Journal of Artificial Intelligence and Data Mining (JAIDM), x(x): xxx-xxx, xxxx

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

A Novel Hybrid Deep Learning Model with Receptive Field-Enhanced Skip Connections and Adaptive Loss for Medical Image Segmentation

Mahdi Zarrin* and Haniyeh Nikkhah

Faculty of Electrical and Computer Engineering, University of Tabriz, Iran.

Article Info

Article History:
Received 10 July 2025
Revised 19 August 2025
Accepted 15 September 2025

DOI:10.22044/jadm.2025.16467.2771

Keywords:
Medical Image Segmentation; Loss Function; Data Imbalance; Skin Lesion; Receptive Field Enhancement.

**Corresponding author: mahdi.zarrin1401@ms.tabrizu.ac.ir (M. Zarrin).*

Abstract

Medical image analysis, crucial for disease diagnosis and treatment, often suffers from the challenge of class imbalance, where normal tissue significantly outweighs abnormal regions in terms of area. Furthermore, varying class ratios across different images within a dataset complicate the application of uniform loss adjustments. To address these issues and advance automated segmentation, this study proposes a novel deep learning model that integrates the strengths of YOLO Version 8's efficient feature extraction modules (SPPF and C2F) within a U-shaped architecture enhanced by a Receptive Field Enhancement (RFE) module. The RFE module, acting as an advanced skip connection, strategically fuses multi-scale features from corresponding and subsequent encoder layers processed through SPPF and C2F to enrich feature transfer and improve receptive field. To specifically tackle the class imbalance and the diversity of class distributions across images, we introduce a novel Adapt Exponential Loss function. This pixel-level loss dynamically adjusts class weights for each image based on its individual lesion-to-total-pixel ratio (k). We evaluated our proposed model and loss function on challenging skin lesion datasets: ISIC 2018, ISIC 2017, and PH2. Our method achieved significant segmentation performance with IoU scores of 86.47%, 85.67%, and 93.13%, and Dice scores of 91.63%, 90.19%, and 96.02% on ISIC 2018, ISIC 2017, and PH2, respectively, demonstrating its effectiveness in accurately delineating skin lesions despite class imbalance and varying lesion proportions. This work contributes a robust framework for medical image segmentation, facilitating more reliable diagnostic tools in dermatology.

1. Introduction

Medical images play a crucial role in diagnosing, treating, and preventing diseases. Analyzing these images through segmentation and localization of lesion areas is critical, and deep learning has emerged as a powerful tool in this domain. By leveraging deep learning, medical images can be automatically analyzed to extract meaningful features, identify specific structures, and diagnose diseases. While widely applied in imaging modalities such as X-rays, CT scans, and MRIs, deep learning's potential in natural images, such as skin images, has also been recognized [1]. In skin cancer diagnosis, the primary goal is to segment and differentiate malignant lesions from other skin regions [2]. This study focuses specifically on image segmentation, a key method in medical image analysis.

Image segmentation, which divides images into discrete regions corresponding to tissues, organs, or lesions, has become essential in medical image processing [3]. Deep learning architectures such as convolutional neural networks (CNNs), including U-Net [4], SegNet [5], and fully convolutional networks (FCNs) [6], have shown great success in

identifying structures of interest within medical images. Advanced architectures such as U-Net++ [7], Attention U-Net [8], 3D U-Net [9], ResU-Net [11] and further improve segmentation results.

Recently, YOLO (You Only Look Once) [14] and ResNet [15] architectures have also demonstrated strong potential in medical image segmentation. YOLO, renowned for its real-time object detection, enhances segmentation accuracy and processing speed [16]. ResNet, with its residual connections, excels in handling deep networks by mitigating the vanishing gradient problem, enabling precise feature extraction [17, 18]. Combining YOLO's object detection capabilities with ResNet's depth and accuracy creates a robust framework for medical image segmentation, enhancing both speed and precision [19, 20].

However, segmentation in medical images faces challenges such as data complexity, noise, interpretive limitations, and class imbalance [21]. Class imbalance, where background pixels outnumber lesion pixels, poses a significant obstacle to accurate anomaly segmentation. Additionally, the intricate and subtle nature of anomalies makes feature extraction difficult. To address these challenges, specialized loss functions [20, 22] and modules like attention mechanisms [23] have been developed.

This study aims to improve skin disease analysis and diagnosis using natural medical images by integrating YOLO-V8 and ResNet. A new custom module, inspired by the SPPF and C2F modules used in YOLO-V8, is introduced to enhance information extraction from medical images. This module combines YOLO-V8's strengths in segmentation with ResNet's robust feature extraction, improving accuracy and efficiency in dermatological assessments. To tackle class imbalance, a novel adaptive weighted loss function is proposed. This loss function calculates the ratio of lesion pixels to total pixels for each image and assigns distinct weights to lesion and background regions. This approach ensures accurate segmentation, particularly for small lesions where imbalance is most problematic. Key contributions of this study are as follows:

- Model Architecture: A U-shaped deep learning model leveraging the ResNet backbone is designed. Custom modules inspired by YOLO-V8's SPPF and C2F components are incorporated to enhance the receptive field and capture critical details effectively.
- Adaptive Weighted Loss Function: A refined loss function, building on our previous work (ExpUNET [3]), dynamically adjusts weights based on the proportion of lesion pixels in each image, addressing data imbalance and improving segmentation precision, particularly for small lesions.
- Comprehensive Validation: Extensive experiments are conducted using datasets such as ISIC 2018, ISIC 2017, and PH2 to validate the model's effectiveness in segmenting lesions from natural skin images.

By addressing segmentation challenges and improving model performance, this research aims to advance medical image analysis and support informed decision-making in dermatology.

2. Related works

2.1. Medical Image Segmentation (Lesion Segmentation Focused on ISIC 2018, ISIC 2017, and PH2)

In medical image segmentation for skin lesions, many studies have advanced automated diagnostic techniques, often using the ISIC 2017, ISIC 2018, and PH2 datasets [25-27] as benchmarks. These datasets have been pivotal in developing and evaluating segmentation algorithms for identifying skin lesions, including melanoma. Below, we review key works that utilize these datasets, illustrating the progress in lesion segmentation and how our approach builds upon existing methods.

The authors in [28] proposed a UNet-based model with multi-scale residual connections, tested on ISIC 2017, achieving an intersection-over-union (IoU) of 76.40%. Al-Masni et al. (2018) [29] introduced a full-resolution convolutional network (FrCN) tested on ISBI 2017 and PH2, achieving Jaccard indices of 77.11% and 84.79%, respectively. A deep learning ensemble method using Mask R-CNN and DeepLabV3+ was proposed in [30], with IoU scores of 79.30 and 80.00 on ISIC 2017 and PH2. Baghersalimi et al. (2019) [31] introduced DermoNet, a fully convolutional network for skin lesion segmentation, tested on ISBI 2016, ISBI 2017, and PH2, with Jaccard coefficients of 85.30% for PH2 and 78.30% for ISIC 2017. Abraham and Khan (2019) [34] addressed data imbalance using a generalized focal loss based on the Tversky index, achieving an IoU of 74.80% on ISIC 2018. A transductive segmentation approach [35] improved generalization, achieving an IoU of 83.00% on ISIC 2018. Other works include a modified conditional generative adversarial network (cGAN) with multi-scale input for enhanced lesion segmentation, achieving IoU scores of 77.20% on ISIC 2018 and 78.94% on ISBI 2017 [37].

Table 1. Summary of deep learning models for skin lesion segmentation and their performance, evaluated using the Intersection over Union (IoU) metric across different datasets. Abbreviations: Seg. = segmentation, IoU = Intersection over Union, con. = connection

Reference	Architecture Modules	Segmentation Loss	IoU		
			ISIC 2017	ISIC 2018	PH2
Venkatesh et al. (2018) [28]	residual con. skip con.	Jaccard	76.40%	-	-
Al-Masni et al. (2018) [29]	Pyramid pooling	CE	77.11%	-	84.79%
Goyal et al. (2019)[30]	dilated conv. parallel m.s. conv. separable conv.	-	79.30%	-	80.00%
Baghersalimi et al. (2019) [33]	skip con. residual con. dense con.	Tanimoto	78.30%	-	85.30%
Abraham and khan (2019) [34]	skip con. image pyramid attention	TV Focal	-	74.80%	-
Cui et al. (2019) [35]	dilated conv. parallel m.s. conv. separable conv	-	-	83.00%	-
Singh et al. (2019) [36]	skip con. residual con. factorized conv. attention mod. GAN	CE ℓ1 EPE	78.94%	77.20%	-
Saini et al. (2019) [37]	skip con. Multi-task	Dice	84.40%	80.90%	89.10%
Nathan and Kansal (2020) [38]	skip con. residual con.	CE Dice	78.28%	83.25%	83.84%
Jin et al. (2021) [39]	skip con. residual con. attention mod	Dice Focal	80.00%	79.40%	-
Saini et al. (2021) [40]	pyramid pooling residual con. skip con. dilated conv. attention mod	Dice	85.00%	84.00%	87.00%
Chen et al. (2022) [41]	skip con. attention mod. recurrent net	CE	80.36%	-	86.15%
Liu et al. (2022) [42]	skip con. dilated conv. multi-scale pyramid pooling Transformer	CE	80.19	83.79	-
Kaur et al. (2022) [43]	-	Tversky	77.80%	85.20%	88.80%
Jiacheng Ruan et al. (2023) [44]	Hadamard Product Attention. Group Aggregation Bridge. multiple attention	CE Dice	79.81%	80.94%	-
Sun et al. (2023)	Multi-scale bridge module. Global-local channel spatial attention module. Scale-aware deep supervision module.	Softdice BCE	79.26%	84.18%	86.86%

Saini et al. (2019) [38] proposed a Faster R-CNN-based method, yielding IoU scores of 89.10% on PH2 and 80.90% on ISIC 2018. The Lesion Net [39] improved segmentation by incorporating a coordinate convolutional layer, achieving test accuracies of 83.84% on PH2 and 83.25% on ISIC 2018. Jin et al. (2021) [40] introduced the Cascade Knowledge Diffusion Network (CKDNet), showing superior performance compared to state-of-the-art methods on ISIC 2017 and 2018. Saini et al. (2021) [41] proposed a CNN model with multiple encoders, achieving 87.0% accuracy on PH2. Chen et al. (2022) [42] developed O-Net, a recurrent attentional CNN for skin lesion segmentation, achieving 86.5% accuracy on PH2. IASTrans-Net [43] improved segmentation accuracy with a multi-scale feature extraction module, outperforming existing methods on ISIC 2017 and 2018. EGE-UNet [44] reduced computational demand by incorporating multi-axis attention and achieved improved performance on ISIC 2017 and 2018. MSCA-Net [45] used multi-scale fusion and attention mechanisms, achieving better results on ISIC 2017, 2018, and PH2. Finally, Qiu et al. (2023) [46] introduced GFANet, achieving IoU scores of 77.75%, 83.66%, and 90.98% on ISIC 2017, ISIC 2018, and PH2, respectively. A summary of the reviewed deep learning-based segmentation methods and their performance across ISIC 2017, ISIC 2018, and PH2 datasets is presented in Table 1.

