applied sciences

MDPI

Article

An Improved Multi-Scale Feature Fusion for Skin Lesion Segmentation

Luzhou Liu, Xiaoxia Zhang *, Yingwei Li and Zhinan Xu

School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China; 21085400582@ustl.edu.cn (Y.L.); xuzhinan@ustl.edu.cn (Z.X.)
* Correspondence: zhangxiaoxia@ustl.edu.cn

Abstract: Accurate segmentation of skin lesions is still a challenging task for automatic diagnostic systems because of the significant shape variations and blurred boundaries of the lesions. This paper proposes a multi-scale convolutional neural network, REDAUNet, based on UNet3+ to enhance network performance for practical applications in skin segmentation. First, the network employs a new encoder module composed of four feature extraction layers through two cross-residual (CR) units. This configuration allows the module to extract deep semantic information while avoiding gradient vanishing problems. Subsequently, a lightweight and efficient channel attention (ECA) module is introduced during the encoder's feature extraction stage. The attention module assigns suitable weights to channels through attention learning and effectively captures inter-channel interaction information. Finally, the densely connected atrous spatial pyramid pooling module (DenseASPP) module is inserted between the encoder and decoder paths. This module integrates dense connections and ASPP, as well as multi-scale information fusion, to recognize lesions of varying sizes. The experimental studies in this paper were constructed on two public skin lesion datasets, namely, ISIC-2018 and ISIC-2017. The experimental results show that our model is more accurate in segmenting lesions of different shapes and achieves state-of-the-art performance in segmentation. In comparison to UNet3+, the proposed REDAUNet model shows improvements of 2.01%, 4.33%, and 2.68% in Dice, Spec, and mIoU metrics, respectively. These results suggest that REDAUNet is well-suited for skin lesion segmentation and can be effectively employed in computer-aided systems.

Keywords: skin lesions; image segmentation; deep learning; atrous convolution; UNet3+

check for updates

Citation: Liu, L.; Zhang, X.; Li, Y.; Xu, Z. An Improved Multi-Scale Feature Fusion for Skin Lesion Segmentation. *Appl. Sci.* **2023**, *13*, 8512. https://doi.org/10.3390/app13148512

Academic Editor: Jan Egger

Received: 16 June 2023
Revised: 21 July 2023
Accepted: 21 July 2023
Published: 23 July 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Skin cancer is one of the most common types of cancer. Excessive exposure to ultraviolet radiation can cause genetic mutations in superficial skin cells, leading to uncontrolled growth and the development of skin cancer. Melanoma, which arises from melanocyte cells, grows rapidly and poses a higher risk than non-melanoma skin cancer. Early treatment of patients can greatly reduce the mortality rate of melanoma [1]. In the early stages of skin disease detection, dermatologists conduct a detailed analysis of patient information by observing lesion regions and conducting biopsies of affected tissues. However, this segmentation approach is slow and susceptible to personal subjective bias, which can lead to significant structural changes after segmentation. In addition, because lesion regions often have low contrast in terms of surface hair, skin color, capillaries, shape, and texture compared to non-lesion areas, it is difficult for dermatologists to accurately assess the extent of lesions [2,3]. Dermoscopy is a non-invasive detection technique that removes surface reflections from the skin to better visualize deeper layers of the skin. Dermoscopy uses magnification equipment to produce magnified images, so that clinicians can clearly see the originally invisible skin morphology and structure, and help them determine whether the pigmented skin lesions are melanoma or non-melanoma. Figure 1 shows some dermoscopy images.

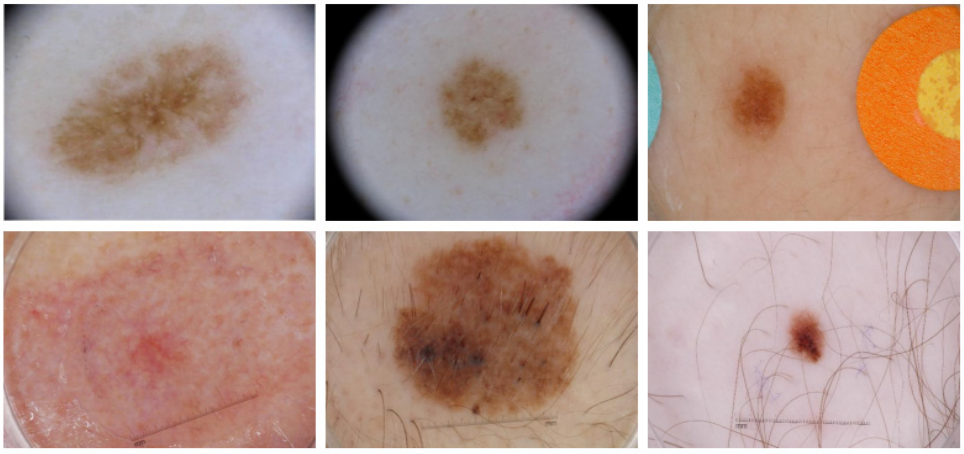


Figure 1. Some dermoscopy images.

Accurate identification of skin lesions in medical images is one of the important research directions in modern medicine, and it is a key step in diagnosis and treating skin cancer [4]. The segmentation method of skin lesions is a difficult task because of significant variations in lesion size, irregular shape and boundaries, complex background information, and blurred edges. Skin lesion segmentation is a necessary step to identify melanoma. The skin lesion areas need to be accurately segmented from surrounding tissues so that doctors can fully observe them, further improving the accuracy of the diagnosis. With an increasing number of skin patients and a limited number of experienced dermatologists, computer-aided diagnosis methods have emerged as an effective way to solve these problems. Accurate segmentation of skin lesions with computer-aided diagnosis can help doctors quickly determine the location of the lesion. This not only significantly reduces the time and segmentation error cost of lesion image segmentation but also promotes further development of related technologies. Currently, classic segmentation algorithms typically perform pixel-level segmentation based on color, texture, and shape features in the image, such as the histogram thresholding method [5], unsupervised clustering method [6], edge-based and region-based method [7], and active contour method [8]. Although traditional skin lesion image segmentation algorithms can achieve a certain degree of segmentation effect, they overly rely on the quality of manually selected features and the introduction of prior information. Meanwhile, due to varying capturing conditions and inherent defects in images, noise interference such as hair, bubbles, and medical measurements may occur during dermoscopy imaging examination. Therefore, accurate extraction of lesions in dermoscopy images is still a significant problem.

In the past few years, the rapid development of deep neural networks has disrupted traditional processing techniques, and deep neural networks have been used in people's lives. In the field of image segmentation, convolutional neural network models have achieved multiple breakthroughs, especially with the emergence of UNet [9], which has brought about a new wave of development in medical image segmentation. This is of great significance in assisting dermatologists to diagnose patients' diseases. UNet uses an encoder–decoder architecture to effectively avoid feature loss during downsampling by applying skip connections. Compared to traditional machine learning methods, the use of deep learning models to assist clinical image segmentation in the field of skin lesion area segmentation has received considerable attention.

Both traditional skin lesion image segmentation algorithms and currently popular deep learning segmentation algorithms aim to achieve fast and accurate segmentation of skin lesion images. Although these methods perform well in skin lesion image processing, they still have many drawbacks. Most existing skin cancer image segmentation models suffer from insufficient feature extraction, inadequate preservation of edge information in the results, and other issues. In the consecutive convolution process, the model may learn more irrelevant features, ultimately leading to poor segmentation performance.

Therefore, extracting low-level and high-level semantic features through networks is equally important for achieving automated skin lesion image segmentation. By doing so, irrelevant or redundant features are avoided during the feature extraction process, and more edge information is retained in the segmentation results.

In conclusion, deep learning algorithms for skin lesion image segmentation are more effective than traditional segmentation algorithms. However, further research and improvement of existing models are needed to improve their robustness and accuracy while suppressing overfitting. According to the current research status, we have proposed a skin lesion segmentation model based on DenseASPP, attention mechanism, and residual learning. The proposed model only adds a few parameters and computational operations, but can highlight the regions of interest related to skin lesion segmentation. We have also conducted experimental evaluations on two datasets, and the results show that the proposed model has advanced performance in skin lesion segmentation. The paper's main contributions are summarized below:

(1) We propose a novel encoder module that utilizes the cross-residual (CR) module to extract features. This module is able to extract deep semantic information to avoid gradient vanishing during network training.
(2) The ECA module is introduced in the feature extraction process of the encoder. This module effectively captures inter-channel interaction information so that the network can learn more lesion characteristics.
(3) The bottleneck layer structure is modified. The redesigned DenseASPP is applied to the bottleneck layer to extend the receptive field via cascading expansion convolution and include more information in the feature map, leading to more exact segmentation of lesions in different sizes.
(4) We propose the REDAUNet network model based on the UNet3+ network. This model is an "end-to-end" architecture and provides better performance for skin lesion segmentation on both the ISIC2017 and ISIC2018 datasets.

2. Related Works

In recent years, neural networks have achieved remarkable success in computer vision tasks. Deep learning segmentation models have overcome the limitations of manual feature segmentation and shown great potential in automatic feature learning. The UNet segmentation network has provided a new way and direction for image segmentation, and its excellent performance has made it an important tool for medical field segmentation tasks. Therefore, researchers in this field have been continuously improving this model, developing faster and more accurate segmentation algorithms to help doctors make accurate diagnoses. Now, there have been many studies on improving the UNet model in the aspect of image analysis.

