Article

DTONet a Lightweight Model for Melanoma Segmentation

Shengnan Hao [1], Hongzan Wang [1], Rui Chen [2], Qinping Liao [2], Zhanlin Ji [1,3], Tao Lyu [2,*] and Li Zhao [4,*]

1 Hebei Key Laboratory of Industrial Intelligent Perception, North China University of Science and Technology, Tangshan 063210, China; haoshengnan@ncst.edu.cn (S.H.); m13930201814@163.com (H.W.); zhanlin.ji@gmail.com (Z.J.)
2 Changgeng Hospital, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China; cra01052@btch.edu.cn (R.C.); lqpa00594@btch.edu.cn (Q.L.)
3 College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China
4 Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China
* Correspondence: lta00821@btch.edu.cn (T.L.); zhaoli@tsinghua.edu.cn (L.Z.)

Abstract: With the further development of neural networks, automatic segmentation techniques for melanoma are becoming increasingly mature, especially under the conditions of abundant hardware resources. This allows for the accuracy of segmentation to be improved by increasing the complexity and computational capacity of the model. However, a new problem arises when it comes to actual applications, as there may not be the high-end hardware available, especially in hospitals and among the general public, who may have limited computing resources. In response to this situation, this paper proposes a lightweight deep learning network that can achieve high segmentation accuracy with minimal resource consumption. We introduce a network called DTONet (double-tailed octave network), which was specifically designed for this purpose. Its computational parameter count is only 30,859, which is 1/256th of the mainstream UNet model. Despite its reduced complexity, DTONet demonstrates superior performance in terms of accuracy, with an IOU improvement over other similar models. To validate the generalization capability of this model, we conducted tests on the PH2 dataset, and the results still outperformed existing models. Therefore, the proposed DTONet network exhibits excellent generalization ability and is sufficiently outstanding.

Keywords: lightweight; segmentation; melanoma

1. Introduction

Melanomas, though perceived by some as surface issues, carry significant underlying risks. The skin, being the body's largest organ, not only plays a protective role in shielding the body from external threats but also serves crucial functions in physiological processes and the immune system. Therefore, any threat to the skin may have profound implications for overall health. The harm caused by skin diseases extends beyond their impact on appearance, affecting the quality of life and mental health of patients [1]. Chronic skin conditions may lead to prolonged pain, itching, and discomfort, significantly impairing daily life and work efficiency. Moreover, the development of some skin diseases may trigger infections and increase the risk of other illnesses. In this context, with the rapid advancement of medical image processing technology, the application of computer-aided diagnostic systems in skin cancer detection has become increasingly important. However, some large image segmentation models have limitations when processing large-scale skin cancer image data, especially in medical scenarios with limited hardware conditions. Because neural networks have advantages over traditional segmentation techniques in dealing with complex problems, we will only discuss neural network methods in this article [2].

Deep learning technology, particularly with convolutional neural networks (CNNs), has achieved remarkable success in skin cancer image segmentation tasks. Models like

U-Net [3], Unet++ [4], and Attention U-net [5], while improving in performance, also demand increasing computational resources. The current deep learning models commonly face challenges such as large parameter counts and high computational requirements, restricting their widespread adoption in practical applications. Hence, this study aims to propose a lightweight deep learning model that maintains high accuracy while reducing computational resource demands, thus enhancing its real-time capabilities. By combining deep learning techniques with lightweight strategies, we aim to provide physicians with a convenient and feasible tool to aid in the skin cancer image segmentation process.

In this paper, we will elaborate on the design and training methods of the proposed lightweight model and demonstrate its effectiveness in skin cancer image segmentation tasks through experiments. Through this research, we hope to bring new perspectives to computer-aided diagnostic technology in the medical field, offering a more reliable and efficient solution for early skin cancer diagnosis. Our approach adopts a structure similar to U-Net but introduces several innovations. Firstly, we utilize two different decoders, unlike the traditional U-shaped structure. This enables us to fully leverage the features obtained during the encoding phase, reducing the loss of information in each upsampling step. Secondly, we introduce three modules: the G_O module, G_ECA module, and ORFB module. Integrating these modules into our model allows us to maintain high accuracy with low parameter counts. Therefore, our entire network structure integrates these modules according to our proposed special structure. While maintaining high accuracy, we have reduced our parameter count to an extremely low standard, making our model less hardware-demanding and faster in terms of processing speed.

2. Background

With the rapid development of deep learning in recent years, an increasing number of sophisticated models have been widely proposed by researchers. In the field of segmentation, the initial focus was on improving accuracy, but inevitably, this led to an increase in the complexity of models. Therefore, in recent years, researchers have not only been focusing on enhancing accuracy but also optimizing the complexity of models. This initiative has greatly accelerated the process of translating research achievements into practical applications, considering that real-world scenarios may not always have high-performance hardware devices.

Researchers have proposed a variety of segmentation methods for skin lesions to improve the accuracy of the models.

Researchers led by H. Wu [6] proposed a network equipped with an adaptive dual attention module (ADAM), which was proposed to capture the boundary continuity and shape irregularity of skin lesions using global average pooling and pixel-level correlation processing. Q. Zhou [7] and colleagues proposed a superpixel-oriented label distribution learning method, which is used to generate superpixels by a simple linear iterative clustering (SLIC) algorithm, and soft tags are combined with hard tags to enhance segmentation performance. Z. Song [8] and colleagues proposed Res-CDD-Net, which used pre-trained ResNeXt50 to extract rich image features, and they introduced channels and spatial attention blocks to highlight lesion areas and used multi-scale capture blocks to deal with lesions of different sizes. VK Singh [9] and collaborators utilized improved conditional generation adversarial networks (CGans), introduced channel attention decomposition modules (Fcas) to enhance feature differentiation between lesions and non-lesions, and employed multi-scale input strategies to improve the segmentation performance. Y. Liu [10] and colleagues proposed FCP-Net, which combines a mixed loss function and game theory interaction to capture context information and fuse multi-scale features through multiple modules to improve the generalization performance. J. Ruan [11] and collaborators proposed a lightweight model consisting of four modules: DGA (dilated gated attention block), IEA (inverted external attention block), CAB (channel attention bridge block), and SAB (spatial attention bridge block) to improve the competitive skin lesion segmentation performance while reducing parameter and computational complexity. Y. Tang [12] and colleagues pro-

