

Comparative Analysis and Adaptation Strategies of Vision Foundation Models in Dermoscopic Image Segmentation

Dermoskopik Görüntü Segmentasyonunda Görsel Temel Modellerin Karşılaştırmalı Analizi ve Adaptasyon Stratejileri

¹Kenan AKYOL , ^{1*}Mesut ŞEKER

¹Dicle Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Sur/Diyarbakır, Türkiye
(*Corresponding Author)

¹kenan.akyol@meb.gov.tr, ²mesut.seker@dicle.edu.tr

Araştırma Makalesi/Research Article

ARTICLE INFO

Article history

Received : 14 August 2026

Accepted : 19 September 2026

Keywords:

Skin Lesion Segmentation, Segment Anything Model, SAM-Adapter, Parameter-Efficient Fine-Tuning, Medical Image Segmentation

ABSTRACT

Accurate segmentation of skin lesions is of critical clinical importance for early melanoma diagnosis; however, the direct application of large foundation models like the Segment Anything Model (SAM) to medical images characterized by low contrast and fuzzy boundaries remains insufficient due to substantial domain gaps. In current study, a parameter-efficient fine-tuning strategy is presented by integrating lightweight adapter modules into a frozen Vision Transformer backbone. The proposed SAM-Adapter approach was comparatively evaluated against the original SAM and MedSAM models on the publicly available ISIC2018 and HAM10000 dermoscopic datasets. Optimized with AdamW, the model achieved the highest Dice (0.9603) and IoU (0.9241) scores along with the lowest HD95 (22.57) value on the HAM10000 dataset. Validated as statistically significant via the Wilcoxon test, this approach minimized computational overhead while demonstrating superior accuracy in boundary delineation. This adapter-based framework is planned for integration into clinical decision-support systems and other medical imaging modalities.

© 2026 Bandırma Onyedi Eylül University, Faculty of Engineering and Natural Science. Published by Dergi Park. All rights reserved.

MAKALE BİLGİSİ

Makale Tarihleri

Gönderim : 14 Ağustos 2026

Kabul : 19 Eylül 2026

Anahtar Kelimeler:

Deri Lezyonu Segmentasyonu, Segment Anything Model, Sam-Adapter, Parametre Verimli İnce Ayar, Tıbbi Görüntü Segmentasyonu

ÖZET

Deri lezyonlarının doğru segmentasyonu, melanomun erken tanısında kritik öneme sahipken; Segment Anything Model (SAM) gibi büyük temel modellerin düşük kontrastlı ve belirsiz sınırlı tıbbi görüntülere doğrudan uygulanması yapısal alan farklılıkları nedeniyle yetersiz kalmaktadır. Bu çalışmada, dondurulmuş bir Vision Transformer omurgasına hafif adaptör modüllerinin entegre edildiği parametre verimli bir ince ayar stratejisi sunulmaktadır. Önerilen SAM-Adapter yöntemi; ISIC2018 ve HAM10000 açık erişimli dermoskopik veri setleri üzerinde orijinal SAM ve MedSAM modelleriyle karşılaştırmalı olarak değerlendirilmiştir. AdamW optimizasyonu ile eğitilen model; HAM10000 veri setinde en yüksek Dice (0.9603) ve IoU (0.9241) skorlarına ulaşırken, en düşük HD95 (22.57) değerini elde etmiştir. Wilcoxon testi ile istatistiksel anlamlılığı doğrulanan bu yaklaşım, hesaplama yükünü en aza indirerek zorlu lezyon sınırlarının belirlenmesinde üstün başarı göstermiştir. Geliştirilen adaptör tabanlı çerçevenin klinik karar destek sistemlerine ve farklı tıbbi görüntüleme modalitelerine entegrasyonu hedeflenmektedir.

© 2026 Bandırma Onyedi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi. Dergi Park tarafından yayınlanmaktadır. Tüm Hakları Saklıdır.

ORCID ID: ¹0009-0001-2533-255X
²0000-0001-9245-6790

1. INTRODUCTION

Skin cancer is one of the most frequently diagnosed malignancies worldwide and remains a major public health concern because of its increasing incidence and mortality [1]. Melanoma represents the most lethal histological subtype owing to its pronounced propensity for metastasis, which makes prompt diagnosis crucial for boosting patient prognosis. As an established non-invasive diagnostic tool, dermoscopy allows clinicians to examine sub-epidermal morphological features invisible to unassisted vision. However, manual reading of dermoscopic imagery is inherently subject to inter-observer variability, high subjectivity, and substantial time demands, regardless of practitioner experience. These operational drawbacks have spurred the adoption of computer-aided diagnosis (CAD) systems, which deliver standardized and reproducible lesion characterization to guide clinical workflows [2]. Accurate lesion segmentation constitutes the cornerstone of CAD systems because it directly influences subsequent diagnostic tasks, including lesion classification, asymmetry and border irregularity analysis, color variegation assessment, and feature extraction [3]. Nevertheless, obtaining reliable lesion segmentation remains highly challenging due to the inherent characteristics of dermoscopic images, such as low contrast between lesions and surrounding skin, irregular and fuzzy lesion boundaries, heterogeneous pigmentation patterns, hair occlusion, illumination variations, and imaging artifacts, which often limit the effectiveness of conventional image segmentation techniques [4].

Progress in deep learning has remarkably enhanced automatic skin lesion delineation. Convolutional neural network (CNN) models, most notably U-Net and its derivatives, have achieved remarkable performance on benchmark datasets [5], while volumetric extensions such as V-Net have further demonstrated the effectiveness of fully convolutional approaches for medical image segmentation [6]. In recent years, Vision Transformer (ViT)-based models have leveraged self-attention to effectively model long-range dependencies, yielding substantially more precise feature representations [7]. Attention mechanisms have also been incorporated into U-Net architectures to selectively emphasize informative features while suppressing irrelevant background regions, further improving segmentation performance in challenging medical imaging tasks [8]. Recent extensions such as Semantic-SAM have further enhanced segmentation capabilities by enabling multi-granularity object recognition and interactive refinement [9]. In addition, deep learning approaches have demonstrated dermatologist-level performance in skin cancer classification tasks [10], highlighting the growing clinical potential of artificial intelligence in dermatology. Despite these advances, most deep learning models require large annotated datasets and often exhibit limited generalization when applied to images acquired under different clinical conditions, imaging devices, or patient populations with diverse skin characteristics.

The advent of foundation models has revolutionized computer vision by offering robust pretrained feature representations that seamlessly adapt to various downstream applications with minimal task-specific training. Notable among these, the Segment Anything Model (SAM) established a prompt-based segmentation paradigm leverages the massive SA-1B dataset, exhibiting remarkable zero-shot generalization across varied natural imagery [11]. Nevertheless, empirical findings indicate that SAM's unadapted application to medical imagery faces hurdles, driven by the marked distributional gap separating natural images from clinical data [12]. SAM's performance may degrade substantially when applied to medical images characterized by low contrast, irregular boundaries, and small lesion sizes, highlighting the need for domain-specific adaptation strategies. Dermoscopic images, in particular, contain subtle lesion boundaries, homogeneous tissue textures, and fine anatomical details that are insufficiently represented during SAM pretraining. Consequently, studies have investigated SAM-based segmentation for dermoscopic images and consistently reported that satisfactory performance cannot be achieved without effective domain adaptation [13].