2.2. Loss Functions for Imbalanced Medical Datasets

Class imbalance in medical datasets [68], where certain disease categories are underrepresented, presents significant challenges for deep learning models, particularly in accurately identifying rare conditions. Traditional loss functions, such as cross-entropy, often fail to address these imbalances, leading to biased predictions. To tackle this issue, various specialized loss functions have been proposed [69]. For instance, Nasalwai et al. (2021) [21] introduced the accelerated Tversky loss (ATL), which improved model performance on imbalanced datasets. Scholz et al. (2024) [48] combined loss functions based on Matthews Correlation Coefficient (MCC) and F1 score with cross-entropy, boosting performance for minority classes. N. Le et al. (2021) [49] proposed the Offset Curves (OsC) loss, which excelled in pixel-wise segmentation for imbalanced data. Mariano Cabezas and Yago Diez (2024) [50] highlighted the importance of foreground weighting to improve lesion segmentation, particularly in brain MRI. Roy et al. (2022) [51] developed the Balanced Weighted Categorical Cross-Entropy (BWCCE) loss for COVID-19 detection in Chest X-rays, achieving high F1 scores and AUC. F. Kofler et al. (2023) [53] introduced the Blob loss to enhance instance-level segmentation for multiple sclerosis and liver tumors, while Sugino et al. (2023) [54] demonstrated that focal weighting improved segmentation accuracy for brain structures. These advancements aim to improve model performance and the detection of rare but critical medical conditions.

3. Methodology

In this section, we provide a detailed description of the network architecture (see Figure 1). Next, we introduce the Receptive Field Enhancement module, which is designed to extract valuable feature representations. Finally, we discuss the loss function, which addresses the imbalance problem in medical images by considering the distribution of each image.

3.1. Proposed network architecture

In this study, we introduce a novel U-shaped architecture built on the ResNet backbone, selected for its effective mitigation of the vanishing gradient problem through residual connections [55]. These connections help preserve essential information, thereby enhancing network performance [17]. Our model follows a traditional encoder-decoder framework with key skip connections between corresponding layers, as depicted in Figure 1, and incorporates four ResNet-101 blocks seamlessly. In the decoder, each stage involves up-sampling using a 2×2 convolution, followed by a convolutional layer, batch normalization (BN), same padding, and the Sigmoid-weighted Linear Unit (SiLU) activation function.
function. The encoder features a Receptive Field Enhancment (RFE) module inspired by YOLOv8 components,which improves feature map decoding by utilizing inputs from different parts of the encoder, leading to better image reconstruction. Finally, the last layer employs a softmax activation function to generate probabilistic predictions.
To further enhance the representational capacity of the proposed network, multi-scale feature learning is implicitly encouraged through the hierarchical structure of the encoder–decoder design. This design choice contributes to improved convergence stability and robustness, particularly when dealing with heterogeneous and high-dimensional medical imaging data.

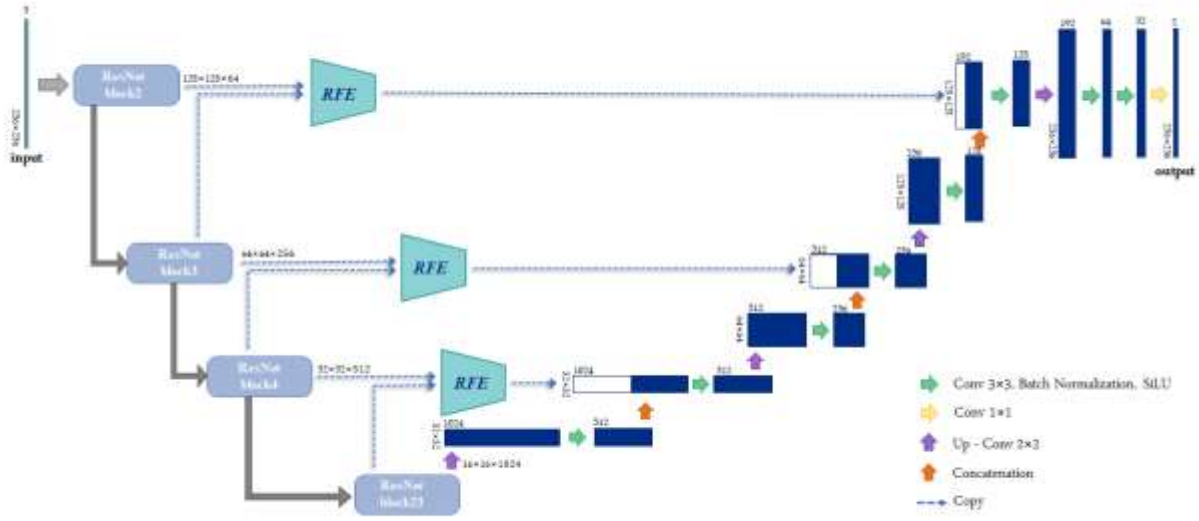

Figure 1. The proposed network architecture for image segmentation uses an encoder-decoder structure with skip connections that transfer feature maps between layers.

3.2. Receptive Field Enhancement (RFE)

To enhance the accuracy of mask prediction in our U-shaped network, we introduce the Receptive Field Enhancement (RFE) module. This module is strategically integrated into the skip connections between the encoder and the decoder pathways. The RFE module is designed to broaden the network's receptive field and extract richer, multi-scale features. It effectively leverages the strengths of the Spatial Pyramid Pooling Fast (SPPF) and the efficient C2F modules, both integral components of the YOLOv8 architecture. By incorporating the RFE into the skip connections, we aim to transfer enhanced feature representations across different scales, facilitating more precise and detailed mask predictions in the decoder.

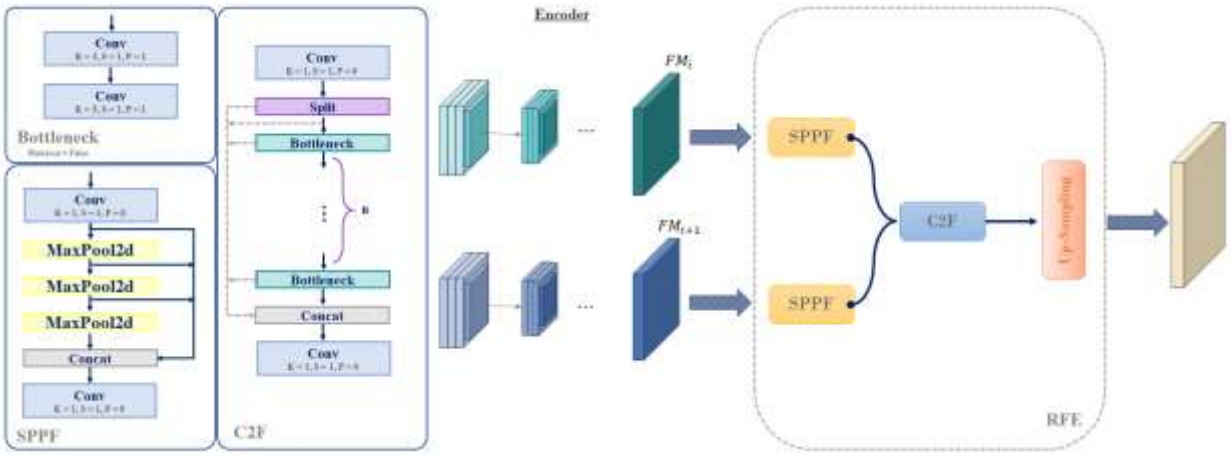

Figure 2. The structure of the Receptive Field Enhancement (RFE) module is designed to boost feature extraction and expand the receptive field.

The C2F module, adapted from YOLOv5's C3 module [56], is designed to reduce computational complexity and enhance gradient performance. It improves feature integration by combining high-level semantic and low-level spatial features, thereby enhancing the detection of small objects. Unlike the C3 module, which utilizes a single DarkNet bottleneck and splits the input into two paths for processing and concatenation [57], the C2F module optimizes the network structure by balancing gradient flow and facilitating multi-scale feature learning. By concatenating outputs from multiple bottlenecks, the C2F module expands the receptive field through feature vector divergence

and multi-level convolution without altering gradient flow. This design significantly boosts feature extraction and object segmentation accuracy, as illustrated in Figure 3.

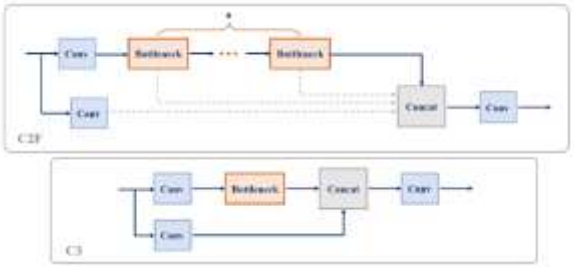

Figure 3. YOLOv5 and YOLOv8 module different.

The proposed module incorporates the Spatial Pyramid Pooling Fast (SPPF) module (Figure 2), a key component of YOLOv8, designed for multi-scale feature extraction using Spatial Pyramid Pooling [58]. The SPPF structure employs multiple spatial pooling levels with varying kernel sizes to capture diverse features [59], enabling effective recognition of both small and large objects [60]. It also enhances computational efficiency, boosting processing speed [67]. Unlike its typical use as a single block in the final feature map of YOLO models, we uniquely utilize two SPPF modules within our skip connections to capture multi-scale information from different encoder layers. In medical image analysis, such as skin lesion segmentation, the SPPF module is beneficial by handling lesions of varying sizes, improving feature extraction and segmentation accuracy. Additionally, its efficiency supports real-time processing in medical applications.

Our proposed module (Figure 1) functions as an advanced skip connection between the encoder and decoder, transferring crucial information using two inputs. It integrates features from both the corresponding encoder layer and the subsequent one (Figure 1). The feature maps from layers i and i+1 are processed through the SPPF module, then passed into the C2F module. The combined features are upsampled to match the i-th decoder layer's dimensions. This approach ensures richer feature transfer, improving network accuracy and performance. Key features include dual-input integration, multi-scale extraction via SPPF, and efficient merging with C2F, reducing complexity while enhancing detection and segmentation, especially for small or complex structures. The impact of our proposed RFE module is comprehensively reported in Table 8, which details its contribution to improving overall segmentation performance.

3.3. The proposed Adapt Exp Loss

An essential aspect of training a semantic segmentation model involves the design of a suitable loss function that effectively quantifies the disparity between the predicted segmentation and the ground truth segmentation. The exponential loss, described in [61], exhibits convex characteristics and undergoes gradual changes during its derivation. In [3], a probabilistic exponential loss function was introduced, drawing inspiration from the exponential loss. This exponential loss function assigns high probabilities to correct class labels for each pixel while penalizing deviations from the ground truth distribution. By minimizing this loss function during the training process, the model acquires the ability to generate accurate and semantically meaningful segmentations. The proposed loss operates at the pixel level, comparing the predicted class probabilities for each pixel with the corresponding ground truth labels. This comparison is conducted on images $X \subset R^{H \times W \times 3}$, which possess pixel-level semantic labels $y \in C^{H \times W}$, where $H, W \text{ and } C$ denote height, width, and the number of classes, respectively. The goal is to train a segmentation model $F : x \to 0,1^{C}$ that can fit on labeled data.

This research modifies our previous loss function [3] to enhance performance on datasets with varying class ratios across images. We propose the Adapt Exp Loss, which dynamically adjusts based on the class ratio in each image. As shown in Figure 4, the distribution of 255-value pixels (representing the lesion class) varies widely, with many images showing small ratios, indicating class imbalance. Since the 255-value class does not always occupy a small portion, a static loss adjustment based on imbalance alone is insufficient. The Adapt Exp Loss provides a more flexible solution to these varying class distributions.

