Lesion

온디바이스 스킨케어 기기용 저장공간에 효율적인 분할 네트워크와 피부 병변 세그멘테이션 응용에 관한 연구

Jeongwon Lee

jeongwonlee0204@gmail.com

15th October, 2025

Contents

Introduction

Skincare Market

- The global beauty industry, comprising skincare, haircare, color cosmetics, fragrance, and toiletries, was led by skincare with a dominant 41% share in 2022 [1], expanding steadily from about USD 142 billion in 2024 to USD 150 billion in 2025 [2].
- In 2021, 6.64 million cases of melanoma and non-melanoma skin cancer occurred worldwide, largely driven by population aging and rising ultraviolet exposure, with growing demand for effective treatment and an expanding dermatology services market [3].

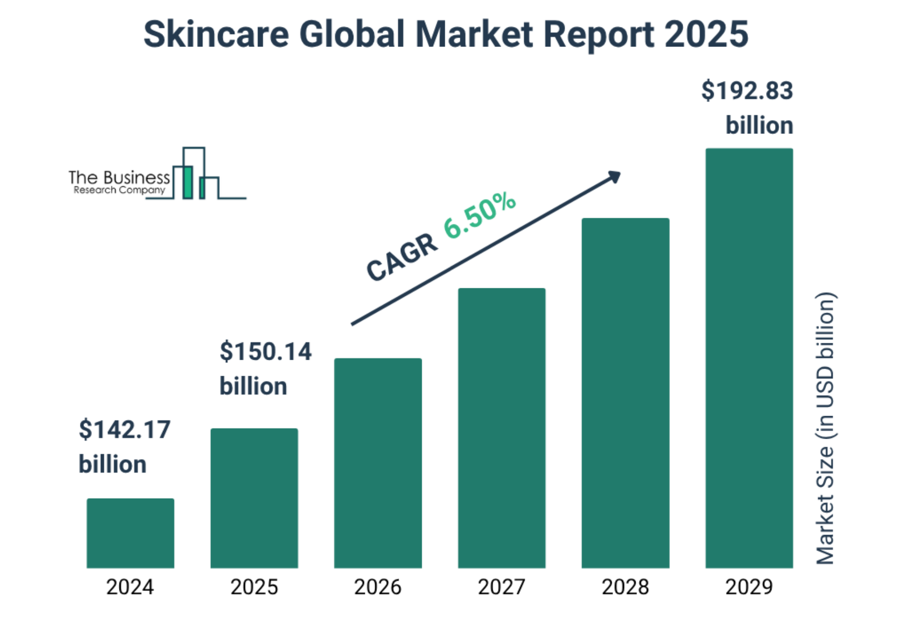

Figure 1. Trends and outlook for the skincare market size [2]

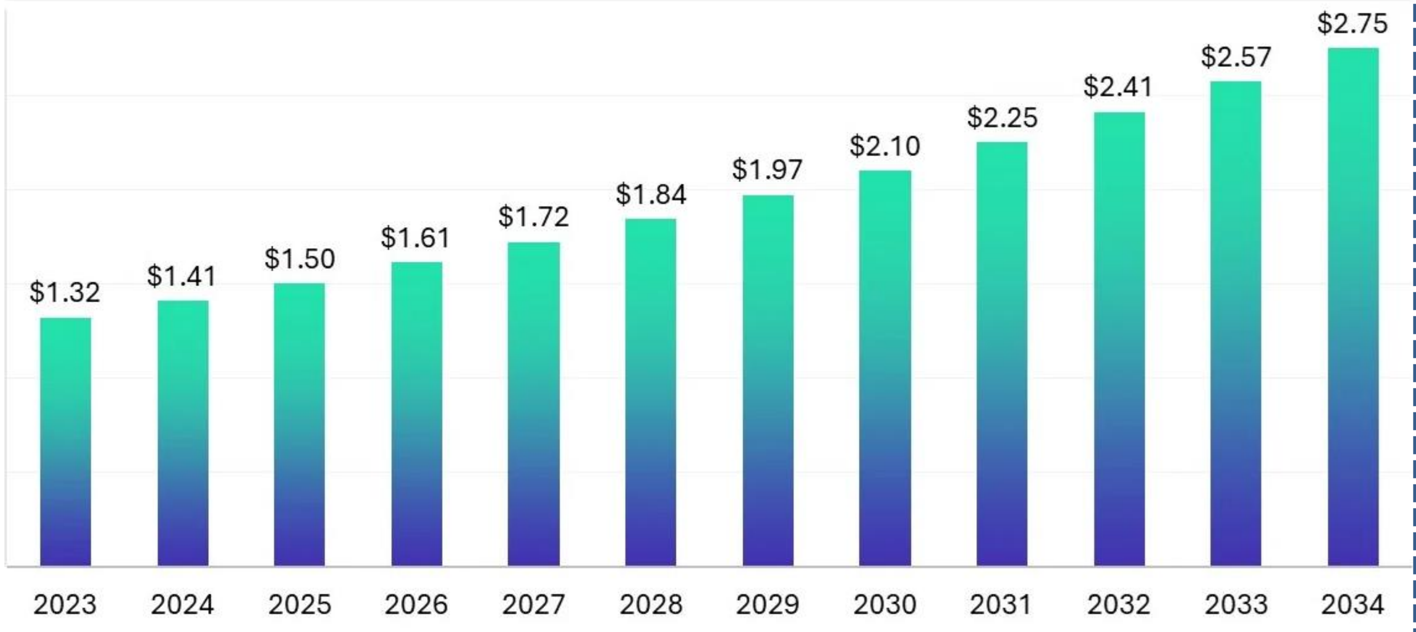

Figure 2. Dermatology market size 2023 to 2034 [4]

Introduction

Segmentation as the Fundamental Task in the Skincare Pipeline

- In skincare and dermatology, pixel-wise lesion segmentation serves as the fundamental task in a pipeline comprising image quality control, region detection, attribute tagging, and risk scoring, producing the reference spatial mask used by subsequent algorithms [5].
- Once the segmentation completes quickly on-device, multiple downstream applications become feasible, including feature extraction for dermoscopic criteria, lesion change tracking, and risk assessment, with automated skin lesion diagnosis being one of them [6].