To address these challenges, several adaptation strategies have been proposed. MedSAM [11] improves SAM through large-scale medical image adaptation, whereas SAM-Adapter [12] introduces lightweight trainable adapter modules into the frozen transformer backbone, enabling parameter-efficient fine-tuning while preserving pretrained representations. Similarly, SkinSAM [14] demonstrated that fine-tuning the SAM encoder on skin lesion datasets significantly improves segmentation performance, while the Medical SAM Adapter [15] extended adapter-based adaptation to various medical imaging modalities. Although these studies have reported encouraging results, comprehensive and systematic comparisons of the original SAM, MedSAM, and SAM-Adapter under identical experimental conditions remain limited for dermoscopic skin lesion segmentation. Furthermore, most existing studies focus on a single adaptation strategy and rarely combine quantitative evaluation, qualitative visual analysis, and rigorous statistical significance testing using multiple benchmark datasets. Consequently, it remains unclear which foundation-model adaptation strategy provides the best balance between segmentation accuracy, robustness, and computational efficiency for dermoscopic skin lesion segmentation under identical experimental conditions.

Motivated by these observations, this study presents a comprehensive comparative evaluation of SAM, MedSAM, and SAM-Adapter using the publicly available ISIC2018 and HAM10000 datasets. All experiments were conducted under identical preprocessing, optimization, and evaluation protocols to ensure a fair and unbiased comparison among the evaluated models. Segmentation accuracy was quantitatively evaluated through the Dice Similarity Coefficient (Dice), Intersection over Union (IoU), and 95th percentile Hausdorff Distance (HD95), with statistical significance validated via the Wilcoxon Signed-Rank Test. Furthermore, qualitative visual inspections were conducted to examine the precision of each model in delineating lesion contours.

The primary contributions of this work are highlighted as follows: First, this work presents a comprehensive and fair comparison of three foundation-model-based segmentation approaches under identical experimental conditions. Second, it thoroughly evaluates a parameter-efficient adapter-based fine-tuning strategy that preserves the pretrained SAM backbone while achieving superior segmentation performance. Third, it integrates quantitative evaluation, qualitative visual analysis, and statistical significance testing to provide robust evidence regarding the effectiveness of adapter-based adaptation for dermoscopic skin lesion segmentation. The structure of the remainder of this paper is arranged as follows: Section 2 details the benchmark datasets, the proposed architecture, the experimental framework, and the assessment metrics. Section 3 reports the quantitative, qualitative, and statistical evaluation outcomes. Section 4 provides a comparative discussion of the findings against state-of-the-art literature while delineating the strengths and constraints of the method. Lastly, Section 5 summarizes the conclusions and highlights potential avenues for future research.

2. MATERIALS AND METHODS

2.1. Datasets

The performance of the proposed SAM-Adapter framework was evaluated using two open-access dermoscopic image repositories: ISIC2018 and HAM10000. These widely established benchmarks offer extensive variability regarding lesion morphology, capture conditions, image resolution, and clinical diagnoses, making them ideal testbeds for assessing the generalization and robustness of deep learning architectures. Specifically, the ISIC2018 dataset comprises 2,594 dermoscopic samples paired with expert-annotated binary ground-truth masks. In addition, the HAM10000 collection encompasses 10,015 high-resolution dermoscopic images spanning seven distinct diagnostic classes gathered across various clinical centers and imaging modalities, thus offering a multi-center evaluation setup for segmentation models [13]. To ensure the reproducibility of the study and establish a fair evaluation framework, a unified data split protocol was adopted across all datasets. Both the HAM10000 and ISIC2018 datasets were partitioned into 80% for training, 10% for validation, and 10% for testing. Specifically, the HAM10000 dataset (10,015 images) was divided into 8,012 training, 1,001 validation, and 1,002 testing images. Similarly, the ISIC2018 dataset (2,594 images) was split into 2,075 training, 259 validation, and 260 independent testing images.

The proposed SAM-Adapter was optimized and tuned using the training and validation subsets. To ensure an unbiased comparative analysis and prevent data leakage, the baseline foundation models (SAM and MedSAM) were evaluated in a zero-shot manner strictly on the hold-out 10% test sets, identical to those used for testing the SAM-Adapter. This strictly isolated testing strategy guarantees that the reported quantitative metrics reflect the true generalization capabilities of all evaluated models on unseen dermoscopic data. To achieve standardized input dimensions across models, all dermoscopic images and their corresponding binary masks were rescaled to 1024×1024 pixels, fully aligning with the spatial resolution expected by SAM's Vision Transformer backbone. Intensity normalization was applied to the image pixels, whereas nearest-neighbor interpolation was strictly utilized for resizing the binary masks to prevent boundary artifacts and preserve precise lesion contours. These preprocessing operations provided uniform data inputs while ensuring anatomical integrity across both benchmark datasets. The primary attributes of the datasets evaluated in this study are outlined in Table 1.

Table 1. Characteristics of the dermoscopic image datasets used throughout the experimental evaluation.

Dataset	Number of Images	Resolution
HAM10000	10,015	1024×1024
ISIC2018	2,594	1024×1024

2.2. Proposed method

The foundation of the proposed method rests on the Segment Anything Model (SAM), an interactive foundation model tailored for broad segmentation tasks. Deviating from standard CNN-based segmentation paradigms, SAM combines a Vision Transformer (ViT) image backbone, a prompt encoder, and a compact mask decoder, which facilitates versatile prompt-based segmentation through points, bounding boxes, or mask inputs [11]. Although SAM displays impressive zero-shot performance, its unadapted deployment on medical images is non-trivial, driven by the substantial distribution shift between natural scenes and specialized medical datasets. Specifically, dermoscopic images feature subtle boundary transitions, homogeneous textures, and fine-grained clinical details that receive insufficient coverage during SAM's pretraining. As a result, parameter-efficient adaptation strategies become indispensable to enhance segmentation efficacy without compromising the model's learned visual knowledge [16].

To tackle this issue, the presented framework leverages a lightweight fine-tuning scheme through the incorporation of trainable adapter modules into the Vision Transformer encoder while leaving the pretrained backbone parameters completely unchanged. Unlike conventional end-to-end transfer learning, only the adapter modules are optimized during training, substantially reducing computational complexity while maintaining the rich semantic representations learned during large-scale pretraining. This adapter-based approach differs from other adaptation strategies such as MedSAM, which improves SAM through large-scale medical image adaptation, and SkinSAM, which fine-tunes the SAM encoder specifically on skin lesion datasets. Similarly, the Medical SAM Adapter

extended adapter-based adaptation to various medical imaging modalities. However, unlike these approaches that require extensive retraining or dataset-specific modifications, the proposed method preserves the pretrained SAM backbone entirely frozen and optimizes only the lightweight adapter modules, offering a more parameter-efficient solution. The overall workflow of the proposed SAM-Adapter framework is illustrated in Figure 1.

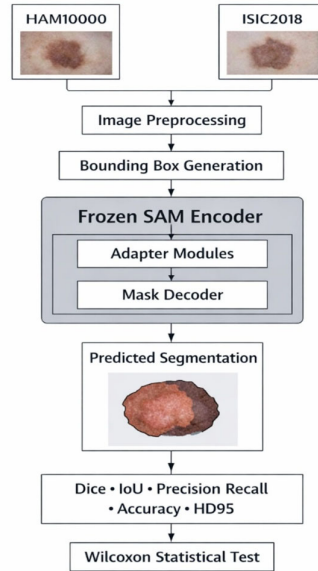


Figure 1. Overall workflow of the proposed SAM-Adapter framework for dermoscopic skin lesion segmentation, illustrating the complete pipeline from image preprocessing and bounding box generation to feature extraction, adapter-based fine-tuning, mask prediction, and quantitative evaluation.

