GDML: Geometric Decoupling Mutual Learning for Robust Skin Lesion Segmentation Using Distance Transforms

Bo Min[a], Kun Wang[a,b,*], Bosen Nie[a], Xuyang Hai[c], Demin Xu[c], Quanrong Wang[a], Yang Yu[d]

[a]*School of Computer and Information, Anhui Polytechnic University, Wuhu, 241000, Anhui, China*
[b]*Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education,Chongqing University, Chongqing, 400044, China*
[c]*School of Electrical Engineering, Anhui Polytechnic University, Wuhu, 241000, Anhui, China*
[d]*School of Biology and Engineering, Guizhou Medical University, Guiyang, 561113, Guizhou, China*

Abstract

Accurate segmentation of skin lesions is crucial for the diagnosis and management of skin diseases. However, this task remains highly challenging due to the structural complexity of lesion interiors and the ambiguity of boundary regions, characterized by low contrast, textural heterogeneity, and blurred interfaces between lesions and healthy skin. To address these critical limitations, we propose a geometric decoupling mutual learning (GDML) model based on distance transforms, which aims to extract deeper geometric representations by explicitly modeling the distinct characteristics of internal and boundary pixels through separate learning branches. The GDML framework employs an information interaction module as its fundamental building block, upon which two parallel and mutually interactive pathways are constructed to facilitate rich cross-modal information exchange. Furthermore, we design a feature enhancement module that bridges the encoder and decoder, dynamically guiding the network's attention toward lesion regions to mine fine-grained geometric features. Additionally, a distance transform-based supervision strategy is introduced to decouple original lesion images into internal-specific and boundary-specific labels. These dual labels respectively guide the learning of internal and boundary pixels: the internal-specific label emphasizes the consistency of intra-lesion features, while the boundary-specific label enforces the continuity and precision of boundary features. Experimental results on three benchmark datasets (ISIC2017, ISIC2018, and PH2) demonstrate that our method outperforms current state-of-the-art approaches in terms of Dice and IoU metrics. The source code is available at https://github.com/minbo20050930-svg/GDML.

Keywords: Skin lesion segmentation, Geometric decoupling, Feature enhancement module, Distance transforms

1. Introduction

Skin lesion segmentation is a pivotal task in medical image segmentation, as it enables clinicians to accurately identify and localize lesion regions. This is particularly critical for the early detection of skin malignancies such as melanoma, where precise segmentation directly supports timely clinical decision-making. Improving segmentation accuracy not only reduces the risk of misdiagnosis and ensures prompt, effective treatment but also significantly enhances patient prognosis. However, this task remains highly challenging due to two core limitations: 1) The complex internal texture and heterogeneous intensity distribution of lesions complicate the separation of target regions from healthy skin. 2) Blurred tissue interfaces and low contrast between lesions and surrounding skin hinder the accurate delineation of boundary contours. Several representative challenging cases are illustrated in Fig. 1 (A).

Advancements in convolutional neural networks (CNNs) have driven remarkable progress in skin lesion segmentation. Since the seminal work of UNet [1], a multitude of UNet-based variants [2–4] have been proposed, demonstrating robust performance across diverse medical image segmentation tasks. These methods excel at capturing discriminative multi-scale features, enabling the segmentation of complex anatomical structures in medical images. Nevertheless, they still struggle to segment lesions with intricate internal textures or ambiguous boundaries, as the coupled learning of internal and boundary features limits the model's ability to address their distinct characteristics. To mitigate boundary ambiguity, recent studies have focused on integrating boundary information into segmentation frameworks [5–7]. While these approaches improve boundary segmentation quality to some extent, direct incorporation of boundary cues often fails to fully exploit the unique properties of internal and boundary features, nor does it capture their latent interdependencies. In contrast to prior work, we aim to design a model that achieves more robust segmentation by explicitly decoupling the learning of complex internal features and fine-grained boundary features.

In this paper, we propose a novel geometric decoupling mutual learning (GDML) framework based on distance transforms to address the limitations of incomplete internal segmen-