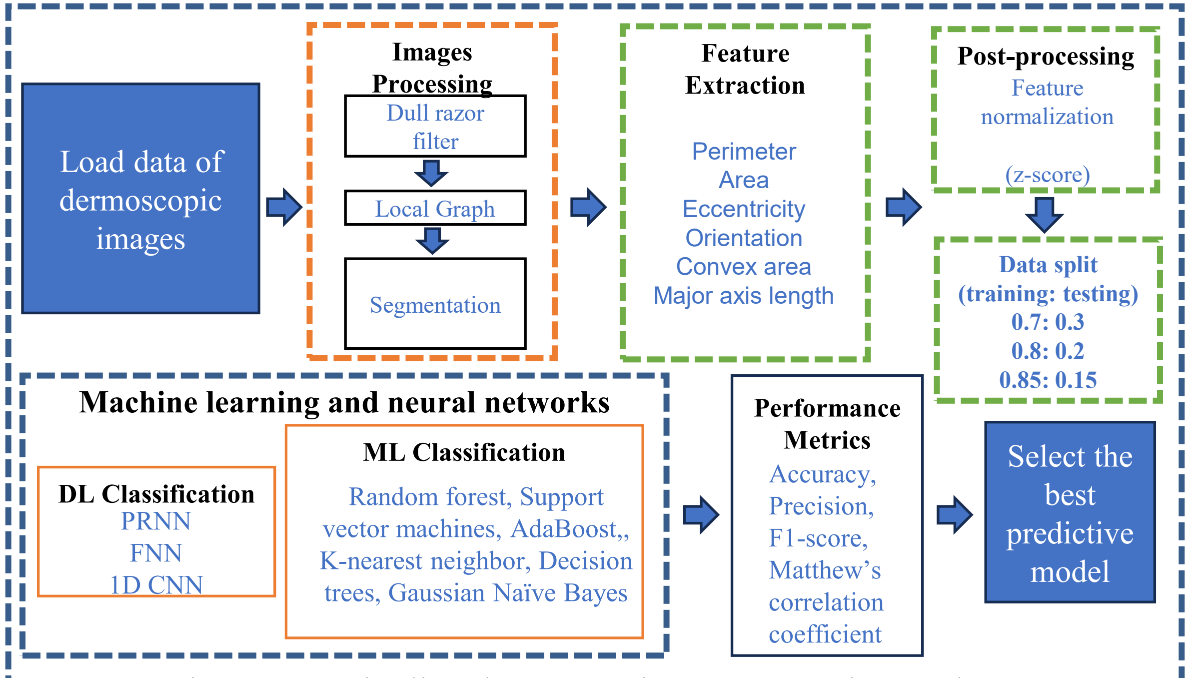

Figure 3. A pipeline demonstrating segmentation as the core operation for downstream tasks [7]

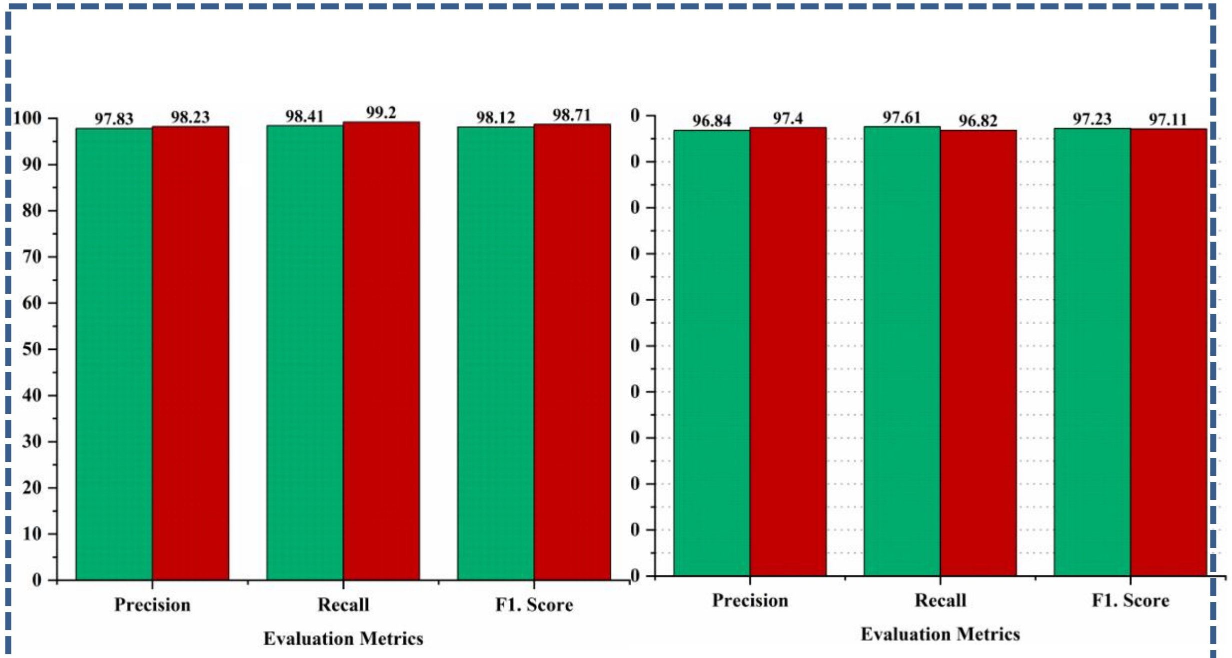

Figure 4. Performance analysis with segmentation and without segmentation process [8]

Introduction

On-Device Constraints

- Skincare devices are generally not equipped with GPUs, and in such cases, computation speed depends on CPU cache efficiency, as limited cache capacity often becomes the bottleneck for running convolutional and memory-intensive segmentation tasks.
- By reducing the model size, the working set of each operation can more effectively reside within the cache, which significantly decreases the frequency of DRAM accesses and enables more efficient and stable inference without latency degradation.

Figure 5. Example of a skin analyzer machine

Item	Specification
Device name	Himirror
CPU (Quad-core)	MediaTek MT8127A
Cache	**128KB (per core : 32KB)**
Motherboard	BM618 mainboard
RAM	DDR3 2 GB
Storage	eMMC 8 GB
Camera pixel	HD-class
Wi-Fi	2.4 GHz 802.11 b/g/n
Power	AC 100-240 V

Table 1. Example of skin analyzer specification (Himirror)

Introduction

[9] Y. Wang, J. Su, Q. Xu and Y. Zhong. "A collaborative learning model for skin lesion segmentation and classification". *Diagnostics*, 13(5):912, 2023.
[10] Y. Yuan, M. Chao and Y.C. Lo. "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance". *IEEE Transactions on Medical Imaging*, 36(9):1876-1886, 2017.

Objective and Procedure

- Skin lesion segmentation takes a dermoscopic image captured by device as input and generate a binary segmentation mask that delineates the lesion from the surrounding skin, providing precise localization and boundary definition for diagnostic analysis [9].
- Dermoscopic images are first preprocessed to remove noise and enhance contrast, then the lesion region is localized, and finally pixel-level segmentation is applied to produce a binary mask that clearly delineates the lesion boundaries for diagnostic analysis [10].