The proposed approach consists of five consecutive stages: image preprocessing, automatic bounding-box prompt generation, feature extraction using the frozen SAM encoder, adapter-based feature learning, and lesion mask prediction. Dermoscopic images from the ISIC2018 and HAM10000 datasets are first resized to 1024×1024 pixels and normalized to provide standardized inputs for the Vision Transformer encoder. During both the training and inference phases, bounding-box prompts were extracted directly from the expert-annotated ground-truth masks (i.e., an oracle bounding-box approach). This experimental design was deliberately chosen to isolate the models' pure segmentation and boundary delineation capabilities from potential localization errors. By providing an optimal prompt, we established the upper-bound segmentation performance of the evaluated foundation models in a fully reproducible pipeline. A distinctive characteristic of the proposed method is that the pretrained SAM backbone remains entirely frozen throughout training. Only the lightweight adapter modules inserted into the transformer blocks are updated, allowing efficient adaptation to dermoscopic image segmentation while preventing catastrophic forgetting of the pretrained visual representations. Each adapter unit is designed around a bottleneck topology that includes a linear down-projection, followed by a *ReLU* activation function and a linear up-projection layer. The updated feature representation is computed through the residual additive connection formulated in Equation (1).

$$F = Z + A_{\theta}(Z) \quad (1)$$

, where Z denotes the frozen transformer features, A_{θ} represents the learnable adapter transformation, and F is the adapted feature representation forwarded to the subsequent transformer block. This residual formulation preserves the semantic information extracted during large-scale pretraining while enabling the adapter modules to learn dermoscopy-specific feature representations.

To train the network, the AdamW optimizer was configured with an initial learning rate set to 1×10^{-4} , utilizing a batch size of 1 over 10 training epochs. These specific hyperparameters were carefully selected based on both hardware constraints and the architectural convergence characteristics of adapter-based fine-tuning. Specifically, processing high-resolution (1024×1024) dermoscopic images through the large-scale Vision Transformer backbone inherently demands a massive GPU memory (VRAM) footprint for storing intermediate activations during the forward and backward passes. Consequently, the batch size was strictly limited to 1 to prevent Out-Of-Memory (OOM) exceptions. Furthermore, the training duration was deliberately restricted to 10 epochs. Because the adapter modules operate within a highly compact trainable parameter space (~ 5.2 million) and leverage the robust, generalized feature representations of the frozen SAM encoder, the network converges rapidly. Empirical observations indicated that extending the training beyond 10 epochs provided negligible improvements in segmentation accuracy while substantially increasing the risk of overfitting on the training distribution. The objective function blends Dice Loss with Binary Cross-Entropy (BCE) Loss in accordance with Equation (2).

$$L = L_{\text{Dice}} + L_{\text{BCE}} \quad (2)$$

, where L_{Dice} maximizes the overlap between predicted and reference masks, whereas L_{BCE} improves pixel-wise classification accuracy. Automatic Mixed Precision (AMP) training was employed to reduce GPU memory consumption and accelerate optimization.

To explicitly demonstrate the computational advantages and parameter efficiency of the proposed framework, a detailed parameter and computational complexity analysis (measured in GFLOPs) was conducted for the ViT-B backbone configuration using a 1024×1024 input resolution. The original SAM architecture consists of 93.74 million parameters and requires approximately 738.01 GFLOPs for the image encoder forward pass. Conventional end-to-end fine-tuning would require calculating gradients and updating all 93.74 million parameters, resulting in massive memory consumption and high computational costs.

In contrast, the proposed SAM-Adapter completely freezes the original image encoder. The injection of bottleneck adapter modules introduces only 1.19 million additional parameters, resulting in a marginal computational overhead of 9.66 GFLOPs (totaling 747.68 GFLOPs). During the optimization phase, only the adapter modules, prompt encoder, and mask decoder are designated as trainable. Consequently, the proposed model contains a total of 94.93 million parameters, of which only 5.25 million are actively trained. By updating merely 5.5% of the total network parameters, the SAM-Adapter dramatically reduces the computational overhead and memory footprint while achieving superior segmentation performance, effectively preventing the catastrophic forgetting of the pretrained foundation model representations.

2.3. Experimental Setup Strategy

The original SAM and MedSAM models were initialized using their publicly available pretrained checkpoints, whereas two dataset-specific adapter models were trained separately for the HAM10000 and ISIC2018 datasets. To ensure a fair comparison and experimental reproducibility, identical preprocessing procedures, optimization settings, input resolution, evaluation metrics, random initialization, and inference protocols were applied to all three segmentation models. Consequently, the observed performance differences can be attributed solely to differences in model architecture rather than variations in experimental settings. The training pipeline adopted for the proposed approach is illustrated in Figure 2.

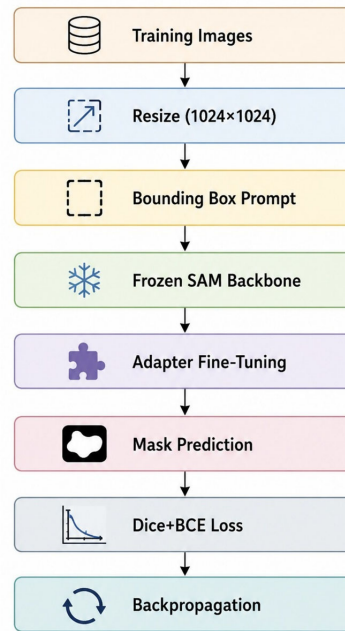


Figure 2. Training pipeline of the proposed SAM-Adapter framework, showing the sequential steps from data loading and preprocessing through feature extraction and optimization to final segmentation mask prediction.

2.4. Evaluation metrics

To comprehensively evaluate the segmentation performance and boundary localization accuracy, three quantitative metrics were employed: the Dice Similarity Coefficient (Dice), Intersection over Union (IoU), and the 95th percentile Hausdorff Distance (HD95). The mathematical formulations, ranges, and clinical interpretations of these metrics are summarized in Table 2. All quantitative metrics were computed directly on the test sets without any post-processing to ensure an unbiased comparison among the evaluated models. In addition, to assess whether the observed performance differences were statistically significant, the non-parametric Wilcoxon Signed-Rank Test was performed on paired test samples, with a significance threshold set at $p < 0.05$.

Table 2. Summary of quantitative evaluation metrics used for segmentation assessment.

Metric	Mathematical Formulation	Value Range	Interpretation / Clinical Focus
Dice Similarity Coefficient (Dice)	$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$	[0,1]	Measures spatial overlap between predicted and reference masks. Higher values indicate superior segmentation overlap.
Intersection over Union (IoU)	$\text{IoU} = \frac{TP}{TP + FP + FN}$	[0,1]	Evaluates the Jaccard overlap index. Provides a stricter measure of mask agreement than Dice.
95th Percentile Hausdorff Distance (HD95)	$\text{HD95} = \max_{95\%} \left(\max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b) \right)$	$[0, \infty)$	Measures the spatial discrepancy between predicted (A) and reference (B) boundaries. Lower values indicate more accurate boundary localization.

Note: TP , FP , and FN denote True Positive, False Positive, and False Negative pixel counts, respectively. $d(a, b)$ represents the Euclidean distance between boundary points.

3. RESULTS

The proposed SAM-Adapter framework was evaluated on the HAM10000 and ISIC2018 dermoscopic image datasets and compared with the original Segment Anything Model (SAM) and MedSAM under identical experimental conditions. Segmentation performance was assessed using the Dice Similarity Coefficient (Dice), Intersection over Union (IoU), Precision, Recall, Accuracy, and the 95th percentile Hausdorff Distance (HD95). In addition, qualitative visual comparisons and statistical significance analysis were performed to provide a comprehensive evaluation of the proposed method.

3.1. Overall quantitative performance

Table 3 summarizes the quantitative performance of the evaluated models on both benchmark datasets. Overall, the proposed SAM-Adapter consistently achieved the highest Dice, IoU, Recall, and Accuracy values while producing the lowest HD95 values, indicating superior overlap accuracy and more precise lesion boundary localization. The performance improvements were consistent across both datasets, demonstrating the effectiveness and robust dataset-specific adaptation capability of the proposed parameter-efficient strategy.

