Article

GSCEU-Net: An End-to-End Lightweight Skin Lesion Segmentation Model with Feature Fusion Based on U-Net Enhancements

Shengnan Hao [1], Haotian Wu [1], Yanyan Jiang [1], Zhanlin Ji [1,2], Li Zhao [3], Linyun Liu [1,*] and Ivan Ganchev [2,4,5,*]

[1] Hebei Key Laboratory of Industrial Intelligent Perception, North China University of Science and Technology, Tangshan 063210, China; haoshengnan@ncst.edu.cn (S.H.); wuhaotian@stu.ncst.edu.cn (H.W.); 13933873356@163.com (Y.J.); zhanlin.ji@ncst.edu.cn (Z.J.)
[2] Telecommunications Research Centre (TRC), University of Limerick, V94 T9PX Limerick, Ireland
[3] Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China; zhaoli@tsinghua.edu.cn
[4] Department of Computer Systems, University of Plovdiv "Paisii Hilendarski", 4000 Plovdiv, Bulgaria
[5] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria
* Correspondence: liujy23@ncst.edu.cn (L.L.); ivan.ganchev@ul.ie (I.G.)

Abstract: Accurate segmentation of lesions can provide strong evidence for early skin cancer diagnosis by doctors, enabling timely treatment of patients and effectively reducing cancer mortality rates. In recent years, some deep learning models have utilized complex modules to improve their performance for skin disease image segmentation. However, limited computational resources have hindered their practical application in clinical environments. To address this challenge, this paper proposes a lightweight model, named GSCEU-Net, which is able to achieve superior skin lesion segmentation performance at a lower cost. GSCEU-Net is based on the U-Net architecture with additional enhancements. Firstly, the partial convolution (PConv) module, proposed by the FasterNet model, is modified to an SConv module, which enables channel segmentation paths of different scales. Secondly, a newly designed Ghost SConv (GSC) module is proposed for incorporation into the model's backbone, where the Separate Convolution (SConv) module is aided by a Multi-Layer Perceptron (MLP) and the output path residuals from the Ghost module. Finally, the Efficient Channel Attention (ECA) mechanism is incorporated at different levels into the decoding part of the model. The segmentation performance of the proposed model is evaluated on two public datasets (ISIC2018 and PH2) and a private dataset. Compared to U-Net, the proposed model achieves an *IoU* improvement of 0.0261 points and a *DSC* improvement of 0.0164 points, while reducing the parameter count by 190 times and the computational complexity by 170 times. Compared to other existing segmentation models, the proposed GSCEU-Net model also demonstrates superiority, along with an advanced balance between the number of parameters, complexity, and segmentation performance.

Keywords: skin lesion segmentation; convolutional neural network; lightweight; attention mechanism

1. Introduction

Skin diseases are diverse, with the most severe and potentially fatal consequence being skin cancer [1], which is a result of a malignant tumor that originates from different tissues and cell types of the skin. It is one of the most common types of cancer and is primarily caused by factors such as excessive exposure to ultraviolet radiation, genetic factors, immune system issues, and carcinogenic substances. According to data from the World Health Organization (WHO), over 3 million people worldwide are diagnosed with non-melanoma skin cancer annually [2], while melanoma affects over 270,000 individuals each year. Basal cell carcinoma is the most common type of non-melanoma skin cancer, accounting for approximately 80% of all skin cancer cases, while melanoma is the deadliest form of skin cancer.

Early diagnosis of skin cancer is of paramount importance. However, timely and accurate analysis of skin lesions is equally crucial, as they might serve as precursors to skin disorders, including malignant melanoma and other skin conditions. Skin lesion segmentation involves the process of delineating the areas of pathology from normal skin regions in medical images, aiding in early detection, diagnosis, and treatment. This task is essential for treatment planning and monitoring disease progression, as well as supporting medical research and education. Therefore, precise skin lesion segmentation not only enhances medical efficiency and patient recovery rates, but also equips medical professionals with valuable tools to better manage skin health.

Currently, skin lesion segmentation methods can be broadly classified into two categories, [3]. The first category consists of traditional machine learning-based image segmentation methods [4], such as thresholding [5,6], region growing [7,8], edge detection [9], and region splitting and merging [8]. However, traditional medical image segmentation methods have limitations. They can be sensitive to noise in images, which may result in discontinuous and inaccurate segmentation results. Additionally, their performance may be limited when dealing with complex textures, shadows, overlapping structures, or blurred boundaries. The second category involves deep learning-based methods. With the advancement of deep learning, significant progress has been made in medical image segmentation. Deep learning-based methods have shown stronger adaptability, robustness, and accuracy compared to traditional machine learning-based methods.

Most existing deep learning models are based on the U-Net model [10], which has an encoder–decoder architecture, due to its simplicity and scalability. Many improved models have been proposed, such as U-Net++ [11], Attention-UNet [12], V-Net [13], and Recurrent Residual U-Net (R2U-Net) [14]. However, previous works still face some challenges. Firstly, previous research tends to introduce more complex modules into U-Net to achieve better performance. However, due to limited memory on mobile medical devices, many models with a large number of parameters cannot be effectively applied in real clinical scenarios. Secondly, medical image segmentation is a layout-specific task [15], where the differences between samples are small, but the differences within each sample can be significant in medical datasets. Based on these considerations [16], this paper proposes a lightweight segmentation model, named GSCEU-Net (https://github.com/1194449282/GSCEU-Net (accessed on 9 August 2023)), which is based on U-Net and utilizes multiple lightweight modules in both the encoding and decoding parts. Extensive experiments, conducted on the ISIC2018 [17] and PH2 [18] public datasets, and a private dataset, demonstrate that the proposed model achieves excellent performance in image segmentation and is also highly competitive in terms of lightweight performance.