Zhou et al. [10] proposed UNet++ to overcome the limitations of same-scale feature map fusion alone in UNet. UNet++ combined "long and short" skip connections to guide the decoder to restore feature maps, and then input more image features to the decoder to complete multi-scale feature fusion in the network. However, UNet++ cannot fully exploit the information contained in the image. Therefore, Huang et al. [11] introduced the UNet3+ network architecture, which combines different levels of information through full-scale skip connections for target matching at different scales. Additionally, it achieves globally based feature representation by implementing deep supervision. Li et al. developed the Attention UNet [12] based on the UNet model, which gradually strengthens the weights of local interest regions through self-attention. The network effectively suppresses activation in irrelevant areas, reduces redundant parts of skip connections, and broadens its application range. Jha et al. proposed DoubleUNet [13], which overlays two UNet architectures to achieve good segmentation results in various medical image segmentation tasks using different modes. This network architecture has shown promising results in improving overall segmentation accuracy. Additionally, Jafari et al. introduced DRUnet [14], a novel segmentation network that integrates ResNet and DenseNet into the encoder–decoder

modules of the UNet model. It achieves high segmentation accuracy with fewer model parameters by exploiting features from both models, thus reducing the negative impact of ResNet's relatively shallow depth while benefiting from DenseNet's rich deep features. Chen et al. [15] proposed TransUNet, which is a network model that combines Transformer to address the problems of UNet in modeling long-range dependencies. This network can greatly improve the segmentation performance of UNet by up-sampling global context features via a decoder and combining them with high-resolution feature maps to achieve accurate segmentation. Gao et al. [16] proposed a hybrid transformer architecture named UTNet. The network applies novel self-attention modules in both the encoder and decoder to extend network operations at different levels. Long-range dependencies at different scales are captured while minimizing computational cost, enhancing medical image segmentation. Lin et al. [17] proposed a segmentation model called DS-TransUNet. Unlike many existing transformer-based solvers, the DS-TransUNet employs a mature dual-scale encoding technique. The mechanism uses a dual-scale encoder to obtain feature representations at different scales, and a good transformer interaction fusion block is designed to efficiently fuse multi-scale information through a self-attentive mechanism. In addition, Ruan et al. [18] proposed a lightweight model called MALUNet, which has the characteristics of a few parameters and low complexity. The model combines four new attention modules with a U-shaped structure, greatly reducing model parameters and computational complexity while improving model performance.

For the improvement of network modules, the current typical modules mainly include the inception module [19], attention mechanism, multi-scale feature extraction module, dense connection module, and deep separable module. In neural networks, feature maps with larger receptive fields contain richer information and can better capture global information. Therefore, Zhao et al. [20] designed a pyramid pooling module to acquire features at different scales and merge them through upsampling and concatenation. However, during the upsampling process, the network may struggle to capture small lesion regions, leading to inaccurate segmentation. Chen et al. [21] proposed using Atrous [22] spatial pyramid pooling to capture multi-scale features. This structure consists of five parallel branches, each processing the input feature maps with different convolutional kernels and cascading them at the output. It can obtain a larger receptive field without increasing the parameters Additionally, to fully utilize the information from different receptive fields, Yang et al. [23] proposed DenseASPP, which shares feature information through a cascading approach. This method differs from the parallel connection approach with atrous convolutions used in ASPP. By utilizing a dense connection of a set of atrous convolutions, it covers a larger range and obtains more multi-scale information without increasing the model size. In recent years, some researchers have introduced the DenseASPP module into their networks and achieved good results. For example, Hu et al. [24] trained a DenseASPP model to learn the location and probability maps of the pancreas. In the segmentation stage, a saliency-aware module that combines saliency maps and image context was introduced in DenseASPP. This method utilizes the prominent pancreas features in coarse segmentation and improves the accuracy in the fine segmentation stage. Xu et al. [25] proposed an automatic mandible segmentation method using a 3D fully convolutional neural network. They incorporated the DenseASPP module into the network to extract dense features at multiple scales. Attention gate (AG) modules were applied in each skip connection to reduce irrelevant background information and focus the network on the segmentation region. The proposed network achieved good segmentation performance and accurate automatic mandible segmentation. Abraham et al. [26] collected information across scales through dense connections in the DenseASPP module, enriching multi-scale contextual features. They combined residual connections between encoder and decoder blocks to learn more robust features and achieve better predictions. Li et al. [27] designed an efficient architecture named PSSAC that improves ASPP and DenseASPP by introducing exponentially increasing scales with a serially connected multiple parallel structure. They explicitly adjusted the receptive field of neurons to capture high-level features at an ultra-dense scale.

This approach addressed the problem of semantic segmentation in large-scale variations and complex scenes.

In some studies, attention mechanisms were introduced to enhance feature extraction by selectively attending to relevant features while ignoring irrelevant ones. Woo et al. [28] proposed a network structure that combines spatial attention and channel attention, enhancing feature extraction without significantly increasing computational complexity or model parameters. Jiang et al. [29] combined channel attention and spatial attention mechanisms based on the UNet architecture and used a weighted cross-entropy loss function, achieving excellent results in skin lesion segmentation datasets. Hu et al. [30] enhanced the resolution of skin lesion segmentation by combining spatial and channel attention mechanisms. The spatial attention pathway captures spatially related features of skin lesions, while the channel attention pathway selectively emphasizes discriminative features in the channel dimension. They also introduced a weighted binary cross-entropy loss function to emphasize foreground lesion regions and achieved state-of-the-art performance in the ISIC2017 challenge. Wang et al. [31] proposed an efficient channel attention (ECA) module specifically designed for deep CNNs, which avoids dimension reduction and effectively captures inter-channel interactions. Alahmadi [32] proposed a multi-scale attention U-shaped network (MSAU-Net) that addresses the challenges of texture and shape variations by improving the traditional U-shaped network with attention mechanisms and incorporating bidirectional convolutional long short-term memory (BDC-LSTM) structures to capture shared discriminative features and suppress less informative features. These methods have achieved promising performance in skin lesion segmentation tasks.

Medical image lesions have complex features such as low contrast, blurry borders, and variable positions, shapes, and sizes. Current network models have difficulty accurately segmenting lesion areas in complex medical image segmentation tasks. Therefore, further research is needed to improve the segmentation performance of network models in such complex segmentation scenarios.

3. Proposed Methods

The UNet network is a widely adopted and efficient network architecture. Its design is relatively simple, making it easy to train and adjust. By using an encoder–decoder structure and skip connections, the UNet network demonstrates strong feature extraction capabilities, multi-scale information-processing abilities, and trainability in image segmentation tasks. In this section, the proposed model based on UNet3+ is presented for dermoscopy image segmentation. Whether based on simple connections or dense connections, UNet and UNet++ cannot fully extract full-scale feature information. In contrast, UNet3+ has redesigned the dense connection method between the encoder and decoder, allowing each decoder to cross all scales to fuse small- and same-scale feature maps from the encoder, and also extract information via the large feature maps generated by the current decoder. This improvement not only makes the model more concise, but also more effectively captures complete semantic information. Therefore, UNet3+ is a promising model to push the development of medical image segmentation in the future.

3.1. Overall Structure of the Network

In this paper, we introduce a new image segmentation model named REDAUNet, which is based on the UNet3+ model. The network model combines the cross-residual module, the efficient channel attention (ECA) mechanism, and the DenseASPP module. Figure 2 shows the overall structure of the network model, which can effectively extract feature information and accurately segment images.

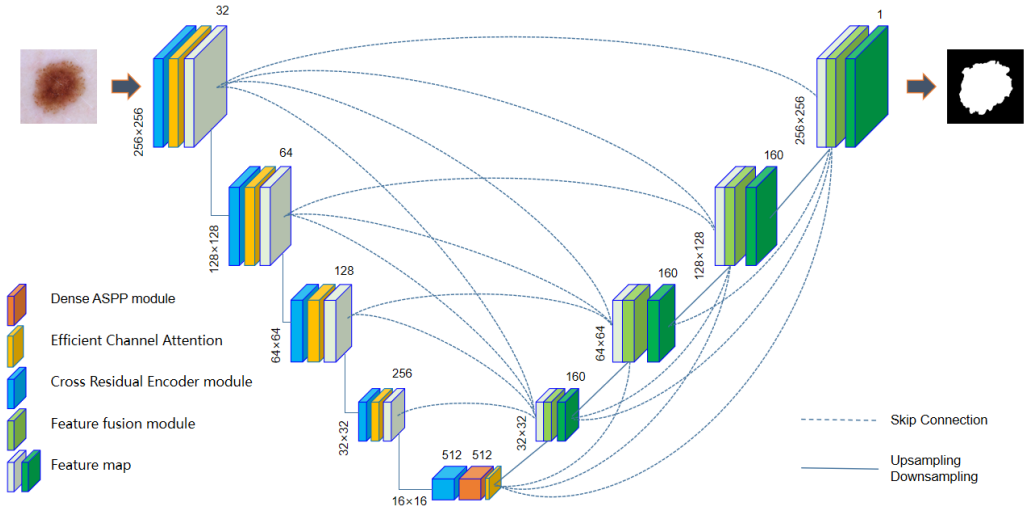


Figure 2. The architecture of the proposed network.