Table 3. Quantitative comparison of SAM, MedSAM, and the proposed SAM-Adapter using Dice, IoU, and HD95 on the HAM10000 and ISIC2018 datasets.

Dataset	Model	Dice $\pm$ SD	IoU $\pm$ SD	HD95 $\pm$ SD
HAM10000	SAM	0.7909 $\pm$ 0.2010	0.6868 $\pm$ 0.2030	54.8230 $\pm$ 40.9686
	MedSAM	0.9031 $\pm$ 0.0528	0.8272 $\pm$ 0.0838	30.8452 $\pm$ 21.2126
	SAM-Adapter	0.9603 $\pm$ 0.0186	0.9241 $\pm$ 0.0338	22.5764 $\pm$ 18.6631
ISIC2018	SAM	0.8173 $\pm$ 0.1439	0.7108 $\pm$ 0.1655	23.5105 $\pm$ 21.0779
	MedSAM	0.8957 $\pm$ 0.0713	0.8179 $\pm$ 0.1057	14.8459 $\pm$ 13.5341
	SAM-Adapter	0.9376 $\pm$ 0.0415	0.8850 $\pm$ 0.0661	14.0449 $\pm$ 13.6580

Table 3 demonstrates that the proposed SAM-Adapter substantially improves segmentation performance compared with both baseline foundation-model approaches. The largest improvements were observed in the Dice and IoU metrics, while the lower HD95 values indicate more accurate lesion boundary localization. Furthermore, the relatively small standard deviations suggest that the proposed approach provides stable and consistent segmentation performance across different lesion images.

As shown in Table 4, MedSAM achieved the highest Precision on both datasets, whereas the proposed SAM-Adapter consistently produced the highest Recall and Accuracy. These results indicate that the proposed approach detects lesion regions more completely while maintaining robust overall segmentation performance. In addition, the relatively low standard deviations demonstrate that the proposed method performs consistently across both benchmark datasets. Notably, the standard deviations reported in Table 3 reveal an important finding regarding model stability. The original SAM exhibits considerably higher variability (SD = 0.201 for Dice on HAM10000), indicating inconsistent performance across different lesion types and imaging conditions. In contrast, the proposed SAM-Adapter achieves substantially lower standard deviations (SD = 0.0186 for Dice on HAM10000), demonstrating not only superior accuracy but also significantly improved robustness and prediction stability across diverse dermoscopic images. This reduced variability is particularly important for clinical deployment, where consistent and reliable segmentation is essential.

3.2. Performance Distribution Analysis

Distribution-based analysis provides complementary information beyond mean performance values by evaluating prediction stability and robustness across the entire test dataset. While average Dice or IoU scores indicate the

overall segmentation performance, they do not reveal the consistency of model predictions or the presence of extreme failure cases. Therefore, violin plots were employed to visualize the distribution of segmentation metrics obtained by SAM, MedSAM and the proposed SAM-Adapter on both benchmark datasets.

Table 4. Comparison of Precision, Recall, and Accuracy of the evaluated segmentation models on the HAM10000 and ISIC2018 datasets.

Dataset	Model	Precision $\pm$ SD	Recall $\pm$ SD	Accuracy $\pm$ SD
HAM10000	SAM	0.9719 $\pm$ 0.0560	0.7078 $\pm$ 0.2163	0.9226 $\pm$ 0.0615
	MedSAM	0.9844 $\pm$ 0.0335	0.8390 $\pm$ 0.0855	0.9567 $\pm$ 0.0344
	SAM-Adapter	0.9751 $\pm$ 0.0254	0.9467 $\pm$ 0.0301	0.9791 $\pm$ 0.0180
ISIC2018	SAM	0.9422 $\pm$ 0.1109	0.7509 $\pm$ 0.1799	0.9307 $\pm$ 0.1032
	MedSAM	0.9764 $\pm$ 0.0489	0.8361 $\pm$ 0.1096	0.9609 $\pm$ 0.0559
	SAM-Adapter	0.9174 $\pm$ 0.0750	0.9641 $\pm$ 0.0349	0.9729 $\pm$ 0.0383

3.2.1 HAM10000 dataset

As shown in Figure 3, the proposed SAM-Adapter consistently exhibits higher Dice and IoU values together with lower HD95 values than the competing approaches. As previously indicated by the standard deviations in Table 4, the proposed framework also provides more stable and robust segmentation performance across different lesion characteristics.

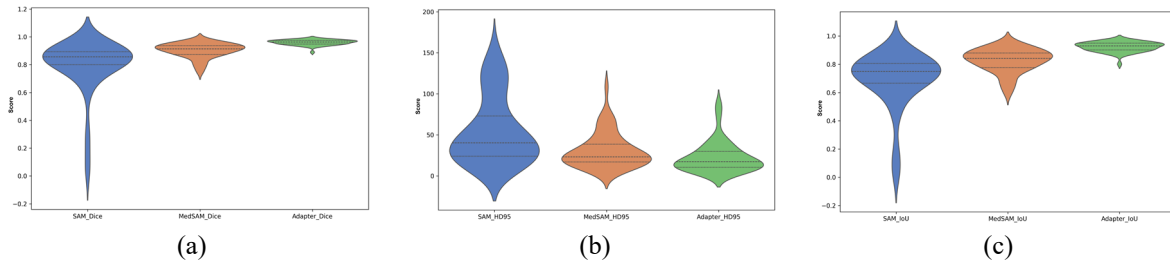


Figure 3. Distribution of quantitative performance metrics on the HAM10000 dataset. (a) Dice score distribution. (b) HD95 distribution. (c) IoU distribution.

A similar trend can be observed for the IoU distributions shown in Figure 3. The proposed approach consistently achieves higher overlap values with reduced variance, indicating improved agreement between predicted lesion masks and expert annotations. Since IoU provides a stricter evaluation of segmentation overlap than the Dice coefficient, the observed improvement further confirms the robustness of the proposed adaptation strategy. Boundary localization performance was further evaluated using the HD95 metric. As illustrated in Figure 3, SAM-Adapter produces the lowest HD95 values among all evaluated models. Lower HD95 values correspond to smaller contour discrepancies between predicted and reference lesion boundaries, demonstrating that the proposed framework more accurately preserves lesion morphology. The compact distribution additionally indicates improved robustness against difficult cases containing irregular lesion contours, hair artifacts, or blurred boundaries. The distributions of Precision, Recall, and Accuracy also demonstrate consistent improvements obtained by the proposed framework. Although the original SAM occasionally achieves high precision, it frequently under-segments lesion regions, resulting in reduced recall values. MedSAM alleviates this limitation to some extent; however, the proposed SAM-Adapter provides a more balanced trade-off between precision and recall, thereby producing the highest overall segmentation accuracy. The reduced variability observed across all evaluation metrics indicates that the proposed approach performs reliably over a wide spectrum of dermoscopic image characteristics.

3.2.2 ISIC2018 dataset

The performance distributions obtained on the ISIC2018 dataset exhibit trends similar to those observed for HAM10000. As illustrated in Figure 4, the proposed SAM-Adapter consistently achieves higher Dice and IoU values while maintaining a considerably narrower distribution than both SAM and MedSAM. These observations demonstrate that the proposed approach preserves its segmentation performance even when evaluated on an independent benchmark dataset acquired under different imaging conditions.