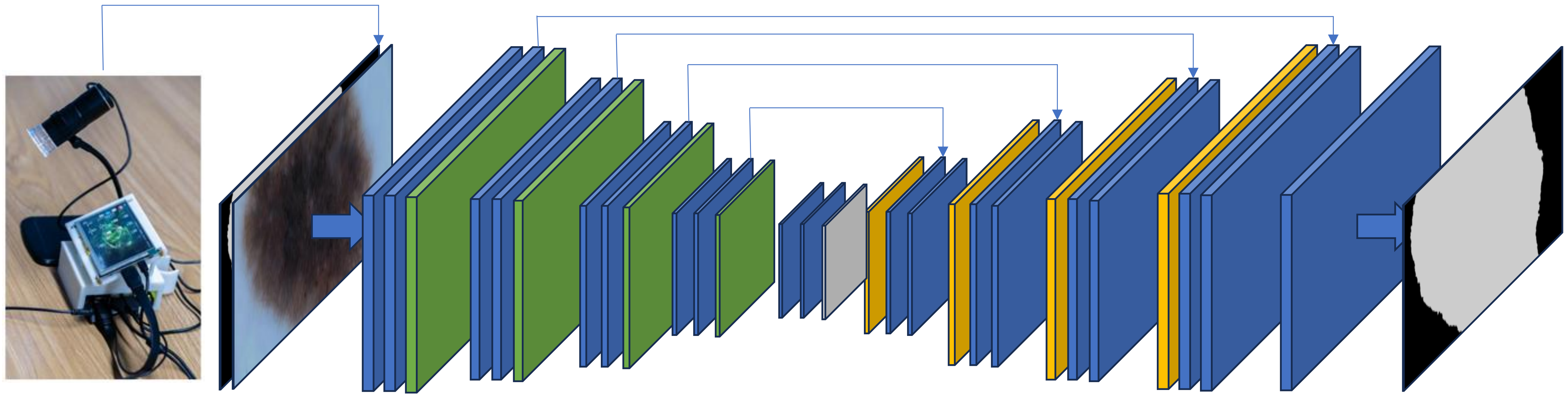

Figure 6. Skin lesion segmentation process

Introduction

Ideal Goal and Challenge

- Skincare devices operate under tight constraints on compute, memory, any on-device segmentation network must be architected with a minimal memory footprint and sufficiently low complexity to run reliably often without dedicated accelerators on such hardware.
- Making a model compact through compression, lean architectural design, often reduces accuracy the central goal is preservation of deployment ready task performance, without accuracy loss, alongside concurrent reduction of model size.

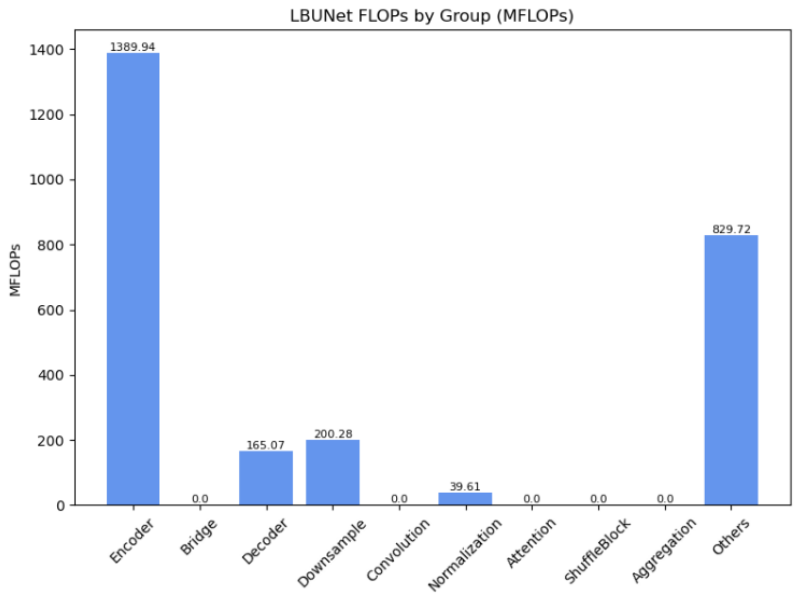

Figure 7. Model computational complexity visualization

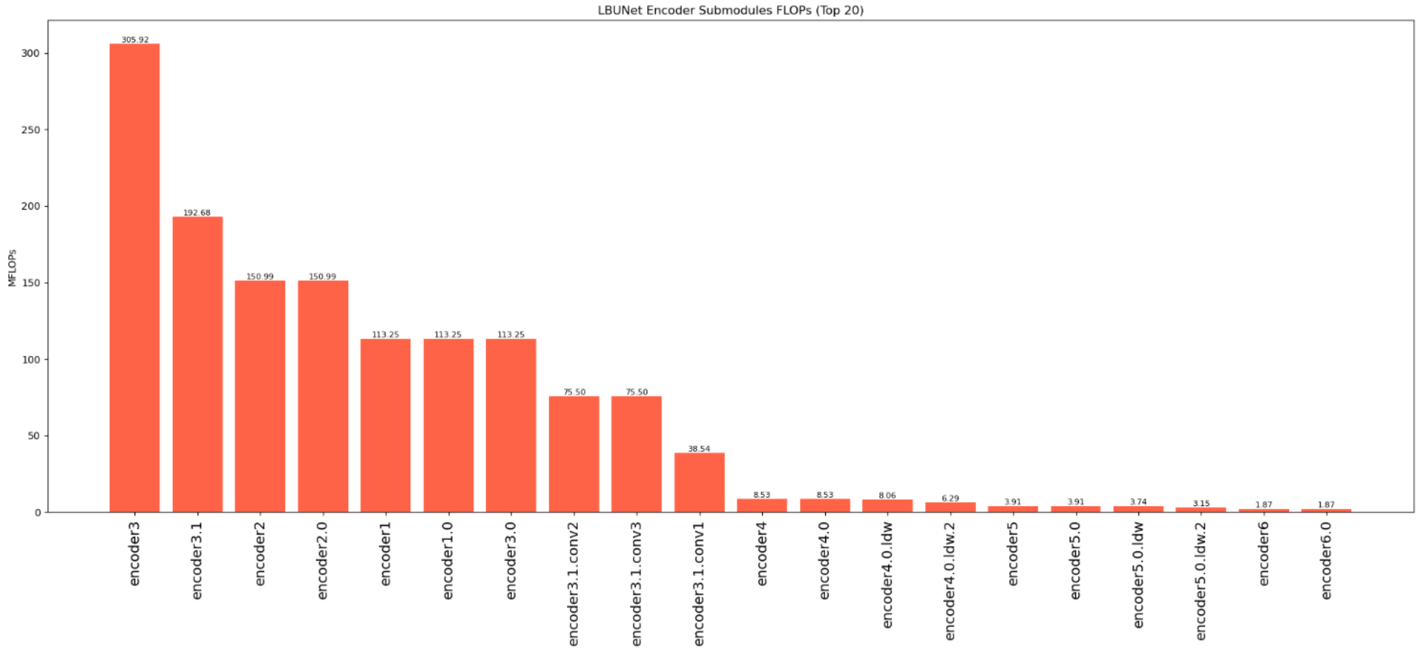

Figure 8. Model encoder submodules computational complexity visualization

Related Work

EGE-UNet (2023) [11]

- EGE-UNet is a UNet-based model enhanced with Group multi-axis Hadamard Product Attention (GHPA) for efficient multi-perspective feature extraction and a Group Aggregation Bridge (GAB) to fuse multi-scale information at different decoding stages.