posed MSCGnet, which uses context information to guide feature learning and introduces context-based attention structures into the decoding path to improve the segmentation accuracy. R. Azad [13] and colleagues proposed TransCeption, which utilizes converters and U-shaped network architectures to enhance feature fusion by redesigning modules and introducing multi-scale feature extraction. J. Mu [14] and colleagues proposed M-CSAFN, which uses a multi-color space adaptive fusion network for PWS segmentation, focusing on the internal differences caused by color heterogeneity. D. Dai [15] and colleagues proposed Ms RED, which uses a multi-scale residual coding fusion module and a decoding fusion module to adaptively fuse multi-scale features and introduces multi-resolution and multi-channel feature fusion modules to enhance feature expression. S. Yang [16] and colleagues designed HMT-Net, combined it with Transformer and MLP, and utilized the attention mechanism of the CTrans module to improve the global understanding of lesions. S. Arshad [17] and colleagues proposed Dermo-Seg, which used ResNet-50 and a mixed loss function to effectively detect the differential structures of lesions. H. Yoon [18] and colleagues proposed a deep-learning-based method for the simultaneous segmentation of wrinkles and pores. Unlike color-based skin analysis, this method relies on the analysis of the morphological structure of the skin. The architecture of the model is U-Net, featuring an encoder–decoder structure. Two types of attention mechanisms were added to the network to focus on important regions. Y. Gulzar [19] and colleagues proposed ViT, which focuses on relevant parts of an image by learning long-range spatial relationships. Y. Tang [20] and colleagues proposed a framework using MS-UNet to improve performance through CIFS and deep monitoring mechanisms. J. Wu [21] and colleagues proposed C-UNet, which incorporates Inception convolutional blocks, etc., and uses Dice loss and conditional random fields for optimization. R. Iranpoor [22] and colleagues proposed ResNet101 as the backbone network in a UNet architecture for accurately segmenting lesions from the healthy parts of the skin. R. Wu [23] and colleagues proposed a Transformer-based UNet model for the complex task of segmenting psoriasis lesions from raw color images.

The following is a brief overview of the current research status on lightweight models.

MSF-Net [24] introduces spatial attention mechanisms through residual connections within convolutional blocks, focusing on key regions. Simultaneously, it incorporates multi-scale dilated convolution (MDC) modules and multi-scale feature fusion (MFF) modules to extract contextual information across scales, adaptively adjusting the receptive field size of feature maps. Yuan C [25] proposed UCM-Net, a novel, efficient, and lightweight solution that integrates multilayer perceptrons (MLPs) and convolutional neural networks (CNNs). UCM-Net has fewer than 50 KB parameters and performs at 0.05 gig operations per second (GLOP). Ijaz H [26] introduced a lightweight encoder–decoder deep learning architecture called MobileUNet and EfficientUNet. The encoders are based on the bottleneck blocks of MobileNetV2 and the MBConv blocks of EfficientNetB0, while the decoder follows the structure of the baseline UNet model. Parallel asymmetric convolution (PAC) modules [27] can also be used to replace the traditional square convolution for feature extraction, or LSR [28] can be introduced in the encoder to reduce the number of parameters. Wei S [29] proposed SRP&PASMLP-Net, which focuses on structural reparameterization and parallel axis displacement multilayer perceptrons (MLPS) for robust segmentation performance and fast inference. They introduced reparameterized multiple convolution (RDC) at an early stage to enrich the feature space. Wang Y [30] designed MAUNext, which emphasizes lightweight backbone network design and incorporates skip connections for multi-scale attention mechanisms. The core modules include multi-scale attention convolution, cooperative neighborhood attention MLP coding, and micro-skip join cross-layer semantic fusion for segmentation tasks.

3. Materials and Methods

In this chapter, we will give an introduction to our network structure, including the overall network structure, each of our innovative structures, and their combination and principle.

3.1. Overall Structure

Our model improves upon the classical encoder–decoder structure, consisting of one encoding stage and two decoding stages. The two decoders, initialized with different weights, decode the results obtained in the decoding stages, resulting in outcomes with different emphases. This significantly enhances the training speed of our model. These outcomes are further fused in the final step to obtain a result closer to the standard answer.

We proposed three modules in total and incorporated them into the new structure, forming our entire network model as shown in the Figure 1.
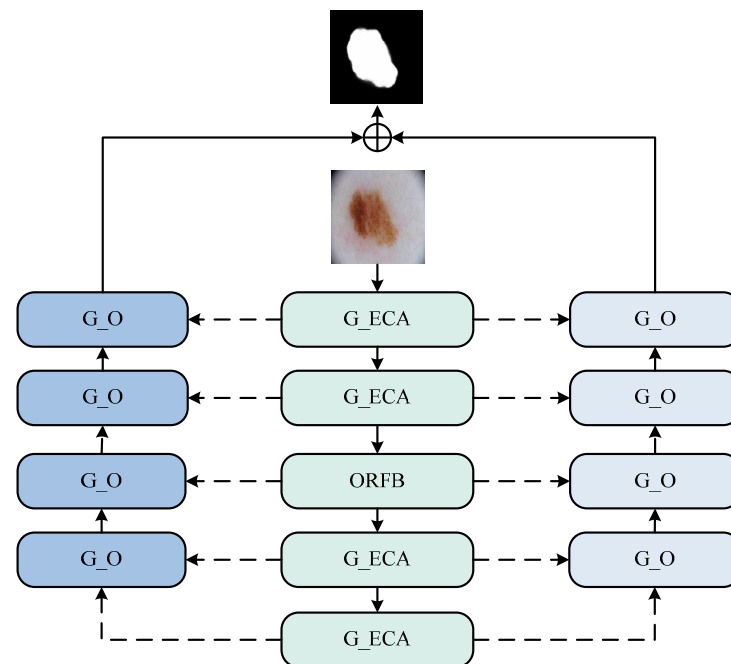

