Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CGMU-Net: A Classification-Guided Cross-Interaction Network for Computer-Assisted Dermoscopy Skin Lesion Segmentation

BingYan Liu
Xiamen University of Technology
MingEn Zhong
zhongmingen@xmut.edu.cn

Xiamen University of Technology
ZhengFeng Li
Xiamen University of Technology
YiJie Tang
Xiamen University of Technology
BinGan Yuan
Tsinghua University

Research Article

Keywords: Dermoscopy, Skin Neoplasms, Image Processing, Computer-Assisted, Deep Learning

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9940905/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

CGMU-Net: A Classification-Guided Cross-Interaction Network for Computer-Assisted Dermoscopy Skin Lesion Segmentation

BingYan Liu[1], MingEn Zhong[1*], ZhengFeng Li[1], YiJie Tang[1], BinGan Yuan[2]

[1]School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024, Fujian, China.

[2]Department of Engineering Physics, Tsinghua University, Beijing 100084, Beijing, China.

*Corresponding author(s). E-mail(s): zhongmingen@xmut.edu.cn;
Contributing authors: 1255591514@qq.com; 1610103896@qq.com; 205214761@qq.com; 1257044697@qq.com;

Abstract

Accurate and effective segmentation of dermoscopic skin lesions is critical for clinical diagnosis. Despite recent advances in deep learning, challenges such as irregular shapes, blurred boundaries, and low contrast remain. To address these issues, we propose CGMU-Net, a classification-guided cross-interaction network. The model integrates a cross-interaction attention fusion module to enhance semantic–boundary feature interaction, a channel-wise feature enhancement module to improve multi-scale consistency, and a classification-guided lesion-aware module to introduce discriminative cues. Experimental results on three public datasets (ISIC 2017, ISIC 2018, and PH2) demonstrate that CGMU-Net achieves Dice scores of 89.21%, 89.91%, and 95.01%, respectively. These results indicate that the proposed method provides accurate and robust lesion segmentation.

Keywords: Dermoscopy, Skin Neoplasms, Image Processing, Computer-Assisted, Deep Learning

1 Introduction

Melanoma is one of the most aggressive forms of skin cancer, originating from melanocytes. In 2025, approximately 105,000 new invasive cases and over 107,000 in situ cases are expected in the United States, with more than 8,400 related deaths [1, 2]. Early detection is therefore essential for improving patient outcomes [3, 4]. However, manual lesion delineation is time-consuming, subjective, and prone to inter-observer variability [5, 6], highlighting the need for reliable computer-assisted analysis.

Deep learning has achieved strong performance in medical image segmentation, yet dermoscopic skin lesion segmentation remains challenging due to variability in lesion scale, indistinct boundaries, low contrast, and artifacts such as hair and small lesions [7]. Existing methods often focus on either semantic representation or boundary refinement, but lack effective mechanisms to jointly exploit semantic context, boundary information, and lesion-aware discriminative cues. In addition, category-level information is rarely incorporated into pixel-level segmentation, limiting unified modeling of lesion localization.

To address these issues, we propose CGMU-Net for dermoscopic skin lesion segmentation, aiming to generate accurate lesion masks for quantitative analysis and melanoma screening. The network emphasizes the interaction among semantic, boundary, and regional features through two key modules: a cross-interaction attention fusion module (CIAFM) for bidirectional semantic–boundary feature regulation, and a channel-wise feature enhancement and excitation module (CFEEM) for improving regional consistency via multi-scale aggregation. A classification-guided lesion-aware strategy further integrates discriminative cues from an auxiliary classification branch.

The main contributions are as follows:

1. CIAFM enables bidirectional interaction between semantic and boundary features through cross-attention.
2. CFEEM improves lesion-region completeness and continuity via channel reweighting and multi-scale aggregation.
3. Extensive experiments on ISIC 2017, ISIC 2018, and PH2 demonstrate the effectiveness and complementarity of the proposed modules.

2 Related work

2.1 Encoder–Decoder Feature Fusion for Skin Lesion Segmentation

U-Net-based encoder–decoder architectures are widely adopted in skin lesion segmentation because they effectively integrate semantic and spatial information through skip connections [8]. Subsequent studies further improved feature representation using residual connections, dense fusion strategies, and context-aware decoding structures. However, dermoscopic images often exhibit blurred boundaries, low contrast, and irregular lesion morphology. Conventional skip fusion based on direct concatenation cannot effectively model the interaction between semantic and boundary features, leading to inaccurate boundary localization and incomplete lesion regions.

2.2 Attention Mechanisms

Attention mechanisms are widely used to enhance lesion-related representations and suppress background interference. ASCU-Net introduced a triple attention mechanism that combines attention gates, spatial attention, and channel attention to simultaneously capture contextual information and spatial correlations for skin lesion segmentation [9], while SENet improved feature representation through channel-wise recalibration [10]. Transformer-based methods further strengthened global contextual modeling by introducing self-attention mechanisms. Representative models such as Swin-Unet and DAE-Former achieved competitive performance in medical image segmentation tasks [11, 12]. However, existing attention mechanisms mainly emphasize semantic enhancement while insufficiently constraining boundary consistency and regional continuity.

2.3 Multi-scale Context Modeling

Skin lesions exhibit substantial variations in size and morphology, requiring effective multi-scale feature extraction. DeepLab introduced atrous convolution and ASPP to capture contextual information at different receptive fields [13]. Subsequent methods further combined pyramid structures and residual learning to improve multi-scale representation. For example, Ms RED employed multi-scale residual encoding and decoding modules to enhance lesion structure recovery [14]. However, most existing methods statically aggregate multi-scale features and lack adaptive scale selection, while boundary continuity is still insufficiently preserved.

2.4 Global Semantic Guidance

Transformer-based segmentation models improve global contextual representation through self-attention mechanisms. TransUNet combines Transformer encoders with U-Net decoders to integrate global semantic information with local features [15], while UNETR and MedT further enhance long-range feature interaction [16, 17]. However, existing studies generally optimize semantic enhancement, multi-scale modeling, and boundary refinement separately. Few studies simultaneously model semantic guidance, boundary interaction, and regional consistency within a unified framework. Therefore, this study proposes CGMU-Net to jointly perform semantic–boundary interaction, multi-scale feature enhancement, and global semantic guidance.