The main contributions of this paper are reflected in four aspects:

1. The proposed GSCEU-Net model adopts the overall U-shaped encoding–decoding structure of U-Net [10]. However, by reducing the number of channels and by incorporating newly designed Separate Convolution (SConv) and Ghost SConv (GSC) modules, along with an Efficient Channel Attention (ECA) module [19], it is able to attain a light weight, which is reflected in the reduction of the number of model parameters and floating-point operations performed;

2. A newly designed SConv module is proposed as a technical advancement derived from the recent FasterNet's partial convolution (PConv) [20] with additional improvements. It upgrades the PConv replication path to a 1×1 convolution path and dynamically calculates the input channel numbers, thereby extracting spatial features from image regions and accelerating the model training convergence;

3. The upgraded path after SConv convolution is further connected with the Ghost module residuals [21] to form a newly designed GSC module. Multi-Layer Perceptron (MLP) [22] is utilized to absorb hidden layer features, and DropPath is applied to refine features and prevent model overfitting, thus further enhancing the model's generalization ability.

4. The decoding part of the proposed model utilizes the ECA attention mechanism, allocating model weights with minimal complexity cost, so as to optimize the decoding process.

Table 1 shows the abbreviations that are frequently used further on in the paper, along with their full names and descriptions.

Table 1. Frequently used abbreviations with their full names and descriptions.

Abbreviation	Full Name	Description
ECA	Efficient Channel Attention	A lightweight attention mechanism
GSC	Ghost Separate Convolution	A backbone network for feature fusion
MLP	Multi-Layer Perceptron	A neural network with input, hidden, and output layers.
PConv	Partial Convolution	A lightweight partial convolution proposed by FasterNet
SConv	Separate Convolution	An enhanced convolution proposed in this paper for feature extraction
U-Net	U-Net	A convolutional neural network model commonly used for image segmentation

2. Related Work

2.1. Medical Image Segmentation

With the advancement of convolutional neural networks (CNNs) [23] in medical image analysis and processing, deep learning-based segmentation methods have become a hot research topic, due to their ability to automatically learn image features and overcome the limitations of manual feature extraction in traditional methods. One typical end-to-end deep network for image segmentation is the fully CNN (FCN) [24]. FCN uses deconvolutional layers to upsample [25] the feature maps from the last convolutional layer, restoring them to the same size as the input image. This allows for generating predictions for each pixel while preserving the spatial information from the original input image. At the end, the upsampled feature maps are pixel-wise classified to achieve the final image segmentation. Based on FCN, Ronneberger et al. designed a U-Net network [10] specifically for biomedical images, which has been widely applied in medical image segmentation since its introduction. Due to its excellent performance, U-Net and its variants have been extensively used in various subfields of computer vision. U-Net is suitable for medical image segmentation because its architecture combines both low-level and high-level information. The low-level information helps improve accuracy, while the high-level information aids in extracting complex features, [26]. UNeXt [27], proposed by Valanarasu et al., combines MLP [22] with U-Net to achieve reduced parameter count while maintaining segmentation performance. Considering the importance of lightweight models in practical applications and mobile health, this study builds upon previous research and introduces various attention modules to ensure high-performing and efficient medical image segmentation. Recently, Ruan et al. proposed MALUNet [16], a network model that incorporates four attention blocks: Channel Attention Bridge (CAB) block, Spatial Attention Bridge (SAB) block, Dilated Gated Attention (DGA) block, and Inverted External Attention (IEA) block. These attention blocks enable the model to acquire multi-stage and multi-scale information, and significantly reduce the number of channels.

2.2. Skin Lesion Segmentation

Recently, the integration of deep learning and machine learning in the context of skin disease segmentation has gained attention. In 2020, Khan et al. proposed a novel automated system for skin lesion localization and recognition [28], which combines the concepts of deep learning and IcNR-based feature selection and follows three main steps: rapid lesion localization through region-based convolutional neural networks (R-CNN) [29], deep feature extraction, and feature selection using the IcNR method. In 2022, Manzoor et al.

introduced a lightweight skin disease detection approach [30] involving optimal feature fusion. These authors initially segmented images using CNN segmentation and applied filter operations after segmentation to eliminate the noise. Statistical feature extraction employed principal component analysis and the Gray-Level Co-occurrence Matrix (GLCM) algorithm [31], supplemented by deep feature extraction through the AlexNet transfer learning method. ABCD [32] rule features were also extracted. In 2023, Zafar et al. [33] compiled and organized skin disease detection from the perspectives of both machine learning and deep learning. Their study was compared with other studies, highlighting its novelty and comprehensiveness. They outlined the main steps of computer-aided skin cancer diagnosis systems, such as preprocessing, segmentation, feature extraction, selection, and classification. While relevant machine learning algorithms can post-process deep learning results and demonstrate advantages through further feature fusion, such integration may increase the complexity of the entire process. Capable of automatic feature learning from data, the GSCEU-Net model, proposed in this paper, reduces the need for manual feature engineering. Additionally, the GSCEU-Net model's training data comprise only 200 KB, making it suitable for deployment on micro devices, achieving high levels of lightweight efficiency.

2.3. Attention

The attention mechanism [34] is a commonly used method in deep learning to assign different weights to different input information. These weights have the flexibility to be adapted in various scenarios, rendering attention mechanisms exceptionally versatile and resilient.

In the realm of medical image segmentation, Oktay et al. introduced Attention-UNet [12], an innovative network incorporating Attention Gates (AG) for the processing of medical images. It calculates similarity scores between different inputs and performs weighted aggregation based on these similarity scores. The lightweight Squeeze-and-Excitation (SE) attention mechanism [35] is used to enhance model representation capacity. The SE mechanism was initially introduced in image classification tasks and has been widely adopted in some lightweight models to improve their performance. Its core idea is to learn a channel attention weight vector to weight the feature maps of the input. Through the squeeze and excitation steps, the importance of each channel in the input feature maps is reweighted, enhancing the representation and learning capability of crucial information. Woo et al. proposed the Convolutional Block Attention Module (CBAM) [36]. The CBAM attention mechanism combines spatial attention and channel attention by element-wise multiplication of the two. This operation is applied to the input feature maps, resulting in the output of the CBAM attention mechanism. The model can weight the feature maps based on channel- and spatial-attention weights, thereby enhancing the representation and learning capability of important information. Although CBAM belongs to lightweight attention mechanisms, it introduces some additional parameters, which increases the model's parameter count and computational complexity compared to a single attention model.