---

*Corresponding author.
Email addresses:* 3230701322@stu.ahpu.edu.cn (Bo Min), kun.wang@ahpu.edu.cn (Kun Wang), 3230701229@ahpu.edu.cn (Bosen Nie), 3230209230@stu.ahpu.edu.cn (Xuyang Hai), 3230803215@stu.ahpu.edu.cn (Demin Xu), 3240706305@stu.ahpu.edu.cn (Quanrong Wang), yuyang@gmc.edu.cn (Yang Yu)
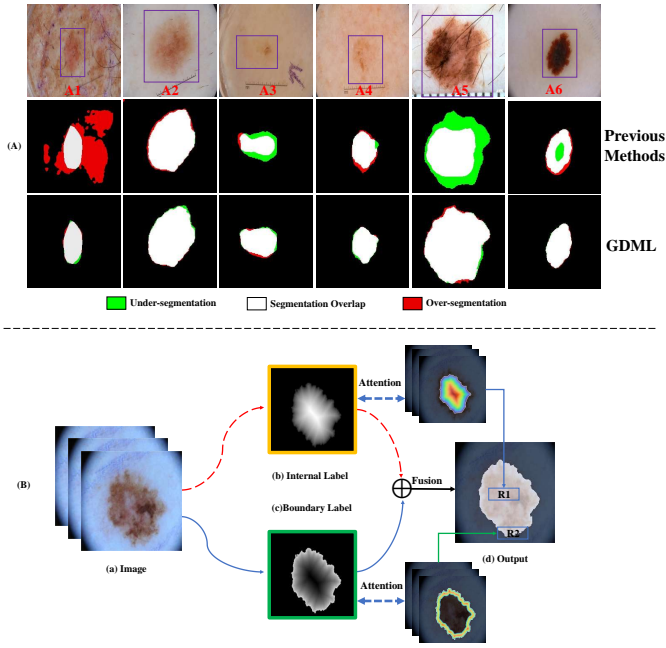
Fig. 1. (A) The challenges in skin lesion segmentation. Red, white, and green indicate the over-segmented regions, the segmentation overlap, and the under-segmented regions, respectively. (A1) Low contrast and inconsistent intensity. (A2) Blurred boundary. (A3) Low contrast; (A4) Low contrast and blurred boundary. (A5–A6) Inconsistent intensity. (B) **Motivation**: Enhancing the integrity of internal features and the continuity of boundary features helps improve segmentation results.

tation and blurred boundaries in existing methods. As shown in Fig. 1 (A), lesion images typically exhibit complex internal structures and ill-defined boundaries. To tackle this, we adopt a geometric decoupling strategy to generate internal-specific and boundary-specific labels from the original lesion images, which respectively emphasize the integrity of internal features (R1) and the continuity of boundary features (R2). The final segmentation result is derived from the synergistic integration of these two complementary feature representations. Specifically, our architecture comprises two parallel, mutually interactive pathways dedicated to capturing internal and boundary features, respectively. These pathways leverage information interaction modules (IIM) to facilitate rich cross-feature exchange during feature extraction. Additionally, we introduce a feature enhancement module (FEM) that dynamically prioritizes lesion regions, strengthening the model's capacity to learn deep geometric representations. Furthermore, we propose a distance transform-based supervision strategy (DTSS) to guide the network: this strategy enables differential learning of complex internal features and fine-grained boundary features while preserving their global consistency. Our main contributions are fourfold:

- We propose a novel geometric decoupling mutual learning framework that explicitly enables differential learning of complex internal features and fine-grained boundary details, addressing the limitations of coupled feature learning in existing methods.

- We propose a feature enhancement module that highlights lesion regions and strengthens geometric feature learning, thereby improving the extraction of internal complex features and boundary detail features.

- We introduce a distance transform-based supervision strategy that guides the network to learn internal and boundary features in a distinct yet mutually consistent manner, ensuring the synergy between intra-lesion integrity and boundary continuity.

- Extensive experiments on three benchmark datasets (ISIC2017, ISIC2018, and PH2) demonstrate that the proposed GDML achieves state-of-the-art (SOTA) performance, validating its superior accuracy and robustness for challenging skin lesion segmentation tasks.