In the figure, the size of the feature map is shown in the left corner of each block, while at the top is the number of channels. The images undergo five downsampling operations to gradually reduce their size and increase the receptive field of the convolutional kernel, thus obtaining more global information. Additionally, the ECA mechanism is adopted to make the model focus more on lesion areas and effectively extract features related to the disease. During the upsampling process, multi-level feature fusion is performed, and the channel limit in the output is 160 to reduce computation. Furthermore, we modified the bottleneck layer between the upsampling and downsampling layers and integrated a newly designed DenseASPP module. The module uses four atrous convolutional layers with different dilation rates to acquire lesion features of different sizes. The detailed segmentation pipeline of the proposed segmentation model includes the following steps:

- Step 1: Input image preprocessing. Preprocess the original image, including resizing, normalization, denoising, etc., to ensure that the input data meet the requirements of the model;
- Step 2: Encoder. The preprocessed image enters the encoder, where a series of convolutional layers and pooling operations are applied to gradually reduce the resolution of the image and extract multi-level feature representations. These feature representations have different semantic information and receptive field sizes, which are used to capture context information at different scales in the image;
- Step 3: Attention mechanism. After each layer of the encoder, an attention mechanism is introduced to add attention weights to the feature maps of each encoder. This enhances the model's focus on key regions. The attention mechanism can dynamically adjust the attention weights based on the relationship between the current pixel and its surrounding pixels, enabling the model to more accurately locate and segment the target region;
- Step 4: DenseASPP. The DenseASPP module is placed between the encoder and decoder. It effectively captures semantic information at different levels and scales in the image by introducing multiple scales of receptive fields and atrous convolutions with different sampling rates. In the decoding stage, it enables more accurate pixel-level predictions. It improves the receptive field of the segmentation model, thereby better understanding the contextual information and semantic content of the image;
- Step 5: Decoder. The feature maps from the encoder are passed to the decoder through upsampling and skip connections. The decoder gradually restores the resolution of the image and fuses it with the corresponding level of encoder features. By using deconvolution and upsampling operations in the decoder, the detailed information of the image is restored, and more accurate segmentation results are predicted by utilizing the semantic features extracted by the encoder through skip connections;

- Step 6: Segmentation probability output. The segmentation prediction generated by the decoder is usually a segmentation mask of the same size as the input image. Each pixel in the image is assigned a probability value indicating its probability of belonging to the target region;
- Step 7: Segmentation image output. Set a probability threshold for the image and set pixels with values greater than the threshold to 255 and pixels with values less than the threshold to 0. Finally, output the processed segmentation result as a segmentation mask.

3.2. Cross-Residual Encoder Module

To improve the fitting ability of a model for different types or stages of skin lesions, complex calculations are required to extract features. Although theoretically, deeper networks have stronger fitting abilities, in practice, increasing the number of layers may lead to network degradation. Shallow features cannot be fully trained, resulting in problems such as gradient vanishing and gradient explosion. Therefore, balancing network depth and training effect is necessary, and appropriate techniques should be used to address these issues to improve model performance. ResNet [33] effectively solved the problem of model degradation in deep neural networks. It introduced a residual structure that allows certain layers in the neural network to skip the next feature extraction layer and connect to other layers, thereby weakening the strong inter-layer connections. The residual structure is shown in Figure 3. The residual structure introduces skip connections, allowing the network to capture subtle differences between the input and output more effectively. As the network depth increases, each residual block can learn additional feature variations, enhancing the expressive power and modeling ability of the network. It helps the network extract richer and more discriminative feature representations, making it well-suited for complex tasks.

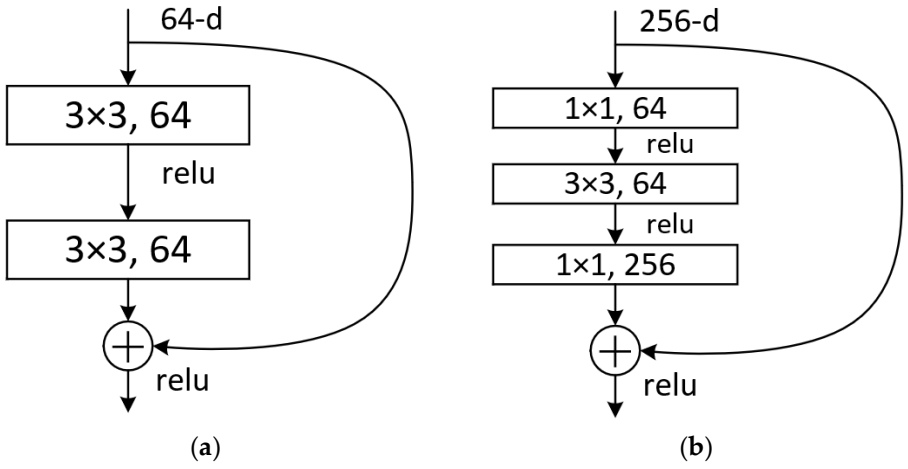


Figure 3. Two different types of residue in ResNet. (**a**) a building block for ResNet34. (**b**) a "bottleneck" building block for ResNet-50/101/152.

The UNet3+ model only used simple convolution and pooling operations during the feature extraction process. However, due to the contextual information lost during the upsampling and downsampling steps, the segmentation accuracy was poor. Despite the widespread effectiveness and popularity of the residual structure across various tasks, its performance improvement may not always be consistent in specific problems and application scenarios. Notably, when there are significant disparities in scale between input and output features, the simplistic addition operation in the residual structure may struggle to effectively align them, resulting in learning difficulties for the network. Therefore, it is crucial to carefully consider the network structure, make appropriate adjustments, and optimize it based on the specific circumstances at hand. In our approach, we introduced cross-residual connections in the encoder module to replace the original plain convolution operation, as shown in Figure 4.

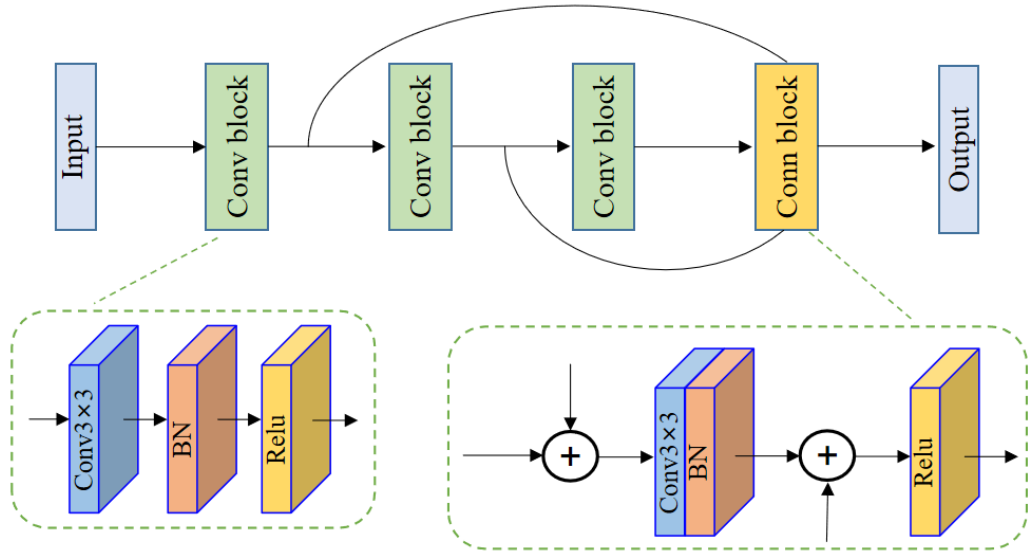


Figure 4. Cross-residual encoder module.

The module consists of three convolution blocks and one connection block. Each convolution block includes a 3 × 3 convolution layer, a batch normalization operation, and an activation function. Each connection block includes two addition operations and one convolution block. The input feature x_0 goes through three convolution blocks to obtain the feature $D_3(y_2)$, which is then combined with the output feature $D_1(x_0)$ of the first convolution block as the input of the fourth convolution layer in the connection block. Next, the output feature $D_2(y_1)$ of the second convolution block is added to the connection block and added to the feature $D_4(y_1 + y_3)$ generated by the fourth convolution operation. The final feature vector X_{out} is obtained by passing through a linear rectification unit. The calculation is shown in Equation (1):

$$\begin{cases} y_1 = D_1(x_0) \\ y_2 = D_2(y_1) \\ y_3 = D_3(y_2) \\ y_4 = D_4(y_1 + y_3) \\ X_{out} = y_2 + y_4 \end{cases} \tag{1}$$

where $D_k(x_i)$ represents the output feature of the input feature x_i by the convolutional block k. During the sampling process, the new network not only effectively preserves the shallow-level feature information of the encoder section to ensure that image details are not lost but also deepens the depth of the network's encoder, improving the network's fitting ability.

3.3. Densely Connected Atrous Spatial Pyramid Pooling Module

In skin lesion image segmentation, both global and local information are equally important. However, due to the different sizes of lesions, it may cause the feature map to shrink. To handle features at different scales, it is necessary to utilize convolutional kernels of different scales to obtain different information, and leverage multi-scale information effectively for skin lesion image segmentation. The traditional parallel ASPP structure consists of five sub-branches, each using a different convolutional kernel to process the input feature map and cascading them at the output. The structure is shown in Figure 5.