2.4. Lightweight CNNs

Lightweight CNNs were developed to meet the requirements of applications with limited resources, high real-time demands, and energy efficiency, aiming to create small, fast, and energy-saving models to reduce costs and meet practical needs. In 2018, Mehta et al. introduced the ESPNet network [37], specifically designed for semantic segmentation tasks. It employs a hierarchical structure comprising a lightweight encoder and a multi-scale decoder, while utilizing techniques such as depth-wise separable convolution, point-wise convolution, and spatial pyramid pooling to efficiently capture multi-scale features and strike a balance between computational cost and performance. It is suitable for mobile devices and embedded systems. In 2019, Wu et al. proposed an adaptable lightweight neural network architecture named FBNet [38], aiming to deliver high performance with

low computational cost. It employs lightweight operations and "ghost modules" to reduce parameter count while maintaining model efficacy. FBNet's flexible design enables customization, based on diverse tasks and resource requirements, making it applicable to various image-processing applications.

In 2020, Han et al. proposed GhostNet [21], which aimed to strike a balance between model efficiency and accuracy by introducing the concept of "ghost" modules. The key idea of GhostNet is to enhance feature-extraction efficiency by incorporating ghost modules into the network. Specifically, a Ghost module employs a technique called "ghost operation" that divides the input feature map into two parts, one of which is called the "ghost feature map". Lightweight convolution operations are applied to the ghost feature map to extract additional features at a lower computational cost.

To produce even faster networks, a new technique called partial convolution (PConv) was introduced in FasterNet [20] in 2023. PConv, shown in Figure 1, minimizes unnecessary computations and memory access to enhance the extraction of spatial features with greater efficiency. It achieves significantly higher running speeds on various devices without compromising the accuracy of various visual tasks. For example, on ImageNet-1k, the smaller FasterNet-T0 outperforms MobileVitXXS [39] by 3.1 times on GPU, 3.1 times on CPU, and 2.5 times on ARM processors, with an accuracy improvement of 2.9%.
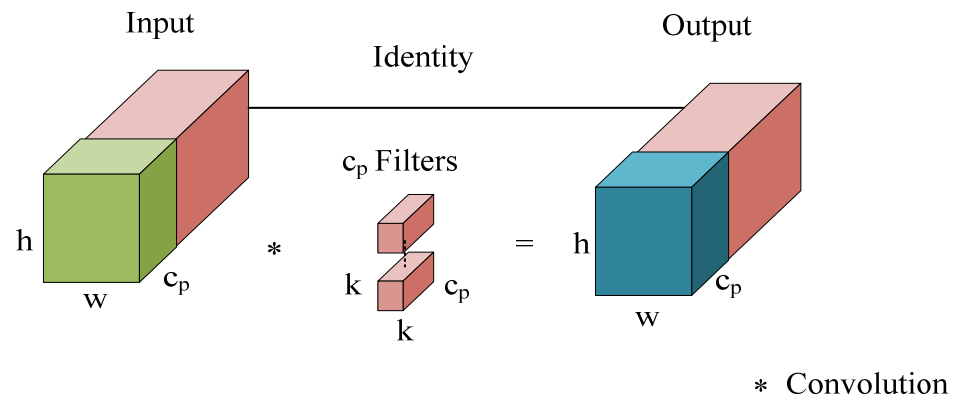

Figure 1. The PConv structure.

3. Proposed Model: GSCEU-Net

3.1. Overall Structure

Figure 2 illustrates the overall structure of the proposed GSCEU-Net model. The model takes RGB skin lesion images as an input and produces a single-channel black-and-white predicted lesion image as an output. GSCEU-Net adopts a U-Net structure [10], consisting of an encoding part and a decoding part. The encoding part gradually reduces the size and number of feature maps, while the decoding part restores the feature maps to their original size through upsampling and skip connections. The presence of skip connections allows the network to utilize high-resolution features (from the encoding part) in the decoding part, thus effectively preserving target boundaries and details. In the proposed model, the GSC module is employed as a key component in the encoding part. The GSC module utilizes 1×1 and 3×3 convolutions to process the input feature maps, extracting and integrating features at different scales. In the decoding part, the Efficient Channel Attention (ECA) mechanism [19] is combined with the GSC module. The ECA mechanism effectively captures channel interactions and enhances important information in the feature maps. After the decoding stage, a 1×1 convolution is employed to map the feature results with four channels to a single-channel predicted image. This design preserves the overall U-Net structure [10], guaranteeing its powerful segmentation capability. Furthermore, by combining lightweight backbone networks and lightweight attention mechanisms, the proposed model represents the most representative features of the image using fewer parameters, which affects positively the model's complexity. Through this design, the proposed model can fully leverage the advantages of U-Net,