The HD95 distributions further support these findings by demonstrating considerably enhanced boundary localization. Compared with the competing approaches, SAM-Adapter exhibits lower median HD95 values together with fewer extreme outliers. This behavior indicates improved robustness when segmenting lesions exhibiting complex morphology, indistinct borders, and heterogeneous pigmentation. Accurate boundary delineation is particularly important for dermoscopic image analysis because small contour deviations may substantially influence subsequent quantitative feature extraction and computer-aided diagnosis. Rather than improving a single evaluation metric, SAM-Adapter consistently enhances all quantitative performance indicators

while simultaneously reducing prediction variability. The consistency observed across both benchmark datasets suggests that the proposed framework consistently achieves high segmentation accuracy when adapted to distinct dermoscopic image distributions.

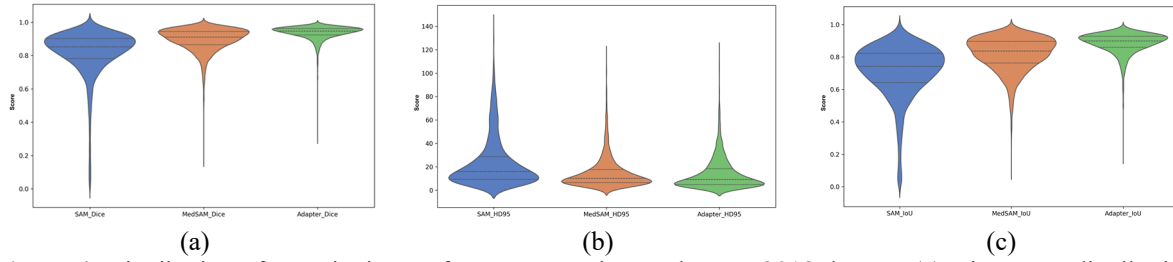


Figure 4. Distribution of quantitative performance metrics on the ISIC2018 dataset: (a) Dice score distribution. (b) HD95 distribution. (c) IoU distribution.

3.3. Qualitative Segmentation Analysis

Quantitative evaluation provides objective evidence regarding segmentation accuracy; however, numerical metrics alone are insufficient to fully characterize model behavior in challenging clinical scenarios. Therefore, qualitative visual comparisons were performed to investigate the ability of the evaluated models to delineate lesion boundaries under different dermoscopic conditions. Representative segmentation examples obtained from the HAM10000 and ISIC2018 datasets are presented in Figures 5–6. The selected examples include lesions exhibiting substantial variations in size, shape, pigmentation, boundary irregularity, and imaging artifacts. These challenging cases provide a comprehensive visual assessment of the robustness and adaptability of the evaluated segmentation models.

3.3.1 Qualitative analysis on HAM10000

Representative segmentation results obtained on the HAM10000 dataset are illustrated in Figures 5 and 6. While Figure 5 presents a direct visual comparison of the segmentation boundaries produced by SAM, MedSAM, and the proposed SAM-Adapter, Figure 6 further compares the segmentation masks generated by the proposed model with the corresponding expert-annotated ground-truth masks. Together, these qualitative results provide a comprehensive visual assessment of the segmentation performance achieved by the evaluated models.

The original SAM generally succeeds in identifying the approximate lesion location; however, inaccurate boundary delineation frequently occurs in lesions with irregular morphology, indistinct borders, and heterogeneous pigmentation. In several representative cases shown in Figure 5, the predicted masks either underestimate peripheral lesion regions or extend beyond the true lesion boundary, indicating limited adaptation of the original foundation model to dermoscopic image characteristics.

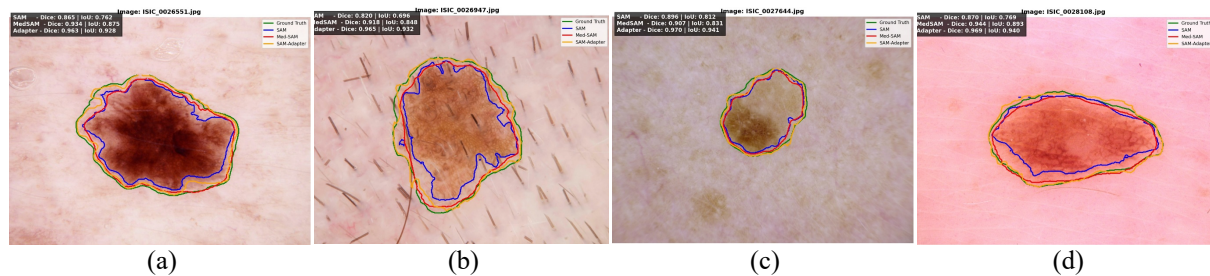


Figure 5. Representative qualitative segmentation results on the HAM10000 dataset. (a–d) Representative lesions showing the segmentation boundaries produced by SAM, MedSAM, and the proposed SAM-Adapter.

MedSAM improves lesion localization by incorporating medical-domain knowledge during pretraining. Consequently, segmentation masks become smoother and more anatomically consistent than those generated by the original SAM. Nevertheless, boundary deviations remain visible in several difficult cases, particularly when lesions exhibit weak contrast against the surrounding healthy tissue or contain highly irregular contours. These observations suggest that although domain-specific pretraining substantially enhances segmentation performance, challenging lesion characteristics continue to affect boundary precision. In contrast, the proposed SAM-Adapter consistently produces segmentation masks that most closely resemble the expert-annotated ground truth.

As illustrated in Figure 6, the predicted lesion contours accurately follow irregular lesion boundaries while simultaneously suppressing false-positive regions outside the lesion. Consistent with the comparative results presented in Figure 5, the proposed framework produces segmentation masks that most closely resemble the expert-annotated ground truth.

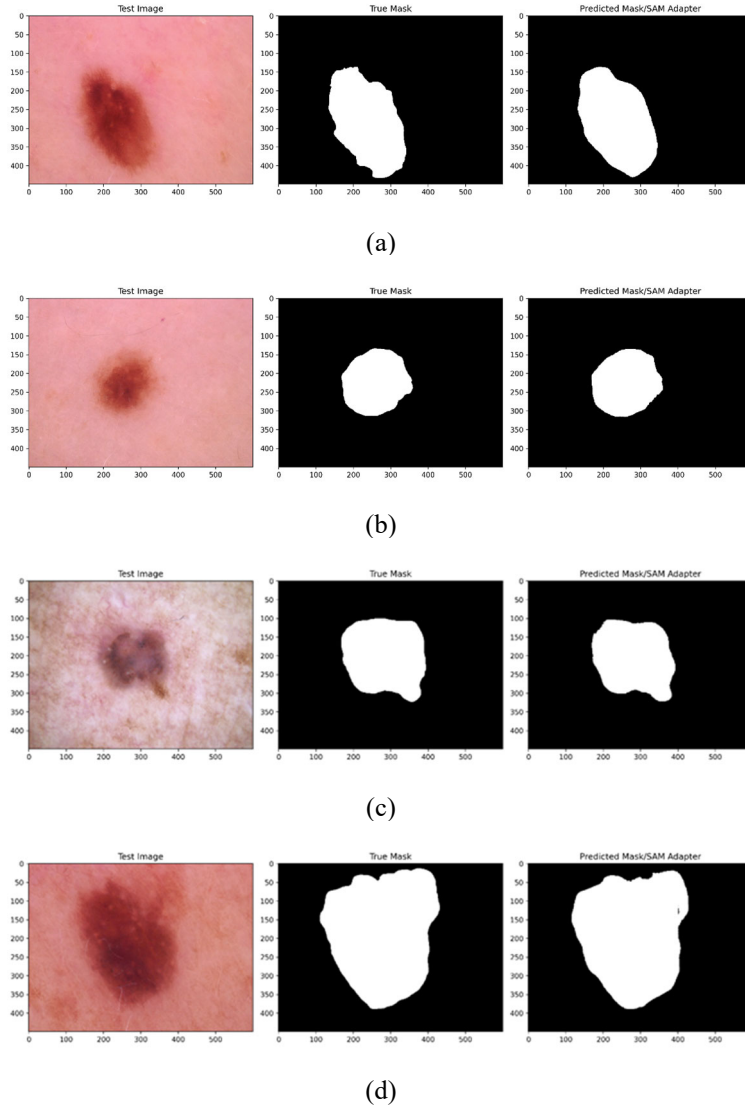


Figure 6. Representative prediction examples obtained using the proposed SAM-Adapter on the HAM10000 dataset. (a–d) Original image, ground-truth mask, and predicted segmentation mask.

