Expert Systems With Applications 285 (2025) 127668

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

EM-Net: Effective and morphology-aware network for skin lesion segmentation

Kaiwen Zhu [a,b,c,1], Yuezhe Yang [c,d,1], Yonglin Chen [a,c,*,1], Ruixi Feng [e], Dongping Chen [d], Bingzhi Fan [d], Nan Liu [d], Ying Li [f], Xuewen Wang [f]

[a] School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei, Anhui, PR China
[b] College of Software Engineering, Southeast University, Nanjing, Jiangsu, PR China
[c] Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, Hefei, Anhui, PR China
[d] School of Artificial Intelligence, Anhui University, Hefei, Anhui, PR China
[e] Stony Brook Institute at Anhui University, Hefei, Anhui, PR China
[f] Department of Plastic Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, PR China

ARTICLE INFO

Keywords:
Skin lesion segmentation
Deep learn network
Domain generalization
Morphology-aware

ABSTRACT

Dermoscopic images are essential for diagnosing various skin diseases, as they enable physicians to observe subepidermal structures, dermal papillae, and deeper tissues otherwise invisible to the naked eye. However, segmenting lesions in these images is challenging due to their irregular boundaries and significant variability in lesion characteristics. To address these challenges, we propose a effective and morphology-aware network that utilizes a hybrid feature extractor combining CNN and ViT architectures. At the same time, we enhanced the proposed segmentation model. Specifically, we propose a boundary delineation component that uses a non-convex optimization function for learning general representations and accurately delineates lesion boundaries, thus enhancing the extraction of details. Additionally, we also introduce an adaptive segmentation strategy through the integration of the few-shot domain generalization module to improve the model's generalization across different datasets. Validation on multiple publicly available dermoscopic image datasets, including ISIC, PH^2, PAD-UFES-20, and the University of Waterloo skin cancer database, demonstrates that our method achieves state-of-the-art performance with significant improvements in Dice, Acc, Pre, IoU, and Re. These results confirm the robustness and adaptability of our model. The code is available at: https://github.com/Bean-Young/EM-Net.

1. Introduction

Skin cancer is one of the most common and life-threatening cancers, with its incidence rising due to factors like greenhouse gas emissions and ozone layer depletion. This alarming trend underscores the critical importance of early detection, which studies show can reduce mortality rates by up to 97% (Crosby et al., 2022; Sethanan et al., 2023). Dermoscopy images play a vital role in early diagnosis and treatment. However, traditional analysis of these images relies heavily on clinicians' manual judgment, which is time-consuming, expertise-dependent, and prone to subjective biases. Advances in digital image processing, particularly in medical image segmentation, offer promising solutions (Dong et al., 2024; He, Wang et al., 2023).

Skin lesion segmentation methods can be broadly classified into four categories: (1) edge detection (Rajab et al., 2004), (2) threshold segmentation (Yogarajah et al., 2010), (3) active contour models (Riaz et al., 2018), and (4) deep learning-based approaches (Sikkandar et al., 2021). Edge detection and threshold segmentation utilize grayscale variations to delineate lesion boundaries. While computationally efficient, these methods struggle with low-contrast images and noise, making them less effective for lesions with blurred boundaries or subtle grayscale differences. Active contour models iteratively adjust contour points by balancing smoothness and edge attraction forces. Despite their improved precision, these methods depend heavily on initial contour settings and involve high computational complexity,as illustrated in Fig. 2.

* Corresponding author at: School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei, Anhui, PR China.
E-mail addresses: zhukaiwen2003@126.com (K. Zhu), yangyuezhe@gmail.com (Y. Yang), chenyonglin@ahjzu.edu.cn (Y. Chen), freshalways1024@163.com (R. Feng), wa2214134@stu.ahu.edu.cn (D. Chen), wa2214138@stu.ahu.edu.cn (B. Fan), wa2214138@stu.ahu.edu.cn (N. Liu), 22118173@zju.edu.cn (Y. Li), wangxuewen@zju.edu.cn (X. Wang).
[1] Equal Contribution.

https://doi.org/10.1016/j.eswa.2025.127668
Received 20 August 2024; Received in revised form 17 February 2025; Accepted 8 April 2025
Available online 30 April 2025

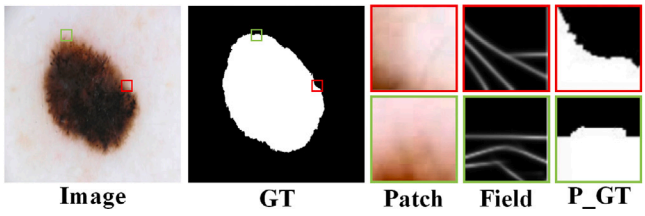

Fig. 1. Illustration of MM capturing morphological features in skin lesion images. The top and bottom images show the position of the red and green boxes, respectively. The P_GT represents the ground truth of this Patch.

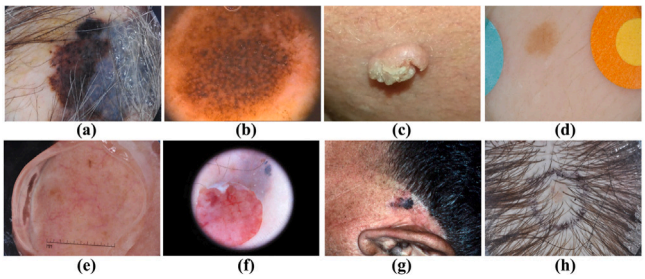

Fig. 2. The insufficient precision of segmentation results in skin lesion analysis, caused by (a) blurry boundaries, (b) variable lesion sizes, (c) shadow interference in imaging, (d) noise interference, (e) indistinct features, (f) feature variability, (g) unusual lesion locations, and (h) hair coverage, constitutes the main challenge in the task of skin lesion segmentation.

Deep learning-based methods have emerged as a robust alternative, leveraging advanced algorithms and parallel computing to deliver superior accuracy and reliability. The U-Net architecture, introduced by Ronneberger et al. (2015), is a benchmark in medical image segmentation. By integrating low- and high-resolution image features through skip connections, U-Net achieves exceptional performance. However, its reliance on convolutional operations limits its ability to capture global and long-range semantic information (Xu et al., 2023). Addressing this limitation, Transformer architectures (Vaswani et al., 2017) offer new possibilities for enhancing segmentation accuracy.

The Vision Transformer (ViT), developed by Dosovitskiy et al. (2020), leverages self-attention mechanisms to capture long-range dependencies, making it highly effective for skin lesion segmentation. Hybrid models like TransUNet (Chen et al., 2021), which combine ViT with convolutional networks, have shown significant promise. Research by Gulzar and Khan (2022) demonstrates that TransUNet outperforms purely convolutional models in segmenting skin lesions. However, challenges such as low contrast, irregular lesion boundaries, and hair occlusions persist. ViT's patch-based processing and lack of local inductive bias limit its ability to preserve fine-grained texture details, reducing its effectiveness for pixel-level segmentation.

Two critical challenges in deep learning-based segmentation are insufficient capture of local texture details and discrepancies in data distribution between source and target domains. Mathematical morphology methods (Verbin & Zickler, 2021) have been explored to address the first challenge, offering pixel-level precision in extracting shape, texture, and color features. Chatterjee et al. (2015) successfully integrated these methods for melanoma recognition. However, their standalone application to complex segmentation tasks is limited, necessitating integration with deep learning models.

The second challenge arises from the assumption that training (source) and testing (target) datasets share similar distributions. Variations in imaging technologies, equipment, and skin lesion characteristics often invalidate this assumption, adversely affecting model performance on cross-domain data. Domain adaptation techniques (Guan & Liu, 2021) have gained attention for addressing this issue. Adversarial domain adaptation methods (Scannell et al., 2021), in particular, have proven effective in handling minimal annotations across diverse data sources. However, significant domain shifts due to equipment differences, diverse data sources, and rare lesion samples pose ongoing challenges.

To address these limitations, this paper introduces a novel deep learning-based method for skin lesion segmentation. Our approach tackles the challenges of local texture capture and domain distribution discrepancies through advanced techniques, including adversarial domain generalization. This work aims to achieve robust and generalized segmentation of skin lesions, overcoming existing barriers to accuracy and reliability in clinical applications.

The main contributions of this paper are as follows:

- We have developed a contour information capturer named Morphology-aware Module (MM) that is highly sensitive to the morphological properties of dermoscopy images. This module uses non-convex optimization to effectively detect image features like edges and corners, as illustrated in Fig. 1.
- Our model introduces an adaptive segmentation strategy through the integration of the Few-shot Domain Generalization (FDG) module. Through the adversarial iterative optimization strategy, the module is able to fit different data distributions with minimal target domain data, thereby enhancing its generalization across diverse domains.
- Our model leverages a hybrid encoder–decoder architecture that integrates CNNs with Vision Transformers. In the feature encoding stage, CNNs are utilized to extract feature maps from input images. These feature maps are then refined using a Transformer module, effectively reducing noise. The decoding phase employs a cascading up-sampling module that progressively restores detailed features, thereby achieving precise lesion segmentation.
- We provide a set of segmentation masks derived from the PAD-UFES-20 dataset, which includes 30 dermatological images captured via cellphone, each exhibiting unique characteristics. These images, segmented under the supervision of dermatologists, are made available on our Github.

The structure of the paper is as follows: Section 2 reviews existing deep learning approaches for skin lesion segmentation. Section 3 details the architecture and methodology of the proposed network. Section 4 describes the experimental setup and results. Section 5 provides an in-depth discussion of the findings. Finally, Section 6 summarizes the study and its key contributions.

2. Related work

This section reviews deep learning models for skin lesion segmentation, categorizing them into two primary groups: methods based on convolutional neural networks (CNNs) and vision transformers (ViTs), and approaches leveraging image morphology information. Additionally, we explore the role of domain adaptation strategies in improving segmentation accuracy, providing a comprehensive overview of current techniques, their potential, and the challenges they face in this domain.

2.1. Skin lesion segmentation method based on CNN and vision transformer

Deep learning methods utilizing CNNs have significantly advanced skin lesion segmentation. However, these CNN-based methods often struggle to capture global features, leading to inaccuracies in segmentation boundaries.