To address this variability, the Adapt Exp Loss computes adaptive weighting factors at the image level, allowing the loss function to respond dynamically to the proportion of lesion pixels present in each sample. This adaptive strategy improves optimization stability and enhances sensitivity to small lesions while maintaining robust performance on images with more balanced class distributions, leading to more consistent segmentation accuracy across the dataset.

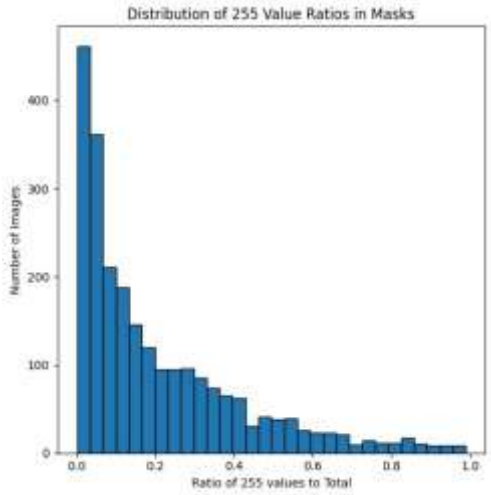

Figure 4. Histogram showing the ratio of lesion class pixels (255-value) to total pixels across dataset images, highlighting significant variability in lesion proportions and the need for a loss function that adapts dynamically to these variations.

The first challenge arises in imbalanced datasets, where the ratio between the majority and minority classes is typically greater than 1. A uniform adjustment is often applied across all input images, but, as shown in Figure 5, not all images exhibit the same conditions. In some images, the majority class may dominate, while in others, the minority class may prevail. To address this, a dynamic approach is needed, where weights are adjusted individually for each image.

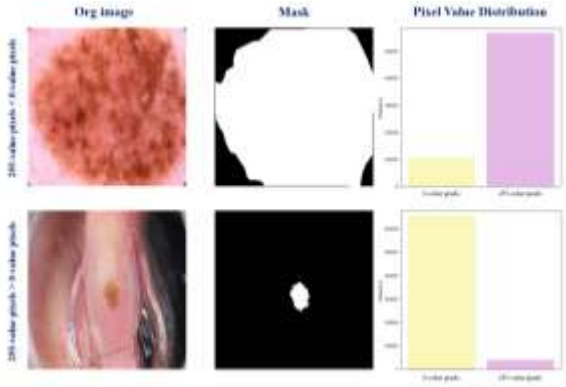

Figure 5. This figure illustrates the ratios of lesion class pixels to background class pixels across two different images. The observed variability in these ratios indicates that a uniform weight assignment for class contributions would be inadequate.

Secondly, applying the same weight to all images is inappropriate, as the ratio between the classes varies across images. As shown in Figure 6, even when the ratio of the first class to the second class is higher in all images, this ratio still differs from image to image. This highlights the need for a loss function that adjusts class weights for each individual image, allowing the function to adapt to the unique class distribution of each input.

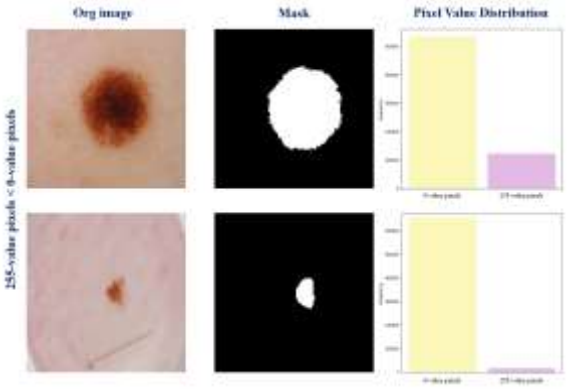

Figure 6. This graph presents an analysis of the first-to-second class ratios across various images, revealing substantial fluctuations in class distributions. The data demonstrates that, even in instances where the lesion class is predominant, the ratio continues to exhibit variability.

To implement this change, a logical weighting method is applied to each class within every image. First, the value of k is calculated for each image, representing the ratio of lesion-class pixels to the total number of pixels. Based on k, a weight is assigned to each component of the loss function for class 0 and class 1. The resulting loss function F is then computed using equation (3).

$$k = \frac{Number\ of\ \ pixels\ in\ lesion\ class}{Total\ number\ of\ \ pixels\ in\ the\ image} \tag{1}$$

This gives $k \in [0,1]$,where:

- $k = 0$ means there are no lesion (class 1) pixels in the image.
- $k = 1$ means the entire image is covered by lesion (class 1) pixels.

$$\lambda_1 = \frac{1}{k+\varepsilon},\ \lambda_2 = \frac{1}{1-k+\varepsilon} \tag{2}$$

In the formulated equation, the first term is multiplied by λ_1 and signifies the loss incurred by positive samples $(y=1)$. Conversely, the second term is multiplied by λ_2 and represents the loss associated with negative samples $(y=0)$

λ_1 is meant to adjust the importance of the lesion class (class 1) pixels. When the lesion pixels are sparse (i.e., when k is small), $\lambda_1 = \frac{1}{k+\varepsilon}$ becomes large, amplifying the loss associated with those pixels. This ensures that even though there are fewer lesion pixels, their contribution to the total loss is significant.

λ_2 performs a similar role for background pixels. When k is large (i.e., most pixels are lesion pixels), $\lambda_2 = \frac{1}{1-k+\varepsilon}$ increases, ensuring that the

contribution of background pixels is not neglected when they are the minority class.

$$L_{AdaptExp} = \lambda_1 y exp\left(-P_{F(x)}\right) + \lambda_2 (1-y) exp\left(-\left(1-P_{F(x)}\right)\right) \quad (3)$$

The variable y refers to the ground truth label associated with a pixel, which can assume the values 0 (background) or 1 (foreground). Its purpose lies in the selection of the appropriate term from the loss function, based on the class label assigned to the pixel. The term $(1-y)$ is employed to quantify the contribution of background or negative samples. It represents the complementary class label for the pixel, such that when $y=0$ (background), $(1-y)=1$, and when $y=1$ (foreground), $(1-y)=0$. The function $exp()$ denotes the exponential function. Moreover, $P(F(x))$ and $(1-P(F(x)))$ correspond to the probabilities assigned to a given pixel belonging to class $y=1$ or $y=0$, respectively.

We provide a mathematical demonstration of the convexity of our proposed loss function to highlight its theoretical soundness and optimization stability.

Theorem: Suppose function $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable over an open domain. Then f is convex if its second derivative satisfies $f''(x) >= 0$ for all x in its domain.

Proof: We prove that both terms in our proposed Adapt Exp Loss function are convex. We denote the loss function (Eq. (3)).

First Term Convexity

We compute the first derivative of the first term:

$$\frac{dT_1}{d\theta} = -\lambda_1 y \frac{dP_{F(x)}}{d\theta} exp\left(-P_{F(x)}\right) \quad (4)$$

Taking the second derivative:

$$\frac{d^2T_1}{d\theta^2} = \lambda_1 y\left(\frac{d^2P_{F(x)}}{d\theta^2} exp\left(-P_{F(x)}\right) + \left(\frac{dP_{F(x)}}{d\theta}\right)^2 exp\left(-P_{F(x)}\right)\right) \quad (5)$$

Since $\lambda_1 > 0$, $y \in 0,1$, and $\exp\left(-P_F(x)\right) > 0$ this term is always non-negative, hence convex.

Second Term Convexity

Similarly, for the second term:

$$\frac{dT_2}{d\theta} = -\lambda_2 (1-y) \frac{dP_{F(x)}}{d\theta} exp\left(-\left(1-P_{F(x)}\right)\right) \quad (6)$$

Taking the second derivative:

$$\frac{d^2T_2}{d\theta^2} = \lambda_2 (1-y)\left(\frac{d^2P_{F(x)}}{d\theta^2} exp\left(-\left(1-P_{F(x)}\right)\right) + \left(\frac{dP_{F(x)}}{d\theta}\right)^2 exp\left(-\left(1-P_{F(x)}\right)\right)\right) \quad (7)$$

Again, since $\lambda_2 > 0$, $(1-y) \in 0,1$, and $exp\left(-\left(1-P_F(x)\right)\right) > 0$, this term is also non-negative and hence convex.

Since both second derivatives are non-negative, their sum $\frac{d^2L_{AdaptExp}}{d\theta^2} \geq 0$ is also non-negative. Therefore, the Adapt Exp Loss is a convex function.

4. Experiments

This section is dedicated to presenting the experimental setups, which cover various aspects such as the datasets utilized, implementation details including parameter configurations, and the evaluation metrics employed.

Table 2. Overview of datasets and data partitions.

Dataset	Total Images	Training Images	Validation Images	Testing Images	Image Details
ISIC 2018	2594	1815	259	520	RGB skin lesion images
ISIC 2017	2000	1400	150	600	RGB skin lesion images
PH2	200	140	20	40	8-bit RGB, 768×560 resolution

4.1. Datasets

We conducted extensive experiments using three publicly available skin lesion segmentation datasets: ISIC 2017 [25], ISIC 2018 [26], and PH2 [27]. The ISIC 2017 and ISIC 2018 datasets were sourced from the International Skin Imaging Collaboration (ISIC), while the PH2 dataset was provided by the Dermatology Service of Hospital Pedro Hispano in Matosinhos, Portugal. Table 2 provides a summary of the datasets and their respective partitions. This systematic data partitioning ensures a thorough evaluation of our

method's robustness and adaptability across diverse datasets. Figure 7 highlights examples of the lesion images from ISIC 2018, ISIC 2017, and PH2, providing a visual representation of the datasets used in this study.

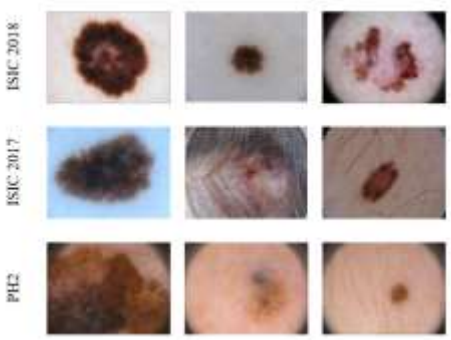

Figure 7. skin lesion samples.

4.2. Implementation details

We carried out our experiments using the Tensorflow framework, running the models on an RTX 3070Ti GPU. For training, we set a batch size of 8 and resized each image to 256 × 256 pixels. To boost model performance, we applied SiLU [62] activation after batch normalization. The Adam optimizer was used for network optimization, and we adopted the "poly" learning rate schedule for the ISIC 2018 and PH2 datasets, as outlined in [60]. Specifically, the learning rate at each step was computed as the initial rate multiplied by $(1-\frac{iter}{max-iter})^{power}$, with the initial rate and power set to 0.01 and 0.9, respectively. To mitigate overfitting, we applied a weight decay of 0.0001 during training. We also initialized the backbone network layers with pre-trained ImageNet weights [63], a strategy that improves accuracy in computer vision tasks. Table 3 presents the hyperparameters tuned during model training. Hyperparameter optimization is crucial, as selecting the right values can significantly impact the model's performance and generalization.

Table 3. Hyperparameters used for training the model.