3.3.2 Qualitative analysis on ISIC2018

Representative segmentation results obtained on the ISIC2018 dataset are illustrated in Figures 7 and 8. While Figure 7 presents a direct visual comparison of the segmentation boundaries generated by SAM, MedSAM, and the proposed SAM-Adapter, Figure 8 further compares the segmentation masks produced by the proposed approach with the corresponding expert-annotated ground-truth masks. Together, these qualitative examples provide a comprehensive visual assessment of the segmentation capability of the evaluated foundation-model-based approaches.

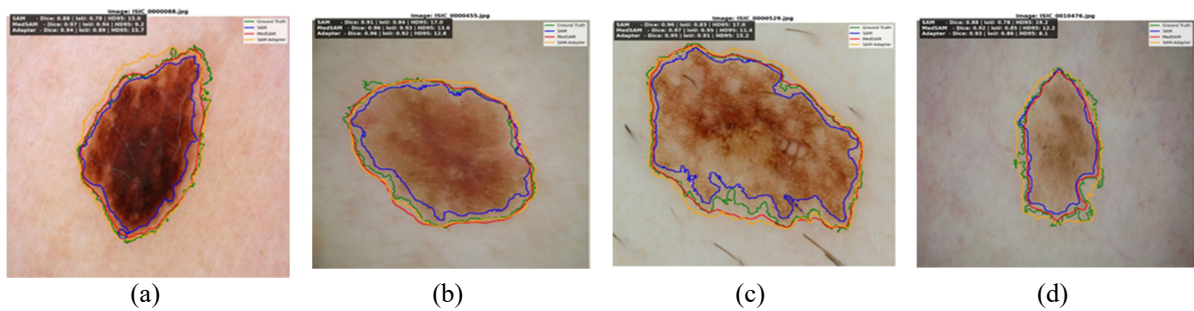


Figure 7. Representative qualitative segmentation results on the ISIC2018 dataset. (a–d) Representative lesions showing the segmentation boundaries produced by SAM, MedSAM, and the proposed SAM-Adapter.

The original SAM frequently produces incomplete segmentation masks in lesions exhibiting low contrast or fuzzy boundaries. Although the overall lesion location is correctly identified, small peripheral regions are often excluded from the predicted mask, resulting in under-segmentation. As illustrated in Figure 7, these boundary inaccuracies become particularly apparent in asymmetric lesions where accurate contour delineation is clinically important. MedSAM considerably improves lesion localization and generally produces smoother segmentation masks than the original SAM. However, small discrepancies remain observable along complex lesion boundaries, especially in images containing gradual intensity transitions between the lesion and surrounding skin. Although medical-domain pretraining enhances segmentation quality, minor deviations from the expert annotations remain visible in several representative cases shown in Figure 7. In contrast, the proposed SAM-Adapter exhibits the highest visual agreement with the expert-annotated ground truth. Across the representative examples, lesion contours generated by the adapter-based framework closely follow the reference boundaries while preserving fine morphological structures. The proposed model also demonstrates improved robustness against imaging artifacts, illumination variations, and heterogeneous pigmentation, producing segmentation masks with fewer discontinuities and reduced boundary leakage compared with the competing approaches.

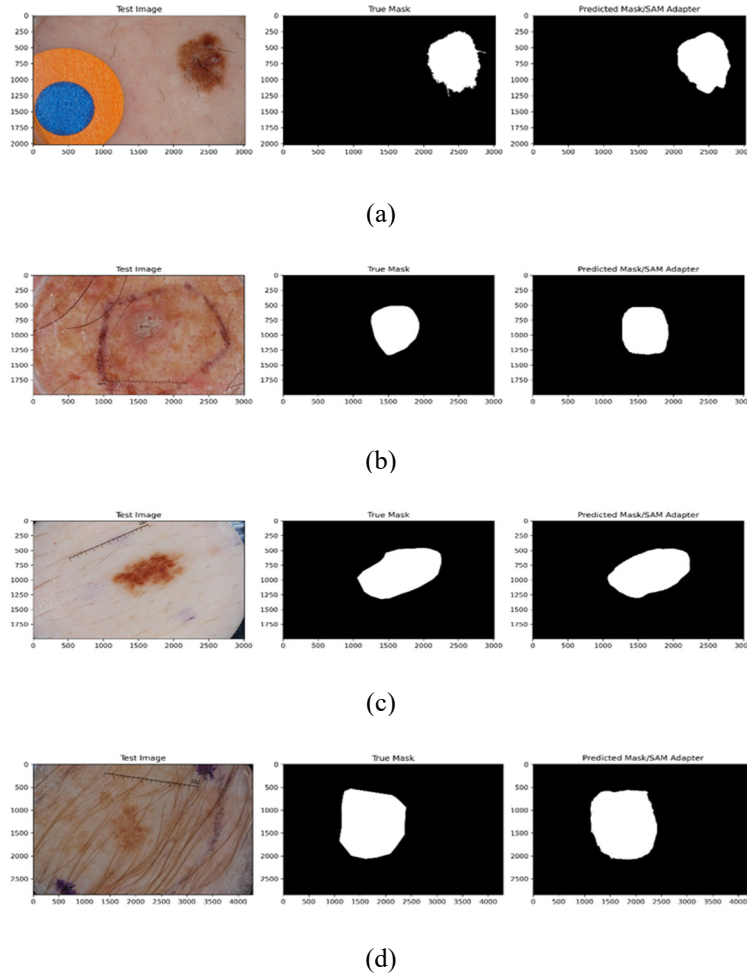


Figure 8. Representative prediction examples obtained using the proposed SAM-Adapter on the ISIC2018 dataset. (a–d) Representative lesion examples. For each example, the columns (from left to right) show the original dermoscopic image, the expert-annotated ground-truth mask, and the segmentation mask predicted by the proposed SAM-Adapter.

The prediction examples presented in Figure 8 further demonstrate the capability of the proposed SAM-Adapter to generate segmentation masks that closely match the expert annotations across lesions with varying sizes, shapes, and pigmentation patterns. The high visual agreement between the predicted masks and the corresponding ground-truth masks supports the quantitative improvements observed in Dice, IoU, and HD95 metrics, demonstrating that the proposed framework effectively preserves lesion morphology while maintaining accurate boundary delineation.

3.3.3 Visual interpretation of the proposed framework

The qualitative comparisons presented in Figures 5–8 reveal an important characteristic of the proposed model. Rather than simply increasing quantitative overlap metrics, adapter-based fine-tuning substantially improves the

anatomical plausibility of predicted lesion boundaries. Moreover, the consistent visual improvements observed across both datasets indicate that the proposed SAM-Adapter adapts effectively to lesions exhibiting diverse shapes, sizes, pigmentation patterns, and boundary complexities. Accurate preservation of lesion morphology is particularly important in dermoscopic image analysis because clinically relevant diagnostic criteria, including asymmetry, border irregularity, and lesion size, directly depend on reliable boundary delineation. Consequently, even relatively small improvements in contour localization may positively influence subsequent computer-aided diagnosis systems. Overall, the qualitative observations strongly support the quantitative findings presented in the previous section and demonstrate that lightweight adapter-based fine-tuning effectively transfers the general visual knowledge of the pretrained Segment Anything Model to the domain of dermoscopic skin lesion segmentation.