To address this limitation, transformer-based architectures have been introduced due to their ability to model long-range dependencies. Vision Transformer (ViT) (Dosovitskiy et al., 2020) demonstrated an

effective balance between speed and accuracy for image classification but exhibited limitations in dense visual tasks. However, transformer-based models often lack the ability to capture local information, leading to a loss of detail in extracting skin lesion textures.

Recent advances have sought to combine the strengths of CNNs and transformers. EAAC-Net (Fan et al., 2024) employs an adaptive attention mechanism and convolutional fusion to improve local and global feature extraction. IEA-Net (Peng & Fan, 2024) integrates a dual attention mechanism to better handle complex scenarios. Rema-Net (Yang et al., 2023), a multi-attention CNN, reduces parameters by 40% compared to U-Net while maintaining high segmentation accuracy through streamlined spatial and reverse attention mechanisms. Despite these hybrid approaches, challenges such as low contrast, irregular lesion shapes, and occlusions like hair continue to hinder pixel-level precision. Our proposed hybrid structure aims to address these issues by leveraging the complementary strengths of CNNs and ViTs.

2.2. Skin lesion segmentation method based on image morphology information

Fuzzy boundaries and irregular lesion contours present significant challenges in dermatological image segmentation. Recent studies (Tong et al., 2021; Wang et al., 2021; Wu et al., 2020) have enhanced segmentation accuracy by focusing on boundary attention and correction techniques. ADFFNet (He, Li et al., 2023) introduces a Boundary Refinement (BR) module that uses global attention to merge semantic and detail features, achieving pixel-level segmentation. DGCU-Net (Ramadan & Aly, 2022) integrates gradient information to improve boundary and texture accuracy. XBound-Former (Wang et al., 2023) addresses regional variations and boundaries using a cross-scale, boundary-aware mechanism.

To overcome the limitations of traditional boundary attention, researchers have turned to mathematical modeling. Verbin and Zickler (2021) applied non-convex optimization for image analysis, while Polansky et al. (2024) refined this approach to address weak boundary detection and limited training data effectively. Inspired by these efforts, we incorporated non-convex optimization techniques into our model to improve segmentation accuracy under challenging conditions, such as low contrast, irregular boundaries, and hair occlusions.

2.3. Skin lesion segmentation method based on domain adaptation and generalization

Medical image segmentation faces significant challenges due to domain shifts caused by variations in imaging devices, lighting, and patient demographics (Tzeng et al., 2015). While deep learning models perform well on specific datasets, their performance often deteriorates in new domains. Domain adaptation methods have been developed to address this issue. For instance, domain-adversarial learning has been used to train domain-invariant U-Net models for robust cardiac structure segmentation across different MRI scanners. Li et al. (2021) proposed a framework based on generative adversarial networks (GANs) to enhance pixel-level tasks, demonstrating strong performance in both in-domain and out-of-domain segmentation tasks.

Domain generalization methods have recently gained traction in skin lesion segmentation. EPVT (Yan et al., 2023) integrates domain-specific and shared prompts within a vision transformer to improve performance across diverse environments. Wang et al. (2022) introduced a cross-domain few-shot segmentation framework to address rare disease segmentation using limited data. By leveraging meta-training, this framework improves generalization from common to rare skin diseases.

Despite these advancements, significant domain shifts in skin lesion images, caused by lighting variations, skin color differences, and device inconsistencies, continue to challenge model robustness. Enhancing domain generalization techniques to address these unique characteristics could significantly improve segmentation performance and reliability.

3. Method

3.1. Overview

This section presents a novel model for skin lesion segmentation, as illustrated in Fig. 3. The segmentation process for dermoscopic images of lesion tissues is organized into three stages: feature encoding and decoding, morphology-aware refinement, and few-shot domain generalization.

In the encoding phase, the model employs a parallel CNN-Transformer hybrid encoder. The CNN extracts shallow image features, which are then processed by the Transformer to capture global features using a self-attention mechanism. During decoding, a cascaded up-sampler is used to recover the segmentation mask of the lesion region. This up-sampler comprises multiple modules that progressively refine and reconstruct the lesion's contours and boundaries.

To enhance the accuracy of morphological contour detection and ensure precise segmentation, the model incorporates non-convex optimization techniques. These include image contour node extraction, color normalization, and node field constraints, all of which improve the decoder's ability to interpret pixel-level information within the lesion region.

Additionally, the model integrates a domain generalization learning framework to address variability across different datasets. This framework enhances the model's capability to adapt to new domains, thereby improving its robustness and generalization in cross-domain segmentation tasks.

Further details on the network's architecture, the morphology-aware module, and the few-shot domain generalization module are provided in Subsections B, C, and D, respectively.

3.2. Main module in the network

3.2.1. Parallel CNN-transformer hybrid encoder

Although the combination of a visual Transformer with a plain upsampling module has enhanced image segmentation performance, it fails to meet the accuracy requirements for skin lesion segmentation in clinical practice. This occurs because the resolution $\frac{H}{P} \times \frac{W}{P}$ is significantly smaller than the original $H \times W$, resulting in the loss of crucial texture details such as skin lesion shapes and boundaries. To address this loss of detail, this study proposes a parallel CNN-Transformer hybrid architecture as an encoder, designed to enhance the extraction of skin lesion features. This architecture effectively recovers and refines complex features through a parallel CNN-Transformer structure, leveraging high-resolution feature maps from CNNs in the decoding stage to enhance the upsampling process. Furthermore, the hybrid encoder melds the CNN's ability to capture local details with the Transformer's proficiency in processing global information, yielding superior performance over either a standalone CNN or a sole Transformer encoder.

In the encoding stage, the dual parallel CNN serves as a feature extractor, generating the feature mapping of the input image, with the captured shallow local features being reshaped and fed into the ViT. Consequently, in the Patch embedding stage of the ViT, the extracted 1×1 Patch from the CNN feature mapping is embedded instead of the original image. Meanwhile, the constructed bridge connection module integrates the features extracted from the main path-branch CNNs and links them to the upsampler of the corresponding layer through double-attention jump connections.

3.2.2. Bridge fusion modules

In the feature fusion stage, three types of bridge connection modules

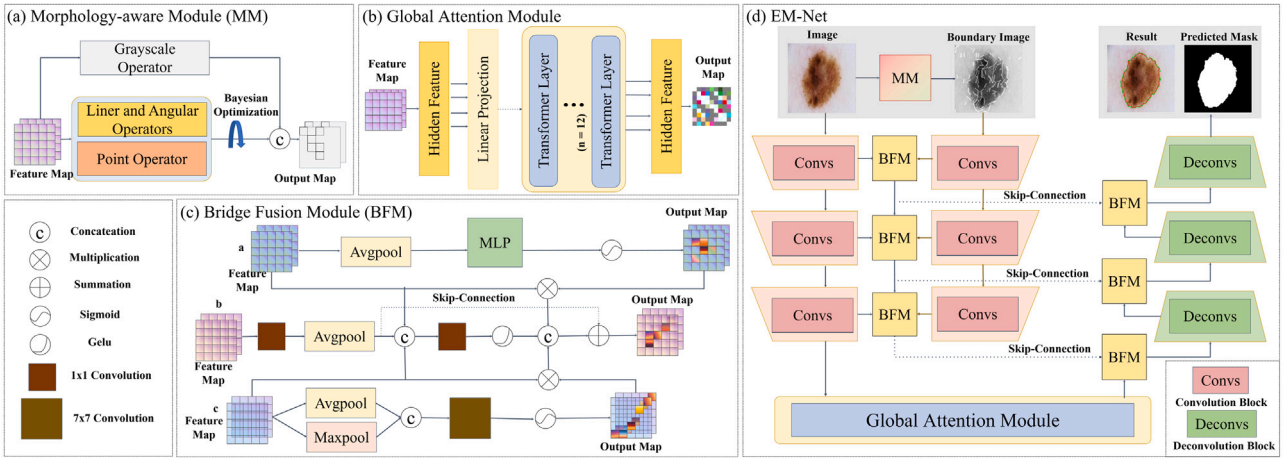

Fig. 3. Figure (a) represents the Morphology-aware Module, Figure (b) represents the Global Attention Module, Figure (c) represents the Bridge Fusion Module, and Figure (d) is the main framework of our net. The proposed net comprises two paths. Initially, the input image is processed by the Morphology-aware Module to generate the boundary information image. Both are then inputted into the parallel encoder, which produces shallow semantic features. Subsequently, the original image undergoes further processing in the Global Attention Module, enhancing the extraction of high-level semantic information. The net integrates these semantic layers by cascading the Up-sampling module with jump connections from the shallow semantics to generate predictive image features. These features are then inputted into the Few-shot Domain Generalization Module, which employs domain feature extractor iterative learning. Finally, the activation function is applied to these predictive features to produce the final segmentation mask.

are integrated: the Weight Multiplication Layer Module, the Three-channel Bridge Fusion Module, and the Two-channel Bridge Fusion Module.

The Weight Multiplication Layer Module activates image features in the first layer using the formula:

$$F = (F_{\mathrm{m}} \times F_{\mathrm{b}}) \times w + b, \tag{1}$$

where F_{m} and F_{b} represent original and obtained through the MM module feature vectors, respectively; w is the weight parameter; and b is the bias. This layer enhances lesion detection by emphasizing critical features and suppressing noise, thereby improving the model's performance and generalization capabilities.

The structural details of the Bridge Fusion Module are shown in Fig. 3(c). Unlike the Three-channel Bridge Fusion Module, the Two-channel Bridge Fusion Module does not have Branch b. Bridge Fusion Module integrates shift window self-attention from the Transformer with capabilities from boundary detail extraction. Together with the Channel Attention and Spatial Attention modules, it enhances semantic and local feature representations. This fusion process, aimed at capturing both global and local information efficiently, the outputs from the channel, spatial, and boundary fusion processes, combined to form the integrated feature.