Hyperparameters	Value
Activation	Softmax
Backbone	ResNet101
Batch size	8
Epochs	50
Learning rate	0.001 and 0.01
Optimizer	Adam

4.3. Evaluation metrics

We employed a diverse set of evaluation metrics to ensure a precise and reliable assessment of our segmentation methodology. Comparative analysis across these metrics provided a deeper understanding of the strengths and limitations of our approach. Below, we describe the metrics used.

- Intersection over Union (IoU):

This metric measures the overlap between the predicted segmentation map (A) and the ground truth (B). It is calculated as the intersection area of A and B divided by their union area, providing a measure of alignment between predictions and the ground truth. An IoU of 1 indicates a perfect match, while 0 means no overlap.

The formula for IoU is:

$$IoU = \frac{A \cap B}{A \cup B} \tag{8}$$

$$mIoU = \frac{1}{n}\sum_{i=1}^{N} IoU_i \tag{9}$$

where N is the number of images, IoU_i represents the IoU score for image i.

- Dice Score:

The Dice score, also known as the Dice coefficient, is another common metric for evaluating segmentation performance. It quantifies the similarity between the predicted segmentation and the ground truth. The Dice score is calculated as twice the intersection of the segmented regions, divided by the sum of their sizes. A Dice score of 1 indicates a perfect match between the segmentations.

The formula for the Dice score is:

$$Dice = \frac{2 * A \cap B}{A \cup B} \tag{10}$$

$$mDice = \frac{1}{n}\sum_{i=1}^{N} Dice_i \tag{11}$$

where N is the number of images, $Dice_i$ represents the $Dice$ score for image i.

Due to our reliance on the results reported in article [64], and in order to effectively utilize and compare with those results, we are required to use the IoU^* Star and $Dice^*$ Star metrics.

$$IoU^* = \frac{TP}{TP + FP + FN} \tag{12}$$

$$mIoU^* = \frac{TP_{sum}}{TP_{sum} + FP_{sum} + FN_{sum}} \tag{13}$$

$$Dice^* = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{14}$$

$$mDice^* = \frac{2 \times TP_{sum}}{2 \times TP_{sum} + FP_{sum} + FN_{sum}} \tag{15}$$

where TP_{sum} represents the total number of true positive pixels for images, FP_{sum} represents the total number of false positive pixels for images and FN_{sum} represents the total number of false negative pixels for images.

5. Results

This section is divided into two subsections for clarity. The first subsection presents the numerical and statistical results from our experiments, while the second subsection provides a detailed comparison and discussion of the qualitative findings.

5.1. Quantitative results

We evaluate the performance of our proposed method by comparing it with state-of-the-art techniques across all three datasets to demonstrate its competitiveness. Additionally, we benchmark the effectiveness of our custom loss function against other alternatives and analyze the contribution of each individual component in our approach to the overall performance.

5.1.1. Comparison with state-of-the-art methods

In this section, we compare our approach with existing state-of-the-art (SOTA) methods across three datasets, referencing the results from [64]. To ensure a fair comparison, no post-processing optimization strategies were applied during the evaluation.

A. Results on ISIC 2018

Thirteen segmentation methods, including TransUnet [65], SANet [32], SwinNet [24], TransFuse-S [66], UNet [4], AttUNet [8], MedT [19], ConvUNext [18], UNext-S [13], MALUNet [12], EGE-UNet [44], UCM-Net [64], ExpUNET [3], and our proposed method, are compared and evaluated in terms of mean Intersection over Union (mIoU) and mean Dice coefficient (mDice) on the ISIC 2018 dataset, as shown in Table 4. Our approach outperformed all other models on both mIoU and mDice skill metrics. The second-best performer was ExpUNET [3], followed by TransUnet [65] and SwinNet [24]. In contrast, UNet [4] and AttUNet [8] were the least effective of all the methods compared. The results for all compared methods in Table 4 are based on the reported values from UCM-Net [64]. To ensure a fair comparison, the results for our proposed method and our previous work, ExpUNET [3], were also obtained under the same experimental conditions as the base paper.

The superior performance observed on the ISIC 2018 dataset can be attributed to the ability of the proposed method to effectively capture both global contextual information and fine-grained boundary details. The integration of the enhanced encoder–decoder architecture with the adaptive loss function enables the model to handle significant variations in lesion size, shape, and contrast, which are common challenges in dermoscopic images. These results indicate that the proposed approach is particularly well suited for large-scale datasets with high intra-class variability, such as ISIC 2018.

B. Results on ISIC 2017

In addition to the results on ISIC 2018, we also summarize the results of different segmentation networks on the ISIC 2017 dataset in Table 5, including TransUnet, SANet, and SwinNet, as well as our proposed method. Our method outperformed all other techniques on each of the metrics we were able to evaluate. For a fair comparison, the results for all methods shown in Table 5 are based on the reported values from the base paper, UCM-Net [64]. To ensure a consistent review and comparison, our proposed method and our previous work, ExpUNET [3], were also trained and tested under the same experimental conditions as those reported in UCM-Net [64]. The next best result was provided by ExpUNET, with EGE-UNet and UCM-Net showing competitive results but not approaching our method. This assessment highlights the robustness of our approach to segment skin lesions on the ISIC 2017 dataset under different performance measures.

The consistent improvements observed on the ISIC 2017 dataset further emphasize the strong generalization capability of the proposed method when applied to data with varying image characteristics, acquisition conditions, and levels of segmentation difficulty. Despite differences in color distribution, lesion morphology, background artifacts, and annotation variability, the model maintains stable and accurate performance, indicating that it learns transferable and discriminative feature representations rather than overfitting to dataset-specific patterns. This robustness can be attributed to the proposed architecture's ability to effectively integrate multi-scale contextual information while preserving fine structural details essential for precise lesion boundary delineation. Moreover, the adaptive loss formulation dynamically balances region-based accuracy and boundary awareness, enabling the model to remain sensitive to subtle lesion variations while being resilient to noise and class imbalance.

Table 4. Comparative prediction results on the ISIC 2018.

Methods	Year	mIoU	mDice	mIoU*	mDice*
TransUnet	2021	82.00 ± 0.152	89.11 ± 0.031	81.68 ± 0.511	89.92 ± 0.010
SANet	2021	80.37 ± 0.124	87.87 ± 0.114	79.39 ± 0.135	88.51 ± 0.084
SwinNet	2021	81.41 ± 0.069	88.58 ± 0.019	80.72 ± 0.069	89.33 ± 0.042
TransFuse-S	2021	81.69 ± 0.128	88.82 ± 0.134	81.29 ± 0.084	89.68 ± 0.070
UNet	2015	79.86 ± 0.075	87.57 ± 0.085	78.27 ± 0.300	87.81 ± 0.188
AttUNet	2018	80.05 ± 0.079	87.62 ± 0.078	78.38 ± 0.151	87.88 ± 0.095
MedT	2021	80.34 ± 0.034	87.77 ± 0.107	79.29 ± 0.411	88.45 ± 0.251
ConvUNeXt	2022	80.51 ± 0.043	87.99 ± 0.049	78.71 ± 0.128	88.09 ± 0.080
UNeXt-S	2022	80.70 ± 0.226	88.17 ± 0.194	79.26 ± 0.497	88.43 ± 0.309
MALUNet	2022	80.95 ± 0.393	88.25 ± 0.315	79.99 ± 0.644	88.88 ± 0.398
EGE-UNet	2023	81.16 ± 0.104	88.36 ± 0.086	80.28 ± 0.363	89.06 ± 0.223
UCM-Net	2023	81.26 ± 0.030	88.48 ± 0.109	80.85 ± 0.251	89.35 ± 0.111
ExpUNET	2024	85.12 ± 0.390	90.98 ± 0.176	84.32 ± 0.207	91.14 ± 0.109
Our Method	2024	**86.47 ± 0.183**	**91.63 ± 0.086**	**85.13 ± 0.157**	**92.37 ± 0.420**

Table 5. Comparative prediction results on the ISIC 2017 dataset.

Methods	Year	mIoU	mDice	mIoU*	mDice*
TransUnet	2021	81.12 ± 0.152	88.38 ± 0.157	80.02 ± 0.011	88.90 ± 0.113
SANet	2021	79.34 ± 0.023	87.04 ± 0.155	79.34 ± 0.102	88.15 ± 0.060
SwinNet	2021	80.51 ± 0.102	87.81 ± 0.231	80.48 ± 0.134	89.19 ± 0.073
TransFuse-S	2021	80.73 ± 0.312	88.03 ± 0.111	79.87 ± 0.109	88.81 ± 0.058
UNet	2015	78.12 ± 0.175	85.97 ± 0.196	76.42 ± 0.381	86.63 ± 0.245
AttUNet	2018	78.45 ± 0.113	86.22 ± 0.124	77.14 ± 0.097	87.10 ± 0.062
MedT	2021	79.05 ± 0.231	86.59 ± 0.125	77.61 ± 0.121	87.40 ± 0.401
ConvUNeXt	2022	78.78 ± 0.362	86.21 ± 0.267	76.98 ± 0.490	86.99 ± 0.313
UNeXt-S	2022	78.79 ± 0.234	86.30 ± 0.140	77.48 ± 0.466	87.31 ± 0.296
MALUNet	2022	79.97 ± 0.389	87.23 ± 0.345	79.11 ± 0.345	88.34 ± 0.215
EGE-UNet	2023	80.22 ± 0.237	87.24 ± 0.149	79.71 ± 0.411	88.71 ± 0.255
UCM-Net	2023	80.71 ± 0.345	87.66 ± 0.221	79.29 ± 0.188	88.45 ± 0.117
ExpUNET	2024	83.60 ± 0.193	89.15 ± 0.320	82.43 ± 0.207	89.77 ± 0.048
Our Method	2024	**85.67 ± 0.237**	**90.19 ± 0.173**	**83.86 ± 0.046**	**90.98 ± 0.392**

C. Results on PH2

Table 6 presents a performance comparison of various segmentation methods on the PH2 dataset. This evaluation includes TransUnet, SANet, SwinNet, TransFuse-S, UNet, AttUNet, MedT, ConvUNeXt, UNeXt-S, MALUNet, EGE-UNet, UCM-Net, ExpUNET, and our proposed method. Our approach achieved the highest scores across all evaluation metrics, demonstrating its superior performance. To maintain the integrity of our comparison, the results for the competing models were adopted directly from the UCM-Net [64]were

publication. We then trained and evaluated both our novel method, in conjunction with our previous work, ExpUNET [3], which employed the exact same experimental protocol as UCM-Net [64], ensured an equitable and accurate assessment. While methods such as TransFuse-S and ExpUNET delivered highly competitive results, and others like TransUnet and SANet also showed strong performance, none were able to match the effectiveness of our approach. This detailed comparison underscores the strong efficacy of our method for skin lesion segmentation on the PH2 dataset, as evidenced by its leading performance on multiple parameters.

Table 6. Comparative prediction results on the PH2 dataset.