3 Method

3.1 Overall architecture

In this study, we propose CGMU-Net, a skin lesion segmentation framework based on an encoder–decoder architecture, as illustrated in **Fig. 1**. The proposed network is designed to address challenging dermoscopic scenarios, including blurred lesion boundaries, irregular structures, and low-contrast regions.

The encoder progressively extracts hierarchical semantic features from the input image, while the decoder gradually restores spatial resolution to generate pixel-level segmentation results. To enhance feature interaction during decoding, a Cross-Interaction Attention Fusion Module (CIAFM) is introduced to establish interaction between semantic features and boundary-sensitive information, thereby improving lesion boundary localization and structural continuity. In addition, a Channel Feature Enhancement and Expansion Module (CFEEM) is designed to strengthen lesion-related feature representation and improve multi-scale contextual modeling. Furthermore, a classification-guided semantic refinement strategy is incorporated to provide global lesion-aware guidance for segmentation optimization. By integrating semantic–boundary interaction, multi-scale feature enhancement, and global semantic guidance within a unified framework, CGMU-Net achieves more accurate lesion localization and boundary recovery.

3.2 Cross-Interaction Attention Fusion Module

To improve the interaction between semantic representation and boundary-sensitive information, a Cross-Interaction Attention Fusion Module (CIAFM) is designed, as illustrated in **Fig. 2**. In dermoscopic images, lesion regions often exhibit blurred boundaries and irregular structures. Conventional skip fusion based on direct concatenation cannot effectively coordinate semantic consistency and edge preservation. High-level semantic features usually enhance lesion localization but may weaken boundary responses, while shallow spatial features preserve detailed structures yet are sensitive to background noise.

To address this issue, CIAFM establishes bidirectional interaction between semantic features and boundary-aware representations during feature fusion. Specifically, the decoder feature and encoder feature are first projected into semantic and boundary branches, respectively. The semantic branch focuses on lesion region representation, whereas the boundary branch enhances edge-sensitive responses. Their interaction process can be formulated as:

$$F_b' = F_b + \beta \cdot A(F_b, F_s) \tag{1}$$

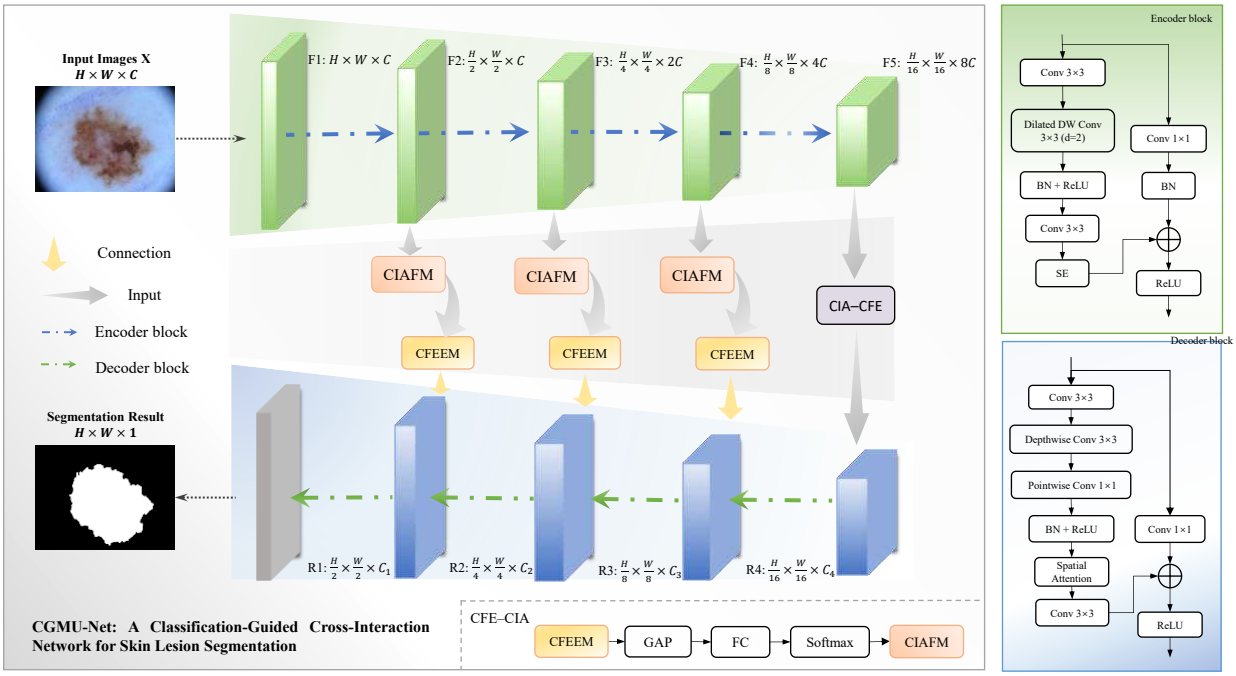


Fig. 1 Overview of CGMU-Net. CIAFM and CFEEM are embedded in encoder stages F_1–F_4 to enhance semantic-boundary interaction and lesion consistency, while CIA-CFE at bottleneck F_5 adds classification-guided modulation. The decoder then produces the final segmentation map.