2. Related work

2.1. Medical image segmentation

In recent years, deep learning-based methods have achieved remarkable success across numerous medical image segmentation tasks, including skin lesion segmentation [8], brain tumor segmentation [9], polyp segmentation [10], and lung nodule segmentation [11]. Among these approaches, UNet [1] stands out for its symmetric encoder–decoder design and skip connections that fuse encoder context with decoder localization, leading to higher segmentation accuracy. Building upon this foundational design, numerous UNet variants have been proposed to address its inherent limitations. For instance, UNet++ [12] introduces nested dense skip pathways to narrow the semantic gap between the encoder and decoder, enhancing feature propagation. AttUNet [2] embeds gating attention mechanisms into skip connections to emphasize task-relevant regions while suppressing background noise. With the rising popularity of Transformers for capturing long-range dependencies, hybrid architectures have emerged as a promising direction. R2UNet[13] embeds recurrent residual convolutional units into each encoder and decoder stage, expanding the effective receptive field and stabilizing optimization, which strengthens context modeling and yields crisper boundaries with higher segmentation accuracy. CENet[14] augments the U-Net backbone with a dense atrous convolution (DAC) block and a residual multi-kernel pooling (RMP) module to enrich multi-scale context and preserve spatial detail, thereby improving boundary localization and overall segmentation accuracy. TransUNet [15] integrates a Vision Transformer (ViT) encoder into the UNet framework, strengthening global feature modeling. SwinUNet [16] replaces traditional convolutional blocks with Swin Transformer modules, enabling efficient full-image context modeling. More recently, state-of-the-art methods have explored diffusion strategies, Vision Mamba and Kolmogorov–Arnold Networks (KANs) to further push performance boundaries. HiDiff [17] leverages diffusion-based refinement to iteratively denoise segmentation masks, improving boundary smoothness and regional consistency. CPFNet [18] combines Global Pyramid Guidance and Scale-Aware Pyramid Fusion modules to inject multi-scale global context into the decoder and refine boundary details. VMUNet [19] incorporates Vision Mamba state-space blocks into a U-shaped architecture,
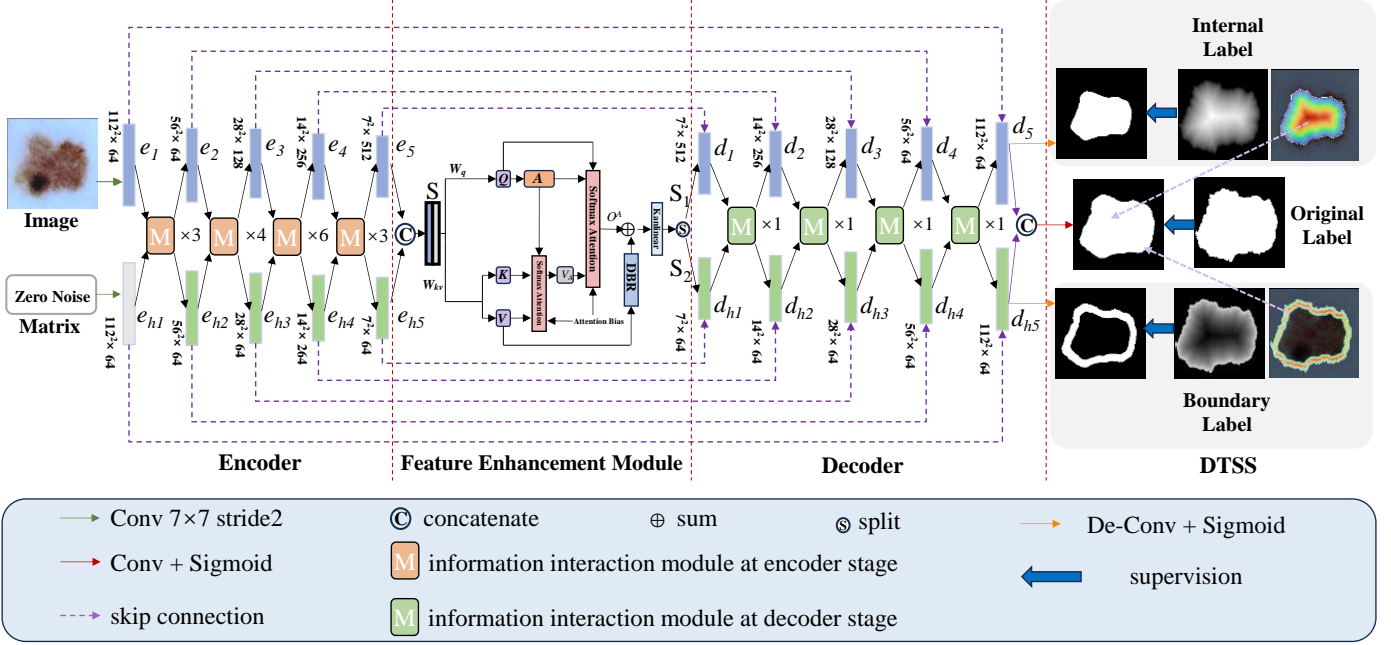
Fig. 2. Architecture of the proposed GDML. GDML constructs two IIM-based paths that respectively focus on internal and boundary features, links the encoder and decoder via FEM to strengthen geometric learning, and employs DTSS to supervise the network for differentiated feature learning.

modeling long-range dependencies with near-linear computational complexity. UKAN [20] integrates KANs into a UNet backbone via convolutional–KAN mixed units, enhancing nonlinear modeling capability and interpretability.

Despite the significant improvements in segmentation accuracy achieved by these methods, a critical limitation remains: they treat all image pixels uniformly during feature learning. This indiscriminate learning paradigm hinders the model's ability to capture fine-grained boundary features and fails to adequately address the challenge of boundary ambiguity in segmentation results, especially for lesions with complex internal textures and ill-defined tissue interface.

2.2. Boundary Information–Based Segmentation Methods

To address boundary ambiguity in medical image segmentation, recent studies have focused on explicitly leveraging boundary information to refine segmentation contours. ET-Net [21] proposes an edge-guided module that captures discriminative edge features in the early encoding stage, effectively preserving local edge details for subsequent boundary refinement. BCDUNet [22] proposes a bi-directional ConvL-STM–based decoder that propagates boundary-aware contextual information across scales, thereby enhancing contour localization and reducing over-segmentation along lesion borders. PraNet [10] designs a parallel reverse-attention and edge-guided decoder that explicitly highlights object boundaries while suppressing background clutter, thereby producing sharper segmentation masks and effectively mitigating boundary artifacts. RefineUNet [3] integrates stage-wise feedback mechanisms and residual attention refinement to narrow the semantic gap between the encoder and decoder, thereby sharpening boundary localization and improving overall segmentation performance. PyDiNet [5] introduces a pyramid dilated module (PDM) to