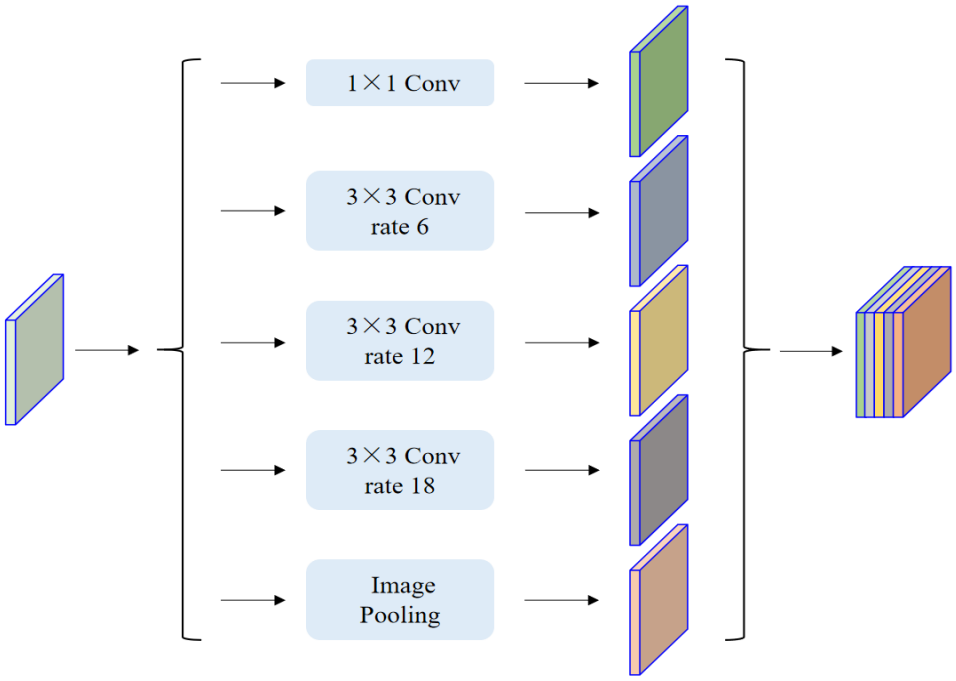


Figure 5. Structure of the ASPP network.

DenseASPP shares feature information through a cascade method that differs from the parallel connection method using atrous convolutions in ASPP. In DenseASPP, the output of each atrous convolution layer needs to be cascaded with the outputs of all previous convolution layers and input feature maps as the input of the next atrous convolution layer. Before each atrous convolution layer, 1×1 convolution is used to integrate feature information. Based on Equation (2), the output of each atrous convolution layer depends on the outputs of all previous layers. This means that layers with smaller dilation rates and those with larger rates depend on each other and form a dense feature pyramid. Through this process, DenseASPP can perceive larger contextual information and use larger filters for feedforward propagation, thus improving the performance of image segmentation tasks.

$$F_{kd_i} = D_{kd_i} C([F_{kd_{i-1}}, F_{kd_{i-2}}, \cdots, F_{kd_0}]) \tag{2}$$

where k is the size of the convolutional kernel, d indicates the expansion rate of the atrous convolution, $[\cdot]$ indicates the cascade operation, $C([\cdot])$ indicates the 1×1 convolution operation, D_{kd_i} represents the atrous convolutional layer of the layer i, and F_{kd_i} is the atrous convolution output of the layer i. Equation (3) shows the size of its receptive field. When K_1 and K_2 represent atrous convolution receptive fields with different expansion rates, respectively, the size of the convolutional kernel receptive fields after the superposition of the two atrous convolutions is shown in Equation (4).

$$K_d = k + (k-1) \times (d-1) \tag{3}$$

$$K = K_1 + K_2 - 1 \tag{4}$$

In order to solve the problem of atrous convolutions with excessive dilation rates potentially extracting many unimportant features, this paper proposes the use of four scales of atrous convolutions with dilations of 3, 5, 7, and 9. Using Equations (3) and (4), it is found that the combination of these four convolutional layers can form a structure with a receptive field size of 51. This structure extends the range of the receptive field yet does not reduce the local dimensionality, enabling the acquisition of feature information of various sizes. The structure is presented in Figure 6.

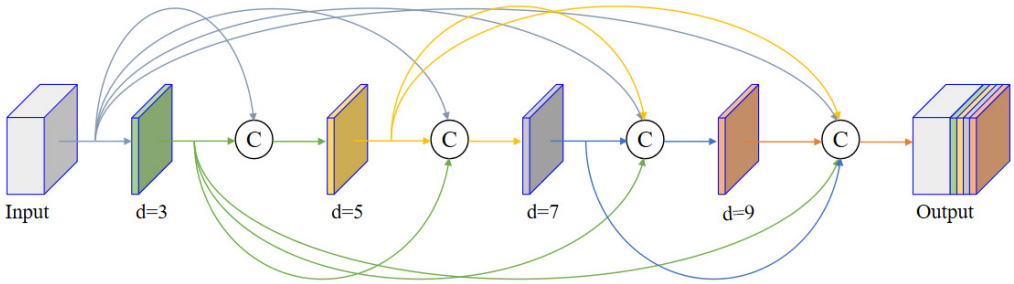


Figure 6. The structure of DenseASPP.

3.4. Efficient Channel Attention Module

In the process of extracting image features from neural networks, it is inevitable to encounter situations where some feature layers have a strong effect on classification or detection results, while others have relatively weak effects. An attention mechanism is a technology that simulates human physiological perception. When humans observe images, the visual system selects and focuses on analyzing important parts of the image. Drawing on this idea, an attention mechanism applies channel weights to feature maps and adaptively assigns different weights to different feature maps based on specific tasks, highlighting the feature maps that have a greater impact on the results and ignoring redundant information that is irrelevant to the results. In recent years, the attention mechanism has become an important component of convolutional neural networks, achieving significant performance improvements in computer vision tasks.

The SE attention [34] mechanism can adaptively recalibrate channel feature responses by explicitly modeling the interdependence between channels. ECA improves SE attention by better learning interdependencies between channels and effectively increasing performance with very few additional parameters. Specifically, ECA replaces the MLP in the excitation step of the SE block with 1D convolution and proposes a local cross-channel interaction strategy to avoid information loss. Additionally, ECA introduces an adaptive method for selecting the size of the 1D convolutional kernel, which automatically selects the optimal receptive field size based on the size and number of the current feature map. This allows ECA to achieve significant performance gains while maintaining high dimensionality and implementing cross-channel interaction, all with minimal added parameters. Figure 7 shows the architecture of the ECA. First, global pooling is performed on features from each channel without dimensionality reduction. Then, using a 1D convolutional kernel with a receptive field of k, each channel and its k nearest neighbor channels interact across channels, followed by allocating a new normalized weight to each channel via a Sigmoid activation function, and applying it to the previous feature map.

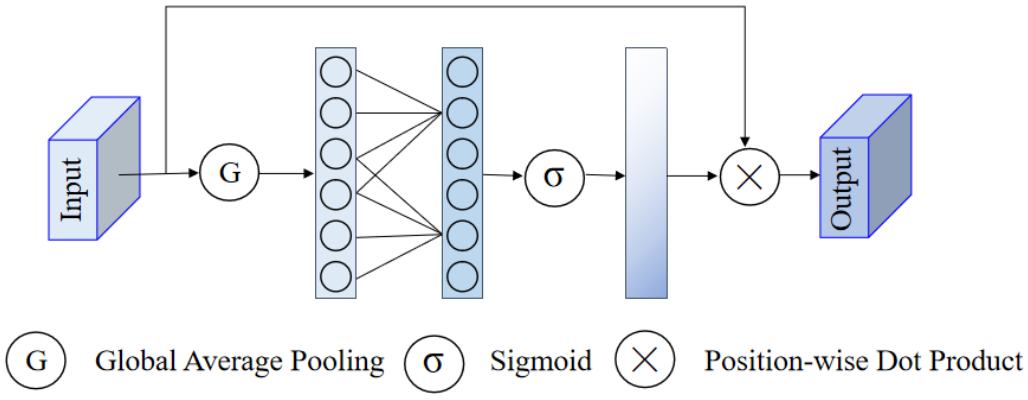


Figure 7. Efficient channel attention module.

The ECA module intends to capture local cross-channel interactions appropriately, so it utilizes a matrix W_k to learn channel attention, and it involves $k \times C$ parameters and avoids complete independence among different groups. Consider interactions between y_i and its k neighbors via sharing the same learning parameters across all channels, as shown in Equation (5):

$$w_i = \sigma(\sum_{j=1}^{k} w^j y_i^j), y_i^j \in \Omega_i^k \tag{5}$$

where Ω_i^k is the set of k adjacent channels of y_i. This approach can be implemented quickly with a 1D convolution of kernel size k, as shown in Equation (6).

$$w = \sigma(CID_k(y)) \tag{6}$$

where *CID* indicates 1D convolution. To determine the weights of the generated channels, it is necessary to first determine the coverage of the interactions, i.e., the convolution kernel is k, and k is adaptively determined by mapping the channel dimensions C, as shown in Equation (7).

$$k = \psi(C) = \left| \frac{\log_2(C) + b}{\gamma} \right|_{odd} \tag{7}$$

where $|t|_{odd}$ denotes the odd number to which t is closest. In all our experiments, we set γ and b to 2 and 1, respectively. By mapping ψ, using nonlinear mapping, the range of interactions is larger for the higher-dimensional channels and smaller for the lower-dimensional channels.

The ECA module provides a method to avoid a dimensional reduction in channel operations, which enables easy information interaction between channels and brings significant performance improvements with almost no extra computational cost. To improve the segmentation ability of the network, this study adopts the ECA module and adds attention learning after each encoder module to further optimize the network performance.