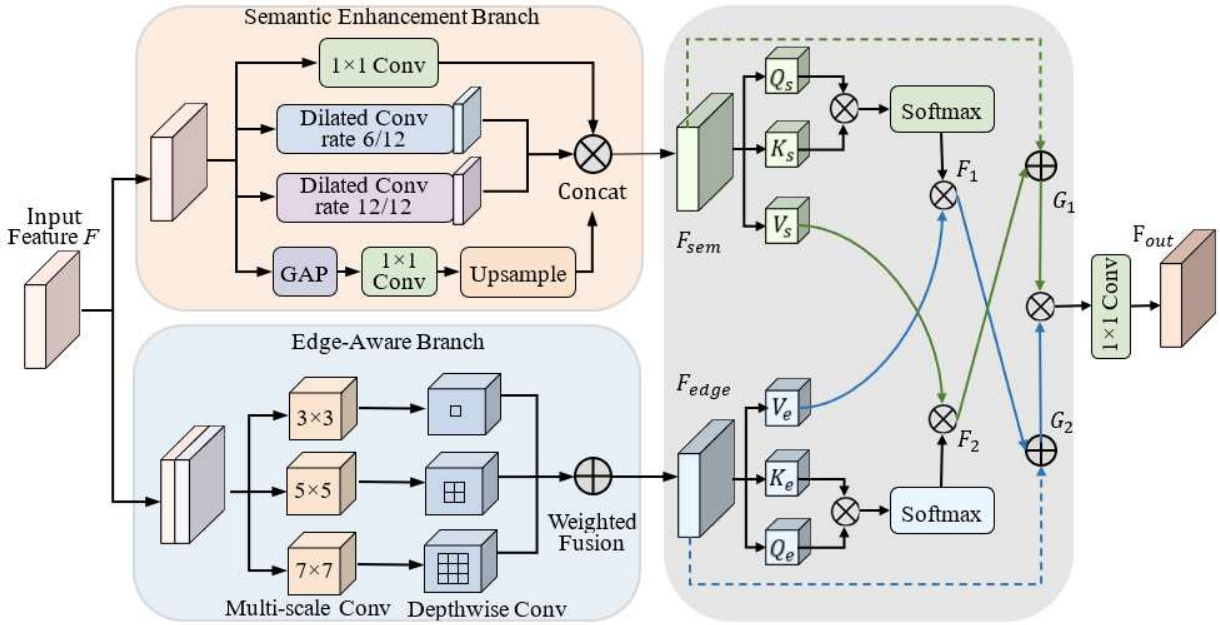


Fig. 2 The framework of the CIAFM module. The semantic branch captures multi-scale context and the EACB-based edge-aware branch extracts fine-grained boundary features; cross-interaction attention fuses them to produce semantically enhanced and boundary-preserved representations.

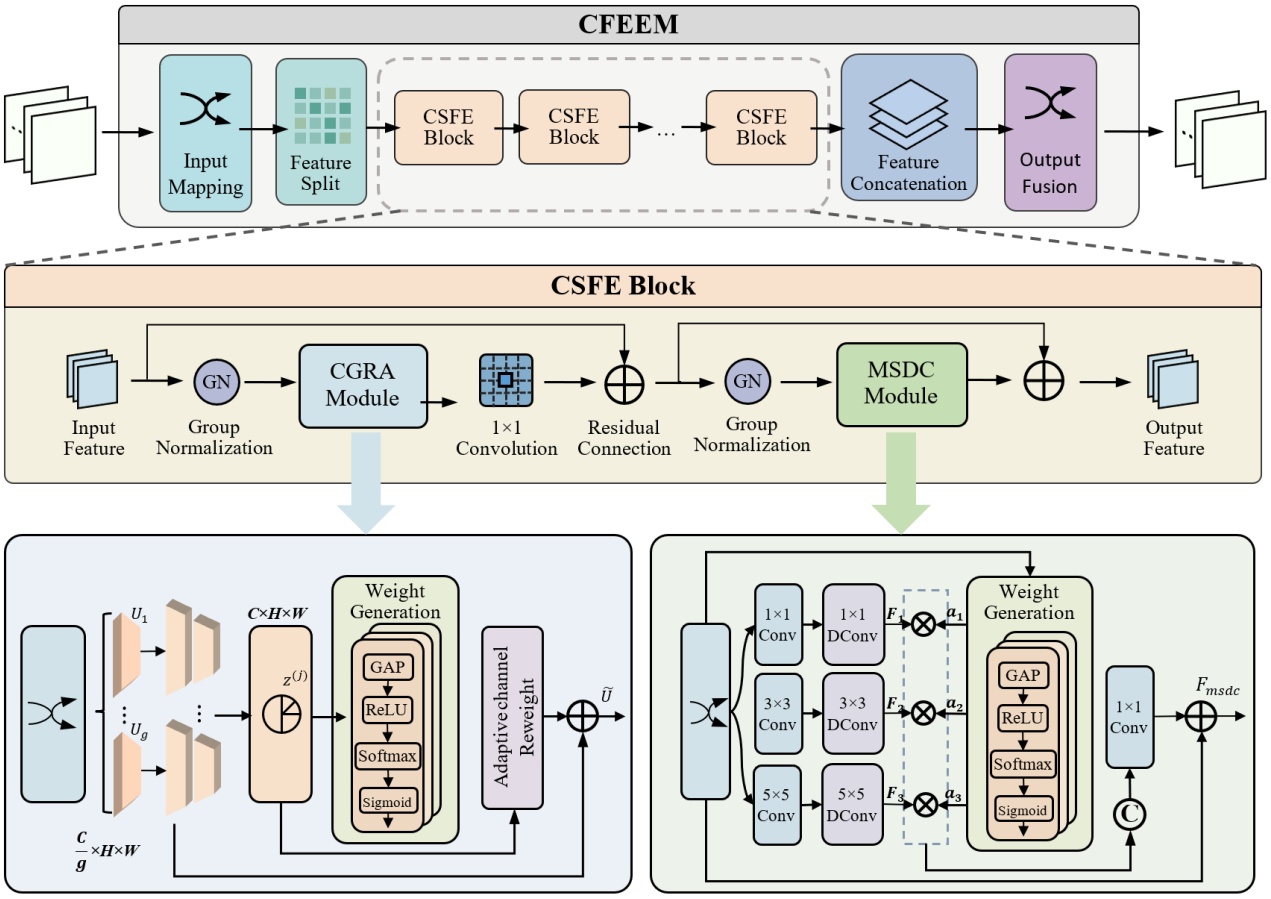


Fig. 3 The framework of the CFEEM module. The input features are mapped and split, processed by N cascaded CSFE-Blocks, and fused; each block integrates global-context guided channel reweighting and multi-scale aggregation to enhance lesion-region integrity and continuity.

where F_s and F_b denote semantic and boundary features, respectively, $A(\cdot)$ represents the cross-interaction attention operation, and α and β are adaptive weighting factors.