enhance the model's sensitivity to subtle lesion variations, facilitating more precise boundary detection for irregular lesions. XBoundFormer [6] embeds cross-scale boundary modeling into a Transformer architecture, utilizing boundary-specific tokens and cross-scale feature interactions to propagate fine-grained contour cues, which significantly boosts edge localization accuracy. LBUNet [7] proposes a lightweight U-Net with a boundary guidance branch and feature-fusion strategy that injects edge cues into the decoder, sharpening contours with few parameters and FLOPs. LiteMambaBound [23] leverages efficient Mamba-based state-space blocks and couples a boundary-aware prediction head with a normalized active contour loss, enabling compact mask generation and crisp boundary delineation under resource-constrained scenarios. Despite these advancements, direct integration of boundary information remains suboptimal due to two key limitations: First, the uneven spatial distribution of boundary pixels can introduce noise into feature learning, degrading the robustness of the model. Second, overemphasis on boundary cues often compromises the integrity of internal lesion regions, leading to incomplete segmentations or spurious holes in the predicted masks. In contrast, our work adopts a geometric label decoupling strategy to explicitly separate the learning of internal and boundary pixels. This design ensures that the model simultaneously enhances the integrity of internal lesion features and the continuity of boundary features, addressing the inherent trade-off between internal segmentation completeness and boundary precision in existing methods.

3. Method

3.1. GDML Architecture

Fig. 2 illustrates in detail the overall architecture of GDML, a dual path network trained with multiple output supervision.
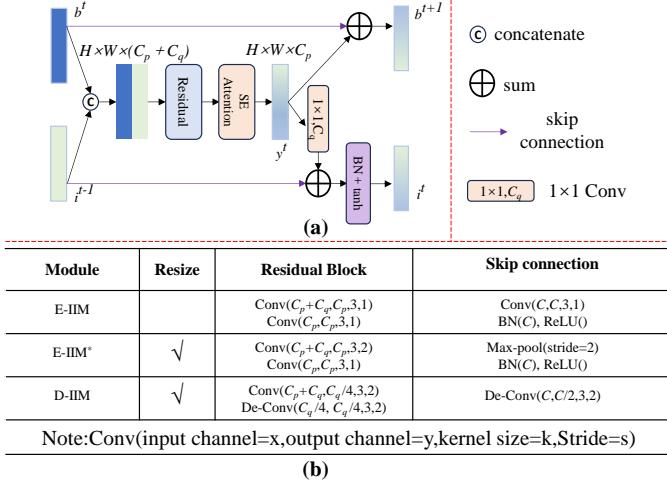
(a)

Module	Resize	Residual Block	Skip connection
E-IIM		$\text{Conv}(C_p+C_q,C_p,3,1)$ $\text{Conv}(C_p,C_p,3,1)$	$\text{Conv}(C,C,3,1)$ $\text{BN}(C), \text{ReLU}()$
E-IIM*	√	$\text{Conv}(C_p+C_q,C_p,3,2)$ $\text{Conv}(C_p,C_p,3,1)$	$\text{Max-pool(stride=2)}$ $\text{BN}(C), \text{ReLU}()$
D-IIM	√	$\text{Conv}(C_p+C_q,C_q/4,3,2)$ $\text{De-Conv}(C_q/4,C_q/4,3,2)$	$\text{De-Conv}(C,C/2,3,2)$
Note:Conv(input channel=x,output channel=y,kernel size=k,Stride=s)			

(b)

Fig. 3. Illustration of the Information Interaction Module.

The encoder and decoder are constructed from IIMs and connected by the FEM, forming two interacting branches that respectively focus on internal features and boundary features. Within each IIM, features from the two branches are mixed to enable effective cross branch information exchange. In the image branch, channel width decreases through the encoder and increases in the decoder to first distill and then recover spatial detail, whereas the matrix branch maintains a fixed width to preserve boundary related information. The encoder contains four IIM-based stages that progressively extract and integrate representations, and at each stage two outputs are generated and passed to the next stage, which finally yield the fused representation S. After further refinement by FEM, S is split along the channel dimension into two subsets, S_1 and S_2, which are fed into the first decoder stage. At each decoder layer, the two branch features are concatenated with the corresponding encoder features through skip connections and then propagated to the next layer, which promotes reuse of high resolution context and stable gradient flow. In the final layer, the outputs from the two paths are fused to obtain the final output. In the final stage, the two branch outputs are fused to generate the segmentation mask, and all three outputs, namely the two branch predictions and the final prediction, are jointly supervised by DTSS.

