Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-20080-7

SEDARU-net: a squeeze-excitation dilated based residual U-Net with attention mechanism for automatic melanoma lesion segmentation

Samira Lafraxo[1] · Mohamed El Ansari[1,2] · Lahcen Koutti[1] · Zakaria Kerkaou[1] · Meryem Souaidi[1]

Received: 23 March 2024 / Revised: 25 July 2024 / Accepted: 10 August 2024

Abstract
One of the most dangerous types of skin cancer, malignant melanoma, must be detected early on in order to receive successful therapy. If melanoma is not diagnosed in a timely manner, it might possibly result in death. According to clinical research, because of their varied hue, texture, and imperceptible borders, these early melanoma indications are very challenging for dermatologists to recognize. Therefore, it's crucial to suggest an automated method that can accurately identify and distinguish between benign and malignant melanoma. Numerous automated methods have been developed by scientists to segment abnormalities from dermoscopic images. On the other hand, conventional models could find it difficult to reliably capture the multi-scale properties, which could result in inconsistent segmentation performance for a variety of object shapes and sizes. Furthermore, models with complicated forms and bounds, such as U-Net and DeepLabV3+, have difficulty properly segmenting tiny, thin, or complex lesions. Thus, we introduce A Squeeze-Excitation Dilated Residual U-Net with Attention Mechanism (SEDARU-Net) in this paper, a novel and automated semantic segmentation network for efficient skin lesion segmentation. To keep spatial information across layers and capture both local and global context, the model is built on U-net combined with dilated convolution. To solve optimization problems, residual blocks are used instead of the basic U-net units. This enhances feature learning, encourages better feature reuse, and allows for the creation of deeper and more robust networks. In order to encourage feature recalibration, global context awareness, and spatial adaptation, each residual block is supplemented with squeeze and excitation units. In addition, The attention gate is also included in the skip connection part of the network to enhance the beneficial channel dimension characteristics and suppress the unreliable background features. According to the results of the experiments on the publicly accessible PH2 dataset, the dice coefficient and intersection over union were determined to be 97.48% and 95.10%, respectively, better than those of current standard methods.

Keywords Melanoma · Segmentation · Squeeze-excitation blocks · Dilatation · Residuals blocks · Attention mechanism

Extended author information available on the last page of the article

Published online: 19 August 2024

1 Introduction

The growth of unchecked abnormal skin cells is known as skin cancer. Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), and Malignant Melanoma (MM) are the three main kinds of skin cancer [40]. Non-Melanoma Skin Cancer refers to basal, squamous, and other kinds of skin cancer that are not melanoma. Compared to the other two (BCC and SCC), melanoma is less prevalent but more harmful. because melanoma spread from one section of the body to another.

According to the Global Cancer Statistics 2020, mela-noma is one of the most fatal skin cancers and one of the malignancies with the fastest global growth rates, causing thousands of deaths annually [53]. The number of persons affected by melanoma has been continuously rising for the past 30 years. In the United States, 207,390 instances of melanoma were detected in 2021, according to statistics [50]. The survival rate of early-stage melanoma is increased by 95% nowadays thanks to the development of the most recent diagnostic tools, such as spectroscopy, compared to 15% for advanced melanoma [7].

Over time, several imaging modalities have been employed to examine the skin. One of the most frequently used imaging methods in dermatology, dermoscopy has improved the accuracy of diagnoses [42]. By using a light magnifying device and immersion fluid, it is a non-invasive imaging technique that makes it possible to see the skin's surface [37]. Dermoscopy has limitations brought on by the human component despite its great value. In the clinics, the physicians employed a few widely used diagnostic tools, including the ABCD (Asymmetry, Border, Color, Different structures) rules, a seven-point checklist, visual methods like laser, and a few more. However, manual interpretation of dermoscopic images takes time and requires clinical training and experience as a dermatologist. Additionally, even a skilled dermatologist might make mistakes when doing a diagnosis [25]. To help the dermatologist make a quick and accurate diagnosis, computer-based diagnostic and analytic procedures are needed.

Computer-aided diagnosis (CAD) technologies have been created to aid dermatologists. Today, at many screening sites and hospitals, CAD has developed into a crucial component of the routine clinical work for the identification of abnormalities in medical images [12]. According to [13] and [21], CAD systems typically include a number of components such as image recording, pre-processing, feature extraction, classification and segmentation. The authors in [26] address the problem of poor contrast in medical images that might impact clinical diagnosis as a pre-processing step. The method divides the images into low-frequency and high-frequency components using shear wavelet modification. Following that, a modified Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is used to contrast-adjust the low-frequency component. To preserve the image's spectral information, the final product is subjected to further processing using a fuzzy contrast enhancement approach. The suggested method may greatly improve picture contrast while maintaining crucial image details, according to experimental results. As stated by Khan et al. [27], feature fusion is a dynamic field that is essential to the last classification stage in the medical field. Better results are obtained when two images or numerous characteristics are combined, either for the purpose of detecting or classifying diseased regions. The identification of salient areas and image quality are both enhanced by the fusion approaches. Additionally, merging medical image improves segmentation accuracy overall, and feature fusion becomes evident in the classification process's last step.

As the characteristics for the classification are generated from the area of interest (ROI) of a segmented mask, segmentation is an essential part of skin cancer diagnostics [28]. Despite the importance of each phase, the segmentation step stands out since it offers visual imagine information. The inaccurate categorization outcome is mostly due to an inadequate segmentation approach. Due to deceiving elements such the color of the lesions, edge information, hair, markers, poor frames, size, blood vessels, and air bubbles, segmentation is the most difficult stage of melanoma identification [41]. Figure 1 depicts some of these challenges.

The majority of conventional approaches for skin lesion segmentation rely on manually hand-crafted parameter determination and image preprocessing. However, due to difficult issues including varied skin lesions, varying forms, and blurred borders, these conventional algorithms perform poorly for segmenting some complicated situations. In contrast, deep learning models can segment dermoscopic images more effectively than conventional algorithms by learning various aspects of the images over time. The segmentation of medical images, such as skin lesions, is where deep learning currently holds a strong position. Deep convolutional neural network-based segmentation techniques have gained great performance with the latest advancements in deep learning [15, 29–31, 34, 35, 52]. For medical image segmentation problems, several researchers have presented the traditional encoder-decoder network design and produced competitive results. In this network topology, the encoder typically extracts image features while the decoder typically outputs the final prediction result and restores the extracted features to the original picture size [44]. When employing deep learning, researchers frequently ran into issues with irrelevant characteristics that lower the specified Deep Learning (DL) model's identification accuracy [49].

For the purpose of autonomously segmenting mela-noma lesions from dermoscopic images, we present in this study a deep learning model called SEDARU-Net. The U-net, dilated convolution, residual blocks supplemented with squeeze and stimulation, and attention mechanism are all used to good effect in the model. The following benefits are provided by this model:

- ✓ Using dilated convolution in the encoder path helps the model keep spatial information across successive layers and better capture local and global context. The U-Net can handle difficult segmentation tasks, where context and small features are crucial for correct segmentation results. This is made possible by the enlarged receptive field and decreased downsampling.
- ✓ By including residual blocks into the decoder path of the U-Net design, it is possible to build deeper and more potent networks while addressing optimization problems and enhancing feature learning and reuse. This result in better segmentation performance, particularly in circumstances where it is important to capture intricate features and complicated spatial connections.
- ✓ By including squeeze and excitation (SE) units in each residual block, the model's segmentation performance was considerably enhanced by feature recalibration, global context awareness, and spatial flexibility.
- ✓ Attention gated is utilized in place of directly concatenating feature maps in skip connections. By using these connections, the model can decide how important each feature map is and regulate how much data is sent from the encoder to the decoder.

The structure of the paper is as follows. The current state of the art is presented in Section 2. The suggested model's approach is demonstrated in Section 3. The results of experiments and comparisons are described in Section 4. Finally, conclusions are drawn in Section 5.

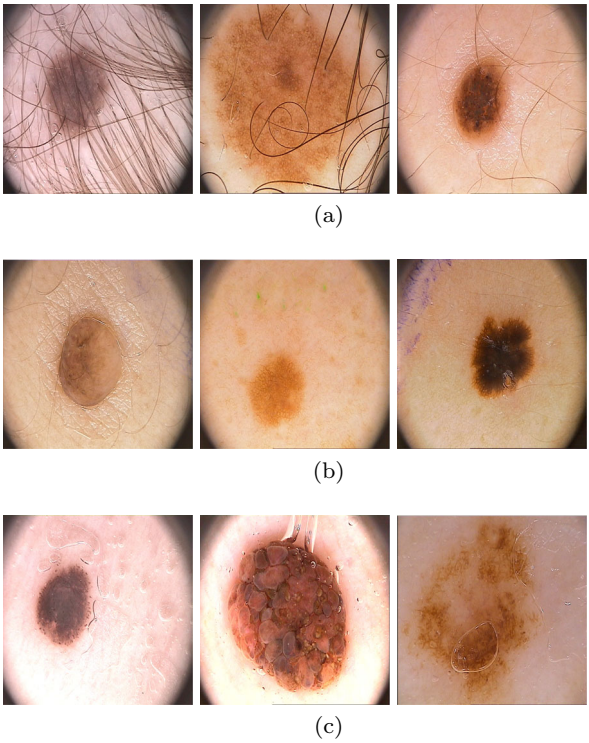

Fig. 1 Different artifacts can be noticed in dermoscopic images. (a) Skin hairs (b) Marker Ink (c) Gel Bubbles

2 Related work

Advanced AI and machine learning techniques, such as deep learning, knowledge graphs, and statistical methods are widely used to tackle real-world problems in domains like operating systems, emotion analysis, drug recommendations, and cybersecurity. The paper [22] proposes using blockchain and generative models to address integrity verification and behavioral classification tasks related to large datasets for smart operating systems. The authorsutilized Deep Convolutional Generative Adversarial Network to extract deep features and classify input properly with an F1-score of 0.973 and a general validation accuracy of 97.08%. Moreover, the notion of blockchain is employed to maintain the authenticity of the dataset and the analytic outcomes. Ali et al. [5] in their work introduce a novel multilingual BERT-based approach to improve the identification of emotions from resource-constrained language data. Dense word embedding representations from pre-trained models are used to represent words in many languages. In addition, BERT has been optimized and pre-trained for the classification task. The work [48] combines knowledge graphs, convolutional neural networks, and sentiment analysis techniques to develop an enhanced drug recommendation system. Through attribute mining, the knowledge graph efficiently captures the relatedness between the user and the object. Authors in [6] explores the use of statistical methods to detect malware and concept drift in AI-based sensor data streams, going beyond just cybersecurity applications.