- Per stage, GHPA computes group axis masks with projections, GAB performs 1×1 fusion on concatenated skips, and an auxiliary head adds classifier weights, resulting in stage-local weight growth that aggregates across scales and enlarges model size.

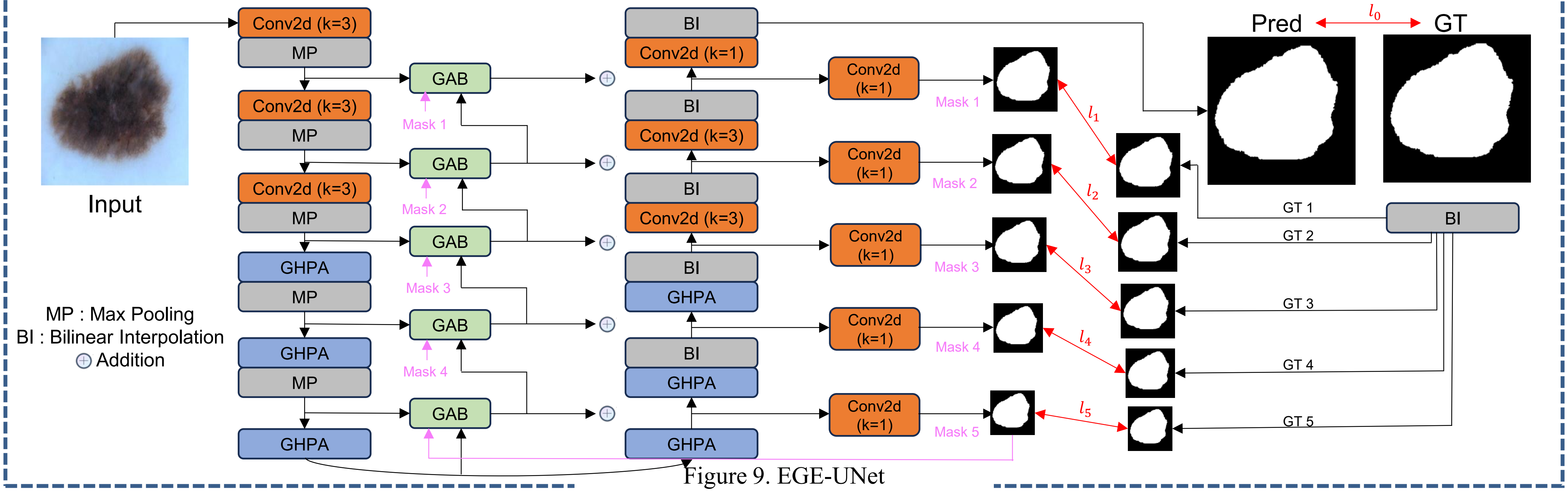

Figure 9. EGE-UNet

[11] J. Ruan, M. Xie, J. Gao, T. Liu and Y. Fu. "EGE-UNet: An Efficient Group Enhanced UNet for Skin Lesion Segmentation". In Proceedings of the *International conference on medical image computing and computer-assisted intervention*. Oct. 2023, pp. 481-490.

Related Work

UltraLight VM-UNet (2025) [12]

- UltraLight VM-UNet is a UNet-based architecture that reduces parameters and computational cost by integrating a Parallel Vision Mamba (PVM) Layer, which employs four parallel Mamba modules processing features with reduced channel dimensions.
- Within the PVM layer, four parallel branches each keep independent weights, and SAB/CAB gating with per-branch convolutions introduces duplicated parameters, which enlarges the model size and fragments the weight layout, reducing cache locality.

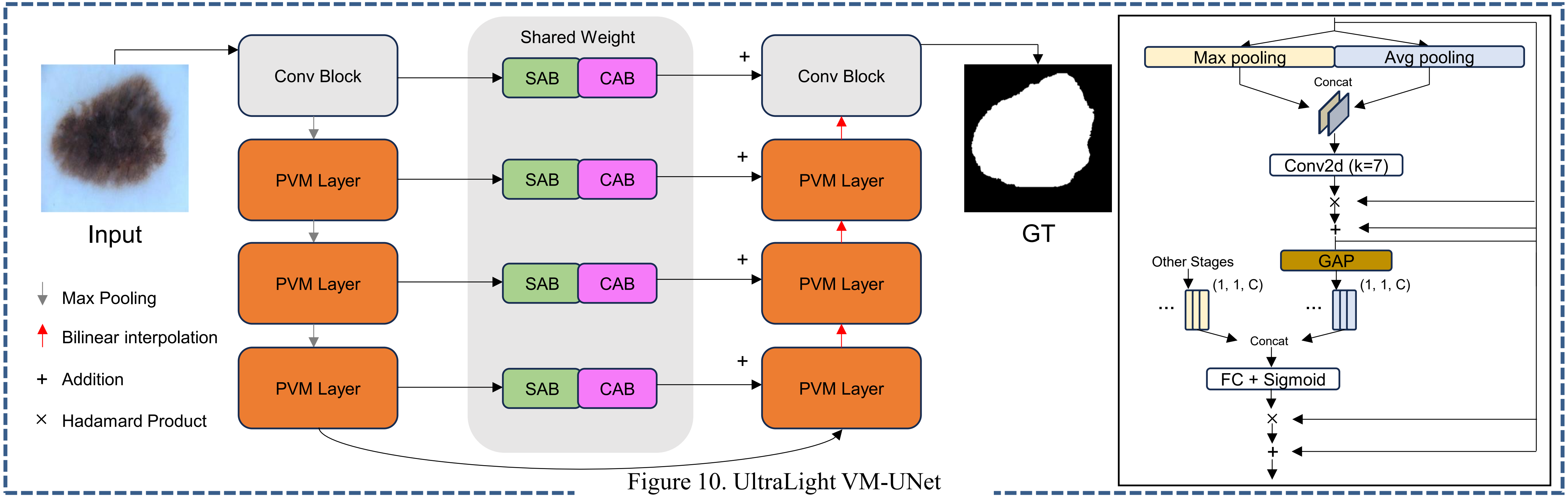

Figure 10. UltraLight VM-UNet

[12] R. Wu, Y. Liu, G. Ning, P. Liang and Q. Chang. "UltraLight VM-UNet: Parallel Vision Mamba significantly reduces parameters for skin lesion segmentation". *Patterns.* 101298, June, 2025.