3.2.3. Cascaded upsampler and skip-connection module

In the upsampling phase of dermoscopic image analysis, a cascade up-sampler (CUP) is employed to sequentially restore image features and generate precise segmentation masks. Initially, the feature sequence $\mathbf{z}_L \in \mathbb{R}^{\frac{HW}{P^2} \times D}$, derived from downsampling, is reshaped to $\frac{H}{P} \times \frac{W}{P} \times D$. The CUP utilizes multiple cascading upsampling modules, each containing a bilinear difference operator, a 3 × 3 convolutional layer, and a ReLU layer. The integration of cascaded upsamplers with a hybrid encoder forms a U-shaped architecture, which enhances feature integration across various resolutions via jump connections and facilitates feature fusion. The comprehensive architecture of the CUP, including the intermediate jump connections, is illustrated in Fig. 3. Additionally, a Bridge Fusion Module is introduced between the encoder and decoder to address the imprecision in segmentation results due to the lack of shallow high-frequency details. This module preserves high-resolution information from high-level feature maps, enabling more accurate restoration of the original image's detailed features and improving the segmentation accuracy of dermoscopic images.

3.3. Morphology-aware module

3.3.1. Description of the field of junctions

The complex combination of texture and detail information in dermoscopic images makes it impossible to simply define the boundary information in the images. In order to be able to recognize multiple boundary elements in dermatological images, we also utilize the "generalized M-junction" structure (i.e., consisting of M corners and one movable vertex) to describe the intersection field, and to achieve a unified characterization of the contours, corners, junctions, and homogeneous regions of the dermoscopic images.

Image Patch Extraction: To address the variability in imaging quality of dermoscopic images arising from different devices and methods, all images are reshaped to a uniform size of 1024 × 1024. Each dermoscopic image I, uniformly resized to 1024 × 1024, is segmented into 64 × 64 patches, collectively denoted as $\mathcal{J}_{64} = \{I_i(\mathbf{x})\}_{i=1}^{256}$. Consequently, each dermatological image is defined as a collection of 256 low-quality image blocks.

Parameterized Boundary Model: The boundary structure of each 64 × 64 image block is characterized by a continuous family of patch types, $\mathcal{P}_{64} = \{\mathbf{u}_\theta(\mathbf{x})\}$, which are parameterized with θ. Unlike natural images, dermoscopic images often feature multiple intersecting boundaries and lines, necessitating the use of additional angular wedges to capture details and boundary information more accurately. The diversity of boundaries and structures in dermoscopic images mandates higher M values to ensure comprehensive description of all pertinent boundaries. Consequently, for $\mathcal{P}_R$, the decagonal intersection model is employed, comprising 10 angular wedges around a vertex. The parameter set $\boldsymbol{\theta} = \left(\phi, x^{(0)}\right) \in \mathbb{R}^{12}$ includes 10 corners $\boldsymbol{\phi} = \left(\phi^{(1)}, \ldots, \phi^{(10)}\right)$ and the vertex position $\mathbf{x}^{(0)} = \left(x^{(0)}, y^{(0)}\right)$.

3.3.2. Non-convex optimization boundary extraction

In the task of skin lesion image segmentation, recognizing the blurred lesion regions and accurately extracting boundaries pose significant challenges. To enhance segmentation accuracy and delineate the fuzzy lesion regions precisely, a non-convex optimization-based image boundary extraction method is introduced, building upon the prior definition.

Definition of Boundary Optimization Function: We optimize the posterior probability by integrating a probabilistic model that combines prior image information with observed data. This model merges global and local image details. The analytical process of interpreting the image

data onto the nodal field is formulated as solving an optimization problem:

$$\max_{\Theta,C} \log p(\Theta) + \log p(\mathbf{C}) + \sum_{i=1}^{N} \log p(I_i|\theta_i, \mathbf{c}_i), \tag{2}$$

where p(Θ) and p(C) represent the spatial consistency terms for the knot parameters $\Theta = (\theta_1, \ldots, \theta_N)$ and the color function $\mathrm{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_N)$, respectively, and $p(I_i|\theta_i, \mathbf{c}_i)$ denotes the likelihood function of the patch I_i given the knot parameters and the color function $\mathbf{c}_i = \left(c_i^{(1)}, \ldots, c_i^{(M)}\right)$. By constructing an optimized objective function as a weighted sum, we achieve the goal of minimizing the discrepancy between the reconstructed image and the original, while ensuring uniformity across individual image patches. Consequently, the analysis of the image's nodal field is transformed into solving a minimization problem:

$$\begin{aligned}\min_{\Theta,C} &\sum_{i=1}^{N}\sum_{j=1}^{M} \int u_{\theta_i}^{(j)}(\mathbf{x}) \left\| c_i^{(j)} - I_i(\mathbf{x}) \right\|^2 d\mathbf{x} \\ &+ \lambda_B \sum_{i=1}^{N} \int \left[B_i^{(\delta)}(\mathbf{x}) - \hat{B}_i^{(\delta)}(\mathbf{x}) \right]^2 d\mathbf{x} \\ &+ \lambda_C \sum_{i=1}^{N}\sum_{i=1}^{M} \int u_{\theta_i}^{(j)}(\mathbf{x}) \left\| c_i^{(j)} - \hat{I}_i(\mathbf{x}) \right\|^2 d\mathbf{x},\end{aligned} \tag{3}$$

where $\hat{I}_i(\mathbf{x})$ denotes the ith patch of the global color map. λ_B and λ_C are parameters controlling the strength of the boundary and color consistency. The $B_i(\mathrm{x})$ denotes the boundary mapping at the ith patch, returning 1 if x qualifies as a boundary position according to θ_i, and 0 otherwise. Additionally, $\hat{B}(\mathrm{x}) = \max_{i\in\{1,\ldots,N\}} B_i(\mathrm{x})$, $B_i(\mathrm{x})$ represents the global boundary mapping as defined by the node field. The $B_i^{(\delta)}(\mathbf{x})$ is a smooth boundary mapping characterized by a falloff width of δ from the exact boundary position. The relaxed global boundary mapping $\hat{B}_i^{(\delta)}(\mathbf{x})$ is computed by averaging the smooth local boundary mappings for each position $\mathbf{x}$ across all patches that contain it.

Boundary Information Extraction: By alternatingly solving Problem (3), the node parameters and colors (Θ, C) are updated, assuming the global mapping $\left(\hat{B}^{(\delta)}, \hat{I}\right)$ is fixed, and subsequently, the global mapping is updated with the expression:

$$c_i^{(j)} = \frac{\int u_{\theta_i}^{(j)}(\mathbf{x}) \left[I_i(\mathbf{x}) + \lambda_C \hat{I}_i(\mathbf{x}) \right] d\mathbf{x}}{(1+\lambda_C) \int u_{\theta_i}^{(j)}(\mathbf{x}) d\mathbf{x}}. \tag{4}$$

Our boundary consistency formula, which ensures alignment between each patch and its overlapping neighbors, suppresses false boundaries by reducing the boundary strength $B_i^{(\delta)}(\mathbf{x})$ for pixels x assigned low scores by their neighbors (as quantified by $\hat{B}_i^{(\delta)}(\mathbf{x})$), and enhances the boundary strength on pixels assigned high scores. Additionally, a smooth consistency strategy is employed such that contours with non-zero curvature are effectively approximated by localized sets of corners with slightly different vertices. The resulting dermoscopic image boundary information effectively characterizes the indicated texture structure and complex biology of the dermatosis and enhances the model's ability to localize more efficiently to the lesion boundary region. Thus far, the rendering obtained by our MM module is shown in Fig. 1.

3.3.3. Fusing image grayscaling information

Given that dermoscopic images feature complex color and interference information that distinguishes them from other medical images, it remains challenging to accurately distinguish and localize lesion areas and interference solely from boundary information. Consequently, we have integrated the visual characteristics of the human eye with dermoscopic imaging features, and incorporated gray scale information from dermoscopic images into the obtained boundary images at a specific ratio.

Overall, the procedure through which the Morphology-aware Module captures the texture details of skin lesions is outlined in Algorithm 1.

Algorithm 1 The Pseudo-code for the MM-Module

Input: Original image I and parameters $(M, J, P, \ldots)$
Output: Generated image I_m

Define various parameters: $\phi^{(j)}, M, x, y, N_{\text{init}}$
Image preprocessing: $I' \leftarrow \text{enhance}(I)$
Segment the image into patches: $\{I_i(\mathbf{x})\}_{i=1}^{256} \leftarrow I'$
Define θ and $\mathbf{c}_i$: $\theta \leftarrow \left(\phi^{(1)}, \ldots, \phi^{(10)}, x^{(0)}, y^{(0)}\right)$
$\mathbf{c}_i \leftarrow \left(c_i^{(1)}, \ldots, c_i^{(M)}\right)$
for $j = 1$ to N **do**
  Optimization of angles: $\phi^{(j)} \leftarrow \underset{\phi}{\operatorname{argmin}}\, \ell_j(\phi)$
  Optimization of vertex: $x^{(0)} \leftarrow \underset{x}{\operatorname{argmin}}\, \ell(\boldsymbol{\phi}, x, y^{(0)})$
  $y^{(0)} \leftarrow \underset{y}{\operatorname{argmin}} \ell(\boldsymbol{\phi}, x^{(0)}, y)$
Splice patches to form an image: $\hat{I} \leftarrow \{\hat{I}_i(\mathbf{x})\}_{i=1}^{256}$
Form the final image $I_m \leftarrow \hat{I} + I \cdot w$

3.4. Few-shot domain generalization module

In publicly available skin lesion image datasets, there are notable differences in characteristics between images of skin lesions from different diseases. In the absence of specialized imaging equipment, such as dermatoscopes, the low-quality images obtained exhibit noticeable domain bias, which may impair the effectiveness of existing skin lesion segmentation networks. This paper introduces a cross-domain generalization strategy for skin lesion image segmentation, utilizing a minimal number of samples. The strategy aims to train a model on a specific dermatoscope image dataset (source domain) to accurately segment images across multiple target domains that have varying feature distributions. The proposed training approach involves joint training with source domain data and a minimal amount of target domain data, followed by testing across multiple target domains. The module sketch is shown in Fig. 4.