Methods	Year	mIoU	mDice	mIoU*	mDice*
TransUnet	2021	90.77 ± 0.030	95.09 ± 0.014	91.89 ± 0.148	95.77 ± 0.080
SANet	2021	89.99 ± 0.349	94.62 ± 0.206	91.06 ± 0.481	95.32 ± 0.264
SwinNet	2021	88.92 ± 0.493	94.03 ± 0.289	89.96 ± 0.388	94.71 ± 0.215
TransFuse-S	2021	91.15 ± 0.131	95.30 ± 0.078	92.16 ± 0.113	95.92 ± 0.061
UNet	2015	89.09 ± 0.224	94.13 ± 0.112	90.08 ± 0.545	94.78 ± 0.302
AttUNet	2018	89.01 ± 0.345	94.06 ± 0.222	89.82 ± 0.371	94.64 ± 0.206
MedT	2021	87.77 ± 0.099	93.35 ± 0.063	88.53 ± 0.338	93.92 ± 0.190
ConvUNeXt	2022	89.46 ± 0.056	94.33 ± 0.035	90.39 ± 0.172	94.95 ± 0.095
UNeXt-S	2022	89.03 ± 0.246	94.09 ± 0.163	89.68± 0.269	94.56 ± 0.149
MALUNet	2022	87.75 ± 0.378	93.35 ± 0.214	88.19 ± 0.329	93.73 ± 0.186
EGE-UNet	2023	88.39 ± 0.303	93.73 ± 0.189	88.93 ± 0.802	94.14 ± 0.450
UCM-Net	2023	89.62 ± 0.221	94.42 ± 0.124	90.51 ± 0.211	94.96 ± 0.116
ExpUNET	2024	91.01 ± 0.234	94.52 ± 0.176	91.86 ± 0.324	94.96 ± 0.109
Our Method	2024	**93.13 ± 0.025**	**96.02 ± 0.293**	**94.46 ± 0.341**	**97.26 ± 0.283**

5.1.2. Comparative Analysis of Architectures, Backbones, and Loss Functions

To comprehensively evaluate our proposed method, we conducted a rigorous comparative analysis against several state-of-the-art architectures, including UNet, Attention UNet, TransUnet, and ResUNet. All models were trained and tested under identical conditions to ensure an impartial assessment of their performance. The results, as presented in Table 7, clearly demonstrate the superiority of our proposed approach. A key objective of our study was to demonstrate the versatility of our method by testing it with different ResNet backbones. We chose to use ResNet-34, ResNet-50, and ResNet-101 to evaluate how our proposed method, which includes a novel loss function and an RFE module, performs with varying backbone complexities. The results conclusively show that the ResNet-101 backbone consistently yielded the highest accuracy, Intersection over Union (IoU), and Dice scores across the board. This superior performance led to our decision to use ResNet-101 as the primary backbone for our final model configuration. Furthermore, a critical finding from this comparative analysis is the confirmed superiority of our proposed loss function. By applying it across different models and comparing its performance against standard loss functions, such as BCE Loss, Dice Loss, and Focal Loss, we were able to provide clear evidence of its effectiveness. The data in Table 7 shows that our proposed loss function consistently improved the performance metrics of every model it was applied to, highlighting its robustness and generalizability.

These results indicate that the performance improvements are not solely attributable to increased model depth, but rather to the synergistic integration of the ResNet-based encoder, the RFE module, and the proposed Adapt Exp Loss In particular, the ability of the proposed loss function to enhance performance irrespective of the underlying network architecture underscores its architecture-agnostic nature, making it a practical and transferable solution for a wide range of medical image segmentation tasks.

Table 7. Quantitative comparison of different models, backbones, and loss functions on the ISIC 2018, ISIC 2017 and PH2 datasets.

Dataset	Model	BCE		DiceLoss		FocalLoss		AdaEXP		GFLOPs
		mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	
ISIC2018	**Trans unet**	80.2	88.65	79.5	87.12	78.4	85.98	81.5	89.65	24.7278
	Unet	79.3	88.4	79.5	88.6	77.2	85.5	79.9	88.71	54.7378
	Attunet	76	85.9	77.95	83.12	70.8	77.8	77.54	86.2	66.6318
	Swinnet	76.5	86.6	75.4	81.5	73.3	80.4	79.51	87.3	30.475
	Ours(resnet34)	80.1	88.34	79.8	88.7	78.32	85.01	81.7	89.2	9.046
	Ours(resnet50)	81.21	88.87	82.14	88.99	79.95	85.65	85.41	90.3	39.701
	Ours(resnet101)	**84.1**	**89.85**	**83.47**	**87.26**	**80.95**	**86.02**	**86.47**	**91.63**	44.576
ISIC2017	**Trans unet**	80.8	88.96	79.8	86.4	77.6	85.64	81.9	89.74	24.7278
	Unet	73.2	80.6	70.6	77.9	65.1	72.6	76.4	85.69	54.7378
	Attunet	70	77.3	69	76.2	65	71.8	73.45	80.8	66.6318
	Swinnet	71.2	78.4	70.1	77.7	69.9	76.3	74.46	81.2	30.475
	Ours(resnet34)	79.1	86.24	78.7	85.8	79.32	85.51	81.4	89.2	9.046
	Ours(resnet50)	81.32	87.65	80.13	85.4	81.95	86.25	84.18	89.9	39.701
	Ours(resnet101)	**83.7**	**89.2**	**82.42**	**88.5**	**80.12**	**85.8**	**85.67**	**90.19**	44.576
PH2	**Trans unet**	90.19	94.58	89.53	92.17	88.32	91.93	91.7	95.65	24.7278
	Unet	89.34	91.37	88.49	91.54	85.21	90	90.8	93.72	54.7378
	Attunet	88.7	91.2	87.91	91.09	85.3	89.6	90.39	93.23	66.6318
	Swinnet	88.1	90.6	87.8	90.6	86.9	91.1	89.51	92.21	30.475
	Ours(resnet34)	87.5	92.7	87.3	91.3	86.32	91.31	88.7	93.4	9.046
	Ours(resnet50)	90.3	93.4	88.14	92.3	87.95	93.5	92.2	94.6	39.701
	Ours(resnet101)	**92.5**	**95.2**	**91.47**	**94.26**	**89.95**	**93.82**	**93.13**	**96.02**	44.576

5.1.3.Comparison with loss functions

Choosing an appropriate loss function is a key consideration for developing effective image segmentation models, as they directly affect performance by considering elements such as the data distribution, data imbalance, and boundary information characteristics that are important to the model design [10]. We compare the proposed Adapt Exp Loss with the commonly applied loss functions, including Binary Cross Entropy (BCE), Dice Loss, Focal Loss and Exponential Loss, across three datasets, i.e., ISIC 2017, ISIC 2018 and PH2. Table 8 illustrates that the proposed Adapt Exp Loss consistently achieved higher performance than the other loss functions in terms of both the Intersection over Union (IoU) and the Dice scores over all datasets considered.

This performance gain highlights the effectiveness of incorporating adaptive mechanisms into the loss formulation, particularly for datasets exhibiting significant variability in lesion size and class distribution. Unlike BCE and Dice Loss, which treat all samples uniformly, and Focal and Exponential Losses, which rely on fixed weighting strategies, the Adapt Exp Loss dynamically responds to image-level class proportions. This adaptability enables more balanced gradient contributions during training, resulting in improved boundary delineation and reduced false negatives. Consequently, the proposed loss demonstrates superior generalization capability across diverse datasets with differing imbalance characteristics. Choosing an appropriate loss function is a key consideration for developing effective image segmentation models, as it directly affects performance by accounting for factors such as data distribution, class imbalance, and boundary information, which are critical to model design [10].

We compare the proposed Adapt Exp Loss with commonly applied loss functions, including Binary Cross Entropy (BCE), Dice Loss, Focal Loss, and

Exponential Loss, across three datasets: ISIC 2017, ISIC 2018, and PH2. As shown in Table 8, the proposed Adapt Exp Loss consistently achieves superior performance compared to the other loss functions in terms of both Intersection over Union (IoU) and Dice scores across all evaluated datasets.

Table 8. Quantitative analysis of proposed method performance using different loss functions on the ISIC 2018, ISIC 2017 and PH2 datasets

Dataset	Loss	IoU	Dice
ISIC2018	**BCE**	84.10	89.85
	DICE	83.47	87.26
	FOCAL	80.95	86.02
	EXP	85.12	90.98
	Ours	**86.47**	**91.63**
ISIC2017	**BCE**	83.21	88.10
	DICE	82.45	87.74
	FOCAL	79.16	85.19
	EXP	84.96	89.93
	Ours	**85.67**	**90.19**
PH2	**BCE**	89.32	91.24
	DICE	88.43	90.34
	FOCAL	84.93	88.62
	EXP	91.74	92.52
	Ours	**93.13**	**96.02**

5.1.4. Comparison of convergence rate of losses

On comparing the convergence rates of the losses using ISIC 2018 (Figure 8), Adapt EXP exhibits better stability and convergence over the other four, yielding optimal loss at epoch 50. The EXP loss also converges reasonably with some small movements. While BCE and Dice losses converge rapidly at the start, they oscillate considerably up until epoch 30, showing signs of instability. Although focal loss decreases sharply at first, it cannot perform well in subsequent epochs and shows a significant level of fluctuation, making it unreliable.

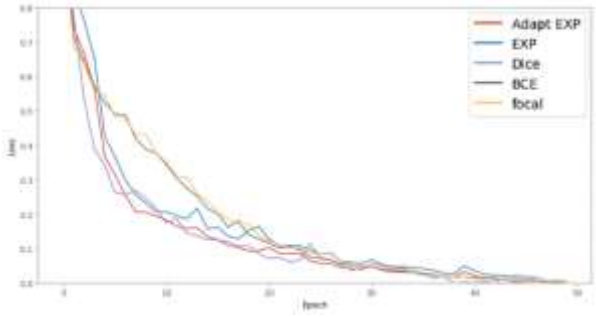

Figure 8. The convergence rate of the loss function through ISIC 2018 dataset, varying with the epoch number

For the ISIC 2017 dataset (Figure 9), Adapt EXP remains the most stable and effective loss function, showing a smooth and consistent decline during training. BCE converges quickly but exhibits fluctuations similar to those in ISIC 2018. EXP follows a steady trajectory but lacks the robustness of Adapt EXP. Dice and Focal losses display early instability with pronounced oscillations, reducing their effectiveness.

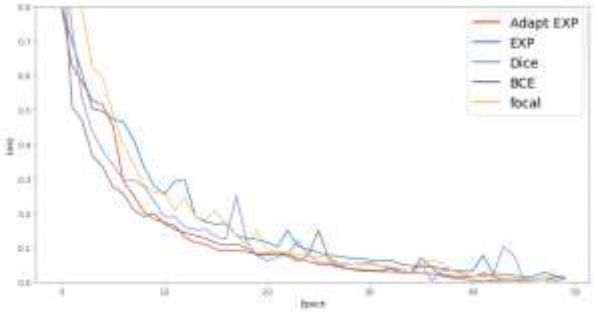

Figure 9. The convergence rate of the loss function through ISIC 2017 dataset, varying with the epoch number.

Figure 10 shows the performance of the following loss functions on the PH2 dataset: Adapt EXP, EXP, BCE, Dice, and Focal losses. Adapt EXP performs best among all, converging smoothly and steadily to minimize loss as early as epoch 50. The EXP loss is consistently dropping but has minor fluctuations. The BCE and Dice losses both quickly converge at first but then start to oscillate significantly, especially between epochs 10 and 30, suggesting long-term instability. Although the focal loss decreases greatly at the beginning, it presents large variations in the later stage and thus reduces its stability.