3.5. Tversky Loss Function

Currently, image segmentation algorithms generally use binary cross-entropy as the loss function. However, when dealing with the problem of extremely irregular and unbalanced foreground–background lesion areas in dermoscopy images, using cross-entropy may lead to poor network training results. Therefore, to solve the data imbalance issue in medical segmentation, the Dice loss function is often introduced. However, this function treats false positives and false negatives equally in terms of detection weight, which may result in high accuracy but low recall in practical application. To address the problem of small and highly imbalanced ROIs in skin lesion segmentation, F_N needs to be weighted higher than F_P for better recall results. Thus, this paper introduces the Tversky loss function [35] to help improve recall rates of results, as shown in Equation (8).

$$Loss = 1 - \frac{T_p}{T_p + \alpha F_P + \beta F_N} \tag{8}$$

The Tversky coefficient is a generalization of the Dice coefficient and the Jaccard coefficient, and sets α and β to 0.3 and 0.7, respectively, to adjust the weights between F_P and F_N and improve the recall when the class is out of balance.

4. Experiments and Results

The proposed model in this study was developed using the Keras 2.3.1 framework on the Windows 10 platform. We used an Intel Xeon CPU E5-2620 v4 @2.1GHz processor, 16 GB memory, and NVIDIA Quadro P600 GPU as hardware support.

4.1. Dataset and Dataset Preprocessing

Extensive experiments were performed to evaluate our proposed mode based on the skin lesion segmentation dataset released by the International Skin Imaging Collaboration

(ISIC) in 2017 and 2018 [36,37], and a wide range of experimental evaluations were conducted on the proposed models. The ISIC2018 dataset includes 2594 RGB skin dermoscopy images that have been labeled with real labels by dermatologists using grayscale binary image-labeling techniques. The label size is the same as the dermoscopy image size, where the values of lesion and non-lesion pixels are set to 255 and 0, respectively. The dataset is divided into training, validation, and testing sets in the ratio 7:1:2. By training the proposed models on the training set and evaluating their effectiveness and robustness on the testing set, we successfully demonstrated the validity of the approach. Additionally, we randomly selected 300 dermoscopy images from the ISIC2017 dataset as an additional test set, as shown in Table 1.

Table 1. Description of the datasets.

Dataset	Images	Size	Train	Valid	Test	Task
ISIC2017	2000	Variable	-	-	300	Binary Segmentation
ISIC2018	2596	Variable	1800	260	536	Binary Segmentation

To overcome the issue of limited sample numbers in medical image training data, data augmentation techniques were used to expand the number of samples and enhance the generalization performance of the models. In the pre-processing stage, hair removal processing was first performed on the images and the input images were uniformly resized to 256 × 256. To optimize the model parameters, we chose the Adam optimizer for stochastic gradient descent and the batch size was 8, with a learning rate of 1×10^{-5}, and conducted 60 rounds of training. In addition, to avoid overfitting, the learning rate became 0.8 times the original after every 10 epochs. At the same time, we adopted an early stopping mechanism; that is, if the loss did not decrease within 10 epochs, the training would automatically stop. Prior to training the network, various data augmentation operations were applied to the training set, such as image rotation with a maximum angle of 10°, random horizontal and vertical offsets of 0.05, and random image magnification and reduction of 0.05. These augmentation methods enriched the dataset and effectively improved the performance of the models.

4.2. Evaluation Metrics

In order to assess the performance of skin disease image segmentation models in clinical settings, it is common practice to use several standard metrics for evaluation. In this experiment, we chose to use five metrics: accuracy (*Acc*), Dice similarity score (*Dice*), specificity (*Spec*), sensitivity (*Sens*), and mean intersection over union (*mIoU*).

Specifically, *Acc* represents the proportion of pixels correctly classified in the segmentation result to the total pixels, *Dice* represents the degree of overlap between the predicted mask and the ground truth mask, *Spec* represents the fraction of pixels classified as negative samples among all negative samples, *Sens* represents the fraction of pixels classified as positive samples among all positive samples, and *mIoU* calculates the intercept ratio between the predicted result and the ground truth result.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

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

$$Spec = \frac{TN}{TN + FP} \tag{11}$$

$$Sens = \frac{TP}{TP + FN} \tag{12}$$

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

where *TP*, *FP*, *TN*, and *FN* represent true positive, false positive, true negative, and false negative. By calculating these five metrics for the predicted masks and ground truth masks of each case in the real dataset and comparing them with other popular models, we can evaluate the performance of the improved skin disease image segmentation model and select the best model for practical applications.

4.3. Ablation Experiments

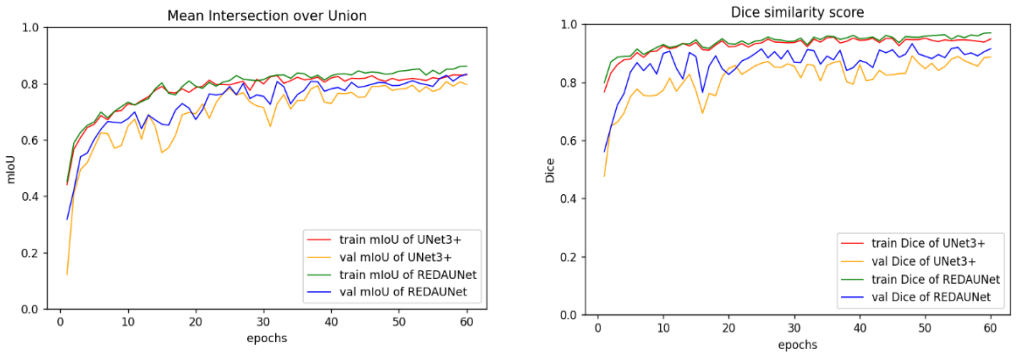
This study aims to explore the contributions of the CR module, ECA attention mechanism, and DenseASPP module to the performance of semantic segmentation networks. A baseline model based on the UNet3+ was established on the ISIC-2018 dataset. We compared the performance improvement effects of different modules with the baseline model. Table 2 summarizes the experimental results. The first comparison is between UNet3+ with residual module and UNet3+. It is observed that UNet3+ with residual module achieves better results compared with UNet3+ in terms of *Dice*, *Spec*, and *mIoU* metrics. The percentage improvements are increased by 0.89%, 3.13%, and 1.17%, respectively. However, the *Sens* value of UNet3+ with a residual module is less than that of UNet3+ because the residual module can pay more attention to features with higher relevance. Then, the second comparison is whether UNet3+ with a residual module adds an ECA attention mechanism. The results in Table 2 indicate that UNet3+ with a residual module and ECA attention mechanism achieved slightly better results by highlighting important features and ignoring irrelevant ones than only UNet3+ with a residual module in terms of the five evaluation metrics. Finally, the third comparison is whether the DenseASPP module is introduced to UNet3+ with a residual module and ECA attention mechanism. As the DenseASPP module can cover multiple ASPP modules and adapt to various lesion sizes to further improve the model performance, experimental results indicate that the proposed module with DenseASPP yields better results, and the percentage improvements of *Acc*, *Dice*, *Spec*, *Sens*, and *mIoU* metrics were 0.31%, 0.45%, 0.51%, 2.06%, and 0.59%, respectively. The results show that skin lesion segmentation performance can be significantly enhanced by improving the way of feature extraction, increasing the attention to relevant features, and acquiring different receptive field features, and the boundary of the segmentation can be made smoother, approaching or even surpassing the ground truth image.

Table 2. Ablation experiment results on the ISIC2018 dataset.

Model	CR	ECA	DenseASPP	*Acc*	*Dice*	*Spec*	*Sens*	*mIoU*
(a)	-	-	-	0.9297	0.8819	0.9341	0.9251	0.8055
(b)	√	-	-	0.9389	0.8908	0.9654	0.884	0.8172
(c)	√	√	-	0.9403	0.8957	0.9723	0.8881	0.8264
(d)	√	√	√	0.9444	0.9020	0.9774	0.9087	0.8323

For a more intuitive understanding of the segmentation effect of each model in the ablation experiment, Figure 8 displays the *mIoU* metric curve and *Dice* coefficient curve of each epoch on the training and validation sets, and Figure 9 shows the segmentation results of four models on the same dataset. Through comparative analysis, we found that the UNet3+ model has significant segmentation errors, particularly with insufficient segmentation performance for detailed parts and even under-segmentation, as shown in Figure 9a. After new residual modules were introduced on the basis of the baseline model, the edge segmentation became smoother, as shown in Figure 9b. Although this mechanism has achieved some improvements, there still exist segmentation deficiencies when the lesion area closely resembles the surrounding area, and the segmentation errors remain substantial compared with the ground truth image. Figure 9c shows that an attention mechanism was added to make the network more focused on lesion-related regions, suppress irrelevant areas, and reduce the probability of erroneous segmentation. To enhance the model's adaptability to lesions of different scales while improving its robustness, this study also includes a DenseASPP module with different expansion rates. Figure 9d illustrates that

the final segmentation boundaries have more similarity to the ground truths, i.e., the precise and true boundaries of a skin lesion. Therefore, the proposed model in this work can increase the ability to segment skin lesions more accurately and better meet practical application needs. Through ablation experiments, the results demonstrate that the proposed model is able to improve the segmentation performance and have a positive effect on lesion segmentation. In Figure 9, it can be observed that the model gradually stabilizes as the iteration number increases. The change curves of the *mIoU* and *Dice* metrics show that the proposed model significantly outperforms others in both training and validation. In addition, the magnitude of change is smaller, which indicates that the proposed model has higher accuracy and greater robustness.



(**a**) (**b**)

Figure 8. Training and validation processes on the ISIC-2018 dataset: (**a**) curves of *mIoU* metric; (**b**) curves of *Dice* metric.