The input to the module is processed by the EM-Net, which is mainly divided into two stages: capturing initial high-level semantic features through downsampling and then generating predicted image features through upsampling. These are the inputs to the feature extractor. The feature extractor obtains high-level features through convolutional layers and a self-attention module, with a domain discriminator that distinguishes the source domain from the target domain by identifying differences between domains.

Specifically, after the features of the two inputs undergo convolution operations, the high-level semantic features are processed through a self-attention mechanism and residual connection structure to capture global context and enhance feature representation, thereby improving the module's performance. The resulting high-level features are then fed into the domain discriminator, which determines whether they belong to the source domain or target domain based on domain labels, resulting in different processing pathways.

For source domain data, the discriminator categorizes the features into seven different classes, which are quantified for numerical analysis and used for optimizing the segmentation network. Therefore, we can define w to represent the specificity of samples within the source domain. The specific loss backpropagation formula is as follows:

$$\theta_S \leftarrow \theta_S - \lambda_S \frac{\partial \mathcal{L}_S}{\partial \theta_S}, \tag{5}$$

$$\theta_S \leftarrow \theta_S + \alpha \lambda_S f(\omega), \tag{6}$$

$$f(\omega) = \alpha(1-\omega), \tag{7}$$

where θ_S, λ_S, and $\mathcal{L}_S$ denote the parameters of the feature maps, the segmentation network weights, and the loss function, respectively; α is

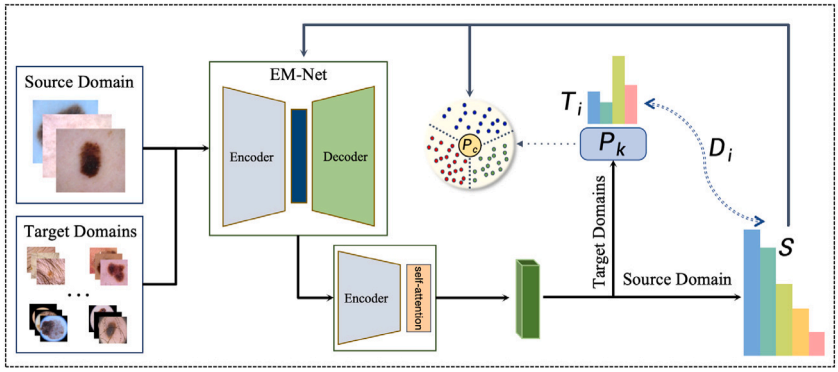

Fig. 4. Few-shot Domain Generalization module: The input to this module consists of source domain data and a minimal amount of data from various target domains. With high-level semantic features and segmentation prediction features obtained through EM-Net, the module outputs domain classification and distribution information acquired through a series of encoders. By assessing the distributional differences between the source and target domains, the resulting loss is backpropagated to the segmentation net.

a constant between 0 and 1, representing the updated network weights each cycle.

On the other hand, for the minimal amount of target domain data, which is extremely valuable, we use the target domain data to directly optimize the segmentation net. By calculating the distribution differences between the target domain data and the source domain data, we can obtain D_i to represent the distribution status among the various domains. The specific formula is as follows:

$$D_i(S,T_i) = \|S - T_i\|_F = \sqrt{\mathrm{tr}((S-T_i)^\top (S-T_i))}, \tag{8}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, S and T_i represent the typical distributions of skin lesion images in the source domain and the ith target domain, respectively.

To ensure that the model can effectively generalize across different target domains while preventing overfitting when handling few-shot tasks, we propose a subspace freezing strategy for few-shot domain generalization. When faced with a new domain's few-shot task, the feature extractor is frozen, meaning that the parameters related to the feature extractor are not updated. The feature space extracted by the feature extractor is learned from a broader domain dataset and has good generalization ability, thus further adjustment is unnecessary in few-shot tasks.

We choose to optimize only the projection subspace P_k specific to the domain and the corresponding domain classifier. Specifically, for each new task T_i, an appropriate subspace P_k needs to be found on the source domain which illustrate the support set S, in order to mitigate the risk of overfitting. This subspace is optimized by constraining the distance to the central subspace P_C and by including a regularization term in the objective function. This approach ensures that the subspace can adapt to the features of the new task while avoiding overfitting to the limited training data.

To this end, a specific optimization objective function is introduced in the paper:

$$\min_{\mathbf{P}_k,\mathbf{D}_i} \mathcal{L}_S(\mathbf{P}_k,\mathbf{D}_i) + \frac{\lambda}{2}\|\mathbf{P}_k - \mathbf{P}_C\|_F^2, s.t. \mathbf{P}_k \in \mathcal{A}, \tag{9}$$

where $\mathcal{A}$ is the set whose elements satisfy the properties of $n \times n$ orthogonal projections.

In summary, we developed Algorithm 2 to achieve few-shot domain generalization for skin lesion segmentation.

Algorithm 2 The Pseudo-code for the domain generalization strategy

Input : Predictive image feature set X, High-level semantic feature set Y, Labels of domains n_i

for each input pair (x, y) in (X, Y) **do**
  $x_{feature} \leftarrow$ ProcessLayers(x, 4 epochs);
  $y_{feature} \leftarrow$ ProcessLayers(y);
  $com_{features} \leftarrow$ Concatenate($x_{feature}, y_{feature}$);
  $High_{features} \leftarrow$ Downsample($com_{features}$);
for F_i in $High_{features}$ **do**
  if n_i is Source Domain **then**
    Mapping to w and backpropagation;
  else if n_i in Target Domains **then**
    $D_i \leftarrow$ L2 norm(F_i, $\mathrm{High}_{\mathrm{source}}$);
    $P_k \leftarrow$ Compute the center of the distribution;
    Optimization and backpropagation;

4. Experiment

4.1. Dataset

4.1.1. ISIC (Codella et al., 2018, 2019; Gutman et al., 2016; Tschandl et al., 2018)

The ISIC dataset, published by the International Skin Imaging Collaboration (ISIC), offers a substantial dermal lesion segmentation resource. ISIC 2016 comprises 900 training images and 379 test images, with ground truth annotations for all. ISIC 2017 contains 2000 training images and 600 test images, also fully annotated. ISIC 2018 includes 2584 training images and 1000 test images. The dataset features a diverse array of skin lesions, primarily focusing on melanoma, with both cancerous and non-cancerous annotations provided.

4.1.2. PH^2 (Mendonça et al., 2013)

The PH^2 dermoscopy image dataset, acquired using the Tuebinger Mole Analyzer System under standardized conditions at Hospital Pedro Hispano, provides 200 dermoscopic images of melanocytic lesions. This dataset includes 80 common nevi, 80 atypical nevi, and 40 melanomas, ensuring uniformity in magnification and acquisition settings. Medical annotations accompany all images, detailing clinical and histological diagnoses, as well as assessments of various dermoscopic criteria.

4.1.3. PAD-UFES-20 (Pacheco et al., 2020)

The PAD-UFES-20 dataset features skin lesion images captured using various smartphone devices, collaboratively compiled by the Dermatology and Surgery Assistance Program of the Federal University of Espírito Santo (UFES), Brazil. This dataset includes 2298 samples representing six different types of skin lesions, such as basal cell carcinoma (BCC), squamous cell carcinoma (SCC), actinic keratosis (ACK), seborrheic keratosis (SEK), Bowen's disease (BOD), melanoma (MEL), and nevi (NEV). Each sample includes a clinical image and up to 22 clinical features, such as patient age, lesion location, Fitzpatrick skin type, and lesion diameter.

4.1.4. University of Waterloo dataset (DermIS, 2012; DermQuest, 2012; Vision and Image Processing Lab, University of Waterloo, 2021)

The University of Waterloo's skin cancer detection study utilizes images from the public databases DermIS and DermQuest, supplemented with manually segmented lesion sites. This dataset consists of 206 images, each with corresponding ground truth annotations, supporting risk assessment for melanoma based on dermatologic photographs.

4.2. Evaluation metrics

We predominantly utilized five established metrics to quantitatively assess skin lesion segmentation performance: Dice coefficient (Dice), Intersection over Union (IoU), Accuracy (Acc), Precision (Pre), and Recall (Re). The calculation formulas are:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{10}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{11}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \tag{12}$$

$$Pre = \frac{TP}{TP + FP}, \tag{13}$$

$$Re = \frac{TP}{TP + FN}, \tag{14}$$

where TP represents true positives (correctly identified lesion pixels), FP represents false positives (non-lesion pixels incorrectly predicted as lesions), TN represents true negatives (correctly identified non-lesion pixels), and FN represents false negatives (actual lesion pixels missed by the model).

4.3. Implementation details

We conducted extensive experiments on the dataset. For the ISIC challenge, we strictly adhered to the training and test sets provided by the competition. For the ISIC 2018 dataset, the largest in volume, we performed self-validation tests, using 2076 of the 2584 training set images for training and the remaining 518 images for testing to confirm our model's feasibility. For the PH^2 dataset, we employed 900 ISIC 2016 training images, randomly selecting 10 for validation and the remaining 190 for testing, demonstrating that our model achieves excellent generalization with minimal target domain data. Similarly, for PAD-UFES-20, we randomly selected 5 images for validation and the remaining 25 for testing. For the University of Waterloo dataset, we selected 10 images for validation and the remaining 196 for testing. This approach demonstrates that our model is highly generalizable, independent of the imaging device and the resolution of the images.

All experiments were conducted on an NVIDIA RTX 3090 GPU with 24 GB of RAM, using Python 3.8 and PyTorch 2.0. The input image size was set to 224 × 224, and the patch size was set to 4. Weights pre-trained on ImageNet were used to initialize model parameters. During training, the batch size is 12, and we used a widely-used SGD optimizer with a momentum of 0.9 and weight decay of 1e−4 for model optimization during backpropagation.