Different computer-based approaches have been created during the past 20 years to address the difficulties associated with segmenting and identifying melanoma [32, 33]. In the literature, several features are utilized to categorize the techniques for skin lesion segmentation

[49]. This number of classes is often defined as four, which includes supervised techniques, edge-based methods, active contour methods, and histogram thresholding methods. Pixel-level characteristics are often used in these techniques. These low-level features' success falls short of expectations. The deep learning concept has entered practically every sector thanks to its astonishing recent success. The direction of research in this field has significantly shifted as it moves toward the field of skin lesion segmentation [41]. Thus, this study is going to examine skin lesion segmentation investigations in two stages: before and after deep learning.

The methods based on hand-craftted features are divided into three groups: edge-based algorithms, region-based algorithms, threshold-based algorithms. Edge-based techniques: which can be manual or automated, search for edge pixels and connect them to create picture contours. Lesion borders are marked using the trackpad in the manual application. As opposed to the manual technique, the automatic one makes use of edge detection methods such the watershed algorithm [10], active contours [14], canny edge detector [45], and multidirection gradient vector flow snake model [9]. This segmentation involves applying an edge filter to the picture, classifying pixels as being near or far from edges based on the filter's output, and assigning those pixels that are not separated by edges to the same class. The approach that uses edges most frequently is active contours.

Region-based methods, assuming that adjacent pixels should have the same value, these algorithms partition pictures into regions or groups of comparable pixels depending on their properties. Each pixel in an area is compared to its neighbors and grouped according to particular circumstances. This class of algorithms includes K-means and fuzzy C-means clustering, mean shift-based gradient vector flow [61], iterative region-based [51], iterative stochastic region-merging [56], and mean shift-based algorithms.

Depending on the threshold estimation methods, threshold-based algorithms can be categorized as point-based or pixel-based segmentation. These algorithms frequently struggle to estimate effective thresholds because of dermoscopic aberrations [17]. Examples of this group of approaches include OTSU [24], histogram estimation [20], morphological operations [3], optimum color channel-based empirical threshold estimation [1], and mean pixel intensity level-based threshold estimation [4]. To solve the threshold estimation problem for image segmentation, the study [55] suggests a deep learning model known as the "Deep Threshold Prediction Network (DTP-Net)". The model predicts the ideal gray-level threshold that optimizes the Dice similarity index between the segmented and ground-truth pictures given grayscale versions of the macro images as input. Out of 11 cutting-edge threshold estimating techniques, the DTP-Net showed the lowest root mean square error in threshold prediction. The model precisely predicted the threshold separating the lesion from the background after being trained to distinguish between the two in the intensity space.

For the purpose of extracting textural information from dermoscopic pictures, Pedro et al. [46] constructed structural co-occurrences matrices (SCM). Compared to other textural aspects, these extracted features offer strong discriminating abilities. On the basis of the ISIC 2016 and 2017 datasets, the evaluation achieved a specificity of more than 90%. The majority of the collected characteristics are categorized using supervised learning methods. The most used classifier for computerized lesion categorization is Support Vector Machine (SVM).

AI models, particularly CNNs, have had great success in many facets of medical imaging because they allow for the end-to-end construction of supervised models without the need for manually extracting features. Very deep residual networks with more than 50 layers were proposed by Yu et al. [58] for a two-stage framework of segmenting and classifying skin lesions. They asserted that the richer and more discriminative characteristics produced by the

deeper networks are used for recognition. The two-stage architecture and very deep networks are computationally costly, despite the fact that the effort produced encouraging results.

For the segmentation of skin lesions, Bi et al. [8] developed multi-stage fully convolutional networks (FCNs). Localized coarse appearance learning was place in the early stages of the multistage, and detailed boundary characteristic learning took place in the latter stages. Additionally, they used a parallel integration strategy to facilitate the merging of the results, which they said improved the detection. Their strategy outscored others in the PH2 dataset (90.66%).

By utilizing 19-layer DCNN, Yuan et al. [59] suggested an end-to-end fully automated technique for segmenting skin lesions. As a measurement, they used the Jaccard Distance and included a loss function. Various parameters, including input size, optimization techniques, augmented strategies, and loss function, were used to compare the outcomes. The best performance was chosen using 5-fold cross-validation with the ISBI training dataset to fine-tune the hyperparameters.

An ensemble CNN strategy for skin lesion segmentation was created by Mahbod et al. [36]. The presented technique combines intra and inter architectures for features abstraction levels, and each architecture is built from a number of pre-trained CNN networks that have been tweaked using provided dermoscopy pictures. Finally, SVM is used to classify the retrieved features after fine-tuning. The ISIC 2017 is used for the experiments. The outcomes were superior to the top-performing methods.

Dermoscopic Skin Network (DSNet), developed by the authors in [17], is a novel automatically generated semantic segmentation network for reliable skin lesion segmentation. They employed depth-wise separable convolution instead of normal convolution to project the learned discriminating features onto the pixel space at various stages of the encoder in order to decrease the number of parameters and make the network lightweight. They also used U-Net and Fully Convolutional Network (FCN8s) to contrast with the suggested DSNet.

In order to accurately segment images of skin lesions, the study in [44] suggests a Gated Fusion Attention Network (GFANet), which creates two progressive relation decoders. A prediction result is created as the initial guide map when the authors fuse numerous tiers of contextual information using a Context information Gated Fusion Decoder (CGFD). Then, it is optimized by a prediction decoder consisting of a shape flow and a final Gated Convolution Fusion (GCF) module, where they iteratively use a set of Channel Reverse Attention (CRA) modules and GCF modules in the shape flow to combine the features of the current layer and the prediction results of the adjacent next layer to gradually extract boundary information. Finally, they employ GCF to combine low-level characteristics from the encoder with the final output of the shape flow in order to hasten network convergence and increase segmentation accuracy.

A convolutional neural network based on position and context information fusion attention, known as PCF-Net, is suggested in the study [23], utilizing UNet as the baseline model. Position and Context Information Aggregation Attention Module (PCFAM), a unique two-branch attention mechanism, is created to aggregate Position and Context information. To extract long-range dependencies, a global context information complementary module (GCCM) was created. To collect multiscale feature data and insert it in the UNet bottleneck, a multi-scale grouped dilated convolution feature extraction module (MSEM) was developed.

There are limitations and challenges which motivated us to introduce the SEDARU-NET architecture in automatic Melanoma lesion segmentation. First, Melanoma lesion has extensive odd shape, completely different size and complex contours that put them out of traditional segmentation models facilities for accurate extraction. The complex nature and richness of properties displayed by melanoma lesions cannot always be satisfactorily encapsulated

within existing models, resulting in the suboptimal performance during segmentation. Second, Precise demarcation of melanoma lesions is critical for early diagnosis, patient care and monitoring disease progression. Lesion segmentation errors or poor tumor delineation may result in misdiagnosis or sub-optimal clinical decision, underlining the necessity of more accurate and better performing lesion area detection. Third, there is wide variation in the color, texture and other visual characteristics of melanoma lesions based on patient factors along with level of progression. Such a large range of variation may be difficult for conventional segmentation models to handle, rendering their performance potentially inconsistent across different types and appearances of lesions. Finally, the segmentation process should be computationally efficient for real-time or near-real time analysis of medical images in a clinical set-up. Although current models for segmentation can be resource-intensive in terms of time and memory, making them difficult to deploy on low-powered clinical devices.

In order to overcome these limitations and challenges, we propose the SEDARU-NET architecture that integrates the following components. Squeeze-Excitation blocks which help the model to learn the transformation of discriminative features to informative features and transformation from weak irrelevant features to informative features which in turn help to segment the features of irregular complex Melanoma lesions. Dilated Convolutions that increase the receptive field of the model that helps us to effectively learn the long-range spatial dependencies and understand the context of the features around the locus. Residual Connections which help us to learn more effective features and that creates an efficient passageway throughout the network for the passage of information which in turn improves the efficiency of the outcome obtained from segmentation. Attention: The attention mechanism helps to highlight more contributing features which in turn helps to effectively segment irregular diverse melanoma lesions with high efficiency. With the incorporation of the above components, we aim to construct an accurate segmentation model SEDARU-NET which is computationally efficient in the greater place for diagnosis and treatment planning of skin cancer. The methodology section will go through our approach in further depth.

3 Proposed methodology

In this section we will outline our approach, SEDARU-Net, depicted in Fig. 2. With substantial improvements to accommodate the characteristics of dermoscopic images and the applications they serve, SEDARU-Net is designed based on the U-net architecture. The SEDARU-Net model employs a deep convolutional neural network design that consists of an encoder and a decoder. The encoder extracts multi-scale features from the input dermoscopic image, while the decoder reconstructs a pixel-wise segmentation map.