while at the same time being able to improve its efficiency and performance, and maintain lightweight characteristics. This enables its better adaptation to the requirements of skin lesion image-segmentation tasks.
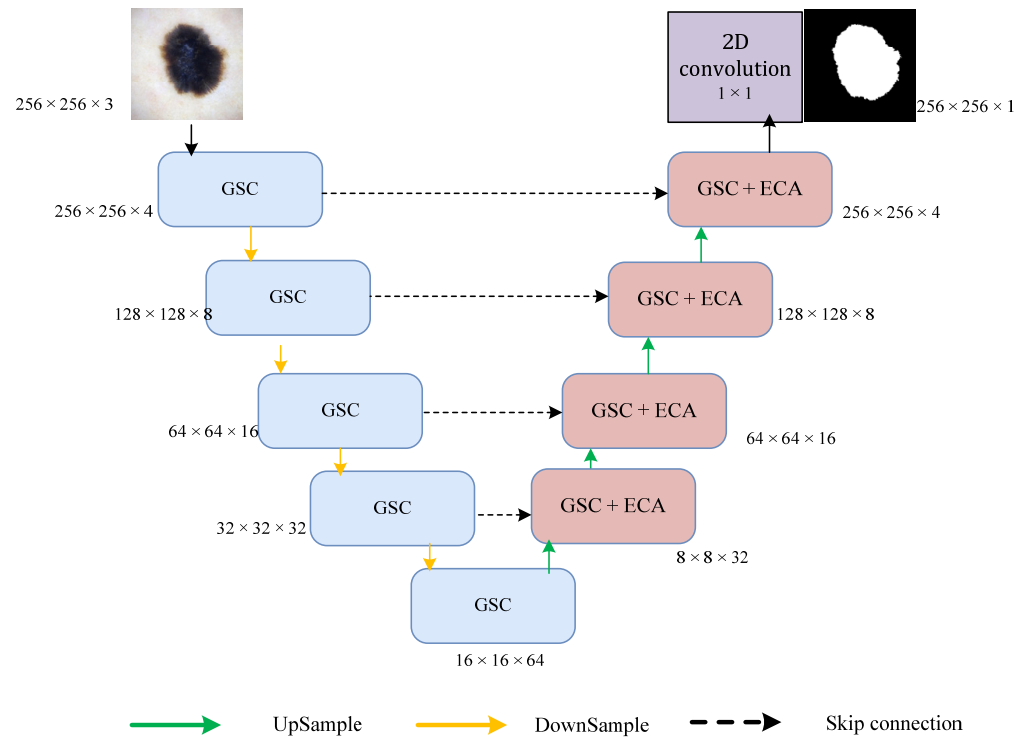

Figure 2. The structure of the proposed GSCEU-Net model.

3.2. Separate Convolution (SConv)

The U-Net architecture [10] comprises a contraction pathway and an expansion pathway, corresponding to increasing and decreasing feature-map depths, i.e., the number of channels. However, our idea of incorporating a PConv module [20] into U-Net is not directly realizable due to the use of a different number of input and output layers. Therefore, a modified PConv module, called the Separate Convolution (SConv) module, is proposed here, as shown in Figure 3 (taking the contracting path as an example, with increasing output channels, and choosing a split value of four for the input channels). A 3 × 3 2D convolution is applied to one quarter of the feature maps, while for the remaining three-quarters, instead of directly concatenating them like in the original PConv, a 1 × 1 2D convolution is performed, followed by concatenation of the results. Finally, the output channels are split, and the processed feature results are mapped to either an enlargement or reduction. By sliding a 3 × 3 convolutional kernel at different positions, the designed SConv module can effectively capture image edges, textures, and other local features, which helps the proposed GSCEU-Net model learn low-level and mid-level features of the images. The 3 × 3 convolution also reduces the number of parameters while maintaining a larger receptive field, enhancing the network's ability to recognize larger skin lesions. The use of a 1 × 1 convolution changes the channel numbers of the feature maps, allowing for decreased model complexity. Additionally, performing convolutions at each position of the feature map allows for the extraction of local spatial features, speeding up the convergence of model training. The SConv operation can be expressed as:

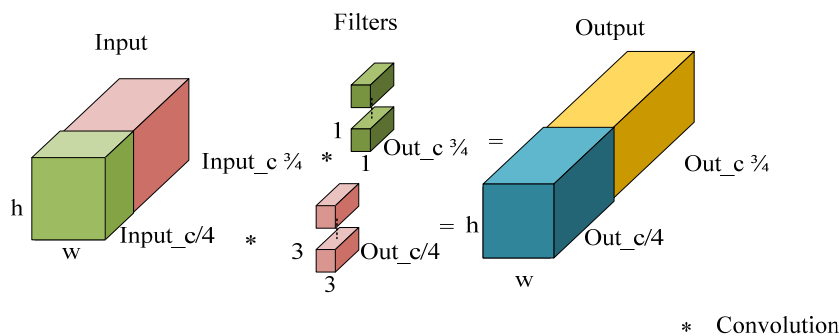$$Y_{SConv} = Cat\{Conv2d_{3\times3}(Split(X/4)); Conv2d_{1\times1}(Split(X\,3/4))\} \tag{1}$$

Figure 3. The structure of the proposed SConv module.

3.3. Ghost Separate Convolution (GSC)

Serving as a foundational component for the proposed GSCEU-Net model, the newly designed GSC module consists of two paths—left and right, as depicted in Figure 4. The left path incorporates a Ghost module [21], which includes two convolutions. The first convolution utilizes a 1×1 kernel, followed by a batch normalization (BN) [40] and a Rectified Linear Unit (ReLU) [41] activation to obtain low-dimensional feature maps. The second convolution employs a 3×3 grouped kernel, followed by BN and ReLU, to enrich the output feature maps from the first convolution. Finally, the results of the two convolutions are concatenated. The right path, consisting of a SConv module followed by a multilayer perceptron (MLP) [22] and a DropPath, which randomly removes excess parameters to prevent model overfitting. The outputs of the left and right paths are added together to produce the final output of the GSC module. The advantage of the Ghost module is that it splits one convolutional layer in a conventional deep neural network into two parts, using fewer parameters to generate more features. This module implements a residual connection network structure at a low computational cost. Both the Ghost module and SConv module have low computational complexity while preserving the respective features. The operation of the Ghost module can be expressed as:

$$Y_{\text{Ghost\_1}} = ReLU(BN(Conv2d_{1\times1}(X))) \tag{2}$$

$$Y_{\text{Ghost}} = Cat\big\{ ReLU\big(BN\big(Group\_Conv2d_{3\times3}(Y_{\text{Ghost\_1}})\big)\big); Y_{\text{Ghost\_1}}\big\} \tag{3}$$
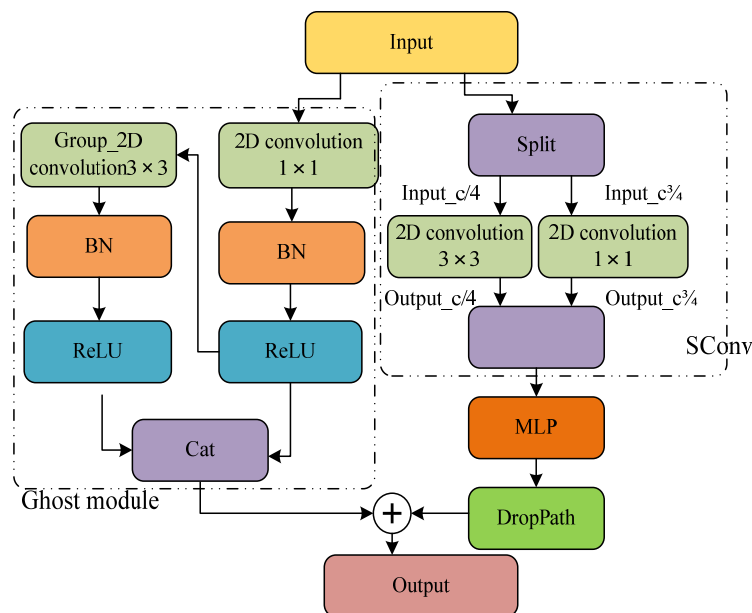

Figure 4. The structure of the proposed GSC module.

The operation of the GSC module can be expressed as:

$$Y_{\text{GSC}} = Y_{\text{Ghost}} + DropPath(MLP(Y_{SConv})) \tag{4}$$

3.4. Efficient Channel Attention (ECA)

ECA [19] is a crucial element in deep learning models for enhancing their representation and performance. It addresses the challenge of dimension reduction by preserving the original dimensions and effectively capturing channel-wise interactions. The ECA module, shown in Figure 5, operates by performing channel-wise global average pooling without dimension reduction. This operation allows the module to capture local cross-channel interactions by considering each channel and its neighboring channels. Specifically, it applies convolutional operations to each channel of the input feature map and normalizes the results using a sigmoid function to generate channel attention weights. These weights determine the importance of each channel in the feature map. Finally, the weighted feature representation is obtained by multiplying the channel attention weights with the input feature map in the channel dimension. Experimental results have demonstrated that the ECA module offers efficiency and effectiveness in model training. It introduces minimal additional parameters and incurs negligible computational overheads, while achieving significant performance gains. Moreover, the design of the ECA module allows for dynamic adjustment of the convolutional kernel values, providing flexibility and adaptability. In the GSCEU-Net model proposed in this paper, a kernel size of $k = 3$ is selected to strike a balance between model complexity and performance. Overall, the ECA module effectively preserves the feature dimensions in the decoding stage, captures channel-wise interactions, and enhances the performance of the proposed model. Its operation can be represented as follows:

$$Y = \sigma(Conv2d_{1\times1}(AvgPool(X))) \otimes X \tag{5}$$

where σ denotes the activation function and $\otimes$ denotes the matrix multiplication.
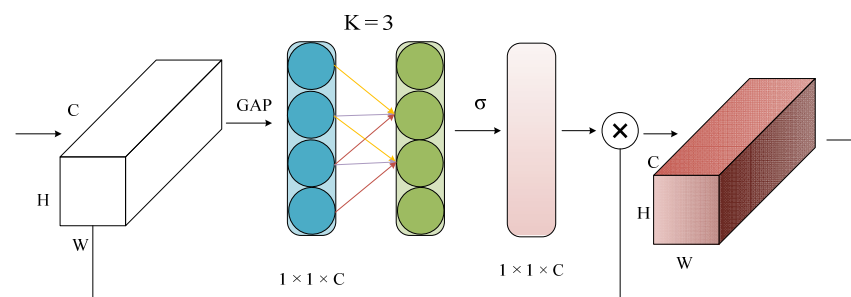

Figure 5. The structure of the ECA module.

4. Experiments and Results

4.1. Datasets and Data Preprocessing

The International Skin Imaging Collaboration Challenge (ISIC2018) public dataset [17], the PH2 public dataset [18], and a private dataset were used in the experiments to both train and assess the performance of the proposed GSCEU-Net model in comparison to other existing segmentation models. Presently, ISIC2018 holds the distinction of being the world's most extensive collection of skin lesion images. It offers meticulously annotated digital images of skin lesions, effectively advancing the development of Computer-Aided Diagnosis (CAD) systems targeted at melanoma and various other forms of skin cancers. The PH2 public dataset was collaboratively gathered by Pedro Hispano Hospital in Matosinhos, Portugal, and the dermatology department at the University of Porto. The private dataset, originating from Peking Union Medical College Hospital, encompasses skin lesion images portraying conditions such as acne and lupus erythematosus.