4.4. Comparisons with state-of-the-art methods

To validate the performance of our proposed model, we conducted comparisons with several representative segmentation networks across public datasets, including ISIC 2016, ISIC 2017, ISIC 2018, PH^2, PAD-UFES-20, and the University of Waterloo. The compared models encompassed various popular segmentation architectures, including CNN-based U-Net (Ronneberger et al., 2015), Att-UNet (Oktay et al., 2018), CE-Net (Gu et al., 2019), CPF-Net (Feng et al., 2020), MS RED (Dai et al., 2022), FAT-Net (Wu et al., 2022), GA-Net (Zhou et al., 2023), CPF-Net (Chen et al., 2024), and ViT coding-based TransUNet (Chen et al., 2021), as well as Swin-UNet (Cao, Wang et al., 2022), alongside specialized dermatological segmentation network ICL-Net (Cao, Yuan et al., 2022). We selected U-Net as the baseline model. All compared models were trained using data from the original publications or under experimental conditions identical to our model.

4.4.1. Analysis of test results on the ISIC dataset

Quantitative comparisons between our model and other skin lesion segmentation methods on ISIC 2016, 2017, and 2018 datasets are presented in Table 1. In particular, Table 2 shows the experimental results comparing our method with other models on the ISIC 2018 Challenge ranking list. Our experiments indicate that our model consistently outperforms others across all test sets.

For the ISIC 2016 dataset, our results demonstrate superiority in IoU, precision, and accuracy. Specifically, our model improved Dice score by 1.60% and accuracy by 4.36% over the baseline. The IoU, a critical metric in segmentation tasks, measures the overlap between predicted and ground truth areas, and in this context, our model excels by an additional 2.99% over the baseline. Notably, while both IoU and Dice coefficient assess segmentation performance, they differ in their formulations: the Dice coefficient focuses more on the similarity between the predicted and true positive regions, while IoU accounts for false positives as well. This distinction is essential for understanding our model's performance in identifying overlapping regions effectively.

The improvement of 1.60% in Dice score indicates a significant enhancement in accurately identifying lesion areas, which can directly benefit clinical decision-making by enabling earlier detection and more effective treatment of skin lesions. Similarly, the 2.99% improvement in IoU suggests that our model better manages false positives, thereby reducing the risk of misdiagnosis and ensuring patient safety.

The ISIC 2017 dataset, known for its complexity, posed no challenge for our method, achieving the highest scores in Dice, accuracy, and IoU. Notably, our model outperformed the baseline by 3.62% in Dice score, showcasing superior generalization abilities.

Regarding the ISIC 2018 dataset, our model adhered to cross-validation methods used by other researchers and achieved excellent results in both training set validation and comparison against publicly available test data. In training set cross-validation, our model secured the highest scores in Dice and IoU. When validating against the test set, we compared our results against all public methods' net test data, achieving state-of-the-art in Dice score and surpassing the previous state-of-the-art in accuracy. Differences in testing environments reported by other researchers may explain slight deviations in IoU compared to the previous optimal model.

In practical application scenarios, the reported improvements (e.g., 1.60% in Dice score and 2.99% in IoU) signify meaningful enhancements in segmentation accuracy, which can greatly impact clinical decision-making. These metrics illustrate our model's effectiveness in real-world tasks, ensuring better detection and treatment planning for skin lesions.

Qualitative comparisons between our approach and others for the ISIC 2016, 2017, and 2018 datasets are illustrated in Fig. 5. Segmentation results are depicted with red lines superimposed on dermoscopy images bordered in green, showing our method's superiority in segmentation accuracy, particularly in challenging scenarios. Our method achieves better outcomes even in extreme cases, indicating its effectiveness in learning skin damage-related features and adapting to diverse segmentation tasks.

4.4.2. Analysis of test results on the PH^2

Table 3 presents a quantitative comparison between our method and other skin lesion segmentation methods on the PH^2 dataset. CE-Net exhibits poor generalization ability on the PH^2 test set, while FAT-Net shows inadequate performance on the ISIC 2016 test set. Our superior performance on the PH^2 dataset, where samples were not visible during model learning, demonstrates satisfactory generalization

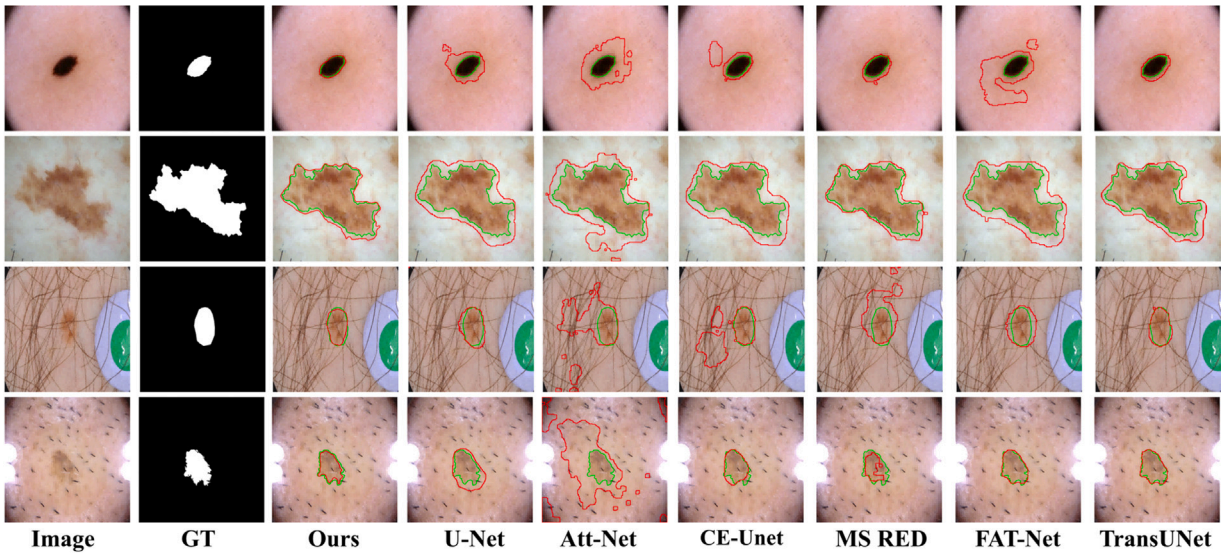

Fig. 5. Visual segmentation performance comparison of our net and some representative methods on the ISIC dataset. Green contours represent the ground truth. Red contours represent the segmentation results of different methods.

Table 1
This study compares our method with alternative approaches using ISIC datasets of varying vintages.

Method	*test*-ISIC 2016					*test*-ISIC 2017					*validation*-ISIC 2018				
	Dice	IoU	Acc	Pre	Re	Dice	IoU	Acc	Pre	Re	Dice	IoU	Acc	Pre	Re
U-Net	90.29	82.58	94.76	91.29	87.72	82.80	73.30	92.30	89.07	79.30	88.43	79.71	93.16	89.46	85.89
Att-UNet	90.83	83.43	94.43	92.12	88.42	83.20	74.40	92.50	87.29	84.11	88.32	79.52	93.17	87.73	87.11
CE-Net	91.80	85.01	95.30	91.84	90.32	85.61	77.54	93.51	89.91	84.52	89.45	81.32	94.03	89.18	87.60
CPF-Net	91.34	84.24	95.18	91.34	89.78	84.70	76.20	93.00	86.67	83.33	88.86	80.45	94.68	88.44	87.57
MS RED	91.43	84.43	95.67	91.60	89.53	84.83	76.32	93.10	87.07	83.74	89.18	80.92	95.03	89.14	87.36
FAT-Net	91.49	84.49	96.07	91.11	**90.41**	85.01	76.92	93.64	87.77	84.52	89.18	80.92	95.18	88.52	87.88
GA-Net	–	–	–	–	–	–	–	–	–	–	90.58	–	95.03	–	–
CPF-Net	–	–	–	–	–	85.83	77.27	93.58	90.88	84.88	90.61	84.25	94.74	89.54	91.27
TransUNet	91.32	84.89	96.21	92.28	89.63	85.51	77.34	93.47	88.56	84.22	89.50	82.61	93.67	88.12	89.50
Swin-UNet	91.27	84.14	96.18	91.39	89.45	83.50	72.28	93.20	84.27	83.14	88.46	79.86	94.45	88.24	87.43
Ours	**91.89**	**85.57**	**96.25**	**95.65**	89.66	**86.42**	**78.37**	**93.97**	**92.17**	**85.33**	**91.47**	**84.78**	**95.35**	**91.28**	**93.15**

Table 2
The comparison experiments of our method with other models on the ISIC 2018 Challenge ranking list.

Method	*test*-ISIC 2018				
	Dice	IoU	Acc	Se	Sp
SCDC (Lei et al., 2020)	88.50	82.40	92.90	95.30	91.10
ACA-Net (Saha et al., 2020)	89.10	81.90	—	94.30	93.20
Deeplabv3+ (Chen et al., 2018)	89.60	82.50	94.20	**96.20**	92.10
SESV-DLab (Xie et al., 2020)	90.20	83.30	94.60	**96.20**	92.50
ICL-Net (Cao, Yuan et al., 2022)	**90.30**	**83.90**	94.40	94.10	92.90
Ours	**90.30**	83.60	**94.70**	92.40	**93.90**

Se: Sensitivity **Sp**: Specificity.

ability, attributed to learning boundaries as general features across different distributions. Compared to the benchmark model, our model achieves a 5.36% higher IoU and surpasses our biggest competitor, ICL-Net, by 1.67%, illustrating our significant advantage in cross-domain dermatology segmentation. The qualitative comparison between our method and others for the PH^2 dataset is shown in Fig. 6.

To provide a more comprehensive assessment of model performance, we conducted a comparative analysis of the characteristics of the ISIC 2016 and PH^2 datasets and their impact on experimental outcomes. The ISIC 2016 dataset has a large sample size and diverse lesion types, including melanoma and non-melanoma cases. This diversity and high-resolution imaging contribute to the robustness of our model, enabling it to excel in complex clinical scenarios. In contrast, the PH^2 dataset, with a smaller sample size and standardized acquisition conditions, offers high image consistency, which likely enhances segmentation performance. However, the limited lesion types and sample size in PH^2 may affect model generalization.

Despite these differences, our model achieved outstanding performance on both datasets, demonstrating its generalization capability across varied imaging conditions and underscoring its potential for real-world clinical applications.