3.2. Information Interaction Module

To enable rich mutual information exchange between the two parallel paths of GDML, we design the IIM illustrated in Fig. 3 (a). The image path feature (b^t) and matrix path feature (i^{t-1}) are first concatenated to form a unified feature representation. The representation is processed by a residual block (f_1) and then by a channel attention block (f_2) to produce discriminative feature (y^t). Following this, one copy of y^t is merged with the previous image feature (b^t) via a skip connection (f_3), as indicated by the purple link in Fig. 3 (a), thereby producing the updated image feature (b^{t+1}). In parallel, another copy of y^t is channel-compressed by a 1×1 convolution h_1, which is subsequently added to the previous matrix path feature (i^{t-1}). Finally, a preactivation unit (h_2), consisting of unshared batch normalization followed by a tanh activation, generates the up-
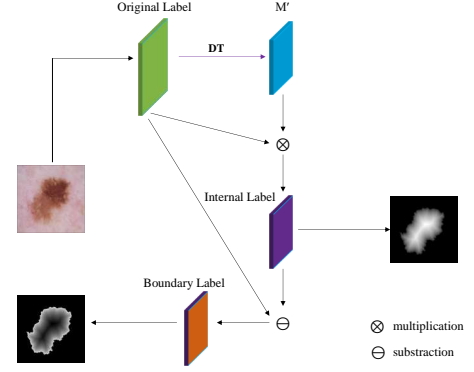

Fig. 4. Illustration of the decoupling of labels based on distance transforms.

dated matrix path feature (i^t). The update can be summarized as:

$$y^t = f_2\big(f_1(\text{Cat}(b^t, i^{t-1}))\big), \qquad (1)$$

$$b^{t+1} = y^t + b^t, \quad i^t = h_2\big(h_1(y^t) + i^{t-1}\big). \qquad (2)$$

IIM is instantiated in three variants for the encoder and the decoder, which mainly differ in the residual blocks and the skip-connection patterns illustrated in Fig. 3 (b). The encoder contains four layers with 3, 4, 6 and 3 IIMs, respectively. At each encoder stage, an E-IIM* first reduces the feature dimensions, followed by multiple E-IIMs to capture increasingly complex feature representations. In the decoder, a D-IIM is applied at each layer to progressively restore the image features.

3.3. Feature Enhancement Module

In medical image segmentation, lesions often have irregular shapes and may overlap with normal tissue, so reinforcing the network's focus on lesion regions is crucial for stronger geometric representations. To achieve this, we propose a FEM that amplifies the representation of critical lesion regions. As shown in Fig. 2, the output feature of the encoder is projected by two separate fully connected layers W_q and W_{kv} to generate the query Q, key K, and value V. Then Q is pooled to obtain an aggregated query feature A, and perform the first Softmax-attention operation involving A, Q, V and a learnable bias B_1, which yields an intermediate attention feature V_A. In the second attention stage, Q serves as the query and A serves as the key, while V_A is used as the value together with another learnable bias B_2, producing an enhanced feature O^A. In parallel, the value V is processed by a lightweight DBR module to reduce computational cost, consisting of a convolution layer, normalization, and an activation function. We then concatenate the DBR output with O^A to integrate complementary information. Finally, a KANLinear layer is placed at the end of FEM to further strengthen feature expressiveness and nonlinearity. The concatenated feature is passed through this KANLinear operator to produce the final FEM output O. The overall computation is formulated as:

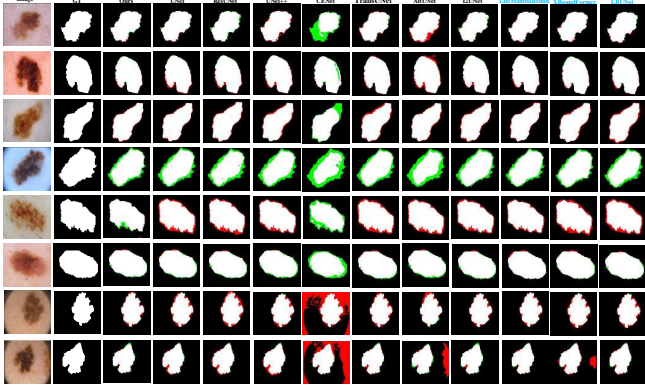$$O = KL(\sigma(QA^T + B_2)\,\sigma(AK^T + B_1)\,V + DBR(V)), \quad (3)$$

Fig. 5. Segmentation results of different methods on the ISIC2017 (rows 1–3), ISIC2018 (rows 4–6), and PH2 (rows 7–8) datasets. Methods shown in blue use boundary information to improve segmentation performance. Best viewed zoomed in.

Table 1. Segmentation methods on three skin-lesion datasets (best results in bold). The three methods below the black line are boundary-information–based methods.