To further enhance the performance on dermoscopic image segmentation tasks, SEDARU-Net incorporates several key modifications to the standard U-net architecture. These include the use of squeeze-and-excitation blocks to adaptively recalibrate the feature maps, feature maps are upsampled with regular dilated convolution for accurately finding information, and the attention gate module is included in the skip connectins to record and screen for high-level features with additional spatial contextual information.

The resulting SEDARU-Net model demonstrates state-of-the-art performance on a range of dermoscopic image segmentation benchmarks, accurately delineating structures such as lesions, blood vessels, and hair follicles. This robust and customized deep learning architecture advances the capabilities of automated dermoscopic image analysis, with important

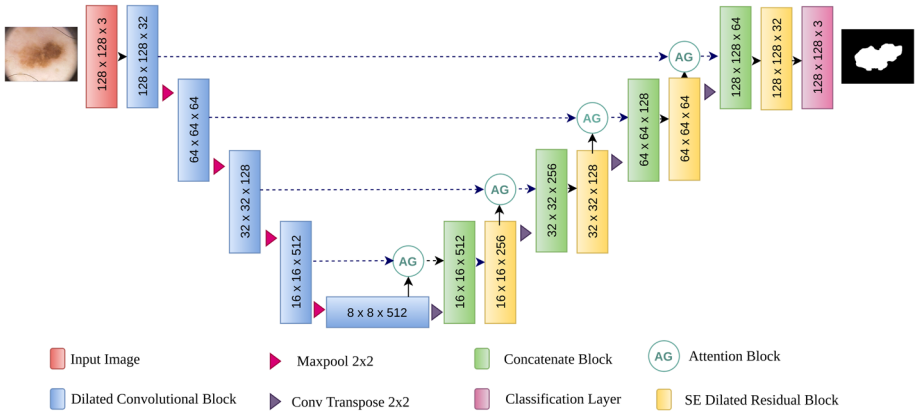

Fig. 2 The proposed SEDARU-Net framework

applications in early skin cancer detection and monitoring. A full description of SEDARU-Net fundamental components is explained as follows.

3.1 Data preprocessing

The original dataset included images with a resolution of 768×560, requiring substantially more expensive processing. As a result, before feeding the segmentation models with input, the images must be scaled. Based on multiple studies, we determined that 128×128 was the right size. The advantages of data augmentation procedures in boosting the amount and quality of training data to accommodate all data changes have been demonstrated by recent studies. In order to expand the dataset and aid in the successful completion of the segmentation task, a variety of data augmentation techniques are applied in our case. These techniques include vertical and horizontal flipping, rotating, and zooming. The majority of the time, image data's pixel values are integers with values ranging from 0 to 255. However, since the neural network algorithm only employs minimal weight values, inputs with big values might make learning more difficult. Data normalization is seen to be a viable solution for this, requiring that each pixel value fall between 0 and 1. Divide all pixel values by 255, the highest pixel value, to get that result. The image can be made more resistant to changes in lighting and other elements that could have an impact on the image's overall brightness and contrast by normalizing it.

3.2 U-Net

Olaf Ronneberger, Philipp Fischer, and Thomas Brox proposed the U-Net [47], a deep learning architecture, in 2015. It is a convolutional neural network (CNN) developed primarily for tasks involving pixel-by-pixel categorization of images and semantic segmentation. The network's U-shaped architecture, which comprises an encoding path and a decoding path, gave rise to the name "U-Net". Similar to a conventional CNN, hierarchical feature representations are captured by downsampling the input image repeatedly via convolutional layers. The downsampled feature maps are upsampled using deconvolutional layers in the decoding path, which is a mirror image of the encoding path and enables accurate localization

of the segmented objects. To provide both global and local context information during the segmentation process, skip connections are also used to concatenate feature maps from the encoding path with the associated upsampled feature maps. U-Net is especially successful in tasks like medical image segmentation, where precise boundary delineation and localization of structures are critical. This is due to the design's ability to let U-Net preserve fine-grained features. Since then, U-Net has gained popularity and influence as an architecture in the field of computer vision, and several iterations have been used to solve various segmentation issues.

3.3 Residual blocks

A crucial part of the ResNet [18] (Residual Network) design are residual units, which are often referred to as residual blocks. They were developed to solve the vanishing gradient issue that might arise during training in very deep neural networks. The network finds it difficult to acquire meaningful representations at deep layers because of the vanishing gradient problem, which might degrade overall performance. Utilizing shortcut connections (skip connections) that go around one or more neural network layers is the fundamental concept of residual units. By allowing the gradient to pass directly through the network, these skip connections make it simpler for the network to pick up valuable characteristics. Instead of attempting to learn the direct mapping, the residual units are meant to learn the residual mapping (the difference between the desired output and the input). A residual unit can be conceptualized mathematically as follows:

Assume that x is the input and that $H(x)$, which is the output that is wanted, reflects the mapping that the residual unit is attempting to learn. The following definition applies to the residual unit's output:

$$H(x) = x + F(x) \tag{1}$$

where x is the residual unit's input. The residual mapping that the residual unit is attempting to learn is called $F(x)$. It shows the discrepancy between the input x and the desired output $H(x)$. A set of convolutional layers with nonlinear activation functions can be used to represent the term $F(x)$. The original information is successfully preserved by the skip connection, which makes sure that the original input x is appended to the output. Figure 3 depicts the difference between the residual block and the regular convolutional block.

3.4 Squeeze and excitation module

By adaptively recalibrating the channel-wise features, the Squeeze-and-Excitation (SE) block is a process used to improve the representational capability of neural networks. Jie Hu, Li Shen, and Gang Sun first mentioned it in the Squeeze-and-Excitation Networks study published in 2018 [19].

Squeezing and exciting are the two fundamental operations that make up the SE block. The exciting operation learns channel-specific weights to adjust the relative relevance of each channel. The squeezing operation is in charge of collecting global information from each channel. Convolutional neural networks' performance has been demonstrated to be considerably improved by the SE block, which can be implemented into a variety of architectures.

1. Squeeze operation To collect each channel's overall statistics, the squeeze operation combines the spatial data from each channel. It involves employing global average pooling for

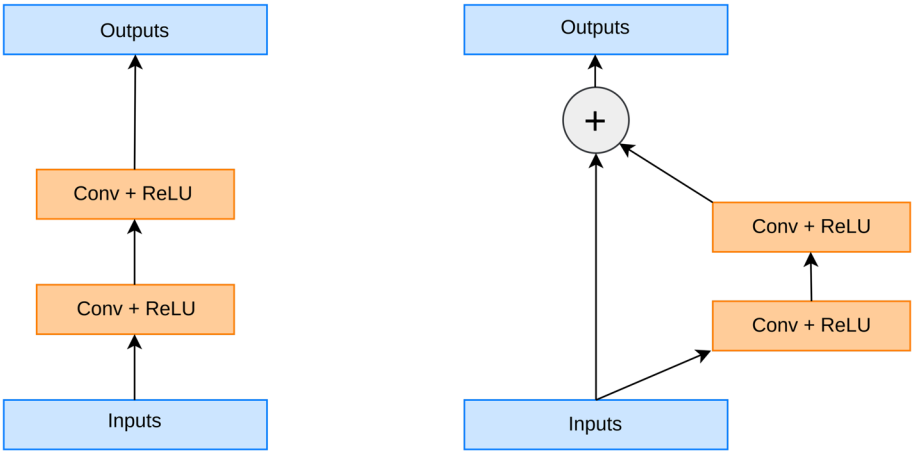

Fig. 3 The distinction between the regular convolutional block and the residual block

each channel's feature map's spatial dimensions (width and height). The spatial dimensions are condensed into a single value per channel via this procedure. Assume for the purposes of this example that the input feature map for the SE block has the dimensions (*C, H, W*), where *C* is the number of channels, *H* is the height, and *W* is the width. This is how the squeezed tensor *Z* is calculated:

$$Z(c) = GlobalAveragePooling(X(c)) \tag{2}$$

For $c = 1$ to C, where *Z(c)* represents the c-th value of the squeezed tensor, indicating the global statistics of the *c-th* channel, and *X(c)* is the *c-th* channel of the input feature map.

2. Excitation operation The excitation operation is in charge of recalibrating the relative relevance of each channel by learning channel-wise weights. It features non-linear activations between two dense, completely linked layers.

Assume that after global average pooling, the compressed tensor Z has the dimensions (C, 1, 1). The following mathematical representation of the excitation operation is possible:

$$S = ReLU(FC1(Z)) \tag{3}$$

where $FC1$ stands for the first dense layer that is completely linked, with weights $W1$ and biases $b1$. ReLU stands for the element-wise Rectified Linear Unit activation function applied to FC1's output. S is the first fully connected layer's output tensor with the dimensions (C, 1).

$$E = Sigmoid(FC2(S)) \tag{4}$$

where $FC2$ stands for the second dense (fully connected) layer, which has weights $W2$ and biases $b2$. Sigmoid represents the element-wise application of the sigmoid activation function to the output of FC2. With dimensions (C, 1), E is the output tensor of the second completely linked layer.

3. Scale and rescale In the third and last stage, the original feature map X is scaled and rescaled using the updated channel-wise weights E. The SE block can now prioritize channels that are more relevant and ignore those that are less crucial. The SE block's output Y is

calculated as follows:

$$Y(c, h, w) = X(c, h, w) * E(c) \tag{5}$$

For $c = 1$ to C and $h, w = 1$ to H, W, where $Y(c, h, w)$ represents the output feature map's $c\text{-}th$ channel value at the coordinates (h, w). $X(c, h, w)$ represents the input feature map's c-th channel value at spatial point (h, w). $E(c)$, which was derived from the excitation procedure, is the weight that has been calibrated for the c-th channel.

The model may focus on more informative characteristics and increase its representation capability by including the SE block into the network design. This results in enhanced performance across a range of tasks. Figure 4 shows a detailed diagram of the Squeeze and Excitation Unit.