Figure 1. The DTO-Net network. (This network consists of an encoding phase and two decoding phases. The encoding phase comprises G_O (ghost octave) and ORFB (octave receptive field block) modules, while the decoding phase consists of the G_ECA (ghost efficient channel attention) module).

3.2. G_O Block

Images contain both smooth variations and detailed information. Similarly, input and output features in convolutional neural networks (CNNs) include high-frequency and low-frequency components. Low-frequency components contain redundant background information, while high-frequency components retain critical target information. To effectively handle this information, there is no need to store low-frequency features at the same resolution as high-frequency information. This way, we can suppress background interference in low-frequency features, improving the network efficiency. This is the idea behind octave convolution.

In this model, we emulate the structure of octave convolution, replacing the convolutions with ghost convolutions. This ensures a high level of information retrieval and filtering performance while further reducing the required parameter count. The module uses a hyperparameter α to control the ratio of high-frequency to low-frequency components.

The goal of this module is to enhance the convolutional neural network's (CNN) perception of information at different scales, thus improving the image recognition performance.

In traditional CNNs, convolution operations use the same kernels to process features at all scales. However, in actual images, features at different levels may have different scales. OctaveConv addresses this by introducing two sub-networks, namely, the high-frequency subnetwork and the low-frequency subnetwork, with each responsible for processing high-frequency and low-frequency information, respectively [31].

The main idea and formula of OctaveConv are as follows:

Input: Assume the input feature map is X, where X_H represents the high-frequency sub-space, and X_L represents the low-frequency sub-space.

Convolution operation: Define two sets of convolution kernels, one for the high-frequency sub-space and one for the low-frequency sub-space. Let the high-frequency convolution kernel be K_H and the low-frequency convolution kernel be K_L. The convolution operation can be expressed as follows:

- High-frequency convolution:

$$Y_H = X_H \times K_H \tag{1}$$

- Low-frequency convolution:

$$Y_L = X_L \times K_L \tag{2}$$

Subspace interaction: Downsample the high-frequency convolution result Y_H and add it element-wise to the low-frequency convolution result Y_L to obtain the final output.

- Output feature map:

$$Y = Y_L + UpSample(Y_H) \tag{3}$$

where *UpSample* represents the upsampling operation.

This design allows the network to use smaller convolution kernels when processing high-frequency information to capture detailed information while using larger convolution kernels when processing low-frequency information to obtain more global information. By introducing two sub-networks, OctaveConv effectively enhances the network's perception of multi-scale information, thereby aiding in improving the image recognition performance.

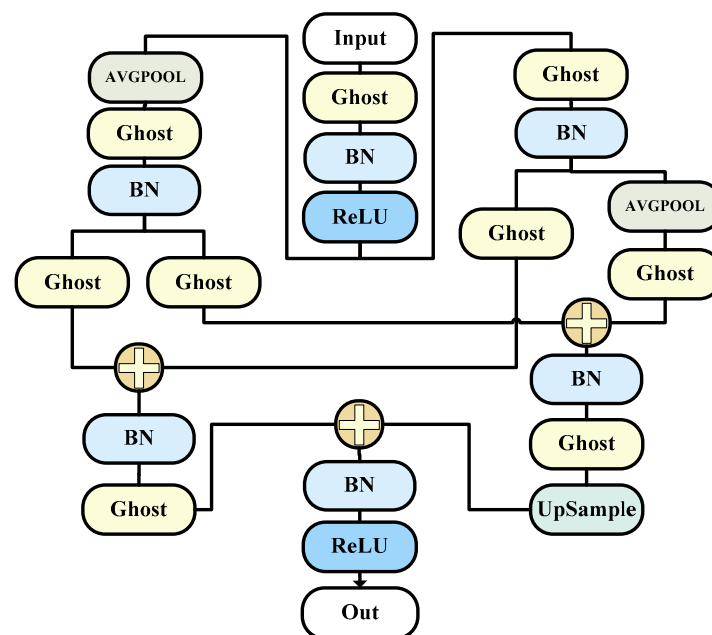The overall structure is as shown in the Figure 2.

Figure 2. The G_O model (this is the specific composition diagram of the corresponding decoder).

3.3. G_ECA Block

We further enhanced module G_O to enable it to serve as an extended encoder, extracting more information. Specifically, we fused module One with efficient channel attention (ECA) [32]. This integration enhances the overall module's perception of inter-channel correlations, facilitating a more focused utilization of specific channel information in our encoding layer. By replacing the entire module with our encoding layer, we enable different layers to improve inter-channel correlations, thereby enhancing the overall network's performance and efficiency.

The ECA module is a mechanism designed to enhance a neural network's attention to channel-wise features. The purpose of this module's design is to increase the model's focus on different channel features, allowing for a more effective capture of important information in the image. The introduction of the ECA module primarily aims to achieve advantages in computational efficiency.

The efficient channel attention (ECA) module is a lightweight attention mechanism designed to enhance a neural network's attention to channel-wise features. Here is a detailed explanation of the ECA module, including some formulaic descriptions.

The input to the ECA (efficient channel attention) module is the feature map *X*, which contains *C* channels, height *H*, and width *W*. The output is the feature map *Y* with the same dimensions.