Method	Pub. Year	ISIC2017		ISIC2018		ISIC2018 → PH2	
		mDice	mIoU	mDice	mIoU	mDice	mIoU
UNet [1]	MICCAI'15	0.919	0.799	0.901	0.769	0.911	0.827
ResUNet [4]	ITME'18	0.918	0.797	0.895	0.764	0.882	0.737
UNet++ [12]	DLMIA'18	0.821	0.743	0.794	0.729	0.789	0.713
CENet [14]	TMI'19	0.911	0.790	0.898	0.764	0.888	0.782
TransUNet [15]	arXiv'21	0.934	0.834	0.921	0.820	0.915	0.829
AttUNet [24]	MICCAI'21	0.713	0.419	0.704	0.442	0.707	0.448
I2UNet [25]	MedIA'24	0.919	0.832	0.916	0.831	0.907	0.825
UKAN [20]	AAAI'25	0.767	0.517	0.693	0.401	0.760	0.529
XBoundFormer [6]	TMI'23	0.936	0.830	0.922	0.808	0.923	0.846
LBUNet [7]	MICCAI'24	0.928	0.815	0.913	0.794	0.910	0.821
LiteMambaBound [23]	Methods'25	0.938	0.832	0.923	0.819	0.921	0.840
Ours	–	**0.947**	**0.844**	**0.938**	**0.835**	**0.925**	**0.873**

where σ denotes the Softmax attention function and KL denotes KANLinear.

3.4. Distance Transform-based Supervision Strategy

To achieve robust segmentation by explicitly distinguishing internal and boundary pixels, we employ DTSS illustrated in Fig. 2. DTSS leverages the decoupling of labels based on distance transforms (DT), as shown in Fig. 4, to provide targeted supervision for our model. Specifically, the input is a binary image M, divided into image of internal regions M_i and image of boundary regions M_b. Each pixel p has a value of 1 if $p \in M_i$ and 0 if $p \in M_b$. The DT computes the distance between pixels using function $f(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$. For internal region pixels, it finds the nearest boundary region pixel and calculates the distance. For boundary region pixels, the distance is set to zero. The distance-transformed image $M'(p)$ is given by:

$$M'(p) = \begin{cases} \min_{q \in M_b} f(p, q), & p \in M_i, \\ 0, & p \in M_b. \end{cases} \quad (4)$$

After obtaining M', we apply a simple linear mapping that rescales the original values to the [0, 1] range which normalizes the distances and makes them comparable across images. Unlike the original image, pixel values in M' depend on their relative positions, with internal pixels having the highest values
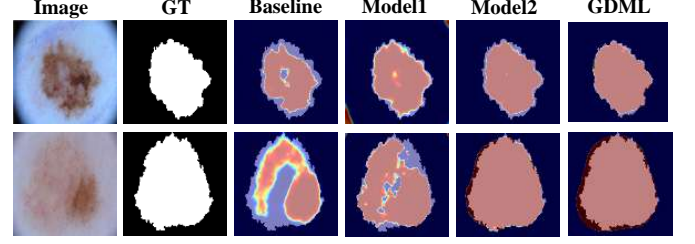

Fig. 6. Attention heat maps of ablation study segmentation results.

Table 2. The ablation results on ISIC2017 and ISIC2018.

Methods	IIM	FEM	DTSS	ISIC2017		ISIC2018	
				mDice	mIoU	mDice	mIoU
Baseline				0.872	0.761	0.861	0.759
Model1	✓			0.922	0.803	0.900	0.787
Model2	✓	✓		0.935	0.815	0.927	0.812
Model3 (Ours)	✓	✓	✓	**0.947**	**0.844**	**0.938**	**0.835**

and boundary pixels having the lowest. Subtracting the internal map from the normalized mask isolates the boundary regions, producing a boundary-specific label that emphasizes pixels farthest from the interior of the lesion. The internal-specific labels L_i and boundary-specific labels L_b are obtained from the original label (Mask) as follows:

$$L_i = \text{Mask} * M', \quad (5)$$

$$L_b = \text{Mask} * (1 - M'). \quad (6)$$

These labels are used for supervision in image path and matrix path, respectively. At the final layer of the GDML, predictions from both paths are fused to produce the final output, using the original label for supervision. By supervising both the image and matrix paths, DTSS encourages the network to learn rich intra–lesion representations and precise boundary representations in a coordinated manner, while the supervision on the fused result enforces global consistency between interior integrity and boundary continuity.

3.5. Loss Function

According to previous research, a major challenge in medical image segmentation is class imbalance [25]. We address it with a joint loss ($\mathcal{L}_{\text{total}}$) that combines Dice loss ($\mathcal{L}_{\text{Dice}}$) [26] and binary cross-entropy loss ($\mathcal{L}_{\text{BCE}}$) [27], supervised by **DTSS**. The formula is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{BCE}}, \quad (7)$$

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{total}_1} + \alpha\,\mathcal{L}_{\text{total}_2} + \beta\,\mathcal{L}_{\text{total}_3}, \quad (8)$$

where $\mathcal{L}_{\text{final}}$ denotes the overall loss in our model, $\mathcal{L}_{\text{total}_1}$, $\mathcal{L}_{\text{total}_2}$, and $\mathcal{L}_{\text{total}_3}$ are the losses computed by **DTSS** that supervise the final output, image path output, and matrix path output, respectively, while α and β are the weighting coefficients for the loss terms of the image path and the matrix path.