After feature interaction, the refined representations are aggregated to generate the final fusion feature:

$$F_{\text{out}} = \text{Conv}([F'_s, F'_b]) \tag{2}$$

3.3 Channel-wise Feature Enhancement and Excitation Module

To further strengthen lesion-related representation capability, a Channel-wise Feature Enhancement and Excitation Module (CFEEM) is proposed, as shown in**Fig. 3**. Since skin lesions exhibit substantial variations in size, texture, and morphology, conventional convolution operations have limited capability in adaptively emphasizing informative lesion features across different receptive fields.

First, global contextual information is aggregated through channel-wise feature compression:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j) \tag{3}$$

where X_c denotes the c-th feature channel, and $H \times W$ represent the spatial dimensions of the feature map. The aggregated channel descriptor is then utilized to generate adaptive channel attention weights:

$$s = \sigma(W_2 \delta(W_1 z)) \tag{4}$$

where W_1 and W_2 denote fully connected transformations, $\delta(\cdot)$ represents the ReLU activation function, and $\sigma(\cdot)$ is the sigmoid function.

Finally, the recalibrated feature representation is obtained by channel-wise excitation:

$$F_{\text{out}} = s \odot X \tag{5}$$

where $\odot$ denotes channel-wise multiplication. In addition, multi-scale contextual features are integrated to improve receptive-field adaptability and strengthen lesion-region continuity. By combining channel enhancement and multi-scale aggregation, CFEEM effectively suppresses redundant background responses and improves discriminative lesion representation for subsequent segmentation prediction.

4 Experimental Results and Analysis

4.1 Experimental Settings

Experiments were conducted on three public dermoscopic datasets: ISIC 2017 (2,150 images), ISIC 2018 (2,694 images), and PH2 (200 images), all with pixel-level annotations. ISIC 2017 and 2018 were split into training, validation, and test sets at a ratio of 7:1:2, while PH2 was divided into 160/20/20 due to its limited size. All images were resized to 512×512 and normalized, with data augmentation including random scaling, flipping, warping, and colour, saturation, and brightness adjustments. The model was implemented in PyTorch and trained using SGD (initial learning rate 0.01, weight decay 5×10^{-4}) with a linear decay schedule. The batch size was 4, and training ran for up to 300 epochs with early stopping (patience 10, threshold 1×10^{-4}). The loss function was a weighted combination of BCE and Dice losses (0.4/0.6).The total loss was defined as:

$$L = 0.4L_{\text{BCE}} + 0.6L_{\text{Dice}} \tag{6}$$

where L_{BCE} measures pixel-wise prediction error, and L_{Dice} penalizes mismatch between the predicted lesion mask and the ground-truth mask. During inference, outputs were passed through a sigmoid function and binarized at 0.5.

4.2 Evaluation Metrics

To comprehensively evaluate segmentation performance, six commonly used metrics were adopted: mean Intersection over Union (mIoU), Dice Similarity Coefficient (DSC), Accuracy (ACC), Sensitivity (Sen), Specificity (Spe), and the 95th percentile Hausdorff Distance (HD95). The first five metrics were calculated from four pixel-level statistics: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Specifically, TP and TN denote correctly classified lesion and background pixels, respectively, whereas FP and FN denote background pixels incorrectly classified as lesion and lesion pixels incorrectly classified as background, respectively. The evaluation metrics are defined as follows:

$$Spe = \frac{TN}{TN + FP} \tag{7}$$

$$Sen = \frac{TP}{TP + FN} \tag{8}$$

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{9}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$mIoU = \frac{1}{C} \sum_{i=1}^{C} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{11}$$

For boundary-based evaluation, let P and G denote the sets of boundary points from the predicted mask and the ground-truth mask, respectively. The directed Hausdorff distance from P to G is defined as:

$$d_H(P, G) = \max_{p \in P} \min_{g \in G} \|p - g\|_2 \tag{12}$$

The symmetric Hausdorff distance is then given by:

$$H(P, G) = \max\big(d_H(P, G),\, d_H(G, P)\big) \tag{13}$$

$$D_{95}(P, G) = \text{Percentile}_{95} \left\{ \min_g \|p - g\|_2,\ p \in P \right\} \tag{14}$$

$$HD_{95}(P, G) = \max\big(D_{95}(P, G),\, D_{95}(G, P)\big) \tag{15}$$

where C denotes the number of classes. Since this study focuses on binary skin lesion segmentation, $C = 2$.

4.3 Comparative experiments

To evaluate the effectiveness of CGMU-Net for dermoscopic skin lesion segmentation, we conducted comparisons with eleven representative segmentation models, including U-Net, U-Net v2 [18], ASCU-Net, TransUNet, Swin-Unet, MedT, MALUNet [19], SCSONet [20], VM-UNet [21], FIF-Net [22], and MHorUNet [23]. To reduce the influence of random initialization and training variability, all experiments were repeated five times, and the results are presented as mean ± standard deviation.

4.3.1 Experimental results on ISIC 2017

The comparative experimental results on the ISIC 2017 dataset are shown in **Table 1**. Compared with other methods, CGMU-Net achieves the best performance in DSC and mIoU, while also obtaining the lowest HD95 value. These results indicate that CGMU-Net effectively improves lesion-region overlap and boundary localization under complex dermoscopic imaging conditions. Higher DSC and mIoU values indicate that the model is capable of producing more complete and continuous lesion region predictions. In addition, the lower HD95 value demonstrates more accurate boundary recovery, reflecting the effectiveness of the proposed method in modeling semantic-boundary interactions.

Fig. 4 presents the qualitative segmentation results on the ISIC 2017 dataset. Existing methods generally suffer from over-segmentation, under-segmentation, and boundary deviation in challenging scenarios. In contrast, the segmentation results generated by CGMU-Net are visually closer to the ground truth in both lesion extent and contour delineation, especially in low-contrast regions and blurred-boundary cases, where the proposed method shows more stable segmentation performance.