3.4 Statistical Comparison

To determine whether the observed differences among the evaluated segmentation models were statistically significant, a paired Wilcoxon Signed-Rank Test was performed for each quantitative evaluation metric. Since segmentation results obtained from SAM, MedSAM, and the proposed SAM-Adapter correspond to the same dermoscopic images, the observations are naturally paired. In addition, preliminary inspection of the metric distributions indicated that the assumption of normality could not be guaranteed. Therefore, the Wilcoxon Signed-Rank Test was selected as an appropriate non-parametric statistical method for paired comparisons.

Table 5. Wilcoxon Signed-Rank test results comparing the segmentation performance of the evaluated models. All p-values indicate statistically significant differences ($p < 0.05$).

Metric	p-value	Significant ($p < 0.05$)
Dice	2.51×10^{-10}	Yes
IoU	1.53×10^{-10}	Yes
HD95	3.10×10^{-8}	Yes

The statistical analysis demonstrated that the performance improvements achieved by the proposed SAM-Adapter were statistically significant across the principal segmentation metrics. In particular, statistically significant differences were observed for the Dice coefficient, IoU, and HD95, confirming that the observed performance gains cannot be attributed to random variation but instead reflect the effectiveness of the proposed parameter-efficient adaptation strategy. The Wilcoxon Signed-Rank Test produced statistically significant differences for all primary evaluation metrics ($p < 0.05$), confirming that the improvements achieved by the proposed SAM-Adapter are statistically significant rather than resulting from random variation. Beyond statistical significance, the observed improvements also demonstrate practical relevance. The simultaneous increase in Dice and IoU together with the reduction in HD95 indicates that the proposed model not only improves regional overlap but also achieves more accurate lesion boundary delineation. Since lesion boundary accuracy directly influences quantitative lesion characterization and subsequent computer-aided diagnostic analysis, these improvements are clinically meaningful rather than merely statistically significant. Overall, the statistical analysis strongly supports the quantitative and qualitative observations presented in the previous sections. The consistency of the obtained results across two independent benchmark datasets further suggests that the proposed adapter-based fine-tuning strategy provides robust and reliable segmentation performance for dermoscopic image analysis.

4. DISCUSSION

The present study comprehensively compared the segmentation performance of the original Segment Anything Model (SAM), MedSAM, and the proposed SAM-Adapter using two publicly available dermoscopic image datasets. Quantitative evaluation, qualitative visual analysis, and statistical significance testing consistently demonstrated that the proposed SAM-Adapter achieved superior segmentation performance under identical experimental conditions. Compared with both baseline foundation-model approaches, the proposed model produced higher Dice and IoU values while simultaneously reducing HD95, indicating more accurate lesion boundary delineation. These findings suggest that parameter-efficient adapter-based fine-tuning effectively transfers the pretrained knowledge of foundation models to dermoscopic skin lesion segmentation [17].

Unlike conventional end-to-end fine-tuning strategies that update the entire network, the proposed model preserves the pretrained SAM backbone and optimizes only lightweight adapter modules. This strategy substantially reduces the number of trainable parameters while maintaining the rich semantic representations learned during large-scale pretraining. Consequently, the proposed method achieves an effective balance between computational efficiency and segmentation accuracy, making it suitable for medical image segmentation tasks where computational resources and annotated training data are often limited.

The quantitative improvements observed in Dice, IoU, and HD95 indicate that adapter-based learning improves both regional overlap and lesion boundary localization. Accurate delineation of lesion boundaries is particularly important for dermoscopic image analysis because clinically relevant diagnostic criteria, including lesion asymmetry, border irregularity, diameter, and morphological shape descriptors, directly depend on reliable segmentation results. Therefore, improved segmentation accuracy may positively influence subsequent computer-aided diagnosis systems and quantitative lesion analysis. Fine-tuning strategies such as SkinSAM [14] have

demonstrated improved performance by updating SAM's encoder parameters on skin lesion datasets. Similarly, the Medical SAM Adapter [15] introduced adapter modules within SAM's architecture for general medical image segmentation, achieving state-of-the-art results across multiple modalities. Table 6 summarizes representative recent studies related to foundation-model-based skin lesion segmentation and compares them with the proposed model.

Table 6. Comparison of the proposed SAM-Adapter framework with representative recent studies on dermoscopic skin lesion segmentation. The proposed study is distinguished by evaluating three foundation-model-based approaches on two public datasets using both quantitative metrics and statistical significance testing.

Study	Dataset(s)	Method	Dice	IoU	Statistical Analysis	Main Contribution
S. Gül et al. (2024) [18]	Public dermoscopic datasets	Semantic SAM	0.651	0.562	No	Demonstrated the advantages and limitations of Semantic SAM for skin lesion segmentation.
Binzagr & Hariri (2024) [13]	ISIC2018, HAM10000, PH2	SAM-Adapter + Vision Transformer	0.943	0.903	Yes	Showed that lightweight adapter tuning improves SAM performance while preserving computational efficiency.
Söylemez (2025) [19]	ISIC2017, ISIC2018, HAM10000	SkinAttn-Net	0.951	0.911	No	Proposed attention-enhanced architecture for accurate lesion boundary delineation.
Current Study	HAM10000, ISIC2018	SAM / MedSAM / SAM-Adapter	0.9603	0.9241	Yes (Wilcoxon)	Comprehensive quantitative and qualitative comparison of foundation-model-based segmentation methods with statistical validation.

As summarized in Table 6, recent studies have consistently demonstrated that domain-specific adaptation considerably improves the segmentation performance of foundation models in medical imaging. Gül et al. (2024) highlighted the limitations of directly applying the original SAM to dermoscopic images and emphasized the necessity of domain adaptation and post-processing for reliable lesion segmentation [18]. Similarly, Binzagr and Hariri (2024) demonstrated that integrating lightweight adapter modules with a frozen SAM encoder significantly improves segmentation accuracy while preserving computational efficiency [13]. Furthermore, Söylemez (2025) reported that attention-enhanced transformer architectures substantially improve lesion boundary delineation by refining feature representations extracted from dermoscopic images [19].

The findings of the present study are consistent with these previous observations while providing a more comprehensive evaluation. Unlike earlier studies that generally investigated a single adaptation strategy, the proposed work compares the original SAM, MedSAM, and SAM-Adapter under identical preprocessing, optimization, and evaluation protocols using two publicly available benchmark datasets. Moreover, the inclusion of quantitative evaluation, qualitative visual analysis, and Wilcoxon Signed-Rank statistical testing provides stronger evidence that the observed performance improvements are not attributable to random variation.

Beyond the quantitative improvements observed in this study, the proposed SAM-Adapter offers important practical implications for computer-aided skin cancer diagnosis. Accurate lesion segmentation constitutes the first and one of the most critical stages of automated dermoscopic image analysis because subsequent diagnostic procedures, including feature extraction, lesion characterization, and classification, directly depend on reliable lesion boundaries. Improved segmentation accuracy therefore contributes not only to higher evaluation metrics but also to more reliable morphological measurements, including lesion asymmetry, border irregularity, diameter, and shape descriptors used in melanoma assessment. Consequently, the proposed method has the potential to improve the robustness and consistency of downstream diagnostic systems while reducing the manual annotation burden for dermatologists.

Another important advantage of the proposed model is its parameter-efficient design. Unlike conventional fine-tuning approaches that require updating the entire network, only lightweight adapter modules are optimized while the pretrained SAM encoder remains frozen. This considerably reduces computational cost, memory consumption, and training time while preserving the strong generalization capability of the pretrained foundation model. Such characteristics make the proposed approach particularly attractive for clinical environments where computational resources and annotated medical datasets are often limited.