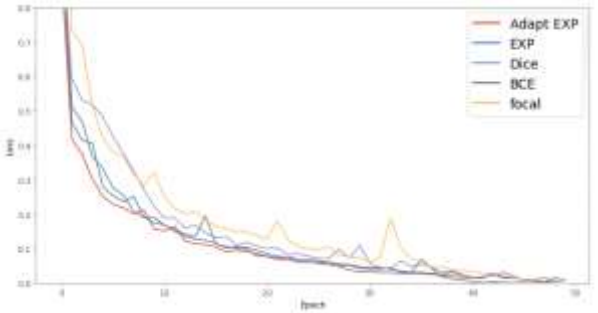

Figure 10. The convergence rate of the loss function through PH2 dataset, varying with the epoch number.

In conclusion, compared with other loss functions, the Adapt EXP loss function yields smooth and stable convergence, achieving the best overall performance on the ISIC 2017, ISIC 2018, and PH2 datasets, respectively. Combining rapid initial loss reduction with long-term stability, this loss maximally minimizes loss without exhibiting the oscillations of the other losses, with reliable convergence by epoch 50. Unlike EXP, BCE, Dice, and Focal losses, which are unstable, especially in the later epochs, Adapt EXP is shown to perform optimally for diverse datasets.

5.1.5. Ablation Study

This section analyzes three key aspects of our method: a systematic evaluation of thecontributions of our proposed components, an assessment of the model's robustness on noisy data, and a 5-fold cross-validation study on the PH2 dataset to confirm its resilience against overfitting.

A. Comparison of Different Components of Our Method:

This section presents an ablation study to evaluate the contributions of our proposed RFE module and loss functions to segmentation performance across the ISIC 2017, ISIC 2018, and PH2 datasets. We compare the baseline model with Binary Cross Entropy (BCE) loss against models enhanced with the RFE module, RFE combined with Exp Loss, and RFE paired with the newly developed Adapt Exp Loss. The results, summarized in Table 7, highlight the incremental gains achieved by each component, measured using Intersection over Union (IoU) and Dice coefficient. Across all three datasets, the integration of the RFE module consistently improved segmentation performance compared to the baseline model. Adding the Exp Loss further enhanced the results, but the most significant improvement was achieved with the Adapt Exp Loss. For the ISIC 2018 dataset, the baseline model achieved an IoU of 84.10 and Dice of 89.85, which increased to 86.47 and 91.63, respectively, when the RFE module and Adapt Exp Loss were combined. Similarly, on the ISIC 2017 dataset, the IoU improved from 83.21 (baseline) to 85.67, and Dice increased from 88.10 to 90.19. The PH2 dataset demonstrated the most substantial gains, with the baseline IoU of 89.32 and Dice of 91.24 rising to 93.13 and 96.02, respectively, using the full configuration. These results underscore the effectiveness of the proposed RFE module and the Adapt Exp Loss in significantly enhancing segmentation accuracy, as shown in Table 9.

B. Model Robustness on Noisy Data:

This section examines our model's robustness by evaluating its performance on noisy data. Given that medical images are highly susceptible to various forms of noise, a robust segmentation method is crucial for real-world clinical applications. Our experiments were conducted on the ISIC 2018, ISIC 2017, and PH2 datasets, which were augmented with Gaussian noise to simulate real-world data imperfections. It is worth noting that our models were trained without any specific data augmentations to handle noise, which underscores their inherent robustness. This highlights the model's ability to maintain reliable segmentation even under noisy conditions.

Table 9. Segmentation performance of the proposed method on the ISIC 2018, ISIC 2017 and PH2 datasets with different configurations.

Dataset	Loss	IoU	Dice
ISIC2018	**Base Line (BCE)**	84.10	89.85
	Base Line (BCE) + RFE	85.08	90.30
	Base Line + RFE + Exp Loss	85.12	90.98
	Base Line + RFE + Adapt Exp Loss	**86.47**	**91.63**
ISIC2017	**Base Line (BCE)**	83.21	88.10
	Base Line (BCE) + RFE	84.53	89.45
	Base Line + RFE + Exp Loss	84.96	89.93
	Base Line + RFE + Adapt Exp Loss	**85.67**	**90.19**
PH2	**Base Line (BCE)**	89.32	91.24
	Base Line (BCE) + RFE	90.18	91.89
	Base Line + RFE + Exp Loss	91.74	92.52
	Base Line + RFE + Adapt Exp Loss	**93.13**	**96.02**

Table 10 summarizes our findings. Despite using ResNet, a relatively older network compared to many modern architectures, our model consistently achieved high accuracy. This demonstrates that our proposed architecture and its unique components (particularly the Receptive Field Enhancement (RFE) module and Adapt Exp Loss) are highly effective at handling noisy inputs. The results also show that as the complexity of the ResNet backbone increases (from ResNet34 to ResNet101), our model's performance improves, indicating its ability to leverage more powerful feature extractors for better segmentation accuracy even under challenging conditions. This confirms the strong efficacy and reliability of our approach on noisy data.

Table 10. The performance of different methods on noisy ISIC 2018, ISIC 2017 and PH2 test set.

Dataset	Model	mIoU	mDice
ISIC 2018	**Ours(resnet34)**	78.4	87.1
	Ours(resnet50)	79.1	87.7
	Ours(resnet101)	83.12	89.95
ISIC 2017	**Ours(resnet34)**	79.4	87.4
	Ours(resnet50)	80.1	87.8
	Ours(resnet101)	81.67	87.19
PH2	**Ours(resnet34)**	86.4	90.1
	Ours(resnet50)	87.1	92.7
	Ours(resnet101)	89.25	93.02

C. Robustness Against Overfitting: A 5-Fold Cross-Validation Study on the PH2 Dataset:

To address the concern regarding potential overfitting on small-scale datasets, we conducted a specific evaluation of our model using a 5-fold

cross-validation strategy on the PH2 dataset, which consists of only 200 images.
As shown in Table 11, the consistency of the results across all folds, with a mean Dice score of 93.13% and a low standard deviation of 0.32, demonstrates the model's robustness and its ability to generalize effectively without overfitting. These findings provide strong evidence that our proposed method is reliable even when trained on limited data, thereby alleviating the reviewer's concern.

Table 11. 5-fold cross-validation results of our proposed method on the PH2 dataset.

Fold	mIoU (%)	mDice (%)
1	93.099	95.724
2	93.118	96.182
3	93.138	95.683
4	93.149	96.215
5	93.162	96.303
Mean	93.133	96.021
St. Dev.	0.025	0.293

5.2. Visualization results

To demonstrate the efficacy of our method for medical image segmentation, particularly for skin images, we provide a visual analysis of segmentation results from the ISIC 2018, ISIC 2017, and PH2 datasets. Figure 11 presents a comparative visualization of our approach against nine other methods, including ExpUNET, UCM-Net, U-Net, Att-UNet, MedT, ConvUNeXt, UNeXt-S, MALUNet, and EGE-UNet. The results for these models are sourced from our base paper, UCM-Net [64]. To ensure a rigorous and fair comparison, we trained and tested both our proposed model and our previous work, ExpUNET [3], under the exact same experimental conditions as that publication. The images used for this visualization were from the test set, thereby ensuring that the models had not previously encountered them. The comparison demonstrates our method's superior performance on these benchmark images.

Figure 12 presents a qualitative comparison of segmentation results across the ISIC 2018 (top row), ISIC 2017 (middle row), and PH2 (bottom row) datasets, illustrating the performance of the proposed model alongside TransUNet, U-Net, Att-UNet, and SwinUNet. In contrast to Figure 11, where results were collected from previously published outputs, all models in this figure were trained and evaluated under the same experimental configuration, including identical training and testing splits, to ensure a fair and controlled comparison. The visual results demonstrate that the proposed method consistently produces more accurate and coherent segmentation masks, particularly in preserving lesion boundaries and capturing fine structural details across datasets with varying levels of difficulty. Competing models frequently exhibit boundary leakage, incomplete lesion coverage, or misclassification of surrounding healthy tissue, especially in cases with low contrast or ambiguous borders. In comparison, our model maintains strong alignment with the ground truth across diverse image characteristics, indicating improved robustness and generalization. These findings further support the effectiveness of the proposed architecture and learning strategy in achieving reliable skin lesion segmentation under standardized evaluation conditions.

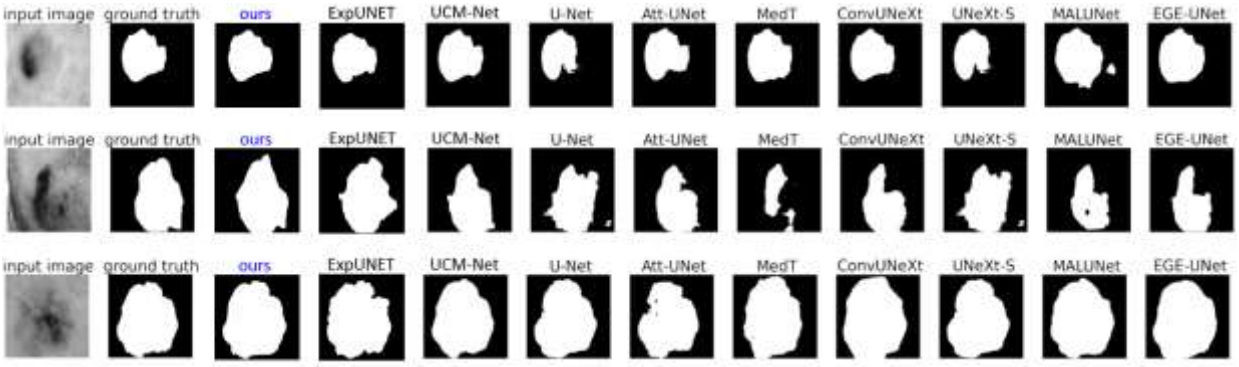

Figure 11. Comparison of segmentation performance between different segmentation networks. (a) Row 1: lesion segmentation on ISIC 2018 dataset. (b) Row 2: lesion segmentation on ISIC 2017. (c) Row 3: lesion segmentation on PH2.

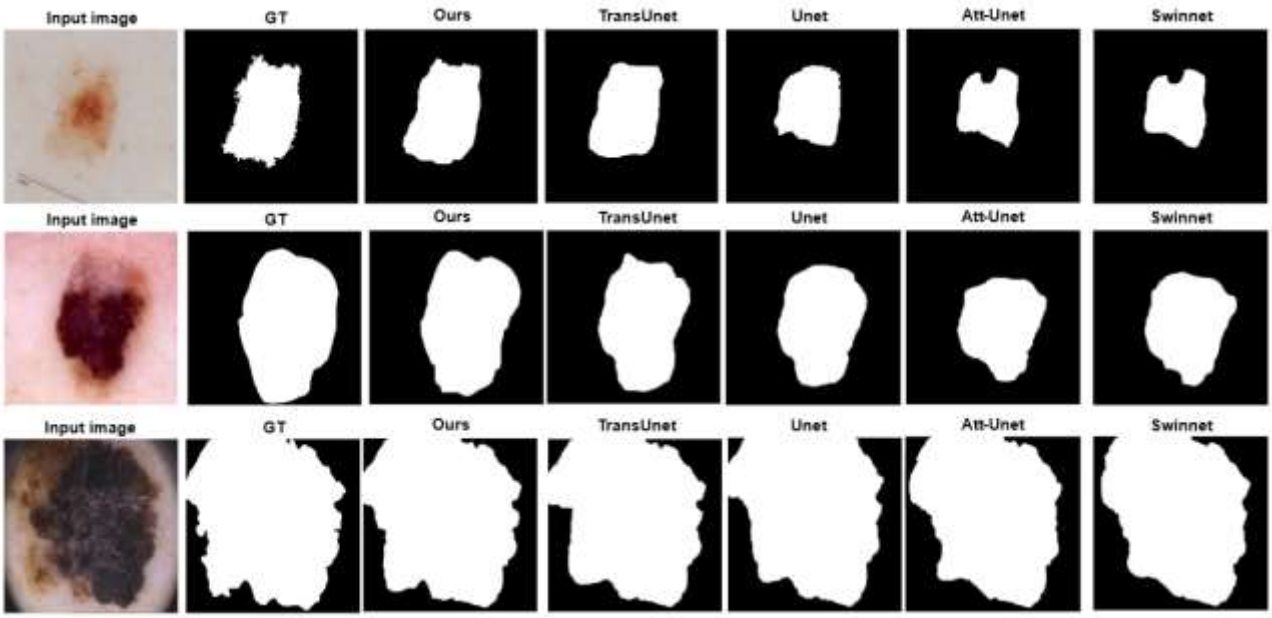

Figure 12. Visual comparison of segmentation results on the ISIC 2018 (top row), ISIC 2017 (middle row), and PH2 (bottom row) datasets. The images show the original image, ground truth, and the output of various models from left to right.