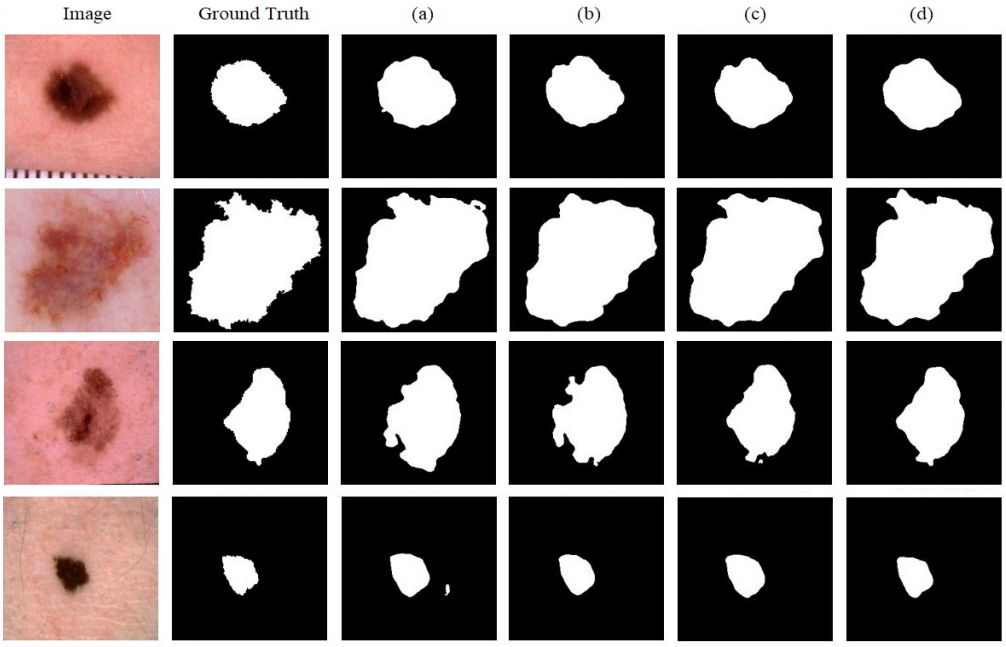


Figure 9. Ablation experiment results of model improvement. (**a**) Segmentation result of UNet3+; (**b**) segmentation result of UNet3+ with cross-residual; (**c**) segmentation result of UNet3+ with cross-residual and ECA; (**d**) segmentation result of REDAUNet.

4.4. Comparative Experiments

4.4.1. Comparison on the ISIC2018 Dataset

To validate the effectiveness of our proposed model, we employed the REDAUNet model on the ISIC2018 dataset and compared it with several other advanced semantic segmentation models, including UNet, Atten UNet, and UNeXt. For each model, we

extensively tested its segmentation performance, and the results are presented in Table 3. Additionally, Figure 10 showcased the segmentation effects more intuitively for image comparison. Based on the results from the ISIC2018 dataset, our proposed REDAUNet model exhibited superior segmentation performance on most metrics compared to the other models, especially in the *Dice* coefficient and *mIoU* metric, which are significantly higher than mainstream network performance. Furthermore, the *Dice* coefficient and *mIoU* metric are 1.75% and 2.44% higher than that of the current best UNeXt model. The Acc and Spec metrics reached 0.9444 and 0.9774, respectively, which improved by 0.07% and 0.91% compared to the second-best segmentation result, achieving excellent results. Although the score of the *Sens* metric is slightly lower than another network structure (UNet3+), it still achieved a good second-best score of 0.9087, which is 1.64% lower than the UNet3+ result. In summary, our REDAUNet model demonstrated more prominent segmentation results on the ISIC2018 dataset, proving the practical application prospects of this method in skin lesion image segmentation.

Table 3. Comparative experimental results on the ISIC2018 dataset.

Model	*Acc*	*Dice*	*Spec*	*Sens*	*mIoU*
UNet [9]	0.9301	0.8742	0.9094	0.8586	0.7786
UNet3+ [11]	0.9297	0.8819	0.9341	0.9251	0.8055
Atten UNet [12]	0.9425	0.8762	0.9605	0.8721	0.7915
UNeXt [38]	0.9437	0.8845	0.9683	0.8726	0.8079
REDAUNet	0.9444	0.9020	0.9774	0.9087	0.8323

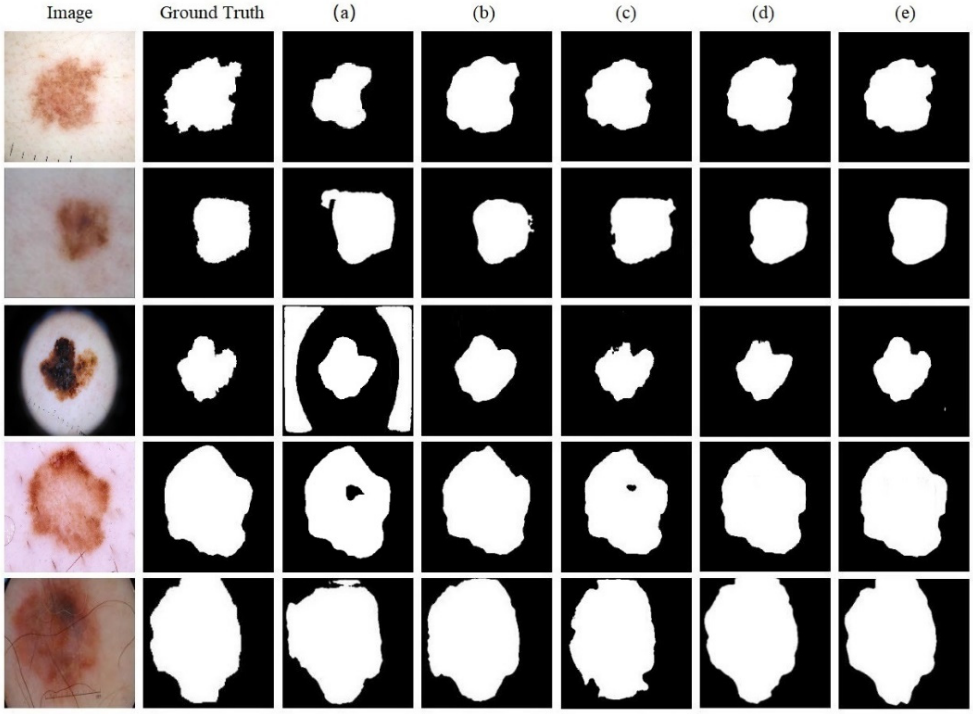


Figure 10. Comparison results on the ISIC2018 dataset. (**a**) Segmentation result from UNet; (**b**) segmentation result from UNet3+; (**c**) segmentation result from Atten UNet; (**d**) segmentation result from UNeXt; (**e**) segmentation result from REDAUNet.

To more accurately evaluate the performance of each model, we conducted a detailed analysis of the output from the selected models. The specific results are shown in Figure 10. For images with clear edges, the similarity of the results of each model to the manually annotated contours is relatively high. However, when the lesion area is close in color to the surrounding skin, the segmentation results of UNet3+, UNeXt, and the proposed model are nearer to the ground truth image, while UNet and Atten UNet models have more areas of

incorrect segmentation, exhibiting obvious differences from the real image, with prominent problems of excessive segmentation or insufficient segmentation, as shown in Figure 10a,c. Although compared to the above two models, UNet3+ and UNeXt perform better, UNet3+ lacks smooth edges for edge-blurred image segmentation results, and UNeXt's handling of irregular edge details is not as good as that of the REDAUNet model. Figure 10e shows that the suggested model can adequately capture the detailed features of the lesion edge, reduce the misclassification of noise pixels, and yield segmentation results that are more similar to the ground truth for skin lesion areas and complex lesion edge colors similar to the ground truth.

Table 4 shows the number of parameters and training duration for each segmentation method, as well as the test time for a single dermoscopic image. Both the training and testing times are measured in seconds. Due to the addition of different modules, there is a small increase in the number of parameters compared to the other models, but the segmentation of the skin lesion region can be performed in less than 1 s, which is still viable in future medical practice.

Table 4. Comparison of different model parameters and operating times.

Model	Params	Train Time/Epoch	Test Time/Single Image
UNet	36.13	41.31	0.07
UNet3+	27.01	56.5	0.14
Atten UNet	42.25	47.42	0.10
UNeXt	4.3	10.68	0.02
REDAUNet	47.77	111	0.16

As an additional comparative analysis of our method with other methods in this specific field, we included the most recent methods from the past two years of research on skin lesion segmentation for quantitative comparison, as shown in Table 5. It should be noted that the dataset used to train the model is the ISIC2018 dataset, and mIoU is considered the most important evaluation metric. Compared with other methods, our method achieves a significant improvement on the ISIC2018 dataset. By calculating *mIoU*, we found that our method outperforms other methods in skin lesion segmentation. Specifically, our method has a greater improvement in both specificity and *mIoU* at the pixel level of the lesion region, proving its better segmentation effect and differentiation ability.

Table 5. Comparison with state-of-the-art models on the ISIC2018 dataset.