Global average pooling: Perform global average pooling on the input feature map, obtaining the average value for each channel. Let the output of the average pooling operation be *Z*, and let the dimensions of *Z* be $C \times 1 \times 1$.

$$Z = AvgPool(X) \tag{4}$$

Here, *AvgPool* represents global average pooling.

Channel attention mapping: Input the output *Z* of the average pooling into a fully connected layer with learned weight parameters, producing channel attention mapping. Let the weight be $W_{attention}$, the bias be $b_{attention}$, and the output of channel attention mapping be:

$$A = \sigma(W_{attention} * Z + b_{attention}) \tag{5}$$

Here, σ represents the sigmoid function.

Channel-wise weighting: Multiply the channel attention mapping A with the input feature map X, weighting the features on each channel. Obtain the final output feature map Y as follows:

$$Y = A \odot X \tag{6}$$

Here, $\odot$ represents element-wise multiplication.

In summary, the ECA module enhances channel attention in a lightweight manner by introducing global average pooling, channel attention mapping, and channel-wise weighting mechanisms. This module is typically embedded in different layers of deep convolutional neural networks to improve the network's focus on different channel features, especially in cases where computational resources are limited.

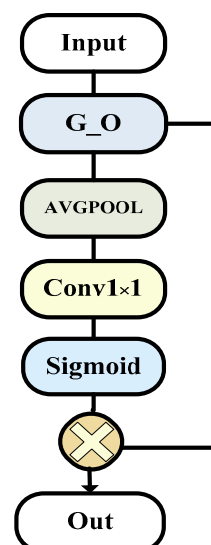The overall structure is as shown in the Figure 3.

Figure 3. The G_ECA model (this is the specific composition of encoder component module G_ECA, which is improved on the basis of G_O module to make it more suitable for encoders).

3.4. ORFB Block

Our module is also an improvement based on the G_O module, making it more capable of unleashing the maximum performance of our network. Our third module combines the features of the first module and the RFB (receptive field block) module to obtain a larger receptive field [32]. This allows our encoding layer to focus on information contained in a wider range of images and then extract it. However, considering that this module will generate a huge number of parameters, we chose to use only one instance of this module after comprehensive consideration. We placed it in the third layer of the encoding layer to ensure maximum accuracy gain while minimizing the increase in parameter count.

The RFB module is a feature extraction module used for object detection. Its design aims to enlarge the receptive field, enhance the perception of targets, and maintain a relatively low computational cost.

The RFB module is a convolutional neural network (CNN) module designed for object detection. Its purpose is to enhance feature extraction capabilities by introducing non-linear operations and multi-scale information fusion. The following provides a detailed introduction to the RFB module, including some formula explanations.

The input to the RFB (receptive field block) module is the feature map X, which contains C channels, height H, and width W. The output is the feature map Y with the same dimensions.

Local feature map processing: Firstly, perform a 1×1 convolution operation on the input feature map, generating a non-linear transformation in the channel dimension. Let the weights be denoted as $W_{non-linear}$ and the bias as $b_{non-linear}$. The output of the non-linear transformation is as follows:

$$U = RELU(W_{non-linear} * X + b_{non-linear}) \tag{7}$$

Here, $*$ represents the convolution operation, and $RELU$ denotes the rectified linear unit.

Multi-scale information fusion: Building upon the non-linear transformation, introduce convolution kernels of multiple scales to perform convolution operations on the feature map. These kernels have different receptive fields, allowing the module to operate on the feature map at different scales. Let the weights of these convolution kernels be as follows:

$$W_{scale}^i \ , i = 1, 2, \ldots, N \tag{8}$$

N is the number of scales. The output of multi-scale information fusion is as follows:

$$V = \sum_{i=1}^{n} W_{scale}^i * U \tag{9}$$

Here, $*$ denotes the convolution operation, and V is the fused feature map.

Channel attention mapping: For the fused feature map V, perform global average pooling to obtain the average value for each channel. Let the weight of the channel attention mapping be $W_{attention}$. The output of the channel attention mapping is as follows:

$$A = \sigma(W_{attention} * (AvgPool(V))) \tag{10}$$

Here, σ represents the sigmoid function.

Channel-wise weighting: Finally, multiply the channel attention mapping A with the fused feature map V, weighting the features on each channel. Obtain the final output feature map Y as follows:

$$Y = A \odot V \tag{11}$$

Here, $\odot$ represents element-wise multiplication.

In summary, the RFB module enhances the perception of targets by introducing non-linear transformation, multi-scale information fusion, and channel attention mechanisms. This module is typically embedded in deep neural networks to extract more representative features, especially achieving good performance in object detection tasks.

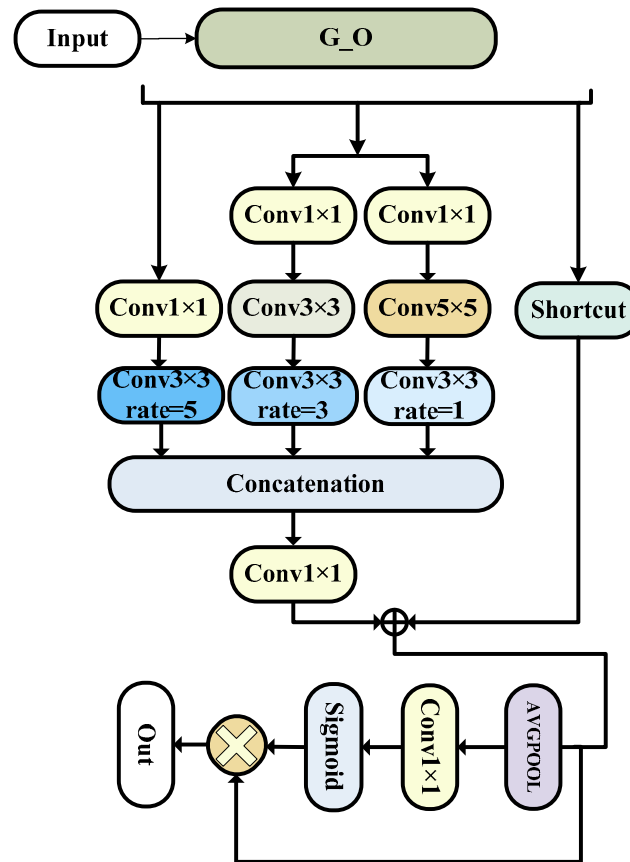The overall structure is as shown in the Figure 4.