Related Work

UCM-Net (2024) [13]

- UCM-Net employs a hybrid architecture combining Multi-Layer Perceptrons (MLP) and Convolutional Neural Networks (CNN) within a UNet framework, utilizing UCM-Net blocks to enhance feature extraction and reduce computational complexity.
- UCM blocks rely on patch-embedding MLP projections and normalizations across block and deblock stages, introducing dense, fully connected weights that increase model size and parameter density compared with lightweight depthwise designs.

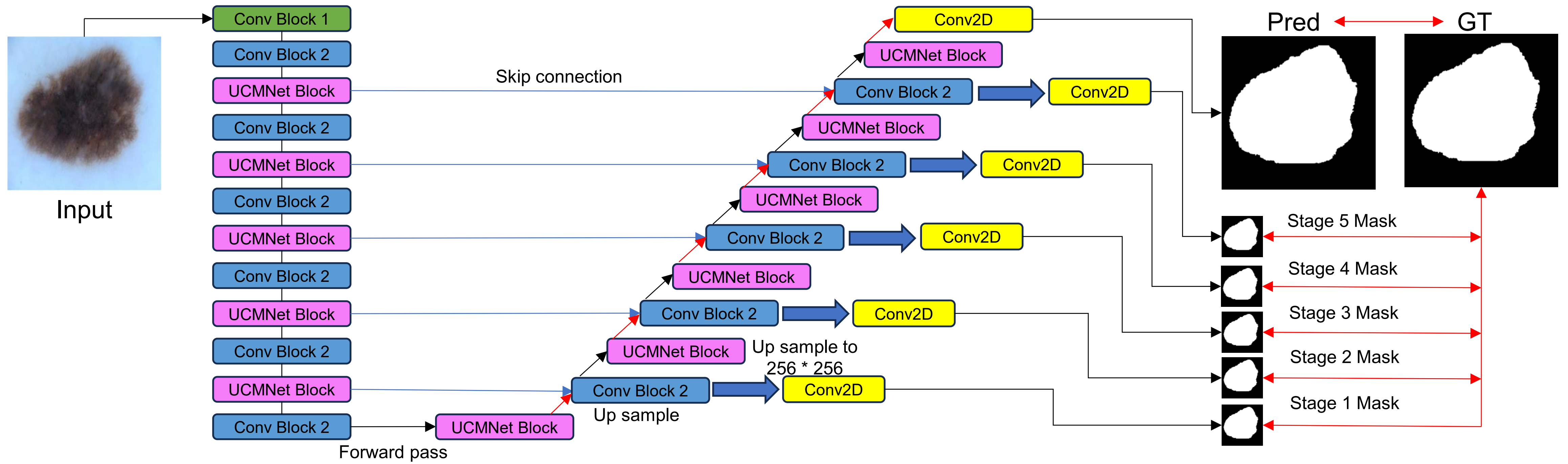

Figure 11. UCM-Net

[13] C. Yuan, D. Zhao and S.S. Agaian. " UCM-Net: A lightweight and efficient solution for skin lesion segmentation using MLP and CNN ". *Biomedical Signal Processing and Control*. 96(1):106573, 2024.

Related Work

Common Drawback of Conventional Methods

- Conventional architectures replicate and accumulate parameters layering stage-local projections and attention with redundant auxiliary heads and retaining parallel branches with independent weights across stages thereby inflating the overall model size.
- This study aims to develop a storage-efficient segmentation model that minimizes weight duplication and fragmentation arising from branches, attention, projections, and auxiliary heads, thereby consolidating parameters into a simple, consistent path across scales.

Table 2. Summary and Drawback

Method	Publication	Summary	Drawback
EGE-UNet [11]	International conference on medical image computing and computer-assisted intervention (MICCAI, 2023)	Employing group multi-axis Hadamard product attention and a group aggregation bridge	Conventional architectures replicate and accumulate parameters layering per-stage projections and attention with auxiliary heads and retaining parallel branches with independent weights across stages thereby inflating the overall model size.
UltraLight VM-UNet [12]	(Patterns, 2025)	Employing parallel Vision Mamba processing and channel-wise parameter optimization	
UCM-Net [13]	(Biomedical Signal Processing and Control, 2024)	Combining MLPs and CNNs within a UNet framework, using specialized UCM-Net blocks to enhance feature extraction	

Proposed Method (Assumed Target Task: Skin Lesion)

Notation and Definition

- Skin lesion dataset is formally defined as $D = \{(I_n, M_n)\}_{n=1}^{N}$, where $I_n \in \mathbb{R}^{H \times W \times 3}$is an RGB dermoscopic image and $M_n \in M$ is the corresponding diagnostic mask. Here N denotes the dataset size and n indexes samples (1, ..., N) to pair each image with its mask.
- Skin lesion segmentation aims to learn a mapping function F: $\mathbb{R}^{(H \times W \times 3)} \rightarrow \{0, 1\}^{(H \times W)}$ that accurately delineates lesion boundaries from dermoscopic images.

Table 3. Symbols and definitions used in the formulation of the skin lesion segmentation task

Symbol	Name	Description
D	Dataset	Overall Dataset
I_n	Skin Lesion Image	n-th Skin Lesion Input Image
M_n	Binary Mask	n-th Binary Ground Truth Mask (1=Lesion Pixel, 0=Background)
N	Number of Dataset	Number of Samples in the Dataset
F	Segmentation Function	Learned function that maps dermoscopic images to segmentation masks

Proposed Method

Procedural Overview

- Proposed model employs a Depthwise Separable Convolution-based U-Net that progressively extracts features in the encoder and restores spatial information through upsampling with simple additive skip connections in the decoder.
- The decoder restores spatial resolution through bilinear upsampling and employs simple additive skip connections that directly combine encoder features with corresponding decoder outputs, with Deep Supervision enhancing multi-scale prediction performance.