Reference	Model	*Acc*	*Dice*	*Spec*	*Sens*	*mIoU*
[39]	CPFNet	0.9496	0.8769	0.9655	0.8953	0.7988
[40]	DAGAN	0.9324	0.8807	0.9588	0.9072	0.8113
[41]	FAT-Net	0.9578	0.8903	0.9699	0.9100	0.8202
[42]	CKDNet	0.9492	0.8779	0.9701	0.9055	0.8041
[43]	TMUNet	0.9603	0.9059	0.9746	0.9038	0.828
Ours	REDAUNet	0.9444	0.902	0.9774	0.9087	0.8323

4.4.2. Comparison on the ISIC2017 Dataset

We conducted experiments on the ISIC2017 dataset to test the performance of the proposed network in terms of generalization and robustness. A total of 300 randomly selected image samples were tested, and the model was compared with other advanced models. Table 6 shows the test results and Figure 11 shows the segmentation results of some lesions. Compared to others, our proposed model showed advantages in three out of five evaluation metrics, indicating higher performance in pixel-level segmentation. The REDAUNet model performed better than other comparative methods in *Dice*, *Spec*, and *mIoU* metrics, surpassing the second-best model by 1.56%, 0.39%, and 2.31%, respectively. The *Acc* and *Sens* metrics are only slightly lower than the best result, with a difference of 0.15% and 2.04%, respectively. Figure 11 shows significant differences in the segmentation

results of various models in lesion areas, especially in some results of UNet, UNet3+, and Atten UNet models, displaying issues of over-segmentation and insufficient generalizability. Although the UNeXt model has high similarity with the REDAUNet model in terms of segmentation effectiveness, their performance differs in detail. Overall, the proposed REDAUNet model shows good performance in segmenting lesions of different sizes and shapes, providing a reliable basis for its generalization performance in different datasets and demonstrating its superiority over some newer skin lesion segmentation algorithms.

Table 6. Comparison of segmentation results on the ISIC2017 dataset.

Model	*Acc*	*Dice*	*Spec*	*Sens*	*mIoU*
UNet [9]	0.9361	0.8555	0.9195	0.8563	0.7667
UNet3+ [11]	0.9419	0.8750	0.9569	0.9175	0.7983
Atten UNet [12]	0.9543	0.8745	0.9718	0.8716	0.7912
UNetXt [38]	0.9586	0.8873	0.9746	0.8723	0.8067
REDAUNet	0.9571	0.9029	0.9785	0.8971	0.8298

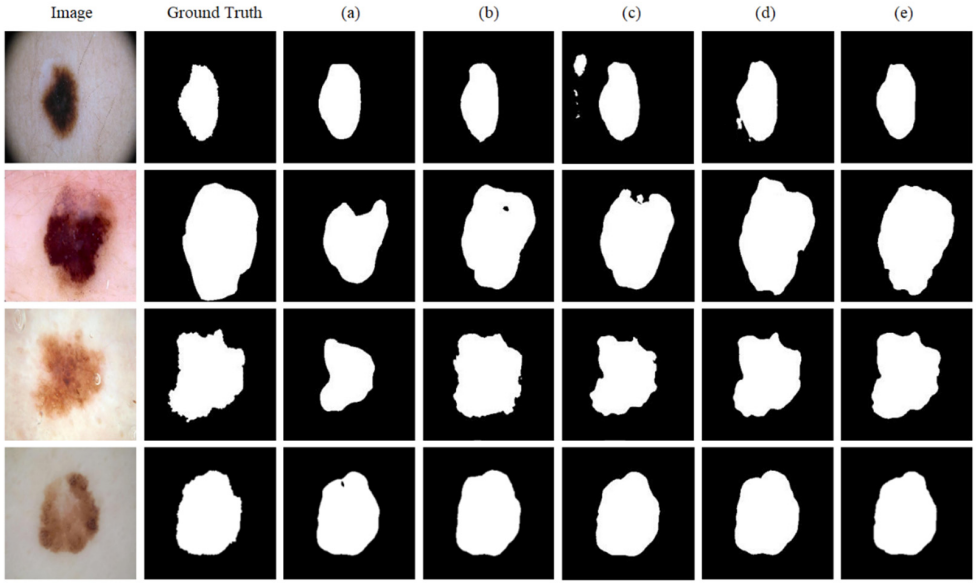


Figure 11. Comparison results on the ISIC2017 dataset. (**a**) Segmentation result from UNet; (**b**) segmentation result from UNet3+; (**c**) segmentation result from Atten UNet; (**d**) segmentation result from UNeXt; (**e**) segmentation result from REDAUNet.

This section described the performance of our proposed model in detail. In order to evaluate the effectiveness of the proposed model for skin lesion segmentation, some experiments were performed. We further improved the segmentation performance of the model by setting up ablation experiments and comparing the evaluation results and segmentation effect map of each module, while also verifying the effectiveness of the proposed model in practice. Furthermore, we evaluated the segmentation performance of the proposed model with representative segmentation models in recent years from all aspects, and conducted a comprehensive analysis of five evaluation metrics (i.e., *Acc*, *Dice*, *Spec*, *Sens*, and *mIoU*), while also comparing the predicted image effects of each model. The results show that the REDAUNet model has good performance on two datasets in terms of all the metrics. This means that the proposed model not only has good segmentation performance in skin lesion image segmentation tasks, but also has good generalization performance. In practical applications, the model can provide effective assistance in the segmentation process of computer-aided systems, achieving more satisfactory segmentation results and further improving the accuracy and efficiency of medical image processing.

We compared the most recent methods of the past two years of research on skin lesion segmentation with our results, as shown in Table 7. Compared with other methods,

our method achieved a balance in metrics on the ISIC2017 dataset. Of the seven models compared, four of the five metrics we measured ranked second, except for the Spec metric, but the Spec metric also outperformed the vast majority of the compared models. In addition, our model has a very small gap with the better-performing FAT-Net, where the mIoU metric, which is the most important among the segmentation metrics, differs by 0.66%. In contrast, our Sens metric is 0.92% higher, indicating a better differentiation ability. All in all, the good performance and segmentation effect of our network model are proved by comparing it with advanced segmentation networks.

Table 7. Comparison with state-of-the-art models on the ISIC2017 dataset.

Reference	Model	*Acc*	*Dice*	*Spec*	*Sens*	*mIoU*
[40]	DAGAN	0.9304	0.8425	0.9716	0.8363	0.7594
[41]	FAT-Net	0.9654	0.9109	0.9847	0.8879	0.8364
[44]	MCGU-Net	0.957	0.8927	0.9855	0.8502	0.8062
[45]	SegNet	0.928	0.8339	0.9222	0.9021	-
[46]	Lesion Net	-	0.8787	0.9608	0.8623	0.7828
[47]	ResUNet++	-	0.8296	-	-	80.03
Ours	REDAUNet	0.9571	0.9029	0.9785	0.8971	0.8298

5. Discussion

For medical image segmentation tasks, better segmentation results help clinical doctors with important clinical diagnosis and treatment assistance. In this study, we improved and extended the Unet3+ model to propose a semantic segmentation network called REDAUnet. We replaced the simple convolutional stacking method in the original encoder with a newly designed residual module to fully extract shallow and deep features of the images. By using the ECA attention mechanism to adaptively adjust the weights of features on each path, the model can focus more on important feature areas, thus improving the accurate segmentation capability of the target. At the same time, we introduced the DenseASPP module between the encoder and decoder to better handle the boundaries and detailed information of the target, enhancing the ability to handle complex scenes.

To achieve effective segmentation of skin lesions, our approach combines image-enhancement techniques to improve the robustness and generalization of the model under different lighting conditions. We conducted detailed experiments and comparative analyses to verify the stability and reliability of our method. In order to verify the effectiveness of each module, we first conducted an ablation experiment on the adopted module. Our Acc, Dice, Spec, and mIoU metrics are 0.9444, 0.9020, 0.9774, and 0.8323, respectively. Compared with previous models, Dice and mIoU metrics of our model increased by 2.01% and 2.68%, indicating that the introduction of modules has significantly improved the performance of the model, as shown in Table 2. We tested our model on two different medical image datasets. First, we compared our model with the other four models under the same environment, as shown in Table 3. Compared with the UNeXt model, our proposed model has a slight improvement in the four indicators, and the improvement in mIoU is 2.44%. In addition, we also compared the advanced networks in the past two years, as shown in Table 5. The proposed model achieved high scores in Spec and mIoU, 0.28% and 0.43% higher than the best model compared, respectively. Experiments on the ISIC2017 dataset show that the proposed model also has good robustness and generalization ability, as shown in Tables 6 and 7. A comprehensive analysis of the segmentation results of REDAUNet on two representative datasets shows that our method can perform well in a variety of complex cases, such as irregular lesion shapes and fuzzy boundaries. However, although the REDAUNet model performs well in semantic segmentation tasks, there are still challenges and room for improvement. First, regarding the issue of a large number of parameters in the Unet3+ model, our model did not reduce the number of parameters. On the contrary, due to the introduction of different modules, there was a slight increase in parameters, requiring higher computational resources. Second, because it contains many

layers of encoders and decoders, its prediction is relatively slow, especially in large-size image segmentation, which increases the computational complexity and time overhead. In addition, the training of the model requires a large number of labeled samples and complex preprocessing of the samples, which may be obstacles in terms of practical applications.

Overall, the REDAUNet model is an effective semantic segmentation network that can improve segmentation performance while maintaining the simplicity of the model structure. However, further optimization and improvement are still needed in the future to meet a wider range of application requirements and challenges.