Figure 4. The ORFB model (this is the specific composition of the encoder component module ORFB, which is improved on the basis of the G_O module, which can improve the accuracy very well, but it will produce too many parameters, so it is only used here).

4. Experiments and Results

4.1. Dataset

In this experiment, we utilized two datasets to validate the effectiveness of our model. The first dataset was ISIC2018, which is a large-scale dataset for dermatopathology research provided by the International Skin Imaging Collaboration (ISIC). We referred to the description and processing method of this dataset by Ali et al. [33], and finally decided to use one of the data enhancement methods of random HueSaturationValue, RandomBrightness, and RandomContrast for our dataset. This dataset comprises skin images from various skin diseases and conditions, including malignant melanoma (melanoma) and benign lesions. We specifically focused on the labeled images of melanoma, dividing them into training, testing, and validation sets in a ratio of 7:1:2.

The second dataset used was PH2, a dermatopathology image dataset developed by the University of Porto in Portugal. This dataset is dedicated to the study of malignant melanoma and benign lesions. PH2 includes macroscopic images from over two-hundred patients, hundreds of digital images, and corresponding pathological images. It provides high-resolution images for in-depth research into the structure and features of skin lesions.

4.2. Evaluation Metrics

In this experiment, we used several metrics to compare with other models, including the following:

IoU (intersection over union) is a metric widely used in segmentation tasks. In semantic segmentation or object detection, it measures the overlap between the region

predicted by the model and the actual region. The IOU value ranges from 0 to 1, where 1 indicates complete overlap and 0 indicates no overlap.

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

The Dice coefficient is another evaluation metric used for segmentation tasks. Similar to IOU, it measures the similarity between the region predicted by the model and the actual region. The Dice coefficient also ranges from 0 to 1, where 1 indicates a perfect match.

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{13}$$

Accuracy (ACC) measures the proportion of correctly predicted pixels among all pixels.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{14}$$

Parameter count refers to the number of weights and biases that a model needs to learn. It is a measure of the size of the model. More parameters usually indicate a more powerful model, but they also require more storage and computational resources. Having too many parameters can lead to overfitting.

Floating-point operations (FLOPs) are the total number of floating-point operations executed during the forward pass of a model. FLOPs serve as a metric for measuring the computational complexity of a model. They are crucial for understanding the computational resources required for inference. Lower FLOPs generally indicate a lighter model, suitable for inference in resource-constrained environments.

4.3. Experimental Setup

In this study, Python version 3.9 was used, and the Windows 10 operating system was used on a computer equipped with a 12th-generation Intel® Core™ i5-13490 CPU and an NVIDIA GeForce RTX 3060 (12 GB memory) (From Pinduoduo, purchased in Tangshan, China).

We set the initial learning rate to 1×10^{-4}, with a minimum learning rate of 1×10^{-5}. The batch size was set to 4, and the number of epochs was 500. The input image dimensions were 256×256, and we used the BCEDiceLoss loss function.

4.4. Results and Analysis

4.4.1. Melanoma Segmentation Comparison with Advanced Models with ISIC2018

To evaluate the performance of our model, we compared it with several mainstream models in the market. We adopted a method of testing our model by randomly inputting images ten times and taking the average to ensure the absence of random occurrences. Subsequently, we obtained Tables 1 and 2. Table 1 presents the results of testing on the ISIC2018 dataset, while Table 2 demonstrates the validation on the PH2 dataset to showcase the generalization performance of our model.

Table 1. Segmentation results of advanced models.

| Models	Parameters	GFLOPS	IoU	Acc	Dice
UNet	7.770	13.780	0.8169	0.9576	0.8838
U-Net++	9.160	34.900	0.8203	0.9587	0.8881
Attention-UNet	8.730	16.740	0.8217	0.9588	0.8863
UNeXt_s [34]	0.300	0.100	0.8057	0.9557	0.8895
MALUNet [11]	0.175	0.083	0.8120	0.9532	0.8924
LW-IRSTNet [35]	0.161	0.301	0.8216	0.9588	0.8854
Ours	0.030	0.126	0.8284	0.9607	0.8845

Table 2. Segmentation results of advanced models in PH2 dataset.

Models	Parameters	GFLOPS	IoU	Acc	Dice
UNet	7.770	13.780	0.8062	0.9276	0.8916
U-Net++	9.160	34.900	0.7929	0.9238	0.8831
Attention-UNet	8.730	16.740	0.8102	0.9303	0.8716
UNeXt_s	0.300	0.100	0.8077	0.9277	0.8900
MALUNet	0.175	0.083	0.8278	0.9351	0.9048
LW-IRSTNet	0.161	0.301	0.8327	0.9386	0.8944
Ours	0.030	0.126	0.8347	0.9388	0.8914

As shown in the table above, it can be observed that the parameter count of our model was significantly lower than all the other models, standing at only 0.03. Although the floating-point operation count was not the lowest, it still ranked among the top. Moving on to performance metrics, our model achieved the highest IOU, leading the second-best by 0.0067. The accuracy was also significantly ahead, reaching a maximum of 0.9607. Additionally, the Dice coefficient was at a medium-to-high level.