Table 1 Statistical comparison of different models on the ISIC2017 dataset. Lower values indicate better performance for HD95. (BOLD indicates The Best And UNDERLINE Indicates The Second Best.)

Model	DSC(%)↑	HD95↓	Sen(%)↑	Spe(%)↑	Acc(%)↑	mIoU(%)↑
U-Net	85.49±0.22	20.85	83.22±0.25	97.58±0.15	95.15±0.18	74.15±0.55
U-Net v2	87.54±0.19	17.98	86.11±0.21	96.46±0.19	95.88±0.20	75.65±0.55
ASCU-Net	86.29±0.10	18.76	85.89±0.24	97.62±0.19	95.36±0.14	76.65±1.25
Trans U-Net	83.95±0.22	17.59	85.32±0.11	**98.59±0.11**	94.95±0.22	72.34±0.35
SwinU-Net	87.36±0.15	20.65	**90.03±0.11**	94.77±0.18	94.59±0.14	71.50±0.32
MedT	88.90±0.17	13.87	89.75±0.15	98.38±0.19	95.79±0.11	79.74±0.35
MALUNet	88.47±0.13	14.70	84.78±0.12	98.47±0.19	96.08±0.08	79.34±0.21
SCOSNet	87.69±0.11	13.77	88.55±0.13	97.72±0.16	95.99±0.11	79.08±0.15
VM-UNet	88.80±0.17	13.85	87.87±0.13	97.98±0.19	**96.28±0.08**	79.86±0.25
FIF-Net	88.96±0.21	13.97	88.75±0.15	98.02±0.21	96.12±0.08	79.64±0.35
Ours	**89.21±0.11**	**13.02**	89.83±0.21	98.51±0.11	95.86±0.11	**80.28±0.15**

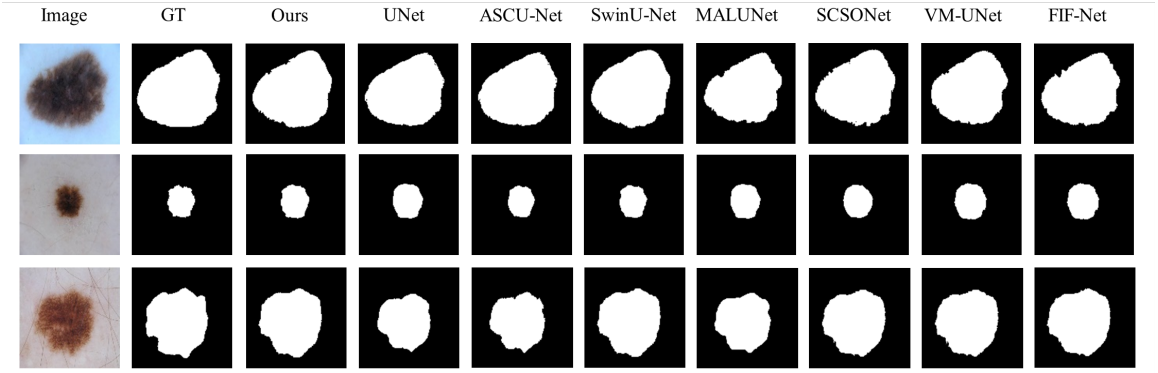


Fig. 4 Segmentation visualization on ISIC 2017 dataset.

4.3.2 Experimental results on ISIC 2018

The comparative experimental results on the ISIC 2018 dataset are shown in **Table 2**. CGMU-Net achieves the best performance in DSC, mIoU, and Sen, while also obtaining competitive results on the remaining evaluation metrics. These results indicate that CGMU-Net can effectively improve lesion-region prediction completeness and reduce missed lesion pixels under conditions with substantial variations in lesion scale and morphology. From the perspective of module effectiveness, the improvement in overlap-based metrics suggests that CGMU-Net enhances multi-scale contextual representation and lesion-region continuity. In addition, the higher Sen value indicates that the classification-guided semantic enhancement strategy provides effective lesion-aware semantic information at the bottleneck, thereby improving the preservation capability for critical lesion regions.

Fig. 5 presents the qualitative segmentation results on the ISIC 2018 dataset. Compared with other methods, CGMU-Net generates more complete lesion regions and more accurate boundary contours under challenging conditions involving complex lesion structures and weak lesion–background contrast. Some comparison methods still exhibit fragmented prediction regions or local boundary misclassification, whereas CGMU-Net maintains better regional continuity.

4.3.3 Experimental results on PH2

The comparative experimental results on the PH2 dataset are shown in **Table 3**. Although the PH2 dataset contains relatively few training samples, CGMU-Net still achieves the best performance in DSC, Sensitivity, and HD95, demonstrating that CGMU-Net maintains stable segmentation capability under small-scale dataset conditions.

Fig. 6 presents the qualitative segmentation results on the PH2 dataset. In challenging scenarios involving small lesion regions and complex background interference, several comparison methods still suffer from local misclassification or incomplete lesion prediction. In contrast, CGMU-Net produces segmentation results that are visually closer to the ground truth in both lesion extent and contour continuity. The classification-guided semantic enhancement strategy further improves the perception capability for inconspicuous lesion regions, thereby enhancing lesion recall and structural continuity.

4.4 Ablation experiments

To validate the effectiveness of the proposed modules, ablation experiments were conducted on the ISIC 2017 dataset using the U-Net as the baseline model. Specifically, CIAFM, CFEEM, and CIA-CFE were progressively introduced into the baseline.

Table 2 Statistical comparison of different models on the ISIC2018 dataset.