6. Conclusions

This paper proposes a fusion multi-scale convolutional neural network based on the UNet3+ structure named REDAUNet. Dermoscopy images are different from the image segmentation of organs such as the brain and lungs, which have clear boundaries and fixed locations. Most skin lesions have fuzzy boundaries with the surrounding area, and especially when the shape of some lesions is complicated, it is difficult to observe the boundary position with human eyes. To address this problem, by introducing a newly designed cross-residual block to replace the simple stacked convolutional blocks in the original encoder, this module not only enhances the ability to extract deep semantic information, but also preserves shallow feature information. Simultaneously, due to the huge variation in the size of skin lesions, some dermoscopy images have less than one-tenth of the lesion area and too many irrelevant areas, an attention mechanism is added to assign a different weight coefficient to each channel so as to enhance relevant feature information and suppress irrelevant information. The DenseASPP module uses feature maps with various expansion rates to obtain multi-scale information. Compared with previous models, the results show that the proposed model has significantly improved the segmentation of skin lesions and has good anti-interference and generalization performance, particularly in terms of Dice and mIoU metrics, which increased by 2.01% and 2.68% with our model. Compared with the UNeXt model, our proposed model has a slight improvement in the four indicators, and the improvement in mIoU is 2.44%. Therefore, the CAD system based on this model has a certain reference value in the medical imaging field. Further research on the model's shortcomings is necessary. Currently, there is a wide range of research and applications for lightweight networks. By replacing the convolutional layer of the model with a separable convolution, the number of parameters and training time of the model will be significantly reduced, but the effect on the model performance needs to be further investigated. In addition, due to the variety of skin lesions, a single model segmentation has low fault tolerance and suffers from excessive bias or variance. It is also helpful to improve the model performance if the results of multiple models are integrated using an integrated model, thus reducing the model error. In the continuous development and innovation of technology, it has become possible to apply research findings to mobile applications. By integrating this research into mobile applications, it is expected to further enhance user experience and meet people's growing needs. In the future, we will further explore the application of other modules and algorithms, strengthen the research on skin lesion images, further enhance the performance and accuracy of semantic segmentation, and promote it to more medical image segmentation tasks.

Author Contributions: Funding acquisition, X.Z.; resources, X.Z.; supervision, X.Z., Y.L. and Z.X.; writing—original draft, L.L.; writing—review and editing, L.L., X.Z., Y.L. and Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Innovation and Entrepreneurship Training Program of the University of Science and Technology Liaoning (grant no. X202210146004).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this paper are public datasets. The ISIC 2017 could be found at https://challenge.isic-archive.com/data/#2017 (accessed on 25 March 2023). The ISIC2018 could be found at https://challenge.isic-archive.com/data/#2018 (accessed on 25 March 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]
2. Binder, M.; Schwarz, M.; Winkler, A.; Steiner, A.; Kaider, A.; Wolff, K.; Pehamberger, H. Epiluminescence microscopy: A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch. Dermatol.* **1995**, *131*, 286–291. [CrossRef]
3. Celebi, M.E.; Iyatomi, H.; Stoecker, W.V.; Moss, R.H.; Rabinovitz, H.S.; Argenziano, G.; Soyer, H.P. Automatic detection of blue-white veil and related structures in dermoscopy images. *Comput. Med. Imaging Graph.* **2008**, *32*, 670–677. [CrossRef]
4. Mathur, P.; Sathishkumar, K.; Chaturvedi, M.; Das, P.; Sudarshan, K.L.; Santhappan, S.; Nallasamy, V.; John, A.; Narasimhan, S.; Roselind, F.S. Cancer Statistics, 2020: Report from National Cancer Registry Programme, India. *JCO Glob. Oncol.* **2020**, *6*, 1063–1075. [CrossRef]
5. Emre Celebi, M.; Wen, Q.; Hwang, S.; Iyatomi, H.; Schaefer, G. Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Ski. Res. Technol.* **2013**, *19*, 252–258. [CrossRef]
6. Suer, S.; Kockara, S.; Mete, M. An improved border detection in dermoscopy images for density based clustering. *BMC Bioinform.* **2011**, *12*, S12. [CrossRef]
7. Abbas, Q.; Celebi, M.E.; Fondón García, I.; Rashid, M. Lesion border detection in der moscopy images using dynamic programming. *Ski. Res. Technol.* **2011**, *17*, 91–100. [CrossRef]
8. Silveira, M.; Nascimento, A.C.; Marques, J.S. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 35–45. [CrossRef]
9. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
10. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
11. Huang, H.; Lin, L.; Tong, R.; Hu, H. Unet 3+: A full scale connected Unet for medical image segmentation. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
12. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [CrossRef]
13. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. In Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28–30 July 2020; pp. 558–564.
14. Jafari, M.; Auer, D.; Francis, S.; Garibaldi, J.; Chen, X. DRU-Net: An Efficient Deep Convolutional Neural Network for Medical Image Segmentation. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1144–1148.
15. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
16. Gao, Y.; Zhou, M.; Metaxas, D.N. Utnet: A hybrid transformer architecture for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; pp. 61–71.
17. Lin, A.; Chen, B.; Xu, J.; Zhang, Z.; Lu, G.; Zhang, D. DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation. In Proceedings of the IEEE Transactions on Instrumentation and Measurement, Xi'an, China, 30 May 2022; pp. 1–15.
18. Ruan, J.; Xiang, S.; Xie, M.; Liu, T.; Fu, Y. MALUNet: A Multi-Attention and Light-weight UNet for Skin Lesion Segmentation. In Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 6–8 December 2022; pp. 1150–1156.
19. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **2020**, *121*, 74–87. [CrossRef] [PubMed]
20. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
21. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Cham, Switzerland, 6 October 2018; pp. 801–818.

22. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *3*, 834–848. [CrossRef]
23. Yang, M.; Yu, K.; Zhang, C.; Li, Z.; Yang, K. DenseASPP for Semantic Segmentation in Street Scenes. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake, UT, USA, 18–23 June 2018; pp. 3684–3692.
24. Hu, P.; Li, X.; Tian, Y.; Tang, T.; Zhou, T. Automatic Pancreas Segmentation in CT Images With Distance-Based Saliency-Aware DenseASPP Network. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1601–1611. [CrossRef] [PubMed]
25. Xu, J.; Liu, J.; Zhang, D.; Zhou, Z.; Jiang, X. Automatic mandible segmentation from CT image using 3D fully convolutional neural network based on DenseASPP and attention gates. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 1785–1794. [CrossRef] [PubMed]
26. Abraham, S.E.; Kovoor, B.C. DenseASPP Enriched Residual Network Towards Visual Saliency Prediction. In Proceedings of the International Conference on Computer Vision and Image Processing, Cham, Switzerland, 24 July 2022; pp. 85–96.
27. Li, Z.; Jiang, j.; Chen, X.; Qi, H.; Li, Q. Superdense-scale network for semantic segmentation. *Neurocomputing* **2022**, *504*, 30–41. [CrossRef]
28. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Cham, Switzerland, 6 October 2018; pp. 3–19.
29. Jiang, Y.; Cao, S.; Tao, S.; Zhang, H. Skin lesion segmentation based on multi-scale attention convolutional neural network. *IEEE Access* **2020**, *8*, 122811–122825. [CrossRef]
30. Hu, K.; Lu, J.; Lee, D.; Xiong, D.; Chen, Z. AS-Net: Attention Synergy Network for skin lesion segmentation. *Expert Syst. Appl.* **2022**, *201*, 117112. [CrossRef]
31. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.
32. Alahmadi, M.D. Multiscale Attention U-Net for Skin Lesion Segmentation. *IEEE Access* **2022**, *10*, 59145–59154. [CrossRef]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Salt Lake City, UT, USA, 29 April 2019; pp. 2011–2023.
35. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Cham, Switzerland, 7 September 2017; pp. 379–387.
36. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv* **2019**, arXiv:1902.03368.
37. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W. 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 Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 168–172.
38. Valanarasu, J.M.J.; Patel, V.M. UNeXt: MLP-Based Rapid Medical Image Segmentation Network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Cham, Switzerland, 16 September 2022; pp. 22–33.
39. Feng, S.; Zhao, H.; Shi, F.; Cheng, X.; Wang, M.; Ma, Y.; Xiang, D.; Zhu, W.; Chen, X. CPFNet: Context pyramid fusion network for medical image segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 3008–3018. [CrossRef]
40. Lei, B.; Xia, Z.; Jiang, F.; Jiang, X.; Ge, Z.; Xu, Y.; Qin, J.; Chen, S.; Wang, T.; Wang, S. Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med. Image Anal.* **2020**, *64*, 101716. [CrossRef]
41. Wu, H.; Chen, S.; Chen, G.; Wang, W.; Lei, B.; Wen, Z. FAT-Net: Feature adaptive Transformers for automated skin lesion segmentation. *Med. Image Anal.* **2022**, *76*, 102327. [CrossRef] [PubMed]
42. Jin, Q.; Cui, H.; Sun, C.; Meng, Z.; Su, R. Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. *Appl. Soft Comput.* **2021**, *99*, 106881. [CrossRef]
43. Reza, A.; Moein, H.; Yuli, W.; Dorit, M. Contextual attention network: Transformer meets U-Net. *arXiv* **2022**, arXiv:2203.01932.
44. Asadi-Aghbolaghi, M.; Azad, R.; Fathy, M.; Escalera, S. Multi-level context gating of embedded collective knowledge for medical image segmentation. *arXiv* **2020**, arXiv:2003.05056.
45. Sahin, N.; Alpaslan, N.; Hanbay, D. Robust optimization of SegNet hyperparameters for skin lesion segmentation. *Multimed. Tools Appl.* **2022**, *81*, 36031–36051. [CrossRef]
46. Nathan, S.; Kansal, P. Lesion net-skin lesion segmentation using coordinate convolution and deep residual units. *arXiv* **2020**, arXiv:2012.14249.
47. Ashraf, H.; Waris, A.; Ghafoor, M.F.; Gilani, S.O.; Niazi, I.K. Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. *Sci. Rep.* **2022**, *12*, 3948. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.