Therefore, from the data shown in this table, we can conclude that our model not only ranked first among all the models in terms of the accuracy of segmentation but also had the lowest number of parameters, which can further reduce the hardware requirements of our model and make it more widely applicable.

The Figure 5 demonstrate a comparative view between our model's predicted results and the actual images. For ease of observation, we overlayed them together.
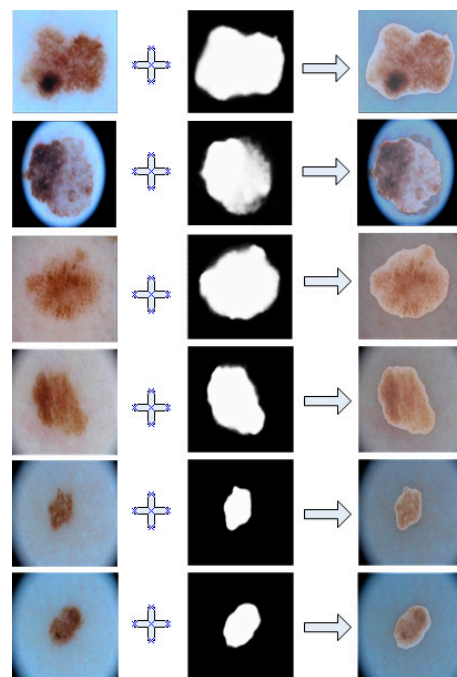
Figure 5. Overlayed image of segmentation results and ground truth in ISIC2018.

As shown in this Figure 5, we can see that the segmented results of our model had a very high coincidence with the real lesions, which indicates be very accurate segmentation of the diseased area.

4.4.2. Melanoma Segmentation Comparison with Advanced Models with PH2

Looking at the performance of our model on the PH2 dataset, it is evident that our IOU and accuracy remained the highest, and the Dice coefficient was also at a leading level. This further validates the generalization performance of our model, demonstrating its ability to achieve excellent results on different datasets.

The Figure 6 shows the performance of our model on the PH2 dataset, showing that the performance was still stable and excellent.
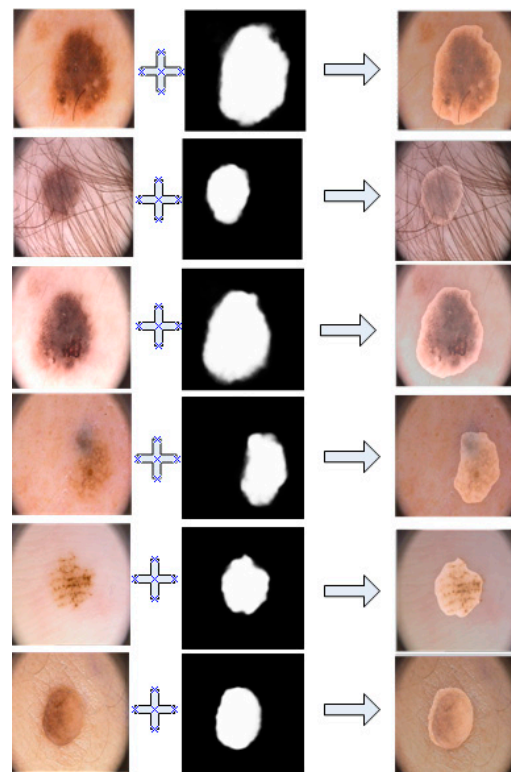

Figure 6. Overlay image of segmentation results and ground truth in PH2.

As shown in the figure above, through our coincidence technique, it is obvious that our segmentation results overlapped perfectly with the lesion area.

In order to better show the segmentation effect of each network, we deliberately captured some of the results and formed a graph with real pictures and the gold standard in Figure 7. From this figure, we can see that each of our models could segment the rough shape, which is enough to prove the excellent performance of the neural network, but in terms of details, it is obvious that our model had better processing performance and was closer to the gold standard.

4.4.3. Cross-Domain Generalization Validation—Breast Ultrasound Dataset

To test the performance of our model in terms of cross-domain generalization, we tested it on a breast ultrasound dataset. The results were compared with those of some other models. The results are shown in Table 3.

Table 3. Segmentation results in breast ultrasound dataset.

Models	IoU	Acc	Dice
UNet	0.6859	0.9642	0.7913
U-Net++	0.6861	0.9625	0.7900
Attention-UNet	0.6875	0.9636	0.7948
Ours	0.6901	0.9633	0.7910

From the data in this table, we can see that the accuracy of our model for breast cancer segmentation is only comparable to that of the classical model in terms of cross-domain generalization. Even though our model is small in terms of the number of parameters, we still have a big gap compared with other advanced models in this field. Therefore, in terms of cross-domain generalization, our model still has huge room for improvement. The