Model	DSC(%)↑	HD95↓	Sen(%)↑	Spe(%)↑	Acc(%)↑	mIoU(%)↑
U-Net	87.15±0.17	21.85	84.40±0.19	97.32±0.17	93.82±0.19	76.84±0.55
U-Net v2	87.62±0.13	20.07	86.20±0.21	96.21±0.16	94.02±0.11	76.93±0.45
ASCU-Net	88.01±0.08	19.36	84.33±0.23	96.85±0.15	**94.80±0.19**	78.64±0.05
Trans U-Net	85.61±0.22	18.19	87.50±0.17	97.90±0.17	93.65±0.22	74.94±0.35
Swin U-Net	86.73±0.32	19.01	86.74±0.23	96.38±0.15	93.07±0.12	75.84±0.65
MedT	88.84±0.12	14.78	88.95±0.17	97.45±0.15	93.85±0.15	79.22±0.18
MALUNet	88.04±0.27	15.60	89.74±0.16	97.79±0.17	94.48±0.27	79.34±0.46
SCSCONet	88.34±0.12	14.76	87.95±0.16	97.56±0.32	94.39±0.12	79.22±0.16
VM-UNet	88.80±0.08	13.85	88.75±0.17	**97.98±0.17**	94.31±0.12	79.86±0.17
FIF-Net	88.94±0.27	**13.28**	88.25±0.14	97.83±0.14	94.67±0.14	79.94±0.46
Ours	**89.91±0.15**	13.32	**90.13±0.12**	97.51±0.22	94.43±0.13	**80.03±0.31**

Table 3 Statistical comparison of different models on the PH2 dataset.

Model	DSC(%)↑	HD95↓	Sen(%)↑	Spe(%)↑	Acc(%)↑	mIoU(%)↑
U-Net	93.78±0.10	19.97	88.66±0.21	95.44±0.11	96.11±0.11	80.14±0.15
U-Net v2	94.02±0.10	18.97	88.11±0.26	93.42±0.15	96.32±0.10	81.18±0.15
ASCU-Net	93.63±0.29	18.34	84.74±0.17	94.94±0.08	96.13±0.19	81.53±0.51
Trans U-Net	93.88±0.10	18.79	86.34±0.15	95.98±0.09	96.23±0.10	81.64±0.35
Swin U-Net	92.92±0.10	23.78	87.03±0.16	93.43±0.14	95.02±0.05	79.04±0.05
MedT	93.63±0.19	17.21	88.63±0.11	95.14±0.13	96.32±0.12	85.95±0.15
MALUNet	94.40±0.10	13.54	84.75±0.08	95.62±0.12	**96.89±0.10**	85.14±0.46
SCSCONet	93.88±0.20	14.52	88.23±0.19	95.44±0.13	96.24±0.13	85.22±0.46
VM-UNet	93.28±0.10	14.27	89.16±0.21	**96.34±0.08**	96.14±0.12	83.95±0.15
FIF-Net	94.60±0.03	13.32	88.93±0.11	95.24±0.11	96.68±0.08	87.24±0.16
Ours	**95.01±0.17**	**13.08**	**89.93±0.13**	95.68±0.08	96.38±0.18	**86.74±0.06**

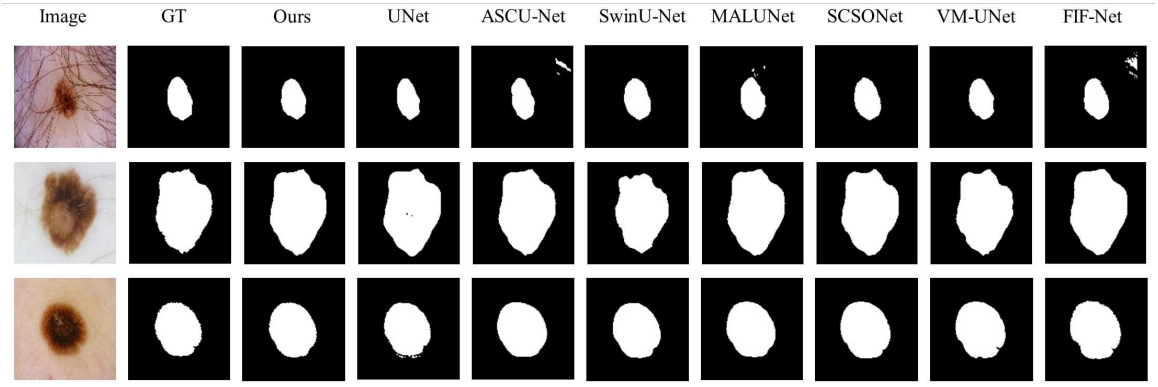


Fig. 5 Segmentation visualization on ISIC 2018 dataset.

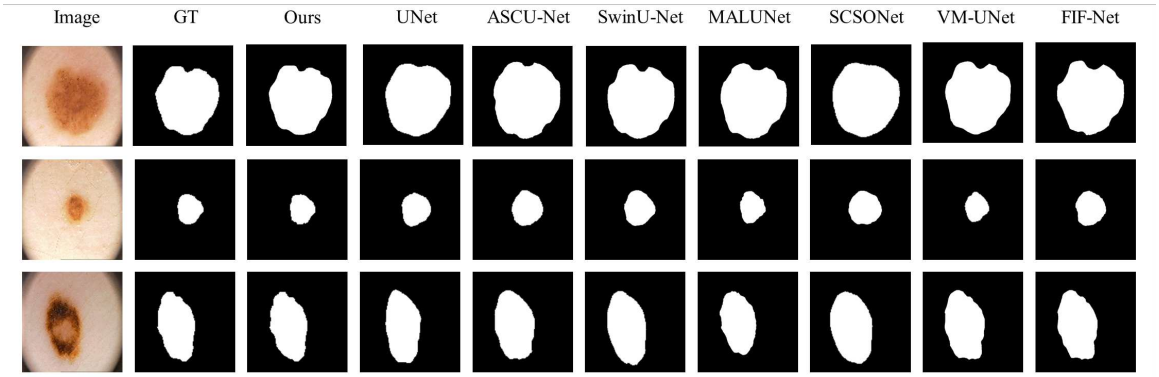


Fig. 6 Segmentation visualization on PH2 dataset.

Table 4 Ablation results of different module configurations on the ISIC 2017 dataset. Bold values highlight the best results.