4. Experiments

4.1. Experimental Details

We evaluate our method on three challenging datasets: ISIC2017 [28], ISIC2018 [29], and PH2 [30]. For ISIC2017

Table 3. The analysis of weighting coefficients.

α	β	ISIC2017		ISIC2018	
		mDice	mIoU	mDice	mIoU
0.2	1.8	0.908	0.772	0.893	0.759
0.4	1.6	0.919	0.822	0.904	0.769
0.6	1.4	0.939	0.842	0.931	0.826
0.8	1.2	0.947	0.844	0.938	0.835
1.0	1.0	0.938	0.840	0.929	0.828
1.2	0.8	0.935	0.836	0.926	0.822
1.4	0.6	0.931	0.831	0.921	0.816
1.6	0.4	0.920	0.818	0.901	0.765
1.8	0.2	0.903	0.761	0.886	0.753

and ISIC2018, we use a 70% training, 10% validation, and 20% testing split. To assess generalization, the model trained on ISIC2018 is tested on the PH2 dataset, which contains 200 images. For consistency, we use the same experimental setup on the experiments of previous methods. All experiments are implemented with PyTorch 2.4.1 on an NVIDIA RTX 3090 GPU. The models use the Adam optimizer for training, with input images resized to 224×224. Training runs for 250 epochs with an initial learning rate of 1e-4 and a batch size of 4, with the weight decay set to 1e-8. During training, data enhancement strategies such as random horizontal flipping, brightness adjustment, and scaling are applied.

4.2. Analysis of Experimental Results

To evaluate GDML's performance, we report the segmentation performance on skin lesions and compare it with eleven other state-of-the-art models, including UNet [1], ResUNet [4], UNet++ [12], CENet [14], TransUNet [15], AttUNet [24], I2UNet [25], UKAN [20], XBoundFormer [6], LBUNet [7] and LiteMambaBound [23], three of which are boundary information–based segmentation methods. As shown in Table 1, GDML achieves the best performance across the two evaluation metrics on three skin lesion datasets. Its Dice and IoU on ISIC2017 and ISIC2018 exceed the second-best results by 0.9% and 1.2%, and by 1.5% and 1.6%, respectively, which strongly demonstrates GDML's superiority. Fig. 5 presents qualitative results. It is evident that our model excels at segmenting lesions with complex interiors and blurred boundaries, showing a clear advantage on such tasks and strong learning and generalization capability.

4.3. Ablation Study on GDML

To quantitatively evaluate the impact of each module and strategy of GDML, we conduct ablation experiments on the ISIC2017 and ISIC2018 datasets by incrementally adding each component. The results are summarized in Table 2 and Fig. 6.

Baseline: We adopt the ResNet-based encoder–decoder backbone from CENet as our baseline and build upon it to design our network.

Model1: The previous feature-extraction module is replaced by IIM and skip connections are between the corresponding layers of the encoder and the decoder. The mDice and mIoU scores increase by 5.0%, 4.2% and by 3.9%, 2.8% on the two datasets, demonstrating the powerful capability of IIM for feature extraction and information exchange.

Model2: We employ our proposed FEM to bridge the encoder and decoder in Model1. Compared with Model1, Model2

increases the two metrics by 1.3%, 1.2% and 2.7%, 2.5% on the two datasets respectively. Fig. 6 further reveals that internal noise is effectively compensated, corroborating the FEM's strong capability of enhancing geometric feature learning.

Model3 (Ours): We modify the output of Model2 to three and apply DTSS to supervise the image path and the matrix path separately. Model3 increases the two metrics by 1.2%, 2.9% and by 1.1%, 2.3% on the two datasets. Fig. 6 shows more complete internal predictions and markedly more robust boundary predictions, validating the effectiveness of DTSS.

4.4. Analysis of Loss Function Coefficients

To better balance the learning of internal and boundary features during training, we adjust the loss-function coefficients for the image and matrix paths (α and β in Eq. (8)). We then conduct ablation studies on α and β to examine their effects on training and performance. As shown in Table 3, the initial values of α and β are 0.2 and 1.8, respectively. We vary them in steps of 0.2 by increasing α and decreasing β. When α equals 0.8 and β equals 1.2, performance peaks on both datasets. Before this point performance rises, and after this point it declines. This suggests that a large gap between α and β, meaning an imbalanced emphasis on learning internal and boundary features during training, harms performance. At the optimum β exceeds α, suggesting that boundary features benefit from stronger supervision or a higher loss weight.

5. Conclusion

In this paper, we present GDML, a framework for robust and accurate skin lesion segmentation that addresses complex internal textures and ambiguous boundaries in medical images. GDML has three components that work together to enhance feature learning and segmentation. IIM enables rich and efficient information exchange between two parallel pathways for internal and boundary feature extraction, producing complementary representations. FEM dynamically prioritizes lesion regions and strengthens the ability to learn discriminative geometric representations from complex images. DTSS adjusts the network output branches to explicitly guide decoupled learning of internal and boundary pixels, yielding deeper task specific representations that support internal integrity and boundary continuity. Experiments on ISIC2017, ISIC2018, and PH2 show the state-of-the-art performance compared with existing methods and confirm robust learning, strong generalization across diverse lesion types, and effective balance between internal completeness and boundary precision. Future work will extend GDML to multimodal medical image segmentation and explore lightweight variants for clinical deployment in resource constrained environments.