4.4.3. Analysis of test results on imaging datasets from different devices

Quantitative comparisons of our method with UNet and TransUNet on the Waterloo and PAD datasets are provided in Table 4. Our model significantly outperforms UNet and TransUNet in terms of IoU, Dice score, accuracy and precision. Since the samples from these datasets were not visible during model training and because they primarily consist of low-definition skin lesion images often captured by cell phones, our superior performance illustrates our model's ability to overcome limitations related to imaging devices and image quality. This success is largely attributed to our effective domain generalization learning.

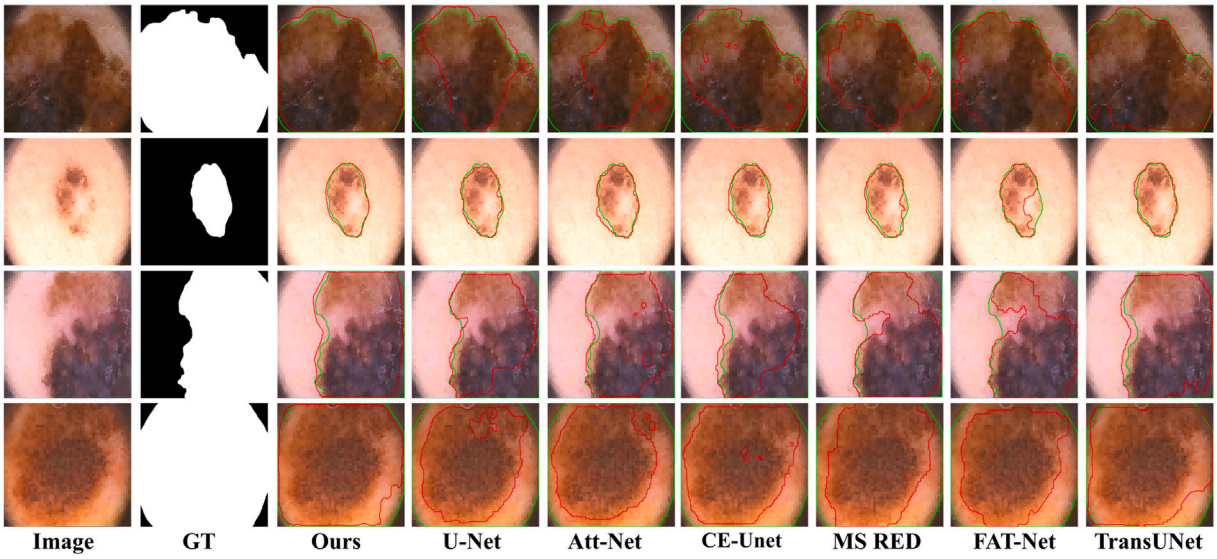

Fig. 6. Visual segmentation performance comparison of our model and some representative methods on the ISIC 2016 + PH^2 dataset.

Table 3
The quantitative evaluation of experiments is conducted on the ISIC 2016+PH^2 dataset.

Method	*test*-PH^2				
	Dice	IoU	Acc	Pre	Re
U-Net	90.56	83.56	94.86	90.30	93.47
Att-UNet	90.29	82.70	94.68	91.35	93.26
CE-Net	90.87	83.90	95.38	93.21	96.01
CPF-Net	91.67	85.48	95.59	91.81	95.90
MS RED	92.65	85.29	95.46	90.38	95.52
FAT-Net	92.21	85.18	95.43	92.85	**96.33**
TransUNet	90.96	83.99	95.42	91.18	95.68
Swin-UNet	92.69	87.08	96.03	91.42	94.87
ICL-Net	92.80	87.25	96.32	—	95.46
Ours	**94.03**	**88.92**	**96.34**	**94.06**	94.57

For the Waterloo dataset, our model achieves a 10.06% higher IoU compared to the baseline model and exceeds TransUNet by 5.62% in IoU. Similarly, on the PAD dataset, our model demonstrates a 5.92% higher IoU relative to the baseline model and surpasses TransUNet by 1.57% in IoU. Consequently, we can conclude that our model significantly outperforms UNet and TransUNet in handling challenging data, including skin lesion images from various imaging devices and those of low quality. The qualitative comparison of our method with others for the Waterloo and PAD datasets is illustrated in Fig. 7.

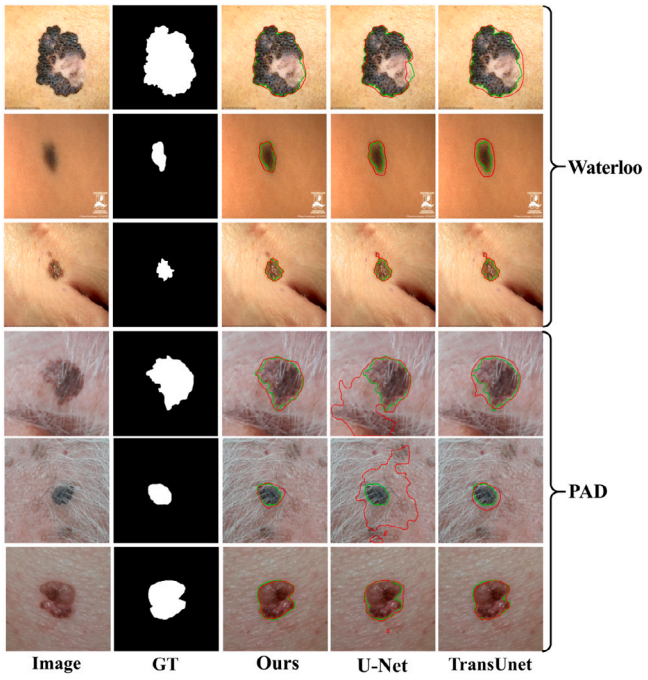

Fig. 7. Visual segmentation performance comparison of our model with representative methods on datasets imaged by mobile phones.

4.5. Ablation study

We conducted ablation studies on our proposed FDG and MM modules using the ISIC 2016 and PH^2 datasets, and an additional ablation study on the BFM module using the ISIC 2016 dataset. The quantitative comparison results for the FDG and MM modules are presented in Table 5, with the qualitative comparison shown in Fig. 8. The quantitative comparison results for the BFM module are provided in Table 6.

4.5.1. Effectiveness of FDG module

This research aims to assess the efficacy of the few-shot domain generalization module. We integrated this module into a model and benchmarked it against our baseline network. On the ISIC 2016 dataset, the model with the FDG module achieved a Dice coefficient of 91.17%, showing notable accuracy improvements. For the PH^2 dataset, the Dice coefficient rose by 0.17%, with accuracy increasing by 0.54% and IoU by 0.46%. The ISIC 2016 tests, where training and testing occurred within the same dataset, presented minimal inter-domain differences, limiting the module's impact due to the small sample size. Conversely, the PH^2 experiments required the model to generalize across domains, having been trained on ISIC 2016 and tested on PH^2. Here, the FDG module significantly boosted performance, emphasizing its value in scenarios involving diverse data distributions. This validates the FDG module's crucial role in enhancing segmentation accuracy in cross-domain settings and illustrates our model's robust generalization capabilities.

4.5.2. Effectiveness of MM module

This series of experiments evaluates the efficacy of the MM module.

Table 4
The study conducts comparative experiments to evaluate the efficacy of the method proposed in this paper against alternative approaches using datasets imaged by mobile phones.

Method	*test*-Waterloo					*test*-PAD				
	Dice	IoU	Acc	Pre	Re	Dice	IoU	Acc	Pre	Re
U-Net	82.71	72.41	96.03	74.78	96.47	87.71	79.45	94.85	82.47	**95.69**
TransUNet	86.51	77.21	97.68	80.12	**96.49**	90.82	83.80	96.98	90.76	92.06
Ours	**90.03**	**82.47**	**98.42**	**94.69**	87.14	**91.88**	**85.37**	**97.41**	**95.69**	89.02

Table 5
Ablation analysis of our net through adding the FDG Module and the MM Module step by step to the baseline model.

Dataset	+FDG	+MM	Dice	IoU	Acc	Pre	Re
ISIC 2016			91.10	84.66	95.90	94.57	89.37
	✓		91.17	84.72	95.99	93.67	90.58
		✓	91.58	85.17	96.04	93.95	**90.92**
	✓	✓	**91.89**	**85.57**	**96.25**	**96.65**	89.66
PH^2			92.60	86.63	95.16	92.02	94.28
	✓		92.87	87.09	95.70	90.31	96.37
		✓	92.99	87.23	95.82	90.04	**96.89**
	✓	✓	**94.03**	**88.92**	**96.30**	**94.06**	94.57

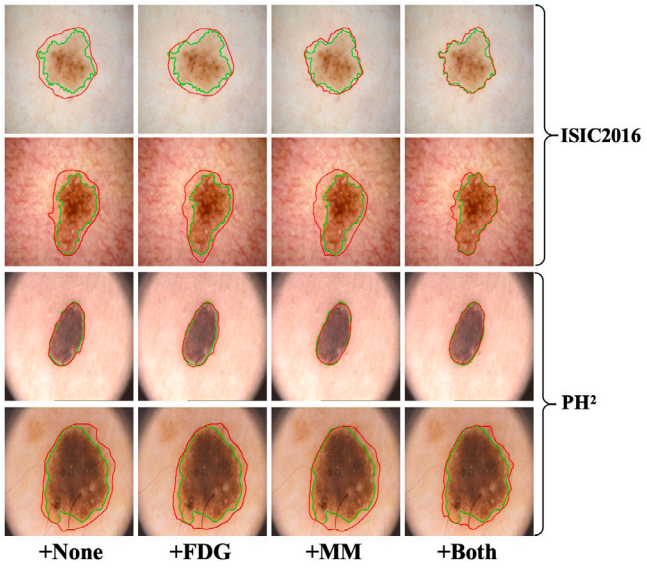

Fig. 8. Results of ablation studies on ISIC 2016 and PH^2 datasets.

We replaced the images processed by the original network's module with uniformly white RGB images, trained the model, and then compared the results with those from the network utilizing the MM module.