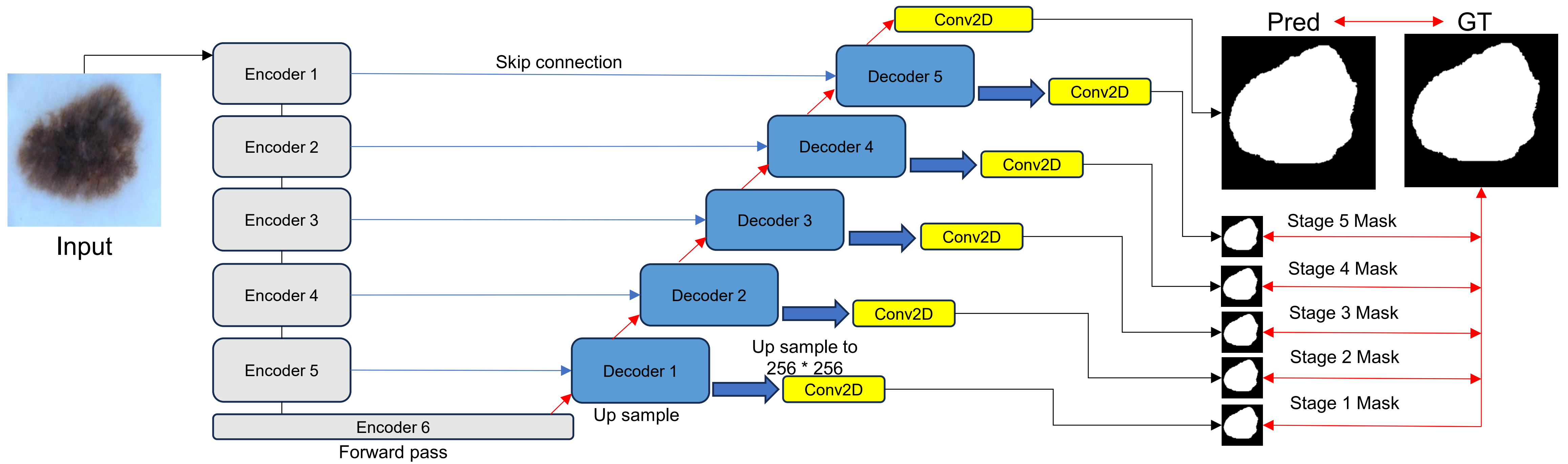

Figure 12. Proposed model architecture

Proposed Method

Details

- Each block performs feature learning through 1×1 Pointwise Conv for channel adjustment followed by 3×3 Depthwise Conv with groups equal to output channels for spatial feature extraction, combined with GroupNorm and GELU activation for training.
- PRCM computes data-dependent per-channel gates by first forming a global context via global average pooling, projecting it onto a learnable low-rank basis to obtain coefficients, and linearly fusing these into channel logits that are squashed with a sigmoid.

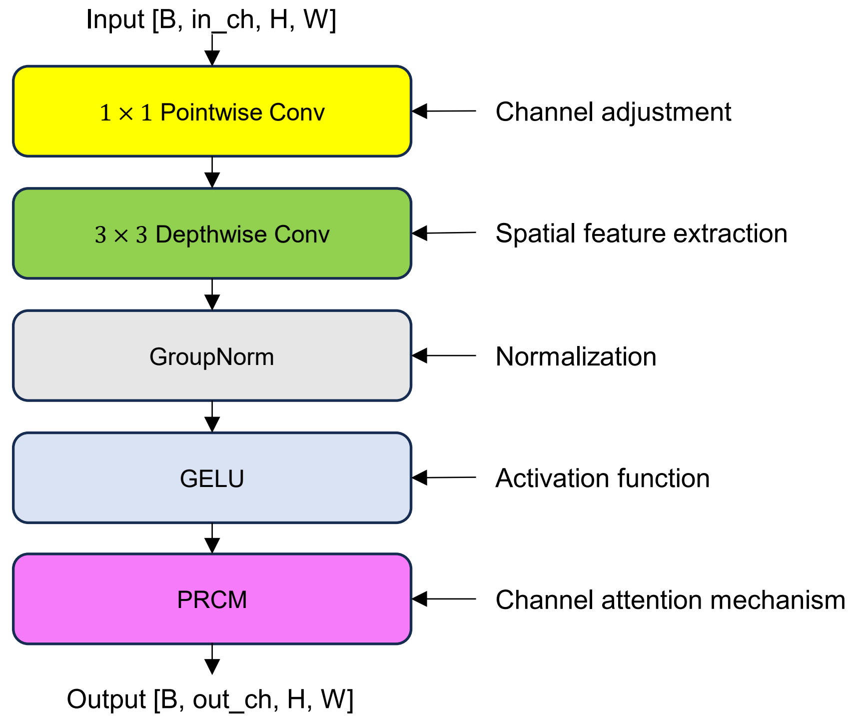

Figure 13. Encoder&Decoder architecture

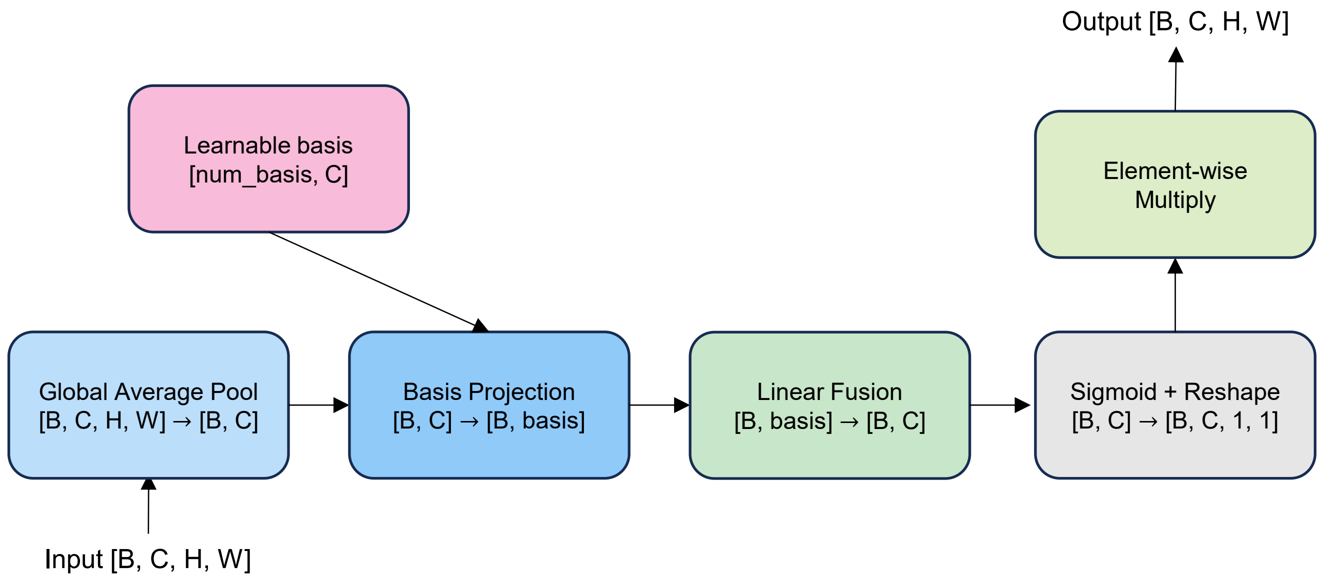

Figure 14. PRCM module

Experimental Settings

Dataset

Dataset	#Images	Image Shape
ISIC 2017	2,150	(256, 256, 3)
PH2	200	(768, 560, 3)