future improvement goal is to create a module that can capture different features while keeping the number of parameters low to deal with this problem.
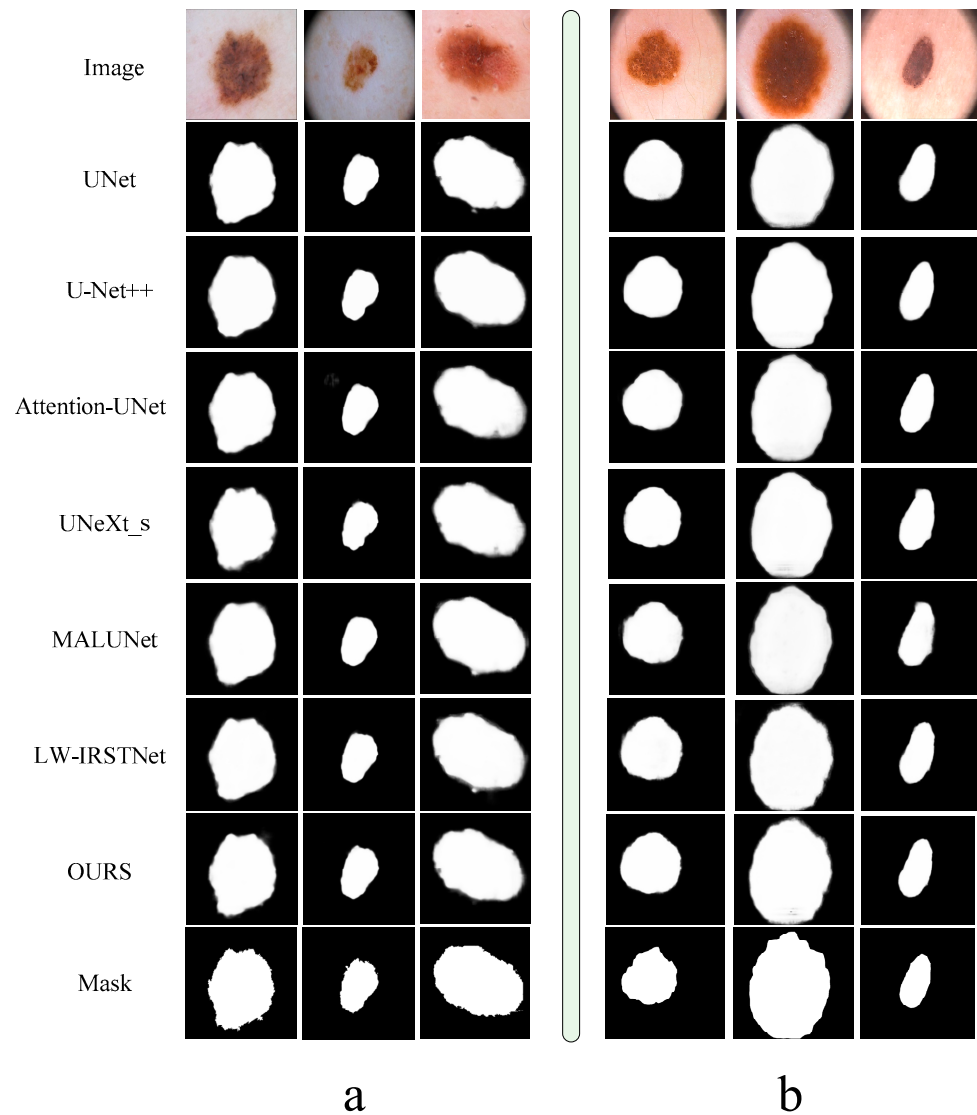

Figure 7. Comparison of segmentation results of each model. The left part (**a**) is the presentation of results in the ISIC2018 dataset, and the right part (**b**) is the presentation of segmentation results in the PH2 dataset.

5. Discussion

Through our research, we successfully proposed and developed a lightweight model focused on the task of melanoma segmentation. The model maintains high accuracy while demonstrating excellent computational efficiency, making it suitable for real-time segmentation in scenarios with limited computing resources. By cleverly introducing lightweight modules, we effectively reduced the model's parameter count and computational burden, making it a viable solution widely applicable in embedded devices and mobile applications.

In various dataset evaluations, our lightweight model showed outstanding performance and demonstrated strong generalization capabilities across different data distributions. This indicates that our model has strong adaptability and can robustly perform melanoma segmentation tasks in diverse clinical scenarios. This achievement provides significant support for advancing the early detection of melanoma in the healthcare sector, with the potential to offer patients more timely and effective treatment.

Future research directions will focus on further improving the lightweight performance of the model to meet the requirements of embedded devices and mobile applications. We plan to achieve an ideal balance between performance and efficiency by delving into the balance of network structure, parameter count, and computational complexity. This will provide valuable experience and guidance for the development of lightweight models in medical image segmentation tasks.

References

1. Zhang, X. Melanoma segmentation based on deep learning. *Comput. Assist. Surg.* **2017**, *22* (Suppl. S1), 267–277. [CrossRef] [PubMed]
2. Ming, G. Analysis and discussion of skin lesion image segmentation methods. *J. Fujian Comput.* **2024**, *40*, 24–29. [CrossRef]
3. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18. Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
4. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018; Proceedings 4. Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
5. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
6. Wu, H.; Pan, J.; Li, Z.; Wen, Z.; Qin, J. Automated skin lesion segmentation via an adaptive dual attention module. *IEEE Trans. Med. Imaging* **2020**, *40*, 357–370. [CrossRef] [PubMed]
7. Zhou, Q.; He, T.; Zou, Y.J.D. Superpixel-oriented label distribution learning for skin lesion segmentation. *Diagnostics* **2022**, *12*, 938. [CrossRef]
8. Song, Z.; Luo, W.; Shi, Q.J.E. Res-CDD-net: A network with multi-scale attention and optimized decoding path for skin lesion segmentation. *Electronics* **2022**, *11*, 2672. [CrossRef]
9. Singh, V.K.; Abdel-Nasser, M.; Rashwan, H.A.; Akram, F.; Pandey, N.; Lalande, A.; Presles, B.; Romani, S.; Puig, D. FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *IEEE Access* **2019**, *7*, 130552–130565. [CrossRef]