To fully contextualize the proposed SAM-Adapter within the broader landscape of medical image segmentation, it is essential to contrast it with standardized traditional architectures such as U-Net and Swin-UNet. Conventional CNN-based models like U-Net, along with purely Transformer-based architectures like Swin-UNet, have long served as the gold standard in medical segmentation. However, they typically rely on training from scratch or limited domain-specific pretraining, requiring large, fully annotated datasets to capture complex lesion morphologies. Furthermore, these traditional architectures often require updating millions of parameters during

full fine-tuning, which increases the risk of overfitting on small clinical datasets and limits their out-of-distribution generalization.

In contrast, the SAM-Adapter paradigm represents a fundamental shift from traditional approaches. By leveraging the massive generalized knowledge of a foundation model pre-trained on over 11 million images (SA-1B), the proposed framework circumvents the heavy data dependency of U-Net and Swin-UNet. Instead of discarding universal visual representations, it seamlessly adapts to the nuanced characteristics of dermoscopic lesions using merely ~5.5% trainable parameters. This positions the adapter-enhanced foundation model as a highly parameter-efficient and robust alternative to traditional fully-trained networks, offering superior dataset-specific adaptation capabilities for dermatological CAD systems where annotated data and computational resources are often constrained.

Despite the promising results, several limitations should be acknowledged. First, the experimental evaluation was conducted exclusively on two publicly available dermoscopic image datasets. Although both HAM10000 and ISIC2018 are widely accepted benchmark datasets, separate adapter models were trained for each dataset in this study. Consequently, the current findings primarily establish high performance within specific dataset distributions rather than true cross-dataset generalization. Further validation using zero-shot cross-dataset evaluation protocols on independent clinical cohorts would provide stronger evidence regarding the absolute generalizability of the proposed approach. Secondly, this study employed ground-truth derived bounding boxes (oracle prompts) during the test phase to strictly evaluate the boundary delineation capacity of the models. While this prevents localization errors from skewing the segmentation metrics, it introduces a potential discrepancy with fully autonomous real-world clinical workflows where ground-truth masks are unavailable. In a practical clinical setting, the proposed SAM-Adapter framework could be deployed in two ways: as a highly efficient interactive tool where dermatologists provide a rough bounding box manually, or as part of a fully automated pipeline by integrating an upstream lightweight object detector (e.g., YOLO) to generate the initial bounding boxes prior to segmentation. Third, only the ViT-B version of SAM was investigated. Evaluating larger backbone architectures, including ViT-L and ViT-H, may further improve segmentation performance at the expense of increased computational complexity. Finally, although segmentation accuracy was extensively analyzed, inference time, computational efficiency, and hardware utilization were not quantitatively compared and should be considered in future studies. Future research may investigate more efficient adapter architectures, adaptive prompt generation techniques, and domain adaptation strategies for multi-center medical image datasets. Another promising direction involves integrating foundation-model-based segmentation with downstream skin lesion classification and clinical decision-support systems to establish a unified computer-aided diagnosis framework. Furthermore, extending the proposed parameter-efficient adaptation strategy to other medical imaging modalities, such as ultrasound, magnetic resonance imaging, computed tomography, and histopathological images, may provide additional evidence regarding its general applicability across diverse medical image segmentation tasks.

5. CONCLUSION

This study presents a comprehensive comparative evaluation of SAM, MedSAM, and the proposed parameter-efficient SAM-Adapter framework for dermoscopic skin lesion segmentation on the HAM10000 and ISIC2018 datasets. Experimental results and Wilcoxon statistical analysis confirmed that SAM-Adapter significantly outperformed baseline foundation models across key evaluation metrics (Dice, IoU, Recall, Accuracy, and HD95), while producing lesion contours that closely matched expert annotations, particularly in challenging cases with irregular boundaries. By freezing the pretrained SAM encoder and optimizing only lightweight adapter modules, the proposed methodology substantially reduces computational complexity without sacrificing segmentation accuracy. Overall, these findings demonstrate that parameter-efficient adaptation is a robust, accurate, and practical strategy for transferring foundation models to domain-specific medical imaging, paving the way for its broader integration into future computer-aided diagnosis systems.

Author Contributions

K. Akyol: Conceptualization, formal analysis, methodology, software, visualization, writing—original draft. M. Şeker: methodology, software, supervision, validation, writing—review and editing.

Acknowledgement

The authors would like to thank the International Skin Imaging Collaboration (ISIC) and the HAM10000 dataset providers for making the dermoscopic datasets publicly available, thereby enabling reproducible research in medical image analysis.

Data Availability

The datasets used in this study are publicly available from the International Skin Imaging Collaboration (ISIC) archive and the HAM10000 dataset repository.

Conflict of Interest

The authors declare that they have no conflict of interest.

REFERENCES

- [1] F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 74, no. 3, pp. 229–263, May 2024.
- [2] H. Naseri and A. A. Safaei, "Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: a systematic literature review," *BMC Cancer*, vol. 25, no. 1, p. 75, 2025.
- [3] L. Mevorach, A. Farcomeni, G. Pellacani, and C. Cantisani, "A Comparison of Skin Lesions' Diagnoses Between AI-Based Image Classification, an Expert Dermatologist, and a Non-Expert," 2025.
- [4] M. Yang and N. I. R. Ruhaiyem, "MCM-UNet: Mamba convolutional mixing network for skin lesion image segmentation," *Biomed. Signal Process. Control*, vol. 112, p. 108791, 2026.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [6] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565–571.
- [7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, and D. Weissenborn, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *ICLR 2021*, 2021. [Online]. Available: <https://arxiv.org/abs/2010.11929>
- [8] Y. Wang et al., "Attention-enhanced U-Net based network for cancerous tissue segmentation," *Biomed. Signal Process. Control*, vol. 106, p. 107728, 2025.
- [9] F. Li et al., "Semantic-SAM: Segment and Recognize Anything at Any Granularity," *arXiv Prepr. arXiv2307.04767*, pp. 1–14, 2023, [Online]. Available: <https://arxiv.org/abs/2307.04767>
- [10] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [11] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nat. Commun.*, vol. 15, no. 1, p. 654, Jan. 2024.
- [12] Z. Su et al., "Adapting SAM to Histopathology Images for Tumor Bud Segmentation in Colorectal Cancer.," *Proc. SPIE--the Int. Soc. Opt. Eng.*, vol. 12933, Feb. 2024.
- [13] F. Binzagr and M. Hariri, "Foundation-Model-Driven Skin Lesion Segmentation and Classification Using SAM-Adapters and Vision Transformers," 2026.
- [14] M. Hu, Y. Li, and X. Yang, "SkinSAM: Empowering Skin Cancer Segmentation with Segment Anything Model," in *CVPR*, 2023. [Online]. Available: <https://arxiv.org/abs/2304.13973>
- [15] J. Wu et al., "Medical SAM adapter: Adapting segment anything model for medical image segmentation," *Med. Image Anal.*, vol. 102, p. 103547, May 2025.
- [16] T. Chen et al., "SAM Fails to Segment Anything? -- SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More," Apr. 2023.
- [17] T. Chen et al., "SAM-Adapter: Adapting Segment Anything in Underperformed Scenes," in 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2023, pp. 3359–3367.
- [18] S. Gül, B. M. Aydın, D. Akgün, R. Öztaş Kara, and G. Çetinel, "Skin Lesion Segmentation with Semantic SAM: Pros and Cons," *Selcuk Univ. J. Eng. Sci. Vol 23, No 3 Selcuk Univ. J. Eng. Sci.*, Dec. 2024, [Online]. Available: <https://sujes.selcuk.edu.tr/sujes/article/view/660>
- [19] Ö. F. Söylemez, "SkinAttn-Net: a multi-level attention-based network for skin lesion segmentation," *Sci. Rep.*, vol. 16, no. 1, p. 3700, 2025.