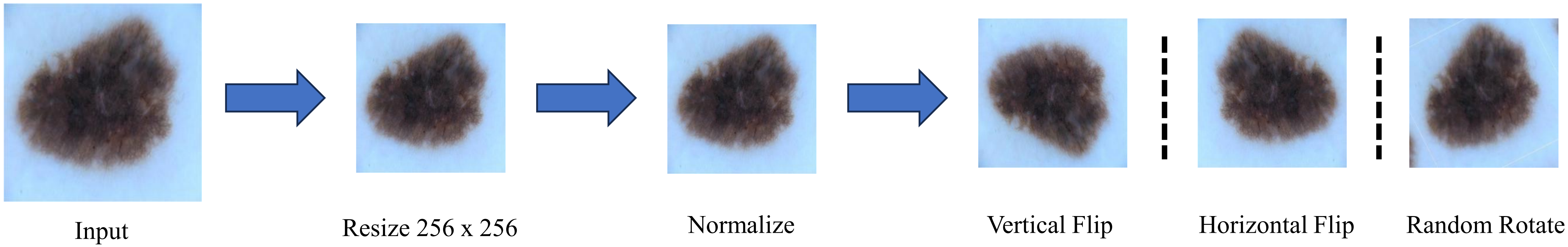

Figure 15. Data preprocessing process

Cross Validation

- Number of Iterations: 10
- Data Split Ratio (Train: Validation: Test) = 0.6 : 0.2 : 0.2

Experimental Settings

Hyperparameter Settings

Method	Batch Size	Epochs	Loss Function	Optimizer	Learning Rate
EGE-UNet UltraLight VM-UNet UCM-Net	8	300	GT_BCEDiceLoss	AdamW	0.001
Our Experiment	8	300	GT_BCEDiceLoss	AdamW	0.001

Evaluation Measures

- *Dice Coefficient*

$$Dice = \frac{2 * |A \cap B|}{|A + B|} = \frac{2 * TP}{(TP + FP) + (TP + FN)}$$

- *IoU*

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{A \cap B}{A \cup B}$$

- *Model Size*
- *Params*
- *Inference Times* (Number of iterations : 50)
- *FPS (Frame Per Second)*

TP : True Positives / TN : True Negatives / FP : False Positives / FN : False Negatives

Experimental Results

Performance Comparison

- The proposed model demonstrates superior performance on the PH2 dataset, achieving the highest mIoU score among all compared methods, and maintains consistent performance across the other ISIC 2017 datasets.
- Proposed Model achieves fewest parameters, and fastest inference among the compared methods with an 88 KB storage footprint, it is suited for deployment on resource-constrained consumer skincare and dermatology devices (Himirror Cache: 128KB).

Table 4. Comparison of Model Size, Params, mIoU, Dice, inference time, FPS on ISIC2017 and PH2

Model	Model Size	Params (K)	ISIC 2017		PH2		Inference Time (ms)	Frame Per Second
			mIoU	Dice	mIoU	Dice		
Proposed	**88KB**	**16.7**	0.8232 ± 0.0093	0.9030 ± 0.0056	**0.8740 ± 0.0070**	**0.9327 ± 0.0040**	**20.78 ± 0.98**	**48.12**
EGE-UNet	299KB	53.3	**0.8241 ± 0.0115**	**0.9035 ± 0.0069**	0.8728 ± 0.0162	0.9319 ± 0.0093	31.46 ± 1.66	31.79
UCM-Net	245KB	47.4	0.8215 ± 0.0141	0.9019 ± 0.0085	0.8723 ± 0.0165	0.9317 ± 0.0094	32.10 ± 2.14	31.15
UltraLight VM-UNet	230KB	49.4	0.8161 ± 0.0135	0.8988 ± 0.0079	0.8626 ± 0.0165	0.9261 ± 0.0096	20.96 ± 1.40	47.71

Experimental Results

In-depth Analysis

- This study fixes the base architecture and training, replaces every 3×3 layer with Standard, Depthwise, or Depthwise Separable, and measures parameter count and accuracy on ISIC2017 and PH2 to isolate the effect of the 3×3 operator on capacity and accuracy.
- Guided by these results, this study adopts depthwise as the core operator with minimal channel alignment and employs PRCM to leverage channel context, preserving efficiency while recovering accuracy through channel mixing without enlarging model size.

Table 5. Effect of convolution type on accuracy–efficiency

Model	Params (K)	ISIC2017		PH2	
		mIoU	Dice	mIoU	Dice
Standard Convolution	112.9	**0.8244 ± 0.0060**	**0.9037 ± 0.0036**	0.8620 ± 0.0035	0.9259 ± 0.0020
Depthwise Convolution	**9.4**	0.7923 ± 0.0035	0.8841 ± 0.0022	0.8305 ± 0.0228	0.9072 ± 0.0136
Depthwise Separable Convolution	21.4	0.8154 ± 0.0044	0.8983 ± 0.0026	**0.8741 ± 0.0123**	**0.9327 ± 0.0070**

Experimental Results

Ablation Study

- To verify the performance stability of the proposed model, an ablation study was conducted by progressively adding key components to the baseline architecture.
- Proposed model is larger than the 77 KB w/o PRCM variant but removing PRCM reduces accuracy below the 230 KB UltraLight VM-UNet, showing that PRCM is a core module that prevents accuracy loss and preserves stability with a small size increase.

Table 6. Ablation study comparing accuracy (mIoU, Dice) and efficiency (Model Size, Params, inference time, FPS)

Method	Model Size	Params (K)	mIoU	Dice	Inference Time (ms)	Frame Per Second
Proposed	88KB	16.7	**0.8740 ± 0.0070**	**0.9327 ± 0.0040**	20.78 ± 0.98	48.12
UltraLight VM-UNet	230KB	49.4	0.8626 ± 0.0165	0.9261 ± 0.0096	20.96 ± 1.40	47.71
w/o PRCM	**77KB**	**15.4**	**0.8583 ± 0.0129**	**0.9237 ± 0.0075**	**18.22 ± 0.78**	**54.88**

Degenerated Accuracy

Experimental Results

Qualitative Analysis

- EGE-UNet leaves a faint boundary blur that softens the lesion edges, UCM-Net produces compact masks but drops small protrusions, and UltraLight VM-UNet over-smooths lesion contours, resulting in rounded shapes and slight under-segmentation.
- Proposed model preserves shape with minimal background over segmentation and aligns with GT, achieving the best IoU/Dice in the first case and competitive results in the second, while maintaining higher boundary fidelity than UCM-Net and UltraLight VM-UNet.