On the ISIC 2016 dataset, the addition of the MM module resulted in a 0.48% increase in the Dice coefficient over the baseline model, affirming its role in enhancing skin lesion boundary detection. The IoU also rose from 84.66% to 85.17%, indicating that the module aids the model in focusing more effectively on high-frequency details while filtering out less relevant, low-frequency information. On the PH^2 dataset, incorporating the MM module led to improvements across all key performance metrics.

These findings underscore that the MM module significantly boosts the model's ability to discern detailed features at the edges of skin lesions, thereby improving overall performance metrics. This enhancement is crucial for the accuracy of skin lesion segmentation tasks.

4.5.3. Effectiveness of BFM module

This study aims to evaluate the effectiveness of Bridge Fusion Modules (BFM) in enhancing segmentation performance. BFM exists in three forms: the Weight Multiplication Layer Module (BFM^*), the Three-channel Bridge Fusion Module (BFM^3), and the Two-channel Bridge Fusion Module (BFM^2). In this research, the baseline model stacks multiple features without fusion. We incrementally incorporated these three modules into the baseline network and conducted extensive comparative evaluations on the ISIC 2016 dataset. The experimental results indicate consistent metric improvements with each module inclusion. Specifically, incorporating BFM^3 yielded a 2.04% increase in Dice coefficient, a Recall of 90.21%, and a 1.11% IoU improvement compared to the baseline. These results demonstrate that BFM^3 enhances network performance and significantly contributes to feature activation during the multi-feature fusion stage. Based on the initial integration of BFM^3, further additions of BFM^* and BFM^2 led to progressive performance enhancements. Specifically, adding BFM^* resulted in minor increases in Dice, IoU, and Accuracy. Following this, the addition of BFM^2 brought a 4.44% increase in Dice, a 1.49% improvement in IoU, a 1.70% increase in Accuracy, and a 4.85% gain in Precision compared to the baseline. Although there was a slight decrease in Recall, this was likely due to the substantial gain in Precision, indicating an overall improvement despite a minor trade-off in sensitivity. These experimental findings validate the effectiveness of combined BFM configurations in feature fusion, demonstrating that integrating various BFM modules amplifies their performance benefits. This underscores BFM's critical role in achieving precise skin lesion segmentation, highlighting the high efficacy of our network.

Table 6
Ablation analysis of our net through adding the BFM Module on ISIC 2016.

BFM^*	BFM^3	BFM^2	Dice	IoU	Acc	Pre	Re
			87.45	84.08	94.55	91.80	89.82
✓			87.39	84.27	94.83	93.16	88.47
	✓		89.49	85.19	95.31	93.72	**90.21**
✓	✓		90.43	85.32	95.97	93.54	89.93
	✓	✓	91.25	85.21	96.00	94.45	90.10
✓	✓	✓	**91.89**	**85.57**	**96.25**	**96.65**	89.66

4.5.4. Effectiveness of the combination of FDG module and MM module

This study evaluates the efficacy of integrating the MM module and the FDG module. We compared the model combining both modules with our proposed backbone network. On the ISIC 2016 dataset, the combination of both modules demonstrated a notable improvement with a Dice coefficient increase of 0.79% to 91.89% and an accuracy enhancement to 96.25%, reflecting improved model predictions. The increase in IoU by 0.91% further indicates enhanced overlap between predicted and actual regions. The performance gains were even more significant on the PH^2 dataset, with increases of 1.43% in Dice coefficient, 1.14% in accuracy, and 2.29% in IoU, peaking at 88.92%.

These results confirm that the combined use of the MM module and FDG module significantly boosts performance across all metrics. The synergy between these modules not only improves boundary prediction accuracy but also enhances the model's adaptability to various data distributions, leading to superior detail capture and robust generalization. This validates the effectiveness of the module combination in refining the accuracy and consistency of segmentation results, thereby optimizing overall model performance.

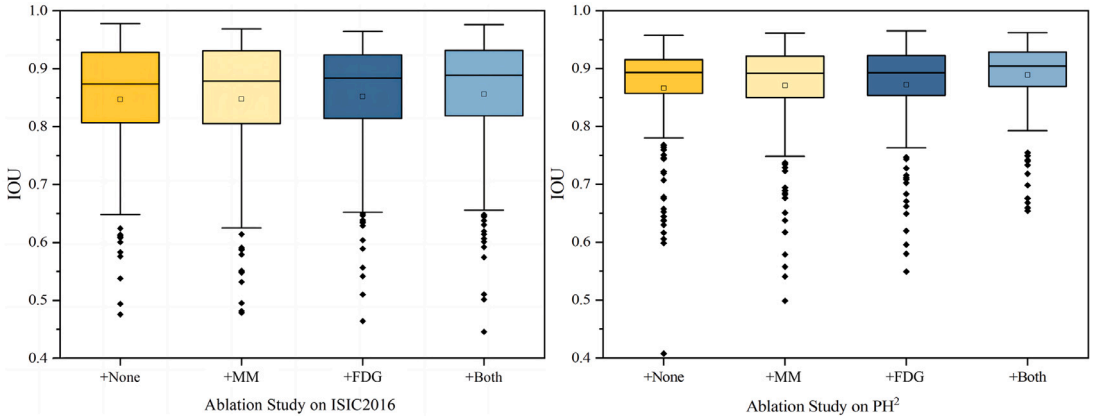

Fig. 9. Quantitative analysis of ISIC 2016 and PH^2 datasets's results.

4.5.5. Quantitative analysis

A quantitative comparison of our ablation experiments is presented in Fig. 9. It shows how additional modules influence model performance. Initially, ISIC 2016 results without modules displayed IoUs between 0.45 and 0.65, indicating unstable segmentation. Adding the FDG module slightly improved the median IoU, but data points remained sparse. Introducing the boundary feature extraction module significantly boosted both the median and first quartile IoUs. Optimal performance was achieved when both modules were used together, yielding higher IoUs, fewer outliers, and tighter data distribution. For PH^2, the baseline median IoU was 0.89, with wide variation and many outliers. Adding the FDG module slightly improved the mean IoU and reduced outliers. The boundary feature extraction module alone raised the mean IoU to 0.8723 and narrowed the data spread. Combining both modules brought the median IoU to a stable 0.91, with data clustered between 0.87 and 0.93, demonstrating improved model generalization and segmentation accuracy.

5. Discussion

5.1. Subjectivity issues in ground truth annotations within public datasets

In the process of performing the dermoscopy image segmentation task, we noticed that there is some subjectivity in the Ground Truth in the public dataset. It primarily due to varied expert interpretations of lesion boundaries. Such inconsistencies introduce non-universal features that can impair the generalization and diagnostic accuracy of deep learning models. Recognizing this challenge is crucial for directing future research and algorithm refinement. There is a pressing need for more standardized and objective evaluation methods. Addressing annotation subjectivity is vital for advancing the precision of dermoscopic image segmentation techniques. Future efforts should focus on innovative methods that better manage these variabilities. We provided some typical examples shown in Fig. 10.

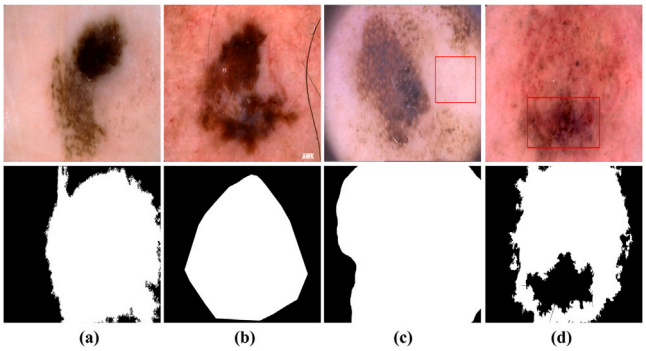

Fig. 10. Controversial data is present in the Open Dermatology dataset. Above is the image. Below is the GroundTruth. The red boxed region denotes the controversial GroundTruth area. It encompasses: (a) overly detailed edge definitions, (b) imprecisely defined edge regions, (c) non-lesional areas erroneously identified as lesional, and (d) controversial lesional regions lacking labels.

5.2. The impact of imaging device variability on segmentation outcomes

In this study, we utilized the PAD-UFES-20 dataset, developed in collaboration between the Federal University of Espírito Santo (UFES, Brazil) and the Program of Assistance in Dermatology and Surgery (PAD), to assess the adaptability and effectiveness of our skin lesion image segmentation model across diverse clinical settings. Unlike conventional high-resolution datasets, the PAD-UFES-20 dataset consists primarily of low-resolution images taken with various smartphone devices. Despite the inherent challenges of low resolution and device variability, which typically hinder image segmentation performance, our model demonstrated significant robustness and adaptability. This performance is particularly pertinent in resource-limited environments where affordable technologies like smartphones can play a crucial role in enhancing healthcare delivery. By effectively utilizing common smartphones for skin lesion screening, our approach could substantially improve global health equity, especially in low-income and remote regions. We plan to continue refining the model to increase its accuracy and reliability in complex clinical scenarios. We presented some typical challenging data in the PAD dataset in Fig. 11.

Fig. 11. Data in PAD that are difficult for neural network models to segment due to imaging equipment and quality.

6. Conclusion

This study presents a high-precision model for segmenting lesions in dermoscopic images. By integrating a hybrid feature extractor that combines CNNs and ViTs, the model effectively captures both spatial and local image features. An adaptive boundary delineation component, based on non-convex optimization, enhances the accuracy of lesion

boundary detection, ensuring detailed and precise segmentation. To further refine the output, texture features are combined with raw image features, enriching the information while preserving fine details in the shallow feature maps. To address challenges in dataset variability, the model incorporates a domain-adaptive adversarial learning strategy. This approach improves generalization by aligning the model with diverse data distribution characteristics, making it more robust across different datasets. Validation on publicly available dermoscopic image datasets, including ISIC, PH^2, PAD-UFES-20, and the University of Waterloo skin cancer database, demonstrates the model's state-of-the-art performance and strong generalization capabilities.