Figures 13 and 14 demonstrate the performance of our model on the ISIC 2018, ISIC 2017, and PH2 datasets, highlighting its effectiveness across varying levels of challenge. In Figure 13, results on less challenging images show that while other models struggled to distinguish abnormal from normal tissues, particularly at the borders due to background-foreground similarity, our method successfully predicted tissue locations and boundaries. This was achieved through an advanced loss function and class-specific weighting, which enhanced performance on datasets with lesions of varying sizes, outperforming other methods. The Hausdorff distance (HD) measure, which assesses the distance between two shapes, further confirms our model's superior performance in accurately segmenting lesion boundaries. The low HD values for our model (6.70 on the ISIC 2018 dataset, 4.32 on the ISIC 2017 dataset, and 5.12 on the PH2 dataset) indicate excellent boundary detection and alignment with the ground truth.

Figure 14 shows segmentation results on more challenging images with less distinct boundaries, where most models struggled to distinguish lesions from surrounding tissue. Unlike others, our method successfully captured significant lesion regions across all datasets. This demonstrates its superior performance, even with noise and ambiguous boundaries, validating its effectiveness for complex skin image segmentation tasks and its ability to capture intricate boundary details. The Hausdorff distance (HD) values for our model (21.43 on the ISIC 2018 dataset, 18.52 on the ISIC 2017 dataset, and 16.97 on the PH2 dataset) further confirm its excellent performance in accurately segmenting lesion boundaries.

Overall, the qualitative results presented in Figures 11, 13, and 14 clearly demonstrate that the proposed method achieves more accurate and visually coherent segmentation outputs compared to competing approaches, particularly in preserving lesion shape and boundary continuity. The consistent alignment between the predicted masks and ground-truth annotations across both simple and highly challenging cases indicates that the proposed architecture and adaptive loss formulation effectively complement each other.

Moreover, these visual observations are consistent with the quantitative improvements reported earlier, reinforcing the reliability of the proposed method across multiple evaluation criteria. By effectively integrating architectural enhancements with an adaptive loss strategy, the proposed approach demonstrates strong potential for practical clinical applications, where precise and reliable lesion delineation is critical for accurate diagnosis and treatment planning.

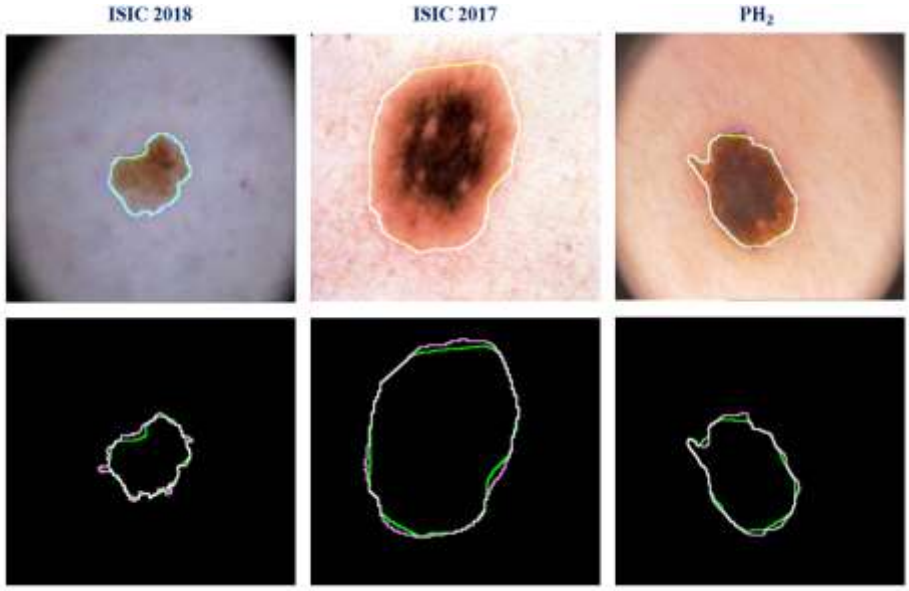

Figure 13. Segmentation performance comparison on less challenging images from ISIC 2018, ISIC 2017, and PH2 datasets, The green area represents the model's prediction, the purple area indicates the ground truth, and the white area shows the region of overlap between the prediction and ground truth.

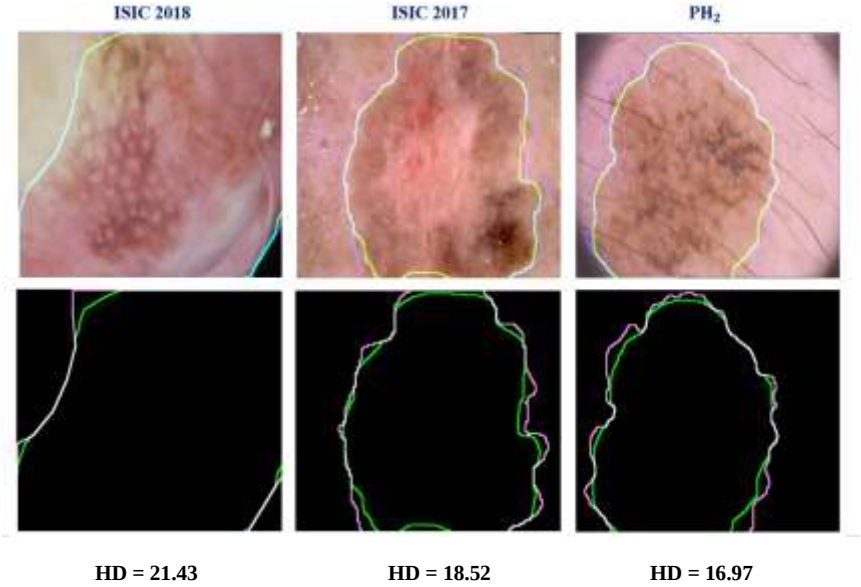

Figure 14. Segmentation performance comparison on more challenging images from ISIC 2018, ISIC 2017, and PH2 datasets. The green area represents the model's prediction, the purple area indicates the ground truth, and the white area shows the region of overlap between the prediction and ground truth.

6. Conclusion

This study presents a novel approach for skin disease segmentation and analysis by combining YOLO version 8 and ResNet architectures. The proposed U-shaped model integrates the ResNet backbone with a newly designed module inspired by the SPPF and C2F modules used in YOLO-V8, enhancing information extraction from medical images. To address the challenge of imbalanced datasets, a new weighted loss function wasdeveloped, assigning distinct weights based on the lesion-to-background pixel ratio in each image. Experiments on ISIC 2018, ISIC 2017, and PH2 datasets demonstrated the model's improved performance in skin condition analysis. Key contributions include the development of an advanced deep learning architecture, a module for enhancing the receptive field, and the introduction

of a weighted loss function for medical image segmentation. Comparative tests against state-of-the-art models showed significant improvements. Future work will focus on refining the model, expanding the loss function to handle multi-class and 3D datasets, and upgrading modules for better feature exploitation, ultimately aiming to support more accurate clinical decision-making.

References

[1] N. Badshah and A. Ahmad, "ResBCU-Net: Deep learning approach for segmentation of skin images," *Biomedical Signal Processing and Control*, vol. 71, p. 103137, 2022.

[2] M. Dildar et al., "Skin cancer detection: a review using deep learning techniques," *International journal of environmental research and public health*, vol. 18, no. 10, p. 5479, 2021.

[3] S. Roshan, J. Tanha, M. Zarrin, A. F. Babaei, H. Nikkhah, and Z. Jafari, "A deep ensemble medical image segmentation with novel sampling method and loss function," *Computers in Biology and Medicine*, vol. 172, p. 108305, 2024.

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

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[7] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J., 2018, September. Unet++: A nested u-net architecture for medical image segmentation. In *International workshop on deep learning in medical image analysis* (pp. 3-11). Cham: Springer International Publishing.

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

[9] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O., 2016, October. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424-432). Cham: Springer International Publishing.

[10] Jadon, S., 2020, October. A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)* (pp. 1-7). IEEE.

[11] Diakogiannis, F.I., Waldner, F., Caccetta, P. and Wu, C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, pp.94-114.

[12] Ruan, J., Xiang, S., Xie, M., Liu, T. and Fu, Y., 2022, December. Malunet: A multi-attention and light-weight unet for skin lesion segmentation. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1150-1156). IEEE.

[13] Valanarasu, J.M.J. and Patel, V.M., 2022, September. Unext: Mlp-based rapid medical image segmentation network. In *International conference on medical image computing and computer-assisted intervention* (pp. 23-33). Cham: Springer Nature Switzerland.

[14] J. Redmon, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[15] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[16] Ragab, M.G., Abdulkadir, S.J., Muneer, A., Alqushaibi, A., Sumiea, E.H., Qureshi, R., Al-Selwi, S.M. and Alhussian, H., 2024. A comprehensive systematic review of YOLO for medical object detection (2018 to 2023). *IEEE Access*, *12*, pp.57815-57836.

[17] Darlow, L.N. and Storkey, A., 2020. What information does a ResNet compress?. *arXiv preprint arXiv:2003.06254*.

[18] Han, Z., Jian, M. and Wang, G.G., 2022. ConvUNeXt: An efficient convolution neural network for medical image segmentation. *Knowledge-based systems*, *253*, p.109512.

[19] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I. and Patel, V.M., 2021, September. Medical transformer: Gated axial-attention for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 36-46). Cham: Springer International Publishing.

[20] Nikkhah, H., Tanha, J., Zarrin, M., Roshan, S. and Kazempour, A., 2025. YM-WML: A new Yolo-based segmentation Model with Weighted Multi-class Loss for medical imaging. *arXiv preprint arXiv:2506.22955*.

[21] Nasalwai, N., Punn, N.S., Sonbhadra, S.K. and Agarwal, S., 2021, May. Addressing the class imbalance problem in medical image segmentation via accelerated

tversky loss function. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 390-402). Cham: Springer International Publishing.