3.5 Dialated convolution

Atrous convolutions, often referred to as dilated convolutions [57], are a kind of convolutional operation that permits an expanded receptive field without significantly increasing the number of parameters. Modern convolutional neural networks frequently employ them to perform tasks like picture segmentation and object recognition. By adding pauses or skips between the kernel parts, the dilated convolution essentially increases the kernel's receptive field.

The following mathematical equations can be used to illustrate the mechanism of dilated convolution:

Consider a 2D dilated convolutional kernel K of size $K\_size \times K\_size$, where K_size is the kernel size, and an input feature map X of size $H \times W$, where H is the height and W is the width. The dilation rate D governs the size of the receptive field and sets the distance between the kernel components.

1. Input Feature Map:

$$X = [x(i, j)] \tag{6}$$

for i = 1 to H and j = 1 to W

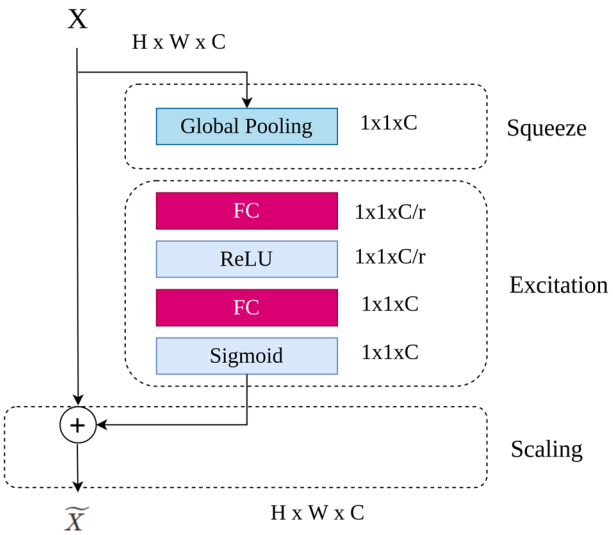

Fig. 4 A detailed diagram of the Squeeze and Excitation Unit

The value at spatial point (i, j) in the input feature map X is denoted by the notation $x(i, j)$.

2. Dilated Convolution Kernel:

$$K = [k(u, v)] \tag{7}$$

for

$$u = -(K\_size - 1)/2 \; to \; (K\_size - 1)/2 \tag{8}$$

and

$$v = -(K\_size - 1)/2 \; to \; (K\_size - 1)/2 \tag{9}$$

Where the value at location (u, v) in the 2D dilated convolution kernel K is represented by the notation $k(u, v)$.

3. Dilated Convolution Operation: By moving the dilated kernel K across the input feature map X, the dilated convolution process calculates the output feature map Y. The convolution procedure is computed as follows for each location (i, j) in the output feature map Y:

$$Y(i, j) = \sum_{u,v} [x(i + D * u, j + D * v) * k(u, v)] \tag{10}$$

The distance between the kernel pieces in this case is governed by the dilation rate D. The dilated convolution operates similarly to the ordinary convolution function when D=1. Without needing to increase the kernel size or the number of parameters, the receptive field grows as D increases due to the spread out nature of the kernel components.

When attempting to capture multi-scale contextual information in pictures, the dilated convolution process is extremely helpful. The network can analyze input data at various resolutions thanks to this, effectively integrating data from a larger input region. As a result, dilated convolutions are frequently employed in deep learning models for semantic segmentation, where accurately segmenting objects in pictures requires collecting both local and global context. Figure 5 depicts the dilated convolution with different dilation rates equal to 1, 2, 3.

3.6 Attention gate

By concentrating on the most important components of the input data, the attention mechanism is a potent tool employed in deep learning to increase the representational capability of models. The model can capture long-range relationships and handle sequential or spatial data more effectively because of the attention mechanism, which enables the model to choose attend to different regions of the input.

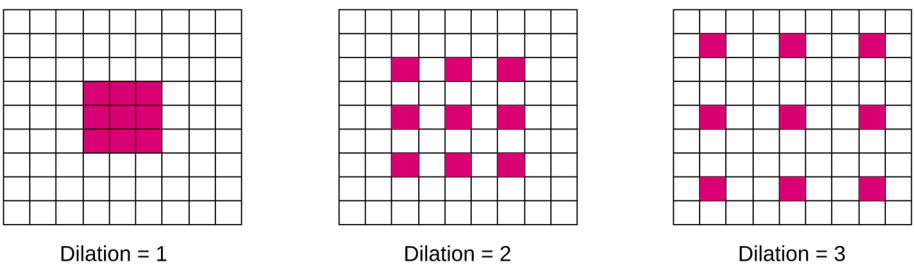

Fig. 5 Dilated convolution with dilation rates equal to 1, 2, 3

The attention gates as introduced by Oktay et al. [39] make use of additive soft attention; an attention gate unit combines the input feature vector with the attention vector to produce a new, weighted feature vector. As shown in Fig. 2, the vectors x^l and g are fed into the attention gate; the vector, g, is retrieved from the network's next lowest layer; as a result, the vector has lower dimensions and more feature representation because data originates from deeper in the network.

Vector x^l in the previous case would be $C_x \times H_x \times W_x$ (filters height width) and vector g would be $C_g \times H_g \times W_g$. Vector x^l undergoes a strided convolution, whereas vector g undergoes a 1×1 convolution. The two vectors are summed elementwise, and as a result, aligned weights get larger while unaligned weights get smaller.

The resultant vector is put through a 1×1 convolution and *ReLU* activation layer, which brings down the dimensions to $H \times W \times 1$. The attention coefficients (weights) are produced by scaling this vector via a sigmoid layer between [0,1], with coefficients closer to 1 denoting more important information. The attention coefficients are upsampled to the x^l vector's original dimensions via trilinear interpolation. The initial x^l vector is element by element multiplied by the attention coefficients α, which scales the vector x^l in accordance with significance. The skip connection then continues to transmit this along normally. Figure 6 illustrated a detailed structure of attention gate module.

3.7 SEDARU-Net architecture

The encoder and decoder parts of our suggested network, SEDARU-Net, are depicted in Fig. 2 as two separate components that together constitute a symmetrical structure. The decoder is in charge of feature location, whereas the encoder is in charge of feature extraction. Five dilated convolutional blocks, four pooling layers, and four SE dilated residual blocks make up the entire architecture. The pooling layer is 2x2, the convolution kernel is 3x3, and the input image is 128x128 pixels. Following a sequence of operations on each feature map, including feature extraction, pooling, and convolution, a binary segmented image of 128x128 pixels is produced.

Dilated convolutional block structure is used in place of the standard unit of the conventional U-Net Fig. 7.b in the network's encoding path Fig. 7.a. The dilation rate regulates the distances between the kernel components, which efficiently broadens the convolutional operation's receptive field without degrading the feature maps' spatial resolution. The network gains the ability to modify the dilated convolution kernels' parameters to capture relevant information during training. The learning process enhances the model's capacity to capture both broad context and tiny details by optimizing the dilation rates and convolutional weights. Our design performs better when dilated convolutions are incorporated into the encoding pro-

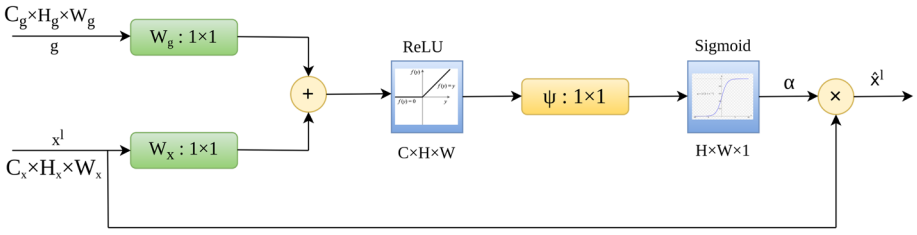

Fig. 6 The structure of attention gate module

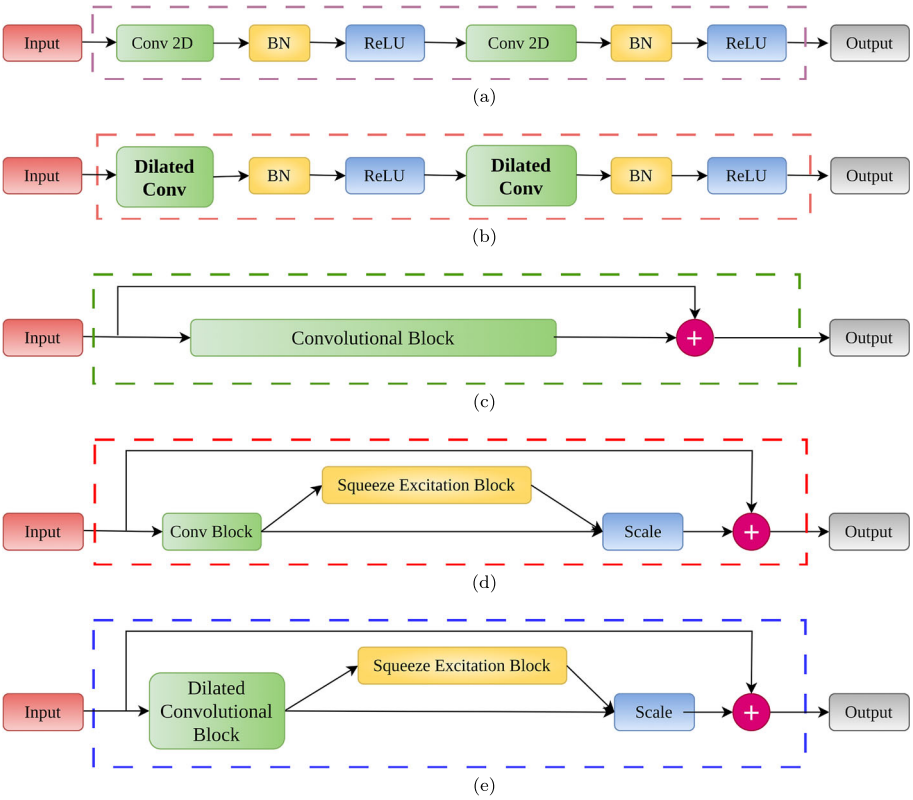

Fig. 7 Different building blocks employed in the paper. (a) U-Net basic building Convolutional block (b) Dilated Convolutional Block (c) Residual Convolutional Block (d) SE Residual Block (e) SE Dilated Residual Block

cess because the receptive field is expanded and more multi-scale contextual data is captured. This leads to greater segmentation accuracy and better handling of complicated images since the network can keep tiny features while still comprehending the image's wider context.