Despite its strengths, the model's complexity and high parameter count may limit its application in resource-constrained environments. Future research will focus on optimizing the model's structure to reduce computational requirements while enhancing its ability to process challenging lesion features. These advancements aim to create more efficient and adaptable tools for dermoscopic image segmentation, ultimately supporting clinical diagnosis and dermatological research more effectively.

CRediT authorship contribution statement

Kaiwen Zhu: Writing – original draft, Writing – review & editing. **Yuezhe Yang:** Writing – original draft, Writing – review & editing. **Yonglin Chen:** Conceptualization, Methodology, Writing – review & editing. **Ruixi Feng:** Visualization. **Dongping Chen:** Visualization. **Bingzhi Fan:** Visualization. **Nan Liu:** Visualization. **Ying Li:** Data curation. **Xuewen Wang:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Natural Science Foundation for the Higher Education Institutions of Anhui Province (Grant No. 2024AH050236).

Data availability

Data will be made available on request.

References

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

Cao, W., Yuan, G., Liu, Q., Peng, C., Xie, J., Yang, X., Ni, X., & Zheng, J. (2022). ICL-Net: Global and local inter-pixel correlations learning network for skin lesion segmentation. *IEEE Journal of Biomedical and Health Informatics*, *27*(1), 145–156.

Chatterjee, S., Dey, D., & Munshi, S. (2015). Mathematical morphology aided shape, texture and colour feature extraction from skin lesion for identification of malignant melanoma. In *2015 international conference on condition assessment techniques in electrical systems* (pp. 200–203). IEEE.

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

Chen, W., Mu, Q., & Qi, J. (2024). TrUNet: Dual-branch network by fusing CNN and Transformer for skin lesion segmentation. *IEEE Access*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision* (pp. 801–818).

Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al. (2018). 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* (pp. 168–172). IEEE.

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

Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., Esener, S., Fitzgerald, R. C., Gambhir, S. S., Kuhn, P., et al. (2022). Early detection of cancer. *Science*, *375*(6586), eaay9040.

Dai, D., Dong, C., Xu, S., Yan, Q., Li, Z., Zhang, C., & Luo, N. (2022). Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation. *Medical Image Analysis*, *75*, Article 102293.

DermIS. (2012). http://www.dermis.net. [Online, Accessed 01 June 2023].

DermQuest. (2012). http://www.dermquest.com. [Online, Accessed 01 June 2023].

Dong, Z., Li, J., & Hua, Z. (2024). Transformer-based multi-attention hybrid networks for skin lesion segmentation. *Expert Systems with Applications*, *244*, Article 123016.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Fan, C., Zhu, Z., Peng, B., Xuan, Z., & Zhu, X. (2024). EAAC-Net: An efficient adaptive attention and convolution fusion network for skin lesion segmentation. *Journal of Imaging Informatics in Medicine*, 1–17.

Feng, S., Zhao, H., Shi, F., Cheng, X., Wang, M., Ma, Y., Xiang, D., Zhu, W., & Chen, X. (2020). CPFNet: Context pyramid fusion network for medical image segmentation. *IEEE Transactions on Medical Imaging*, *39*(10), 3008–3018. http://dx.doi.org/10.1109/TMI.2020.2983721.

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

Guan, H., & Liu, M. (2021). Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, *69*(3), 1173–1185.

Gulzar, Y., & Khan, S. A. (2022). Skin lesion segmentation based on vision transformers and convolutional neural networks—a comparative study. *Applied Sciences*, *12*(12), 5990.

Gutman, D., Codella, N. C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1605.01397.

He, Z., Li, X., Chen, Y., Lv, N., & Cai, Y. (2023). Attention-based dual-path feature fusion network for automatic skin lesion segmentation. *BioData Mining*, *16*(1), 28.

He, X., Wang, Y., Zhao, S., & Chen, X. (2023). Joint segmentation and classification of skin lesions via a multi-task learning convolutional neural network. *Expert Systems with Applications*, *230*, Article 120174.

Lei, B., Xia, Z., Jiang, F., Jiang, X., Ge, Z., Xu, Y., Qin, J., Chen, S., Wang, T., & Wang, S. (2020). Skin lesion segmentation via generative adversarial networks with dual discriminators. *Medical Image Analysis*, *64*, Article 101716.

Li, D., Yang, J., Kreis, K., Torralba, A., & Fidler, S. (2021). Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8300–8311).

Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013). 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* (pp. 5437–5440). IEEE.

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

Pacheco, A. G., Lima, G. R., Salomao, A. S., Krohling, B., Biral, I. P., de Angelo, G. G., Alves Jr, F. C., Esgario, J. G., Simora, A. C., Castro, P. B., et al. (2020). PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, *32*, Article 106221.

Peng, B., & Fan, C. (2024). IEA-Net: Internal and external dual-attention medical segmentation network with high-performance convolutional blocks. *Journal of Imaging Informatics in Medicine*, 1–13.

Polansky, M. G., Herrmann, C., Hur, J., Sun, D., Verbin, D., & Zickler, T. (2024). Boundary attention: Learning to find faint boundaries at any resolution. arXiv preprint arXiv:2401.00935.

Rajab, M., Woolfson, M., & Morgan, S. (2004). Application of region-based segmentation and neural network edge detection to skin lesions. *Computerized Medical Imaging and Graphics*, *28*(1–2), 61–68.

Ramadan, R., & Aly, S. (2022). DGCU–Net: A new dual gradient-color deep convolutional neural network for efficient skin lesion segmentation. *Biomedical Signal Processing and Control*, *77*, Article 103829.

Riaz, F., Naeem, S., Nawaz, R., & Coimbra, M. (2018). Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 489–500.

Ronneberger, O., Fischer, P., & Brox, T. (2015). 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* (pp. 234–241). Springer.

Saha, A., Prasad, P., & Thabit, A. (2020). Leveraging adaptive color augmentation in convolutional neural networks for deep skin lesion segmentation. In *2020 IEEE 17th international symposium on biomedical imaging* (pp. 2014–2017). IEEE.

Scannell, C. M., Chiribiri, A., & Veta, M. (2021). Domain-adversarial learning for multi-centre, multi-vendor, and multi-disease cardiac MR image segmentation. In *Statistical atlases and computational models of the heart. M&Ms and EMIDEC challenges: 11th international workshop, STACOM 2020, held in conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, revised selected papers 11* (pp. 228–237). Springer.

Sethanan, K., Pitakaso, R., Srichok, T., Khonjun, S., Thannipat, P., Wanram, S., Boonmee, C., Gonwirat, S., Enkvetchakul, P., Kaewta, C., et al. (2023). Double AMIS-ensemble deep learning for skin cancer classification. *Expert Systems with Applications*, *234*, Article 121047.

Tong, X., Wei, J., Sun, B., Su, S., Zuo, Z., & Wu, P. (2021). ASCU-Net: attention gate, spatial and channel attention u-net for skin lesion segmentation. *Diagnostics*, *11*(3), 501.

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, *5*(1), 1–9.

Tzeng, E., Hoffman, J., Darrell, T., & Saenko, K. (2015). Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE international conference on computer vision* (pp. 4068–4076).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Verbin, D., & Zickler, T. (2021). Field of junctions: Extracting boundary structure at low snr. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6869–6878).

Vision and Image Processing Lab, University of Waterloo (2021). Skin cancer database. https://uwaterloo.ca/vision-image-processing-lab/research-demos/skin-cancer-detection. (Accessed 01 June 2023).

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

Wang, J., Wei, L., Wang, L., Zhou, Q., Zhu, L., & Qin, J. (2021). Boundary-aware transformers for skin lesion segmentation. In *Medical image computing and computer assisted intervention–mICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part i 24* (pp. 206–216). Springer.

Wang, Y., Xu, Z., Tian, J., Luo, J., Shi, Z., Zhang, Y., Fan, J., & He, Z. (2022). Cross-domain few-shot learning for rare-disease skin lesion segmentation. In *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing* (pp. 1086–1090). IEEE.

Wu, H., Chen, S., Chen, G., Wang, W., Lei, B., & Wen, Z. (2022). FAT-Net: Feature adaptive transformers for automated skin lesion segmentation. *Medical Image Analysis*, *76*, Article 102327.

Wu, H., Pan, J., Li, Z., Wen, Z., & Qin, J. (2020). Automated skin lesion segmentation via an adaptive dual attention module. *IEEE Transactions on Medical Imaging*, *40*(1), 357–370.

Xie, Y., Zhang, J., Lu, H., Shen, C., & Xia, Y. (2020). SESV: Accurate medical image segmentation by predicting and correcting errors. *IEEE Transactions on Medical Imaging*, *40*(1), 286–296.

Xu, G., Zhang, X., He, X., & Wu, X. (2023). Levit-unet: Make faster encoders with transformer for medical image segmentation. In *Chinese conference on pattern recognition and computer vision* (pp. 42–53). Springer.

Yacin Sikkandar, M., Alrasheadi, B. A., Prakash, N., Hemalakshmi, G., Mohanarathinam, A., & Shankar, K. (2021). Deep learning based an automated skin lesion segmentation and intelligent classification model. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 3245–3255.

Yan, S., Liu, C., Yu, Z., Ju, L., Mahapatra, D., Mar, V., Janda, M., Soyer, P., & Ge, Z. (2023). Epvt: Environment-aware prompt vision transformer for domain generalization in skin lesion recognition. In *International conference on medical image computing and computer-assisted intervention* (pp. 249–259). Springer.

Yang, L., Fan, C., Lin, H., & Qiu, Y. (2023). Rema-Net: An efficient multi-attention convolutional neural network for rapid skin lesion segmentation. *Computers in Biology and Medicine*, *159*, Article 106952.

Yogarajah, P., Condell, J., Curran, K., Cheddad, A., & McKevitt, P. (2010). A dynamic threshold approach for skin segmentation in color images. In *2010 IEEE international conference on image processing* (pp. 2225–2228). IEEE.

Zhou, L., Liang, L., & Sheng, X. (2023). GA-Net: Ghost convolution adaptive fusion skin lesion segmentation network. *Computers in Biology and Medicine*, *164*, Article 107273.