The ISIC2018 public dataset comprises 2594 skin microscopy images accompanied by segmentation mask labels. For conducting the experiments, this dataset was partitioned

into training, validation, and test subsets, utilizing a ratio of 7:1:2, respectively. Before commencing the model training, we randomly selected one-third of the training images and programmatically simulated additional random body hair. Moreover, throughout the training phase, we applied various operations such as horizontal and vertical flipping, random adjustments to brightness, Gaussian blurring, mean smoothing filtering, and random hue saturation to the ISIC2018 training subset (as depicted in Figure 6). It is important to note that none of these supplementary operations were applied to the validation and test subsets.
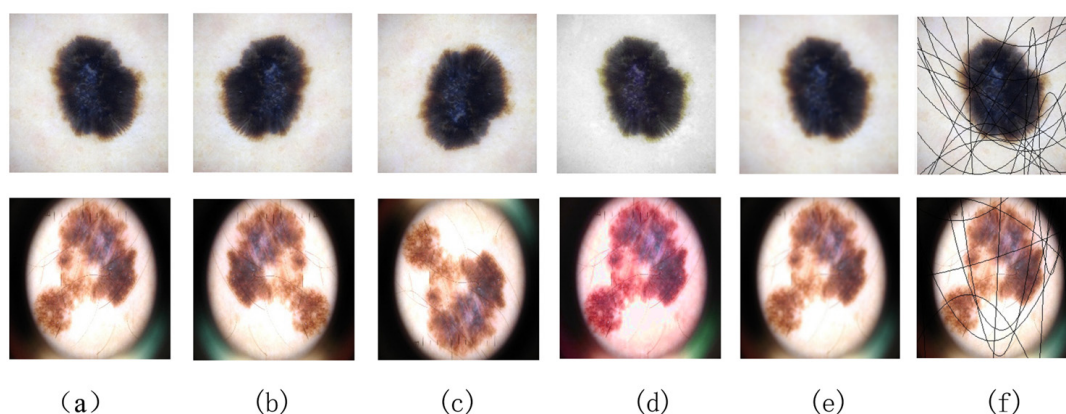

(**a**)   (**b**)   (**c**)   (**d**)   (**e**)   (**f**)

Figure 6. Image preprocessing operations: (**a**) Original image; (**b**) Horizontal flip; (**c**) Vertical flip; (**d**) Hue and saturation adjustment; (**e**) Gaussian blur; (**f**) Introducing random body hair.

The ISIC2018 public dataset is widely used in the field of dermatology and skin cancer research. It consists of high-resolution dermoscopy images collected from various sources, including different anatomical sites and a wide range of skin conditions. Therefore, it is the most compelling dataset among the three datasets used in the experiments. Consequently, the ablation study experiments, outlined in Section 4.5.4, were conducted solely on this dataset.

Consisting of only 200 images, the PH2 dataset was employed as an auxiliary testing set to evaluate the models trained on the ISIC2018 dataset. Encompassing 1010 images, the private dataset was randomly partitioned into training, validation, and test subsets using an 8:1:1 ratio for experimentation purposes.

Details of the utilized datasets are shown in Table 2.

Table 2. Datasets: sizes, splitting, and image resolution.

Dataset	Total Size (Images)	Size of Training Set (Images)	Size of Validation Set (Images)	Size of Testing Set (Images)	Image Resolution (Pixels)
ISIC2018 [17]	2594	1816	259	519	256×256
PH2 [18]	200			200	256×256
Private dataset	1010	808	101	101	256×256

4.2. Experimental Environment

The experiments were conducted by utilizing PyTorch version 1.12.1 [42], in conjunction with Python version 3.10.6, and operating on Ubuntu 22.04. All trials were executed on a computer equipped with a 12th Gen Intel® Core™ i5-12400 CPU, 16 GB of RAM, and an NVIDIA GeForce RTX 3060 with 12 GB VRAM. The training process extended across 500 epochs. To optimize our model, we employed the SGD optimizer [43] with an initial learning rate of 1×10^{-2}, weight decay set to 1×10^{-4}, momentum set to 0.9, and batch size set to 8.

The determination of the number of epochs, which represents complete training iterations on a dataset, strikes a balance between model convergence and avoiding model overfitting. We opted for a value of 500 epochs (c.f. Figure 7). To prevent rapid oscillations and instability, an initial learning rate of 0.01 was chosen, based on empirical observations. Furthermore, a gradually decaying learning-rate schedule was employed to achieve finer convergence towards the end of training. A batch size of 8 was selected to strike a balance between efficient gradient computation and memory utilization. A smaller batch size results in increased parameter variance, whereas a larger batch size strains memory resources.
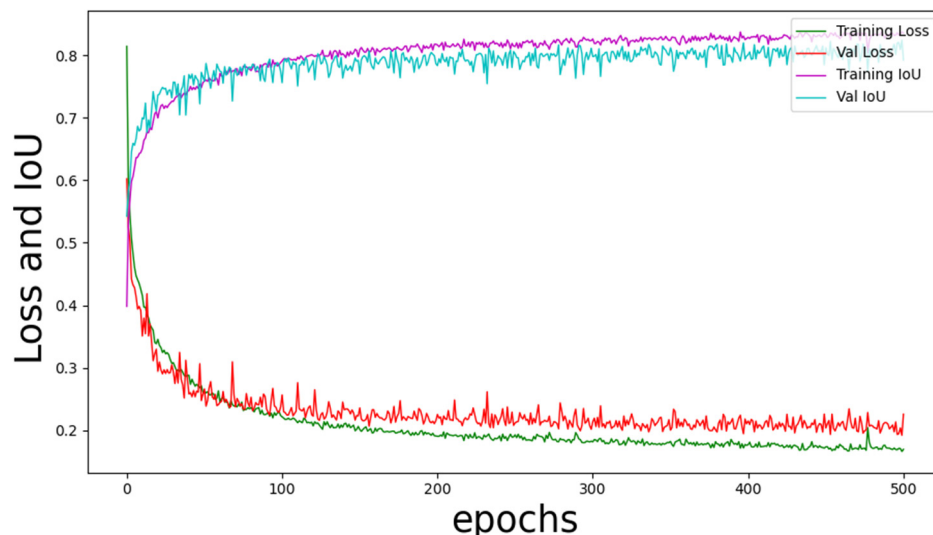

Figure 7. The training process of GSCEU-Net on ISIC2018 dataset.