References

[1] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.

[2] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999 (2018).

[3] D. Lin, Y. Li, T. L. Nwe, S. Dong, Z. M. Oo, Refineu-net: Improved u-net with progressive global feedbacks and residual attention guided local refinement for medical image segmentation, Pattern Recognition Letters 138 (2020) 267–275.

[4] X. Xiao, S. Lian, Z. Luo, S. Li, Weighted res-unet for high-quality retina vessel segmentation, in: 2018 9th international conference on information technology in medicine and education (ITME), IEEE, 2018, pp. 327–331.

[5] M. Gridach, Pydinet: Pyramid dilated network for medical image segmentation, Neural networks 140 (2021) 274–281.

[6] J. Wang, F. Chen, Y. Ma, L. Wang, Z. Fei, J. Shuai, X. Tang, Q. Zhou, J. Qin, Xbound-former: Toward cross-scale boundary modeling in transformers, IEEE Transactions on Medical Imaging 42 (6) (2023) 1735–1745.

[7] J. Xu, L. Tong, Lb-unet: A lightweight boundary-assisted unet for skin lesion segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024, pp. 361–371.

[8] D. Dai, C. Dong, S. Xu, Q. Yan, Z. Li, C. Zhang, N. Luo, Ms red: A novel multi-scale residual encoding and decoding network for skin lesion segmentation, Medical image analysis 75 (2022) 102293.

[9] M. Ali, H. Hu, T. Wu, M. Mansoor, Q. Luo, W. Zheng, N. Jin, Segmentation of mri tumors and pelvic anatomy via cgan-synthesized data and attention-enhanced u-net, Pattern Recognition Letters 187 (2025) 100–106.

[10] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Pranet: Parallel reverse attention network for polyp segmentation, in: International conference on medical image computing and computer-assisted intervention, Springer, 2020, pp. 263–273.

[11] R. Roy, T. Chakraborti, A. S. Chowdhury, A deep learning-shape driven level set synergism for pulmonary nodule segmentation, Pattern Recognition Letters 123 (2019) 31–38.

[12] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4, Springer, 2018, pp. 3–11.

[13] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, V. K. Asari, Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, arXiv preprint arXiv:1802.06955 (2018).

[14] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, Ce-net: Context encoder network for 2d medical image segmentation, IEEE transactions on medical imaging 38 (10) (2019) 2281–2292.

[15] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306 (2021).

[16] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: European conference on computer vision, Springer, 2022, pp. 205–218.

[17] T. Chen, C. Wang, Z. Chen, Y. Lei, H. Shan, Hidiff: Hybrid diffusion framework for medical image segmentation, IEEE Transactions on Medical Imaging (2024).

[18] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, X. Chen, Cpfnet: Context pyramid fusion network for medical image segmentation, IEEE transactions on medical imaging 39 (10) (2020) 3008–3018.

[19] J. Ruan, J. Li, S. Xiang, Vm-unet: Vision mamba unet for medical image segmentation, arXiv preprint arXiv:2402.02491 (2024).

[20] C. Li, X. Liu, W. Li, C. Wang, H. Liu, Y. Liu, Z. Chen, Y. Yuan, U-kan makes strong backbone for medical image segmentation and generation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, 2025, pp. 4652–4660.

[21] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, L. Shao, Et-net: A generic edge-attention guidance network for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, Springer, 2019, pp. 442–450.

[22] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, S. Escalera, Bi-directional convlstm u-net with densley connected convolutions, in: Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019, pp. 0–0.

[23] Q.-H. Ho, T.-T. Tran, V.-T. Pham, et al., Litemamba-bound: A lightweight mamba-based model with boundary-aware and normalized active contour loss for skin lesion segmentation, Methods 235 (2025) 10–25.

[24] S. Wang, L. Li, X. Zhuang, Attu-net: attention u-net for brain tumor segmentation, in: International MICCAI brainlesion workshop, Springer, 2021, pp. 302–311.

[25] D. Dai, C. Dong, Q. Yan, Y. Sun, C. Zhang, Z. Li, S. Xu, I2u-net: A dual-path u-net with rich information interaction for medical image segmentation, Medical Image Analysis 97 (2024) 103241.

[26] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), Ieee, 2016, pp. 565–571.

[27] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436–444.

[28] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, et al., Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), in: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), IEEE, 2018, pp. 168–172.

[29] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, et al., Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), arXiv preprint arXiv:1902.03368 (2019).

[30] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, J. Rozeira, Ph 2-a dermoscopic image database for research and benchmarking, in: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2013, pp. 5437–5440.