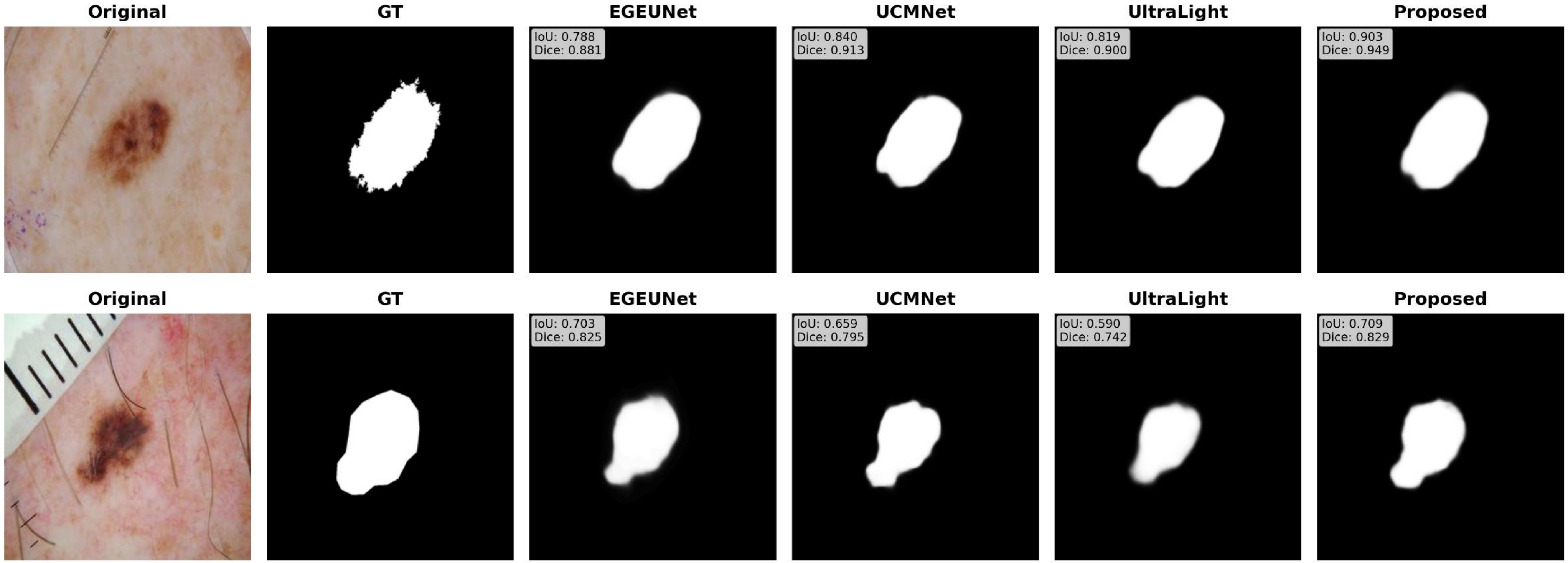

Figure 16. Qualitative comparison of predicted segmentation masks on representative skin lesion images

Conclusion and Contribution

Conclusion

- This study introduces a storage-efficient segmentation network that uses PRCM (Projection-Reconstruction Channel Modulation) to compress global context onto a small basis and modulate channels with a low-rank map, enabling cache-resident attention.
- The proposed model maintains segmentation accuracy and inference time while using about 33% of the storage required by task-matched baselines for skin lesion segmentation under identical training and evaluation settings on ISIC2017 and PH2.
- Experiments showed an advantage in model size, with the proposed network consistently remaining in the tens of kilobytes while recent comparators stayed in the hundreds of kilobytes, enabling cache-resident execution on CPU-only devices.

Contribution

- Journal
  - A Survey of Neural Network Segmentation and Validation on Plant Specimen Images of Korean Violaceae, ***Expert Systems with Applications***, 2026
  - Survey on Replay-based Continual Learning and Empirical Validation on Feasibility in Diverse Edge Devices using Representative Method, ***Mathematics***, 2025
  - EP-Rex: Evidence-Preserving Receptive-Field Expansion for Efficient Crack Segmentation, ***Symmetry***, 2025
- Conferences
  - Justice AI: Implementation and Evaluation of Legal Case Retrieval AI Using Dense Passage Retrieval, ***ICCE***, 2025
  - Proceeding Reducing Computational Overhead in Skin Lesion Segmentation using Dynamic Selective Aggregation U-Net, ***IEIE***, 2025
  - A Survey on Self-Training for Semantic Segmentation, ***IEIE***, 2024
- Others
  - Justice AI: Implementation and Evaluation of Legal Case Retrieval AI Using Dense Passage Retrieval. AI Technology-based Employment and Start-up Workshop, 2024. 1st prize

Thank you

Reference

[1] M. Dube and S. Dube. “Towards sustainable color cosmetics packaging”. *Cosmetics*, 10(5):139, 2023.

[2] The Business Research Company. “Skincare Global Market Report 2025.” The Business Research Company, Jan. 2025, https://www.thebusinessresearchcompany.com/report/skincare-global-market-report#:~:text=What%20Is%20The%20Skincare%20Market,Size%202025%20And%20Growth%20Rate

[3] L. Zhou, Y. Zhong, L. Han, Y. Xie and M. Wan. “Global, regional, and national trends in the burden of melanoma and non-melanoma skin cancer: Insights from the global burden of disease study 1990–2021”. *Scientific Reports*, 15(1): 5996, 2025.

[4] Towards Healthcare. “Dermatology Market Size, Shares, Challenges and Opportunities”, May 2025, https://www.towardshealthcare.com/insights/dermatology-market-sizing

[5] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, Y. Xu, J. Qin, S. Chen, T. Wang and S. Wang. “Skin lesion segmentation via generative adversarial networks with dual discriminators”. *Medical Image Analysis*. 64(1):101716, 2020.

[6] Z. Liu, J. Hu, X. Gong and F. Li. “Skin lesion segmentation with a multiscale input fusion U-Net incorporating Res2-SE and pyramid dilated convolution”. *Scientific Reports*. 15(1):7975, 2025.

[7] S. Moldovanu, M. Miron, C. G. Rusu, K. C. Biswas and L. Moraru. “Refining skin lesions classification performance using geometric features of superpixels”. *Scientific Reports*. 13(1): 11463, 2023.

[8] M. Mateen, S. Hayat, F. Arshad, Y. H. Gu and M. A. AI-Antari. “Hybrid Deep Learning Framework for Melanoma Diagnosis Using Dermoscopic Medical Images”. *Diagnostics*, 14(19): 2242, 2024.

[9] Y. Wang, J. Su, Q. Xu and Y. Zhong. “A collaborative learning model for skin lesion segmentation and classification”. *Diagnostics*, 13(5):912, 2023.

[10] Y. Yuan, M. Chao and Y.C. Lo. “Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance”. *IEEE Transactions on Medical Imaging*, 36(9):1876-1886, 2017.

[11] J. Ruan, M. Xie, J. Gao, T. Liu and Y. Fu. "EGE-UNet: An Efficient Group Enhanced UNet for Skin Lesion Segmentation". *In Proceedings of the International conference on medical image computing and computer-assisted intervention(MICCAI)*. Oct. 2023, pp. 481-490.

[12] R. Wu, Y. Liu, G. Ning, P. Liang and Q. Chang. "UltraLight VM-UNet: Parallel Vision Mamba significantly reduces parameters for skin lesion segmentation". *Patterns*. 101298, June, 2025.

[13] C. Yuan, D. Zhao and S.S. Agaian. " UCM-Net: A lightweight and efficient solution for skin lesion segmentation using MLP and CNN ". *Biomedical Signal Processing and Control*. 96(1):106573, 2024.