4.3. Evaluation Metrics

In the experiments, four commonly used evaluation metrics were utilized to measure the segmentation performance of the compared models, namely the Intersection over Union (*IoU*), Dice Similarity Coefficient (*DSC*), *accuracy*, and *sensitivity*. Among these, the first two are the main evaluation metrics used in image segmentation tasks.

IoU, also referred to as the Jaccard index, stands as a prevalent metric within the realm of semantic segmentation. *IoU* quantifies the degree of overlap between the predicted segmentation and the ground truth, divided by the area encompassing their union. In the experiments, its computation takes the form of:

$$IoU = \frac{TP}{TP + FP + FN} \tag{6}$$

where *TP* (true positives) denotes the count of accurately classified pixels as belonging to an entity (a skin lesion in our context), *FN* (false negatives) corresponds to the number of inaccurately classified pixels as not belonging to an entity, and *FP* (false positives) indicates the count of inaccurately classified pixels as belonging to an entity.

DSC stands as the most extensively employed metric for assessing the performance of image segmentation models. Its computation involves doubling the intersection area between the predicted segmentation and the ground truth, then dividing it by the sum of pixels in both sets, as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{7}$$

Accuracy (*Acc*) is used to evaluate the overall pixel-level segmentation performance, calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where *TN* (true negative) signifies the count of pixels accurately recognized as not belonging to an entity.

Sensitivity (*Sen*) indicates the ratio of accurately segmented skin lesion pixels, calculated as follows:

$$Sen = \frac{TP}{TP + FN} \tag{9}$$

Additionally, in the experiments, the parameter count (represented in millions, M) and computational complexity (measured in billion floating point operations per second, GFLOPS) were also used to compare the models. Parameter count refers to the number of learnable parameters of a model. Large parameter counts generally result in higher memory usage. Using GFLOPS as an estimation of the computational resources required by a model, an inference can be made for the model's efficiency across different hardware platforms.

4.4. Loss Function

In the experiments, a composite loss function was used, particularly the Dice-related composite loss function, which often achieves better segmentation results and stronger performance than a single loss function.

The BCE loss function [44] is a common and intuitive loss function. It is widely used in binary classification tasks and is also suitable for binary segmentation (lesion/non-lesion) in skin disease semantic segmentation. The BCE loss function measures the accuracy of predictions by comparing the similarity between the predicted segmentation result and the ground-truth labels. The gradient computation of the BCE loss function is relatively simple, which aids in the training and optimization of a model. It can be used in conjunction with commonly used optimization algorithms (such as stochastic gradient descent) and exhibits good numerical stability. The advantage of the BCE loss function is that it can be combined with other loss functions to further enhance the model's performance. The BCE loss function is defined in [44] as follows:

$$L_{BCE} = -\sum_i (g_i ln(p_i) + (1 - g_i)ln(1 - p_i)) \tag{10}$$

where g_i represents the segmentation outcome of pixel *i* as determined by a medical professional, and p_i corresponds to the segmentation outcome of pixel *i* generated by the network.

The DSC loss function [45] directly optimizes the evaluation metric of the segmentation model, namely the Dice coefficient. The DSC loss function can better handle the severe class imbalance between positive and negative samples that may exist in skin disease images. The DSC loss function exhibits good smoothness during the optimization process. Compared to other loss functions, such as cross-entropy loss, the DSC loss function can optimize the model parameters more smoothly. This helps in avoiding the issues of gradient vanishing or exploding, thereby stabilizing the training process of the model. By using the DSC loss function for training, it is possible to directly optimize the model's DSC during the training process, thus improving the segmentation performance of the model on the test set. The DSC loss function is defined in [45] as follows:

$$L_{DSC} = 1 - 2\frac{\sum_i g_i p_i}{\sum_i g_i + \sum_i p_i} \tag{11}$$

In order to accelerate the network's convergence, alleviate the issues of gradient vanishing, address class imbalance concerns in the backpropagation process, and achieve effective skin disease segmentation, we amalgamate these two loss functions during model training, as in [46], as follows:

$$L = \frac{1}{2}L_{BCE} + L_{DSC} \tag{12}$$

4.5. Results

The performance of the proposed GSCEU-Net model was benchmarked against well-established medical image segmentation models through comprehensive experiments conducted across the three aforementioned datasets, with the outcomes presented in this subsection. The experimental configurations of the majority of compared models were identical to those used for GSCEU-Net.

4.5.1. ISIC2018 Experiments

Table 3 showcases the comparative segmentation performance results derived from experiments conducted on this dataset (the most outstanding outcome for each metric is highlighted in **bold**). It can be clearly seen that GSCEU-Net achieves a balance between the number of parameters, computational complexity, and segmentation performance, with *IoU*, *DSC*, and *accuracy* evaluation metrics reaching the highest values. The *sensitivity* value is lower than that of MALUNet by only 0.0132 points. The GSCEU-Net's parameter count and GFLOPS also reach the highest values across the compared models, with parameter counts being more than 4 times smaller than that of the second-best model (MALUNet). Figure 7 depicts the training process of the proposed GSCEU-Net model, showing the number of epochs, loss function, and IoU for training and validation. Sample visual contrasts in skin lesion segmentation results attained by various models on this dataset are show-cased in Figure 8.

Table 3. Segmentation-performance-comparison results on the ISIC2018 dataset.