Squeeze and excitation dilated residual learning structures Fig. 7.e are used in place of the typical U-Net's basic unit in the network's decoding process. In contrast, the residual structure augments shortcut connections based on a single forward propagation, enabling the training of deeper networks without degrading performance and the extraction of more discriminative features. Each dilated convolution (dilation = 2) is followed by batch normalization and ReLU activation procedures in the residual unit. Incorporating batch normalization not only lessens the model's sensitivity to starting parameters, but it also partially exerts the regularization effect. Due to its ability to avoid the gradient vanishing problem, it is most frequently employed for activation in the ReLU function. After the output and input are joined together (identity short-cut), it is squeezed and excited (SE) in order to eliminate unnecessary characteristics and excite more useful ones.

In order to improve the network feature combination process by enabling it to concentrate on the most useful features of the feature maps, we also add the attention mechanism in the skip connections. Training teaches the attention mechanisms in the skip connections to give the combined characteristics the proper weights. The attention mechanisms are improved through this learning process to efficiently direct the network's focus on pertinent features.

4 Experimental results

In this section, we provide a brief overview of the dataset, evaluation criteria, and implementation details. We then use several sets of experiments to demonstrate the superiority of the proposed SEDARU-Net over other conventional segmentation models. First, segmentation results of several models are obtained from dermoscopic images on PH2 datasets, and the model performance is examined in light of the findings. Second, we demonstrate how each additional component has an improved effect on the model's ability to segment skin lesions more accurately.

4.1 Datasets

The PH2 and ISIC-2016 datasets, two publicly available benchmark datasets, were used to assess the suggested methodology.

4.1.1 PH2 dataset

The Dermatology Service at Hospital Pedro Hispano in Portugal created the PH2 [38] database. 200 photos of melanocytic lesions with a resolution of 768×560 pixels are included in this collection, comprising 160 images in the Nevi class and 40 images in the Melanoma class. Additionally, for the segmentation procedure, the 200 images ground truth photos are also supplied by qualified medical professionals and made available to the general public [2]. Figure 8 shows some of the dermoscopic images and their corresponding ground truth from the PH2 dataset.

4.1.2 ISIC-2016 dataset

The International Symposium of Biomedical Imaging Collaboration (ISBI) has made these datasets available for the purpose of melanoma segmentation and identification [16]. ISIC-2016 has 379 test samples with matching ground truth test sample masks and 900 training

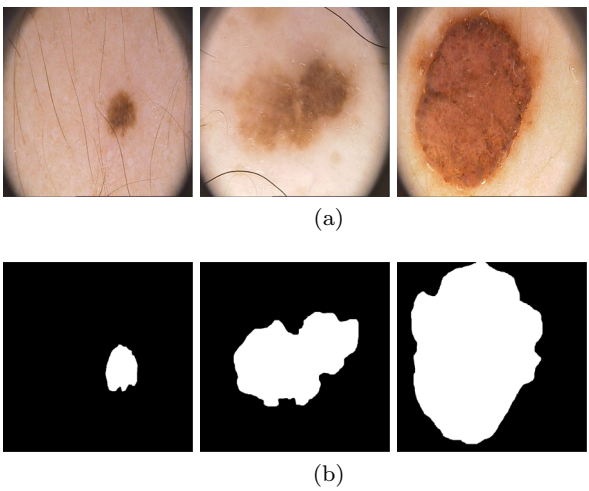

Fig. 8 Example frames from the PH2 publicly available datasets (a) Image (b) Corresponding Ground Truth

samples with matching ground truth masks for the segmentation task. The organizers supply ground facts for training and test samples so that models may be learned and the segmentation approach can be evaluated. Table 1 depicts some of the two datasets details and Fig. 9 shows some of the dermoscopic images and their corresponding ground truth from the ISIC-2016 dataset.

4.2 System implementation

In terms of computational resources, this model runs on an Intel Core i7 machine equipped with a GeForce GTX 1050 (4 GB RAM dedicated to the GPU). The deep learning framework is Python 3.6 with the Keras framework and Tensorflow as the backend, and the operating system is Ubuntu 16.04 LTS with Cuda 8.0.61 installed. We set the batch size to five and the model training epochs to fifty during the network training procedure. We manually conducted several trials using different sets of hyperparameters on the same dataset and with the same model in order to identify the optimal collection of hyperparameters. Those sets were selected on the basis of the findings of the empirical investigation. The initial learning rate was set to 0.0001, the network optimizer used Adam's algorithm, and the loss function to dice coefficient loss. During training, a dropout method with a ratio of 0.5 is employed to keep the network from overfitting. 20% is used for validation and 80% of each dataset is randomly selected for training. More detailed descriptions of the network architecture are depicted in Table 2.

4.3 Evaluation metrics

The performance of the models may be evaluated and compared using several indicators. The Jaccard index, also known as intersection over union (IoU), the Dice coefficient, accuracy, precision, recall, and F1-score are the metrics that are most frequently used for medical image segmentation applications. Following is the formula for calculating these metrics.

Dice coefficient The Dice coefficient (DC) is a comparative statistic for contrasting the pixel-by-pixel outcomes of the segmentation that was anticipated and the ground truth that was provided. Its value range is 0 to 1, with 0 denoting complete spatial overlap and 1 denoting total spatial overlap between two sets of binary segmentation results. The Dice coefficient may be calculated by dividing the total area of A and B by two times the area of their intersection.

$$DiceCoefficient(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{11}$$

Jaccard index The Jaccard index is a common metric for assessing segmentation strategies. As seen in the equation below, it determines how similar the anticipated value (A) and the

Table 1 Distribution of the datasets used in our study

Dataset	Images	Train / Test	Imaging type	Resolution	Color Depth	Aspect Ratio
PH2	200	160 / 40	Dermoscopy	768×500	24-bit RGB	4:3
ISIC-2016	1279	900 / 379	Dermoscopy	500×500 to 3000×4000	24-bit RGB	4:3 to 16:9

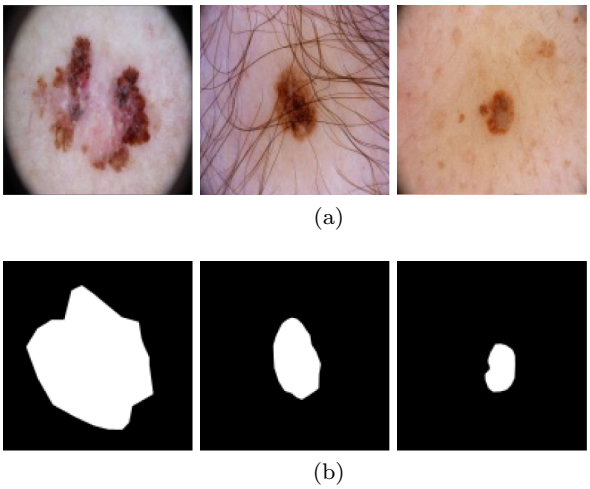

Fig. 9 Example frames from the ISIC-2016 publicly available datasets (a) Image (b) Corresponding Ground Truth

actual value (B) are.

$$JaccardIndex(A, B) = \frac{|A \cap B|}{A \cup B} = \frac{|A \cap B|}{|A| + |B| + |A \cap B|} \tag{12}$$

Accuracy With (TP), (TN), (FP), and (FN) standing for true positive, true negative, false positive, and false negative, respectively, high accuracy means that the pixels were properly identified.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{13}$$

Precision Precision, is a statistic that assesses the percentage of positive cases that were properly detected out of all instances that were anticipated to be positive.

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

Table 2 Network architecture details

Hyperparameter	Value
Input image size	128×128
Filter size	3×3
Dilation rate	2
Filter number	32, 54, 128, 256, 512
Activation function	Sigmoid
Pooling size	2×2
Number of dilated convoltutional blocks	5

Recall The proportion of accurately detected positive events among all really positive instances is measured by recall, also known as sensitivity or true positive rate.

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

F1-score The harmonic mean of recall and accuracy is known as the F1-score. It offers a fair evaluation of a model's performance, accounting for both false positives and erroneous negatives.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{16}$$

4.4 Results and discussions

4.4.1 Ablation study on PH2 dataset

Using an ablation research, we were able to determine which Unet family variation was the best for melanoma detection. Because melanoma borders are unclear and melanoma varies greatly in texture and color, melanoma identification is a difficult task. Thus, we maintained the same experimental configuration and investigated the effect of Unet variations on melanoma localization. The Table 3 compares the performance of several variants of the U-Net family of models, including the standard U-Net, ResU-Net, AttResU-Net, DilatedU-Net, and DAttResU-Net. The metrics reported are the Dice score and Intersection over union. The Dice score is a measure of the overlap between the predicted segmentation and the ground truth, with a value of 1 indicating perfect agreement. Pixel accuracy is the percentage of pixels that are correctly classified. The table shows that the SEDARU-Net model achieves the highest Dice score of 97.48% and IoU of 95.10%, but has the longest one iteration time of 62s. The standard U-Net model has the fastest one iteration time of 12s, but the lowest Dice score IoU among the variants.