[23] Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H. and Yin, S., 2023. Deep learning attention mechanism in medical image analysis: Basics and beyonds. *International Journal of Network Dynamics and Intelligence*, pp.93-116.

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

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

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

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

[28] Venkatesh, G.M., Naresh, Y.G., Little, S. and O'Connor, N.E., 2018, September. A deep residual architecture for skin lesion segmentation. In *International Workshop on Computer-Assisted and Robotic Endoscopy* (pp. 277-284). Cham: Springer International Publishing.

[29] Al-Masni, M.A., Al-Antari, M.A., Choi, M.T., Han, S.M. and Kim, T.S., 2018. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Computer methods and programs in biomedicine*, *162*, pp.221-231.

[30] Goyal, M., Oakley, A., Bansal, P., Dancey, D. and Yap, M.H., 2019. Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *Ieee Access*, *8*, pp.4171-4181.

[31] K He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[32] Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K. and Cui, S., 2021, September. Shallow attention network for polyp segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 699-708). Cham: Springer International Publishing.

[33] Baghersalimi, S., Bozorgtabar, B., Schmid-Saugeon, P., Ekenel, H.K. and Thiran, J.P., 2019. DermoNet: densely linked convolutional neural network for efficient skin lesion segmentation. *EURASIP Journal on Image and Video Processing*, *2019*(1), p.71.

[34] Abraham, N. and Khan, N.M., 2019, April. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)* (pp. 683-687). IEEE.

[35] Cui, Z., Wu, L., Wang, R. and Zheng, W.S., 2019, October. Ensemble transductive learning for skin lesion segmentation. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)* (pp. 572-581). Cham: Springer International Publishing.

[36] Singh, V.K., Abdel-Nasser, M., Rashwan, H.A., Akram, F., Pandey, N., Lalande, A., Presles, B., Romani, S. and Puig, D., 2019. FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *Ieee Access*, *7*, pp.130552-130565.

[37] Saini, S., Gupta, D. and Tiwari, A.K., 2019, December. Detector-segmentor network for skin lesion localization and segmentation. In *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics* (pp. 589-599). Singapore: Springer Singapore.

[38] Nathan, S. and Kansal, P., 2020. Lesion Net--Skin Lesion Segmentation Using Coordinate Convolution and Deep Residual Units. *arXiv preprint arXiv:2012.14249*.

[39] Jin, Q., Cui, H., Sun, C., Meng, Z. and Su, R., 2021. Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. *Applied soft computing*, *99*, p.106881.

[40] Saini, S., Jeon, Y.S. and Feng, M., 2021, April. B-SegNet: branched-SegMentor network for skin lesion segmentation. In *Proceedings of the Conference on Health, Inference, and Learning* (pp. 214-221).

[41] Chen, P., Huang, S. and Yue, Q., 2022. Skin lesion segmentation using recurrent attentional convolutional networks. *IEEE Access*, *10*, pp.94007-94018.

[42] Liu, X., Fan, W. and Zhou, D., 2022, November. Skin lesion segmentation via intensive atrous spatial transformer. In *International Conference on Wireless Algorithms, Systems, and Applications* (pp. 15-26). Cham: Springer Nature Switzerland.

[43] Kaur, R., GholamHosseini, H. and Sinha, R., 2022. Skin lesion segmentation using an improved framework of encoder-decoder based convolutional neural network. *International Journal of Imaging Systems and Technology*, *32*(4), pp.1143-1158.

[44] Ruan, J., Xie, M., Gao, J., Liu, T. and Fu, Y., 2023, October. Ege-unet: an efficient group enhanced unet for skin lesion segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 481-490). Cham: Springer Nature Switzerland.

[45] Sun, Y., Dai, D., Zhang, Q., Wang, Y., Xu, S. and Lian, C., 2023. MSCA-Net: Multi-scale contextual attention network for skin lesion segmentation. *Pattern Recognition*, *139*, p.109524.

[46] Qiu, S., Li, C., Feng, Y., Zuo, S., Liang, H. and Xu, A., 2023. GFANet: Gated fusion attention network for skin lesion segmentation. *Computers in Biology and Medicine*, *155*, p.106462.

[47] Qiu, S., Li, C., Feng, Y., Zuo, S., Liang, H. and Xu, A., 2023. GFANet: Gated fusion attention network for skin lesion segmentation. *Computers in Biology and Medicine*, *155*, p.106462.

[48] Scholz, D., Erdur, A.C., Buchner, J.A., Peeken, J.C., Rueckert, D. and Wiestler, B., 2024. Imbalance-aware loss functions improve medical image classification. *Medical imaging with deep learning*, pp.1341-1356.

[49] Le, N., Le, T., Yamazaki, K., Bui, T., Luu, K. and Savides, M., 2021, January. Offset curves loss for imbalanced problem in medical segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 9189-9195). IEEE.

[50] Cabezas, M. and Diez, Y., 2024. An analysis of loss functions for heavily imbalanced lesion segmentation. *Sensors*, *24*(6), p.1981.

[51] Roy, S., Tyagi, M., Bansal, V. and Jain, V., 2022. Svd-clahe boosting and balanced loss function for covid-19 detection from an imbalanced chest x-ray dataset. *Computers in Biology and Medicine*, *150*, p.106092.

[52] Gökkan, O. and Kuntalp, M., 2022. A new imbalance-aware loss function to be used in a deep neural network for colorectal polyp segmentation. *Computers in Biology and Medicine*, *151*, p.106205.

[53] Kofler, F., Shit, S., Ezhov, I., Fidon, L., Horvath, I., Al-Maskari, R., Li, H.B., Bhatia, H., Loehr, T., Piraud, M. and Erturk, A., 2023, June. Blob loss: Instance imbalance aware loss functions for semantic segmentation. In *International Conference on Information Processing in Medical Imaging* (pp. 755-767). Cham: Springer Nature Switzerland.

[54] Sugino, T., Kawase, T., Onogi, S., Kin, T., Saito, N. and Nakajima, Y., 2021, July. Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks. In *Healthcare* (Vol. 9, No. 8, p. 938). MDPI.

[55] Borawar, L. and Kaur, R., 2023, March. ResNet: Solving vanishing gradient in deep networks. In *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022* (pp. 235-247). Singapore: Springer Nature Singapore.

[56] Liu, G., Hu, Y., Chen, Z., Guo, J. and Ni, P., 2023. Lightweight object detection algorithm for robots with improved YOLOv5. *Engineering Applications of Artificial Intelligence*, *123*, p.106217.

[57] Zhang, Z., 2023. Drone-YOLO: An efficient neural network method for target detection in drone images. *Drones*, *7*(8), p.526.

[58] Sriram, S., Vinayakumar, R., Sowmya, V., Alazab, M. and Soman, K.P., 2020, July. Multi-scale learning based malware variant detection using spatial pyramid pooling network. In *IEEE INFOCOM 2020-IEEE conference on computer communications workshops (INFOCOM WKSHPS)* (pp. 740-745). IEEE.

[59] Wang, X., Gao, H., Jia, Z. and Li, Z., 2023. BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors*, *23*(20), p.8361.

[60] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), pp.834-848.

[61] Schapire, R.E. and Singer, Y., 1998, July. Improved boosting algorithms using confidence-rated predictions. In *Proceedings of the eleventh annual conference on Computational learning theory* (pp. 80-91).

[62] Elfwing, S., Uchibe, E. and Doya, K., 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks*, *107*, pp.3-11.

[63] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

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

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

[66] Zhang, Y., Liu, H. and Hu, Q., 2021, September. Transfuse: Fusing transformers and cnns for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 14-24). Cham: Springer International Publishing.

[67] He, K., Zhang, X., Ren, S. and Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern*

analysis and machine intelligence, *37*(9), pp.1904-1916.

[68] Miri Rostami, S. and Ahmadzadeh, M., 2018. Extracting predictor variables to construct breast cancer survivability model with class imbalance problem. *Journal of AI and Data Mining*, *6*(2), pp.263-276.

[69] A. Kazempour, J. Tanha, S. Roshan, M. Zarrin and H. Nikkhah, "MC-BioCLIPSR: A Mamba-CNN Hybrid Network with BioMedCLIP-Guided Loss for High-Resolution Brain MRI Reconstruction," *2025 15th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, Islamic Republic of, 2025, pp. 1-6, doi: 10.1109/ICCKE68588.2025.11273813.

یک مدل یادگیری عمیق ترکیبی نوین با بهبود اتصالات پرش با تقویت میدان دریافتی و تابع زیان تطبیقی برای بخش‌بندی تصاویر پزشکی

مهدی زرین * و حانیه نیکخواه

گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، ایران.

ارسال ۲۰۲۵/۰۷/۱۰؛ بازنگری ۲۰۲۵/۰۸/۱۹؛ پذیرش ۲۰۲۵/۰۹/۱۵

چکیده:

تحلیل تصاویر پزشکی، که نقشی کلیدی در تشخیص و درمان بیماری‌ها دارد، با چالش عدم توازن کلاس‌ها روبه‌رو است؛ جایی که نواحی بافت سالم از نظر مساحت بسیار بیشتر از نواحی غیرطبیعی‌اند. همچنین، نسبت‌های متغیر کلاس‌ها در تصاویر مختلف استفاده از تنظیمات یکنواخت برای تابع زیان را دشوار می‌کند. برای رفع این مشکلات و ارتقای بخش‌بندی خودکار، این پژوهش مدل یادگیری عمیق نوینی را معرفی می‌کند که در آن ماژول‌های کارآمد استخراج ویژگی در YOLOv8(C2F , SPPF) برای ارائه ماژول افزایش میدان دریافتی (RFE) ادغام شده‌است. ماژول (RFE) در معماری U-شکل به عنوان یک اتصال میان‌گذر پیشرفته عمل کرده و ویژگی‌های چندمقیاسه‌ی حاصل از لایه‌های متناظر و متوالی رمزگذار را که توسط SPPF و C2F پردازش شده‌اند، تلفیق می‌کند تا انتقال ویژگی و میدان دریافتی بهبود یابد. برای مقابله با عدم توازن و تنوع توزیع کلاس‌ها، ما تابع زیان جدیدی به نام Adapt Exponential Lossپیشنهاد می‌کنیم که در سطح پیکسل به‌طور پویا وزن کلاس‌ها را برای هر تصویر بر اساس نسبت پیکسل‌های ضایعه به کل پیکسل‌ها تنظیم می‌کند. مدل و تابع زیان پیشنهادی بر روی مجموعه‌داده‌های ISIC 2018، ISIC 2017و PH2 آزمایش شدند و به ترتیب امتیاز IoU برابر با ۸۶,۴۷٪، ۸۵,۶۷٪ و ۹۳,۱۳٪ و Dice برابر با ۹۱,۶۳٪، ۹۰,۱۹٪ و ۹۶,۰۲٪ به‌دست آمد. نتایج نشان می‌دهد که مدل ما در بخش‌بندی دقیق ضایعات پوستی با وجود عدم توازن و تنوع اندازه‌ی ضایعات عملکردی درخشان دارد. این پژوهش چارچوبی برای بخش‌بندی تصاویر پزشکی ارائه می‌دهد و گامی مؤثر در توسعه‌ی ابزارهای تشخیصی دقیق‌تر در پوست‌شناسی است.

کلمات کلیدی: بخش‌بندی تصاویر پزشکی, تابع زیان, عدم توازن داده, ضایعه پوستی , تقویت میدان دریافتی.