10. Liu, Y.; Zhou, J.; Liu, L.; Zhan, Z.; Hu, Y.; Fu, Y.; Duan, H. FCP-net: A feature-compression-pyramid network guided by game-theoretic interactions for medical image segmentation. *IEEE Trans. Med. Imaging* **2022**, *41*, 1482–1496. [CrossRef] [PubMed]
11. Ruan, J.; Xiang, S.; Xie, M.; Liu, T.; Fu, Y. (Eds.) 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.
12. Tang, Y.; Fang, Z.; Yuan, S.; Xing, Y.; Zhou, J.T.; Yang, F.J.I.A. imscgnet: Iterative multi-scale context-guided segmentation of skin lesion in dermoscopic images. *IEEE Access* **2020**, *8*, 39700–39712. [CrossRef]
13. Azad, R.; Jia, Y.; Aghdam, E.K.; Cohen-Adad, J.; Merhof, D.J. Enhancing Medical Image Segmentation with TransCeption: A Multi-Scale Feature Fusion Approach. *arXiv* **2023**, arXiv:2301.10847.
14. Mu, J.; Lin, Y.; Meng, X.; Fan, J.; Ai, D.; Chen, D.; Qiu, H.; Yang, J.; Gu, Y. M-CSAFN: Multi-color Space Adaptive Fusion Network for Automated Port-wine Stains Segmentation. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3924–3935. [CrossRef] [PubMed]
15. Dai, D.; Dong, C.; Xu, S.; Yan, Q.; Li, Z.; Zhang, C.; Luo, N. Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation. *Med. Image Anal.* **2022**, *75*, 102293. [CrossRef] [PubMed]
16. Yang, S.; Wang, L.J.S. HMT-Net: Transformer and MLP Hybrid Encoder for Skin Disease Segmentation. *Sensors* **2023**, *23*, 3067. [CrossRef] [PubMed]
17. Arshad, S.; Amjad, T.; Hussain, A.; Qureshi, I.; Abbas, Q.J.D. Dermo-Seg: ResNet-UNet Architecture and Hybrid Loss Function for Detection of Differential Patterns to Diagnose Pigmented Skin Lesions. *Diagnostics* **2023**, *13*, 2924. [CrossRef] [PubMed]
18. Yoon, H.; Kim, S.; Lee, J.; Yoo, S.J.D. Deep-Learning-Based Morphological Feature Segmentation for Facial Skin Image Analysis. *Diagnostics* **2023**, *13*, 1894. [CrossRef] [PubMed]
19. Gulzar, Y.; Khan, S.A.J.A.S. Skin lesion segmentation based on vision transformers and convolutional neural networks—A comparative study. *Appl. Sci.* **2022**, *12*, 5990. [CrossRef]
20. Tang, Y.; Yang, F.; Yuan, S. A multi-stage framework with context information fusion structure for skin lesion segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 1407–1410.
21. Wu, J.; Chen, E.Z.; Rong, R.; Li, X.; Xu, D.; Jiang, H. (Eds.) Skin lesion segmentation with C-UNet. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2785–2788.
22. Ali, R.; Hardie, R.C.; Narayanan, B.N.; De Silva, S. (Eds.) Deep learning ensemble methods for skin lesion analysis towards melanoma detection. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019; pp. 311–316.
23. Izadi, S.; Mirikharaji, Z.; Kawahara, J.; Hamarneh, G. (Eds.) Generative adversarial networks to segment skin lesions. In Proceedings of the 2018 IEEE 15Th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 881–884.
24. Shao, D.; Ren, L.; Ma, L.J.B. MSF-Net: A Lightweight Multi-Scale Feature Fusion Network for Skin Lesion Segmentation. *Biomedicines* **2023**, *11*, 1733. [CrossRef]
25. Yuan, C.; Zhao, D.; Agaian, S.S.J. UCM-Net: A Lightweight and Efficient Solution for Skin Lesion Segmentation using MLP and CNN. *arXiv* **2023**, arXiv:2310.09457.
26. Ijaz, H.; Sultan, H.; Altaf, M.; Waris, A. (Eds.) Embedded Skin Lesion Segmentation using Lightweight Encoder-Decoder Architectures. In Proceedings of the 2023 3rd International Conference on Artificial Intelligence (ICAI), Wuhan, China, 17–19 November 2023; pp. 176–181.
27. Ma, Y.; Wu, L.; Gao, Y.; Gao, F.; Zhang, J.; Luo, Z. ULFAC-Net: Ultra-Lightweight Fully Asymmetric Convolutional Network for Skin Lesion Segmentation. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 2886–2897. [CrossRef]
28. Lu, C.; Xu, H.; Wu, M.; Huang, Y. (Eds.) IESBU-Net: A Lightweight Skin Lesion Segmentation UNet with Inner-Module Extension and Skip-Connection Bridge. In Proceedings of the International Conference on Artificial Neural Networks, Crete, Greece, 26–29 September 2023; pp. 115–126.
29. Wei, S.; Chen, H.; Zhao, J.; Zhou, Y.; Yang, Y. SRP&PASMLP-Net: Lightweight skin lesion segmentation network based on structural re-parameterization and parallel axial shift multilayer perceptron. *Int. J. Imaging Syst. Technol.* **2023**, *34*, 22985.
30. Wang, Y.; Wang, J.; Zhou, W.; Liu, Z.; Yang, C.J.P.i.M.; Biology. MAUNext: A lightweight segmentation network for medical images. *Phys. Med. Biol.* **2023**, *68*, ad0a1f. [CrossRef] [PubMed]
31. Chen, Y.; Fan, H.; Xu, B.; Yan, Z.; Kalantidis, Y.; Rohrbach, M.; Yan, S.; Feng, J. (Eds.) Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October 2019–2 November 2019; pp. 3435–3444.
32. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. (Eds.) ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
33. Ali, R.; Hardie, R.C.; De Silva, M.S.; Kebede, T.M.J. Skin lesion segmentation and classification for ISIC 2018 by combining deep CNN and handcrafted features. *arXiv* **2019**, arXiv:1908.05730.

34. Valanarasu, J.M.J.; Patel, V.M. (Eds.) Unext: Mlp-based rapid medical image segmentation network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; pp. 23–33.
35. Kou, R.; Wang, C.; Yu, Y.; Peng, Z.; Yang, M.; Huang, F.; Fu, Q. LW-IRSTNet: Lightweight infrared small target segmentation network and application deployment. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 1–13. [CrossRef]