The proposed SEDARU-Net achieves the highest value on IoU, DC because it is equipped with Dilated convolutional blocks, Residual units, Squeeze-Excitation blocks, and Attention Mechanism to solve the problem of different target sizes, irregular shapes, and blurred boundary of lesions. The AttU-Net's attention mechanism aids the network in concentrating on the

Table 3 Ablation study on PH2 dataset

Model	DC%	IoU%
U-Net	94.24	90.16
ResU-Net	94.92	90.34
AttResU-Net	95.88	92.10
DilatedU-Net	96.05	92.40
DAttResU-Net	97.19	94.54
SEDARU-Net	***97.48***	***95.10***

pertinent object borders, edges, and textures. This is very helpful for segmenting objects with complex shapes. By allowing the model to concentrate on the areas that need tighter delineation, creating sharper object borders, it facilitates precise localisation of objects. By encouraging better gradient flow, addressing the vanishing gradient issue, and aiding the learning of more representative feature hierarchies, residual blocks can improve the performance of a U-Net design. This makes it possible for the network to maintain context and collect precise details. SEDARU-Net performs better than them nevertheless. This may be as a result of the encoding path's performance being improved by the inclusion of dilated convolutions by increasing the receptive field while maintaining the same spatial resolution. As a result, the network can collect more extensive contextual data without sacrificing its capacity to collect small details. These experimental findings show that the proposed SEDARU-Net can perform superbly on tiny datasets as well.

This type of ablation study can be useful for understanding the trade-offs between different U-Net variants and selecting the most appropriate model for a given application, considering factors such as accuracy, speed, and computational resource requirements.

To further evaluate the performance of the proposed SEDARU-Net approach, we conducted other experiments on ISIC-2016 dataset, a skin cancer image dataset that is widely used for method validation and benchmarking. ISIC-2016 was chosen because unlike PH2 dataset which contains a limited number of images, this second dataset has a larger number of images. Our method achieved promising results on the ISIC-2016 dataset, with an accuracy of 92.86%, a dice coefficient of 89.51%, and a jaccard index of 81.13%. The total number of epochs was set to 50. There could be several reasons why our model performs well on the PH2 dataset but not as well on the ISIC2016 dataset. First, the image quality, resolution, and preprocessing steps differ between the two datasets, affecting the model's performance. Moreover, the PH2 dataset may have more homogeneous lesions, while the ISIC-2016 dataset have a more diverse range of lesions, making it more challenging for the model to generalize. In addition, inconsistent and noisy annotations in the ISIC2016 dataset could confuse the model and lead to poorer performance. To address these issues, we experiment with different data augmentation techniques to increase the diversity of the training data. We also tried to adjust the model complexity to better fit the ISIC2016 dataset.

4.4.2 Qualitative results

The results of the ablation experiments are graphically displayed in Fig. 10, which includes seven instances of dermoscopic images, their underlying ground truth, and the segmentation masks for the various suggested models. The dermoscopic image examples in Fig. 10 show several difficult problems, such as the existence of interfering information, such as hair and markers, blurred borders or poor contrast between lesions and background, irregular forms and varied sizes of lesions, and interfering information, such as hazy boundaries or low contrast between lesions and background. The challenge is whether the network can acquire richer characteristics and interpret boundary information more correctly to segment lesions with irregular forms and different sizes. While this is going on, the suggested SEDARU-Net, which includes SE blocks, Attentions gate units, and dilated residual modules, performs better when it comes to segmentation. The suggested extra blocks improve the performance of skin lesion segmentation, according to the analysis of the evaluation metrics data in Table 3 and the presentation of segmentation results in Fig. 10. The best segmentation performance network is created by the efficient combination of these modules.

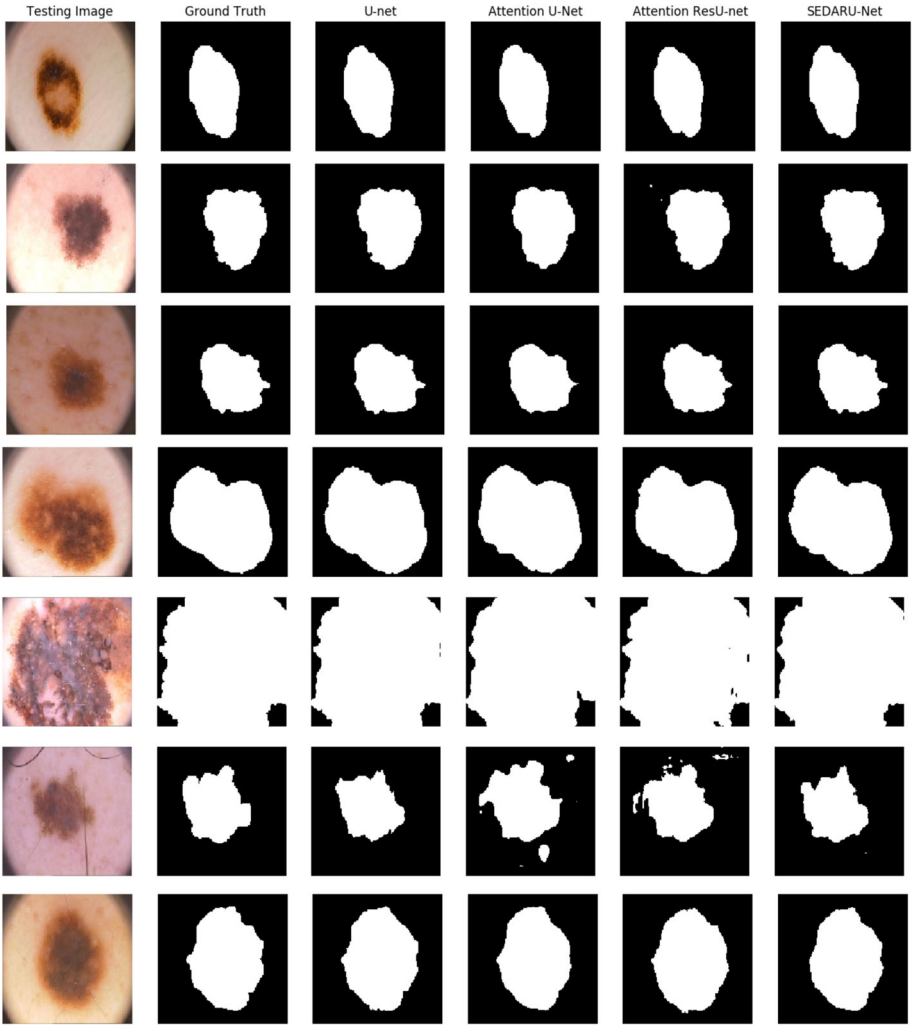

Fig. 10 Segmentation results of different models on the PH2 dataset

4.4.3 Comparison with the state of the art approaches

Additionally, our strategy is contrasted with many current cutting-edge strategies. They are all based on deep learning architecture, and Table 4 summarizes their results using the PH2 dataset. The most used metrics in segmentation tasks, including accuracy, dice coefficient, and Jaccard index, are included in this table along with a brief discussion of each approach. The table demonstrates that, on the PH2 dataset, our suggested architecture virtually gets the greatest metrics of any technique. It generates the greatest ACC at 97.42%, demonstrating that the SEDARU-Net outperforms deep learning-based approaches in terms of performance. The suggested SEDARU-Net outperforms the competing systems, iFCN, iU-Net, and GFANet, on the primary evaluation measure IoU, by 8%, 6.36%, and 4.12%, respectively. It's possible that SEDARU-Net will exceed all of these previous great studies thanks to the four potent components we add to the fundamental U-Net.

Table 4 The comparison of our best method with the previously published approaches on the same dataset

Author	Network	Acc (%)	IoU (%)	DC (%)	Method
Zafar *et al.* [60]	DeepLabv3+	90.47	87.1	93.02	MobileNet + DeepLabv3
Peng *et al.* [43]	-	93.00	85.00	90.00	U-Net + Discrimination Network
Tong *et al.* [54]	ASCU-Net	94.30	86.15	92.12	AG + Spatial and Channel Attention U-Net
Chen *et al.* [11]	O-Net	95.14	86.15	92.12	Recurrent Attentional Convolutional Networks
Ozturk *et al.* [41]	iFCN	96.92	87.10	93.02	FCN + Residual Blocks
Jiang *et al.* [23]	iU-Net	-	88.74	93.80	U-Net + Swin Transformer + CNN
Qiu *et al.* [44]	GFANet	97.09	90.98	95.06	Gated Fusion Attention Network
Proposed method	***SEDARU-Net***	***97.42***	***95.10***	***97.48***	***SEDARU-Net***

The bold entries indicate the results of the proposed model

Using dilated convolutions instead of the usual convolutional blocks (Dilated Convolution Blocks) in U-Net has different benifits; First, dilated convolutions have the nice property of increasing convolutional filters receptive field without adding parameters. This allows the model to understand more information from a wider spatial context within the input image. We also can use dilated convolutions with different dilation rates in the U-Net architecture to extract features at multiple scales efficiently. This contributes to model ability of capturing both local and global contextual information that is essential for accurate segmentation of complex structures, skin lesions in this work. Second, the encoder part of the standard U-net uses a series of max-pooling or strided convolutions to reduce spatial resolution gradually on feature maps. Using dilated convolutions here instead of these kernels allows the encoder to maintain a high spatial resolution so that more detail could be preserved at a pixel level and, thusly, improve segmentation accuracy (particularly for smaller or complex lesions). Third, dilated convolutions (when increasing the dilation factor) can reach a significantly wider range without requiring an additional number of extra layers in the network, as compared to using standalone deeper stack of standard convolutional operations. This may allow us to create more efficient models that have less parameters and perform faster inference which is a desirable property in real world applications. Finally, because the dilation operation creates holes (missing information) in the feature maps, dilated convolutions can create boundary artifacts that don't align with boundaries/margins of objects and invalidate predictions. That is, it can demand additional techniques such as padding the image or special treatment in boundary regions to guarantee a good segmentation near its borders.