Network	DSC(%)↑	HD95(mm)↓	Sen(%)↑	Spe(%)↑	Acc(%)↑	mIoU(%)↑
Baseline	85.49	20.85	83.22	97.58	95.15	74.15
Baseline + CIAFM	87.32	15.10	86.16	97.60	95.20	77.49
Baseline + CFEEM	87.87	17.83	85.57	97.65	95.31	76.13
Baseline + CIAFM + CFEEM	88.91	14.43	89.94	97.82	95.69	80.15
CGMU-Net	89.21	13.02	90.03	98.07	95.86	80.28

The experimental settings include: Baseline (without any proposed modules), CIAFM only, CFEEM only, CIAFM + CFEEM, and the full CGMU-Net. The quantitative results are presented in **Table 4**.

From the results, it can be observed that introducing CIAFM alone leads to significant improvements in both region-overlap metrics and boundary-related metrics, with a particularly notable reduction in HD95. This indicates that CIAFM effectively enhances lesion localization and boundary contour recovery. Similarly, incorporating CFEEM alone also improves segmentation performance, with more pronounced gains in DSC and mIoU, demonstrating its capability in enhancing lesion region continuity and multi-scale contextual representation. When CIAFM and CFEEM are jointly applied, the model further outperforms either module used individually, suggesting that the two modules are complementary in terms of boundary-sensitive representation and structural continuity. With the further introduction of the proposed module, the complete CGMU-Net achieves the best performance across most evaluation metrics, indicating that lesion-aware semantic guidance can further improve boundary recovery, reduce missed lesion regions, and alleviate fragmentation in segmentation results.

The qualitative results are illustrated in **Figure 7**. The baseline model suffers from blurred boundaries, incomplete regions, and fragmented segmentation results in complex scenarios. After introducing CIAFM, the model produces more concentrated responses around lesion boundaries, thereby improving boundary localization. In contrast, CFEEM focuses more on enhancing the spatial continuity of lesion regions and reducing over-segmentation. When both modules are combined, the model generates clearer and more complete lesion regions. The full CGMU-Net produces segmentation results that are closer to the ground truth, especially in challenging cases such as irregular lesions, low contrast, and complex background interference.

4.5 Computational Complexity Analysis

To evaluate the practical feasibility of CGMU-Net, computational complexity was analyzed in terms of parameters and FLOPs under a 512×512 input. As shown in **Table 5**, segmentation performance is reported alongside complexity to assess the efficiency–accuracy trade-off.

CGMU-Net contains 9.62M parameters and 5.09G FLOPs. When compared to heavier models like Trans-U-Net, Swin-Unet, and VM-UNet, it achieves superior performance in DSC, HD95, and mIoU while incurring significantly lower computational cost. Furthermore, compared to the standard U-Net, CGMU-Net slightly increases the parameter count but reduces the FLOPs. This results in improved region overlap and boundary accuracy, indicating that its performance gains stem from more effective feature representation rather than a larger model scale.

Although MALUNet has the lowest complexity, its inferior performance suggests limited capacity for modeling complex lesions. In contrast, CGMU-Net achieves the best overall performance with moderate complexity, demonstrating a better efficiency–accuracy balance.

This balance is important for clinical applications, where both accuracy and efficiency are required. While not the most lightweight, CGMU-Net provides a more practical trade-off. Further evaluation of inference latency on real-world hardware is still needed.

5 Conclusion

This study proposes CGMU-Net, a classification-guided cross-interaction network for dermoscopic skin lesion segmentation. The framework is designed to address challenges such as lesion-scale variation, blurred boundaries, low contrast, and occlusion. Specifically, CIAFM enhances semantic–boundary interaction, CFEEM improves multi-scale region representation, and the classification-guided strategy introduces lesion-aware refinement at the bottleneck.

Experiments on ISIC 2017, ISIC 2018, and PH2 demonstrate that CGMU-Net achieves competitive performance under a unified protocol. Ablation studies further confirm that CIAFM, CFEEM, and classification-guided refinement provide complementary gains in boundary delineation, region continuity, and semantic discrimination.

Overall, CGMU-Net offers an effective solution for dermoscopic lesion segmentation. However, current validation is limited to public datasets and binary tasks. Future work will focus on external clinical validation, cross-domain generalization, computational efficiency, and extension to more diverse medical imaging scenarios.

Author Contributions BingYan Liu: Writing – original draft, Resources, Methodology; Mingen

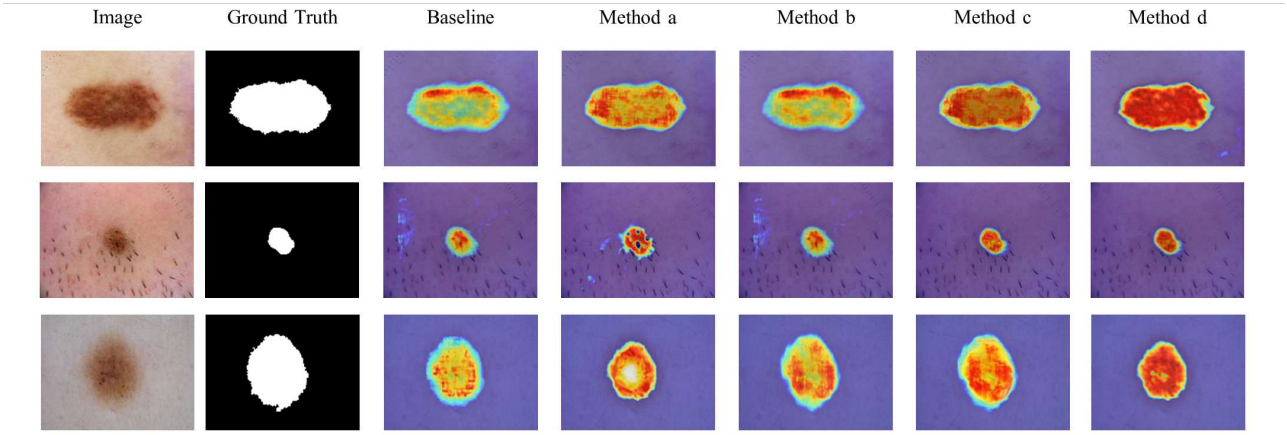


Fig. 7 Visual comparison of module ablation experiment on ISIC2017 dataset

Table 5 Computational complexity and segmentation performance comparison on the ISIC 2017 dataset. (BOLD indicates The Best And UNDERLINE indicates The Second Best.)