Moreover, integrating SE blocks into the U-Net architecture can help the U-Net model learn more discriminative and relevant features for the segmentation task by adaptively rescaling the feature channels. It can also enhance the model's ability to adaptively recalibrate the feature representations, leading to improved segmentation accuracy, increased model expressivity, and better generalization capabilities. However, the specific performance gains will depend on the complexity of the segmentation task, the characteristics of the dataset, and the overall design of the U-Net model. It can enhance the overall representational power of the U-Net architecture, allowing it to better capture the intricate patterns and relationships in the input data. The channel-wise attention mechanism introduced by SE blocks can help the U-Net model generalize better to unseen data. By focusing on the most important features, the model can become more robust to variations in the input data, leading to improved segmentation performance on diverse skin lesion samples.

The residual connections help the model learn more effective features by facilitating the flow of information across different layers, improving the overall segmentation performance. When the attention mechanism helps the model focus on the most relevant features, further enhancing its ability to segment irregular and diverse melanoma lesions.

By incorporating all these components in one model, the SEDARU-NET architecture aims to address the limitations of existing segmentation models (which integrate just one component attention mechanism or residuals blocks) and provide a more robust, precise, and computationally efficient solution for automatic melanoma lesion segmentation, which is crucial for early diagnosis and effective treatment planning.

4.4.4 Failure cases analysis

However even if our architecture could produce positive results, there are still certain failure cases, as Fig. 11 shows. The majority of failure cases exhibit irregular borders, small lesions, lesions with ambiguous boundaries, lesions with pigment variation, and lesions with irregular

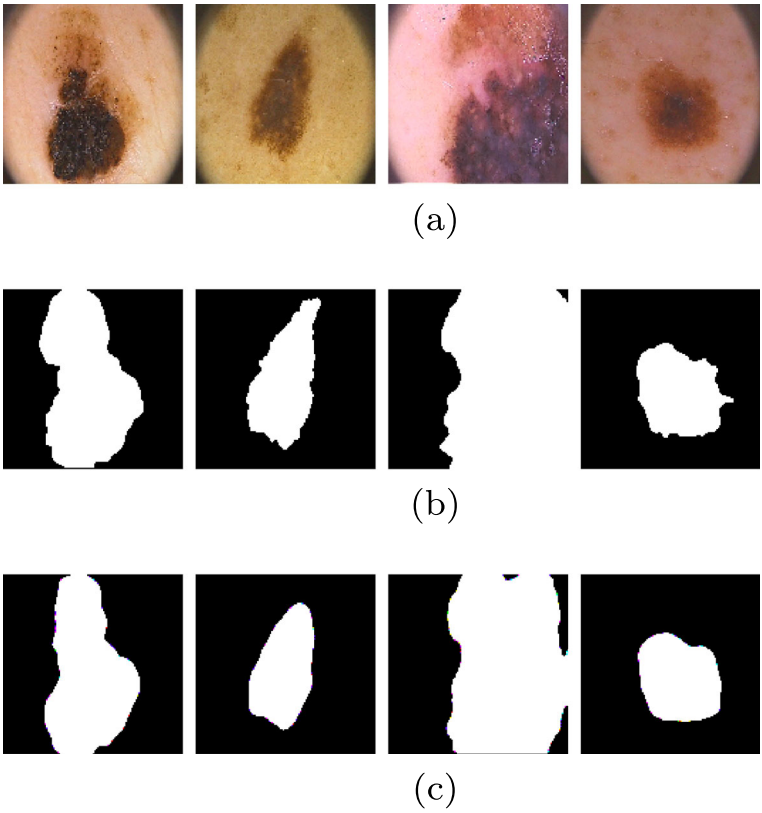

Fig. 11 Failure case analysis (a) Original Image (b) Corresponding Ground Truth (c) Predicted Mask

textures. The model struggled to accurately segment lesions with highly irregular or ill-defined borders. These lesions often exhibited complex shapes and textures, which challenged the model's ability to capture the full extent of the melanoma. In addition, smaller melanoma lesions, especially those less than 5 mm in diameter, were more prone to segmentation errors. The model tended to either over-segment or under-segment these smaller targets, leading to reduced accuracy. Moreover, certain lesions had boundaries that were difficult to distinguish from the surrounding healthy skin, either due to subtle color variations or the presence of hair, moles, or other skin features. The model often failed to delineate the precise boundaries of these ambiguous lesions. Melanoma lesions with significant intra-lesional pigment variation, such as those with both dark and light regions, were challenging for the model to segment accurately. The model tended to either miss parts of the lesion or include surrounding healthy skin. Lesions with highly irregular surface textures, such as those with nodular or verrucous features, were more likely to be misclassified or poorly segmented by the model.

Based on the analysis of the failure cases, we have identified several potential avenues for improving the performance of the melanoma segmentation model. First, leveraging contextual information, such as the anatomical location of the lesion or the patient's age and medical history, could help the model better distinguish between melanoma and other skin lesions with similar visual characteristics. Second, utilizing a multi-scale feature extraction approach, such as a feature pyramid network or a hierarchical attention mechanism, could enable the

model to better capture both local and global features, improving its ability to segment lesions with varying sizes and complexities. Third, expanding the training dataset to include a more diverse set of melanoma lesions, including those with irregular borders, small sizes, and varied pigment distributions, could help the model generalize better to a wider range of lesion characteristics.

5 Conclusion

In this study, we present SEDARU-Net, a straightforward yet reliable deep learning model, which suggests that performance improvements are possible when cutting-edge U-net is combined with revolutionary concepts like attention mechanism, residual blocks, squeeze and excitation units, and dilated convolution. The mean dice coefficient and mean intersection over union were found to be 98.74% and 97.51%, respectively, exceeding state-of-the-art techniques despite the challenges of dermoscopic images with poor contrast and unpredictability in shape and size.

The attention gates in the skip connections, which enable U-Net to dynamically change its feature fusion process and concentrate on the most crucial information at various scales, are integrated to accomplish this outperformance. This improves the model's capacity to manage different segmentation difficulties, promote feature reuse, and successfully integrate contextual data. Additionally, residual blocks were included to aid with feature propagation, allowing low-level information to move smoothly throughout the network without the need for a complex architectural structure. Additionally, squeeze and excitation blocks now incorporate a reduction of less significant characteristics to provide additional discriminative features. Finally, the U-Net can capture additional context and long-range relationships by employing dilated convolutions.

The existing SEDARU-Net system has shown some promise, but there is still need for improvement. Future studies could concentrate on innovative deep learning methods like generative adversarial networks (GANs) for enhancing data and various Transformers designs like BERT and GPT for improved feature extraction. To improve its diagnostic capabilities, the system may use data from other medical imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI) investigations. Future studies may focus on enhancing technology for real-time usage, which would provide instantaneous feedback and, maybe, more quicker diagnosis and treatment. Long-term patient outcomes will probably be improved by further study and development of CADx systems for the early detection of skin cancer.

Acknowledgements This work was partially supported by the Ministry of National Education, Vocational Training, Higher Education and Scientific Research, The Ministry of Industry, Trade and Green and Digital Economy, Digital Development Agency (ADD) and National Center for Scientific and Technical Research (CNRST). Project number: ALKHAWARIZMI/2020/20.

Author Contributions Methodology and Funding Acquisition, (S. LAFRAXO & M. EL ANSARI & Z. KERKAOU); Project Administration, (M. EL ANSARI & L. KOUTTI); Writing: Original Draft Preparation, (S. LAFRAXO & M. SOUAIDI); Writing: Review and Editing (M. EL ANSARI); All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The dataset analysed during the current study is available in the [http://www.fc.up.pt/addi/] repository.

Declarations

Conflicts of interest/Competing Interests The authors declare that they have no conflict of interest. The authors have no relevant financial or non-financial interests to disclose.

Consent to publish The participant has consented to the submission of the case report to the journal.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

1. Abbas AA, Guo X, Tan WH, Jalab HA (2014) Combined spline and b-spline for an improved automatic skin lesion segmentation in dermoscopic images using optimal color channel. J Med Syst 38:1–8
2. ADDI-Project (2003) PH2 database. http://www.fc.up.pt/addi/
3. Akram T, Khan MA, Sharif M, Yasmin M (2018) Skin lesion segmentation and recognition using multichannel saliency estimation and m-svm on selected serially fused features. J Ambient Intell Humaniz Comput 1–20
4. Al-abayechia AAA, Guoa X, Tana WH, Jalabc HA (2014) Automatic skin lesion segmentation with optimal colour channel from dermoscopic images. Sci Asia 40(1):1–7
5. Ali N, Tubaishat A, Al-Obeidat F, Shabaz M, Waqas M, Halim Z, Rida I, Anwar S (2023) Towards enhanced identification of emotion from resource-constrained language through a novel multilingual bert approach. ACM Trans Asian Low-Resour Lang Inf Process
6. Amin M, Al-Obeidat F, Tubaishat A, Shah B, Anwar S, Tanveer TA (2023) Cyber security and beyond: Detecting malware and concept drift in ai-based sensor data streams using statistical techniques. Comput Electr Eng 108:108702
7. Balch CM, Gershenwald JE, Sj Soong, Thompson JF, Atkins MB, Byrd DR, Buzaid AC, Cochran AJ, Coit DG, Ding S et al (2009) Final version of 2009 ajcc melanoma staging and classification. J Clin Oncol 27(36):6199
8. Bi L, Kim J, Ahn E, Kumar A, Fulham M, Feng D (2017) Dermoscopic image segmentation via multistage fully convolutional networks. IEEE Trans Biomed Eng 64(9):2065–2074
9. Cavalcanti PG, Scharcanski J (2013) Macroscopic pigmented skin lesion segmentation and its influence on lesion classification and diagnosis. In: Color medical image analysis, Springer, pp 15–39
10. Chakkaravarthy AP, Chandrasekar A (2018) An automatic segmentation of skin lesion from dermoscopy images using watershed segmentation. 2018 International Conference on Recent Trends in Electrical. Control and Communication (RTECC), IEEE, pp 15–18
11. Chen P, Huang S, Yue Q (2022) Skin lesion segmentation using recurrent attentional convolutional networks. IEEE Access 10:94007–94018
12. Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imag Grap 31(4-5):198–211
13. Fan H, Xie F, Li Y, Jiang Z, Liu J (2017) Automatic segmentation of dermoscopy images using saliency combined with otsu threshold. Comput Biol Med 85:75–85
14. Gangwar K (2021) Study on different skin lesion segmentation techniques and their comparisons. In: 2021 IEEE international conference on imaging systems and techniques (IST), IEEE, pp 1–6
15. Garbaz A, Lafraxo S, Charfi S, El Ansari M, Koutti L (2022) Bleeding classification in wireless capsule endoscopy images based on inception-resnet-v2 and cnns. In: 2022 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), IEEE, pp 1–6
16. Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A (2016) Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). arXiv:1605.01397
17. Hasan MK, Dahal L, Samarakoon PN, Tushar FI, Martí R (2020) Dsnet: Automatic dermoscopic skin lesion segmentation. Comput Biol Med 120:103738
18. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

19. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
20. Jaisakthi S, Chandrabose A, Mirunalini P (2017) Automatic skin lesion segmentation using semi-supervised learning technique. arXiv:1703.04301
21. Jalalian A, Mashohor S, Mahmud R, Karasfi B, Saripan MIB, Ramli ARB (2017) Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. EXCLI J 16:113
22. Jan S, Musa S, Ali T, Nauman M, Anwar S, Ali Tanveer T, Shah B (2021) Integrity verification and behavioral classification of a large dataset applications pertaining smart os via blockchain and generative models. Expert Syst 38(4):e12611
23. Jiang Y, Dong J, Zhang Y, Cheng T, Lin X, Liang J (2023) Pcf-net: Position and context information fusion attention convolutional neural network for skin lesion segmentation. Heliyon 9(3):e13942
24. Joseph S, Olugbara OO (2022) Preprocessing effects on performance of skin lesion saliency segmentation. Diagnostics 12(2):344
25. Khan MA, Lali IU, Rehman A, Ishaq M, Sharif M, Saba T, Zahoor S, Akram T (2019) Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. Microsc Res Tech 82(6):909–922
26. Khan SA, Hussain S, Yang S (2020) Contrast enhancement of low-contrast medical images using modified contrast limited adaptive histogram equalization. J Med Imaging Health Inform 10(8):1795–1803
27. Khan SA, Khan MA, Song OY, Nazir M (2020) Medical imaging fusion techniques: a survey benchmark analysis, open challenges and recommendations. J Med Imaging Health Inform 10(11):2523–2531
28. Korotkov K, Garcia R (2012) Computerized analysis of pigmented skin lesions: a review. Artif Intell Med 56(2):69–90
29. Lafraxo S, El Ansari M (2020a) Gastronet: Abnormalities recognition in gastrointestinal tract through endoscopic imagery using deep learning techniques. In: 2020 8th International conference on wireless networks and mobile communications (WINCOM), IEEE, pp 1–5
30. Lafraxo S, El Ansari M (2020b) Regularized convolutional neural network for pneumonia detection trough chest x-rays. In: International conference on advanced intelligent systems for sustainable development, Springer, pp 887–896
31. Lafraxo S, El Ansari M (2021) Covinet: Automated covid-19 detection from x-rays using deep learning techniques. In: 2020 6th IEEE congress on information science and technology (CiSt), IEEE, pp 489–494
32. Lafraxo S, Ansari ME, Koutti L (2022a) Melanoma lesion recognition using deep convolutional neural network and global average pooling. In: 2022 5th International conference on advanced communication technologies and networking (CommNet), pp 1–6, https://doi.org/10.1109/CommNet56067.2022.9993899
33. Lafraxo S, El Ansari M, Charfi S (2022b) Melanet: an effective deep learning framework for melanoma detection using dermoscopic images. Multimed Tools Appl 1–25
34. Lafraxo S, El Ansari M, Koutti L (2023a) Computer-aided system for bleeding detection in wce images based on cnn-gru network. Multimed Tools Appl 1–26
35. Lafraxo S, Souaidi M, El Ansari M, Koutti L (2023) Semantic segmentation of digestive abnormalities from wce images by using attresu-net architecture. Life 13(3):719
36. Mahbod A, Schaefer G, Ellinger I, Ecker R, Pitiot A, Wang C (2019) Fusing fine-tuned deep features for skin lesion classification. Comput Med Imag Grap 71:19–29
37. Mayer J (1997) Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. Med J Aust 167(4):206–210
38. Mendonça T, Ferreira PM, Marques JS, Marcal AR, Rozeira J (2013) Ph 2-a dermoscopic image database for research and benchmarking. In: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, pp 5437–5440
39. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, et al. (2018) Attention u-net: Learning where to look for the pancreas. arXiv:1804.03999
40. Orthaber K, Pristovnik M, Skok K, Perić B, Maver U, et al. (2017) Skin cancer and its treatment: novel treatment approaches with emphasis on nanotechnology. J Nanomater 2017
41. Öztürk Ş, Özkaya U (2020) Skin lesion segmentation with improved convolutional neural network. J Digit Imaging 33:958–970
42. Pellacani G, Seidenari S (2002) Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized-light videomicroscopy. Clinics in Dermatology 20(3):222–227
43. Peng Y, Wang N, Wang Y, Wang M (2019) Segmentation of dermoscopy image using adversarial networks. Multimed Tools Appl 78:10965–10981
44. Qiu S, Li C, Feng Y, Zuo S, Liang H, Xu A (2023) Gfanet: Gated fusion attention network for skin lesion segmentation. Comput Biol Med 155:106462

45. Rashid Sheykhahmad F, Razmjooy N, Ramezani M (2015) A novel method for skin lesion segmentation. Int J Inf Secur Syst Manag 4(2):458–466
46. Rebouças Filho PP, Peixoto SA, da Nóbrega RVM, Hemanth DJ, Medeiros AG, Sangaiah AK, de Albuquerque VHC (2018) Automatic histologically-closer classification of skin lesions. Comput Med Imag Graph 68:40–54
47. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer, pp 234–241
48. Saadat H, Shah B, Halim Z, Anwar S (2022) Knowledge graph-based convolutional network coupled with sentiment analysis towards enhanced drug recommendation. IEEE/ACM Trans Comput Biol Bioinform
49. Saba T, Khan MA, Rehman A, Marie-Sainte SL (2019) Region extraction and classification of skin cancer: A heterogeneous framework of deep cnn features fusion and reduction. J Med Syst 43(9):289
50. Siegel RL, Miller KD, Fuchs HE, Jemal A et al (2021) (2021) Cancer statistics. Ca Cancer J Clin 71(1):7–33
51. Singh L, Janghel RR, Sahu SP (2021) Slicaco: An automated novel hybrid approach for dermatoscopic melanocytic skin lesion segmentation. Int J Imag Syst Technol 31(4):1817–1833
52. Souaidi M, Lafraxo S, Kerkaou Z, El Ansari M, Koutti L (2023) A multiscale polyp detection approach for gi tract images based on improved densenet and single-shot multibox detector. Diagnostics 13(4):733
53. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3):209–249
54. Tong X, Wei J, Sun B, Su S, Zuo Z, Wu P (2021) Ascu-net: attention gate, spatial and channel attention u-net for skin lesion segmentation. Diagnostics 11(3):501
55. Venugopal V, Joseph J, Das MV, Nath MK (2022) Dtp-net: A convolutional neural network model to predict threshold for localizing the lesions on dermatological macro-images. Comput Biol Med 148:105852
56. Wong A, Scharcanski J, Fieguth P (2011) Automatic skin lesion segmentation via iterative stochastic region merging. IEEE Trans Inform Technol Biomed 15(6):929–936
57. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122
58. Yu L, Chen H, Dou Q, Qin J, Heng PA (2016) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging 36(4):994–1004
59. Yuan Y, Chao M, Lo YC (2017) Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE Trans Med Imaging 36(9):1876–1886
60. Zafar M, Amin J, Sharif M, Anjum MA, Mallah GA, Kadry S (2023) Deeplabv3+-based segmentation and best features selection using slime mould algorithm for multi-class skin lesion classification. Mathematics 11(2):364
61. Zhou H, Schaefer G, Celebi ME, Lin F, Liu T (2011) Gradient vector flow with mean shift for skin lesion segmentation. Comput Med Imag Graph 35(2):121–127

Authors and Affiliations

Samira Lafraxo[1] · Mohamed El Ansari[1,2] · Lahcen Koutti[1] · Zakaria Kerkaou[1] · Meryem Souaidi[1]

✉ Samira Lafraxo
samira.lafraxo@edu.uiz.ac.ma

Mohamed El Ansari
melansari@gmail.com

Lahcen Koutti
l.koutti@uiz.ac.ma

Zakaria Kerkaou
kerkaou.zakaria@gmail.com

Meryem Souaidi
souaidi.meryem@gmail.com

1 LabSIV, Department of Computer Science, Faculty of Sciences, Ibnou Zohr University, Agadir 80000, Morocco

2 Department of Computer Sciences Faculty of Sciences, Informatics and Applications Laboratory, Moulay Ismail University, Meknès, Morocco