Method	Para (M)↓	FLOPs(G)↓	DSC (%)↑	HD95↓	mIoU(%)↑
U-Net	7.7683	13.7571	85.49±0.22	20.85	74.15±0.55
Trans U-Net	25.2124	6.5543	83.95±0.22	17.59	72.34±0.35
Swin-Unet	41.3476	15.6929	87.36±0.15	20.65	71.50±0.32
MALUNet	**0.1772**	**0.0852**	88.47±0.13	14.70	79.34±0.21
VM-UNet	27.4342	4.1102	88.80±0.17	13.85	79.86±0.25
CGMU-Net	9.6213	5.0876	**89.21±0.11**	**13.02**	**80.28±0.15**

Zhong: Writing – review & editing, Supervision; ZhengFeng Li: Writing – review & editing, Visualization; YiJie Tang: Writing – review & editing, Validation; BinGan Yuan: Writing – review & editing, Formal analysis.

Funding This investigation was not funded by a particular grant from any public, commercial, or charitable funding body.

Data Availability Data will be made available on request.

Declarations

Competing interests The authors declare no competing interests.

References

[1] Galambus, J., Smythe, C., Cockerell, C.J.: Pathology of melanoma: New concepts (2025 update). Dermatologic Clinics **43**(3), 409–418 (2025) https://doi.org/10.1016/j.det.2025.03.008

[2] Siegel, R.L., Kratzer, T.B., Giaquinto, A.N., Sung, H., Jemal, A.: Cancer statistics, 2025. Ca **75**(1), 10 (2025) https://doi.org/10.3322/caac.21871

[3] Ma, H., Xu, C., Nie, C., Han, J., Li, Y., Liu, C.: DBE-Net: dual boundary-guided attention exploration network for polyp segmentation. Diagnostics **13**(5), 896 (2023) https://doi.org/10.1200/GO.20.00122

[4] Meedeniya, D., De Silva, S., Gamage, L., Isuranga, U.: Skin cancer identification utilizing deep learning: A survey. IET Image Processing **18**(13), 3731–3749 (2024) https://doi.org/10.1049/ipr2.13219

[5] Mahbod, A., Tschandl, P., Langs, G., Ecker, R., Ellinger, I.: The effects of skin lesion segmentation on the performance of dermatoscopic image classification. Computer Methods and Programs in Biomedicine **197**, 105725 (2020) https://doi.org/10.1016/j.cmpb.2020.105725

[6] Zafar, K., Gilani, S.O., Waris, A., Ahmed, A., Jamil, M., Khan, M.N., Kashif, A.S.: Skin lesion segmentation from dermoscopic images using convolutional neural network. Sensors **20**(6), 1601 (2020) https://doi.org/10.3390/s20061601

[7] Chen, S., Wang, Z., Shi, J., Liu, B., Yu, N.: A multi-task framework with feature passing module for skin lesion classification and segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1126–1129. IEEE, ??? (2018). https://doi.org/10.1109/ISBI.2018.8363769

[8] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image

segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241. Springer, Munich, Germany (2015). https://doi.org/10.1007/978-3-319-24574-4_28

[9] Tong, X., Wei, J., Sun, B., *et al.*: ASCU-Net: Attention gate, spatial and channel attention U-net for skin lesion segmentation. Diagnostics **11**(3), 501 (2021) https://doi.org/10.3390/diagnostics11030501

[10] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141. IEEE, Salt Lake City, UT, USA (2018). https://doi.org/10.1109/CVPR.2018.00745

[11] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision, pp. 205–218. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-25066-8_9

[12] Azad, R., Arimond, R., Aghdam, E.K., *et al.*: Dae-former: Dual attention-guided efficient transformer for medical image segmentation. In: International Workshop on Predictive Intelligence in Medicine, pp. 83–95. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-46005-0_8

[13] Chen, L.C., Papandreou, G., Kokkinos, I., *et al.*: DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(4), 834–848 (2018) https://doi.org/10.1109/TPAMI.2017.2699184

[14] Dai, D., Dong, C., Xu, S., *et al.*: Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation. Medical Image Analysis **75**, 102293 (2022) https://doi.org/10.1016/j.media.2021.102293

[15] Chen, J., Lu, Y., Yu, Q., et al.: TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation (2021)

[16] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., *et al.*: UNETR: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 574–584 (2022). https://doi.org/10.1109/WACV51458.2022.00181

[17] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., Patel, V.M.: Medical transformer: Gated axial-attention for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 36–46. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_4

[18] Peng, Y., Chen, D.Z., Sonka, M.: 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). https://doi.org/10.1109/ISBI60581.2025.10980742

[19] Ruan, J., Xiang, S., Xie, M., *et al.*: MALUNet: A multi-attention and lightweight U-net for skin lesion segmentation. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1150–1156. IEEE, ??? (2022). https://doi.org/10.1109/BIBM55620.2022.9995040

[20] Chen, H., Li, Z., Huang, X., Peng, Z., Deng, Y., Tang, L., Yin, L.: SCSONet: Spatial-channel synergistic optimization net for skin lesion segmentation. Frontiers in Physics **12**, 1388364 (2024) https://doi.org/10.3389/fphy.2024.1388364

[21] Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. ACM Transactions on Multimedia Computing, Communications and Applications (2024) https://doi.org/10.1145/3767748

[22] Xin, J., Su, N., Zhang, S., Shen, Q., Zhou, F., Wang, Z.: Fif-Net: A Feature Interaction Fusion Network for Skin Lesion Segmentation. Available at SSRN 5212592 (2024). https://doi.org/10.2139/ssrn.5212592

[23] Wu, R., Liang, P., Huang, X., *et al.*: MHORUNet: High-order spatial interaction U-net for skin lesion segmentation. Biomedical Signal Processing and Control **88**, 105517 (2024) https://doi.org/10.1016/j.bspc.2023.105517