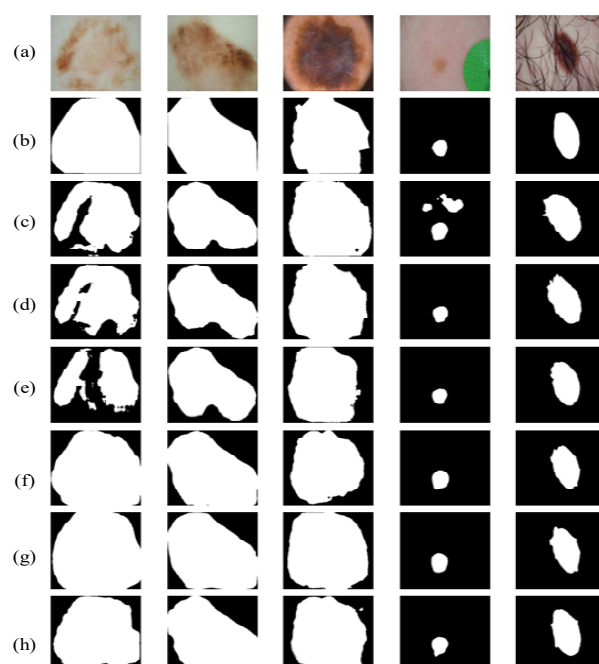| Model	Parameters (Million)	GFLOPS	IoU	DSC	Acc	Sen
U-Net	7.770	13.780	0.7887	0.8784	0.9508	0.8760
U-Net++	9.160	34.900	0.7952	0.8824	0.9528	0.8732
Attention-UNet	8.730	16.740	0.7967	0.8833	0.9533	0.8591
UNeXt_S	0.300	0.100	0.8057	0.8895	0.9557	0.8586
MALUNet	0.175	0.083	0.8120	0.8924	0.9532	**0.8875**
GSCEU-Net (*proposed model*)	**0.041**	**0.081**	**0.8145**	**0.8948**	**0.9571**	0.8743

Figure 8. Segmentation results on the ISIC2018 dataset: (**a**) original image; (**b**) ground truth; (**c**) U-Net result; (**d**) U-Net++ result; (**e**) Attention-Unet result; (**f**) UNeXt_S result; (**g**) MALUNet result; (**h**) GSCEU-Net result.

4.5.2. PH2 Experiments

To assess the segmentation ability of a trained model on a novel dataset and validate its ability to generalize and exhibit robustness, additional experiments were conducted on the publicly available PH2 dataset, comprising a mere 200 images.

Table 4 illustrates the results of the segmentation performance comparison yielded by these experiments (the most outstanding outcome for each metric is highlighted in **bold**). Similarly to ISIC2018, the proposed GSCEU-Net model outperforms all mainstream models with scores that are 0.0163, 0.0092, and 0.0083 points higher than the first runner-up (MALUNet) in terms of *IoU*, *DSC*, and *accuracy*, respectively. Furthermore, in terms of *sensitivity*, GSCEU-Net, although not the first, is only 0.0011 points behind the winner (MALUNet).

Table 4. Segmentation-performance-comparison results on the PH2 dataset.

Model	Parameters (Million)	GFLOPS	IoU	DSC	Acc	Sen
U-Net	7.770	13.780	0.8062	0.8916	0.9276	0.9224
U-Net++	9.160	34.900	0.7929	0.8831	0.9238	0.8909
Attention-UNet	8.730	16.740	0.7458	0.8505	0.9090	0.8241
UNeXt_S	0.300	0.100	0.8077	0.8900	0.9277	0.8874
MALUNet	0.175	0.083	0.8278	0.9048	0.9351	**0.9484**
GSCEU-Net (*proposed model*)	**0.041**	**0.081**	**0.8441**	**0.9140**	**0.9434**	0.9473

Comparative results of skin lesion segmentation achieved by different models on this dataset are illustrated in Figure 9. Evidently, GSCEU-Net delivers marginally superior lesion segmentation outcomes in contrast to other models. Notably, GSCEU-Net excels in effectively segmenting larger lesions, whereas U-Net frequently struggles to encompass the entirety of lesion areas and displays discernible divergences in shape when matched against ground-truth images. These findings affirm that the supplementary modules incorporated into U-Net effectively enhance segmentation performance and significantly contribute to the model's commendable generalization capabilities.
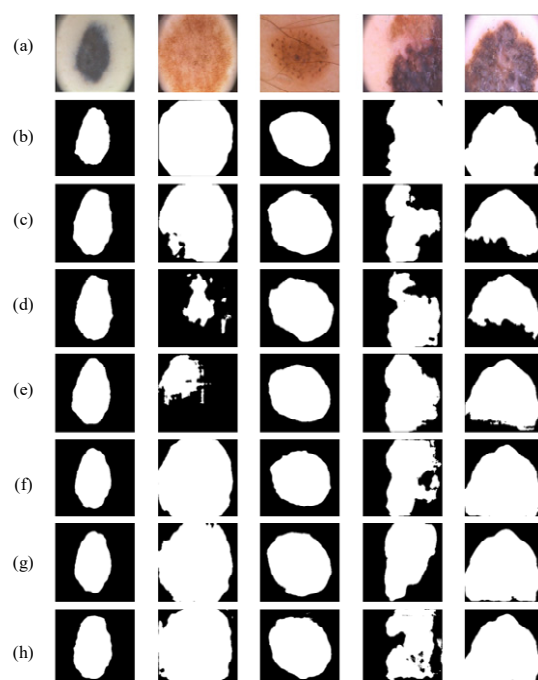

Figure 9. Segmentation results on the PH2 dataset: (**a**) original image; (**b**) ground truth; (**c**) U-Net result; (**d**) U-Net++ result; (**e**) Attention-Unet result; (**f**) UNeXt_S result; (**g**) MALUNet result; (**h**) GSCEU-Net result.

4.5.3. Private Dataset Experiments

Next, experiments were carried out on the private dataset. In contrast to the ISIC2018 dataset, this collection of data comprises fewer images, leading to reduced training duration and quicker network convergence. However, the lesions in this dataset are shallower and have blurred edges, making segmentation more difficult.

Table 5 showcases the outcomes of the segmentation performance comparison attained through experimentation on this dataset (the most outstanding outcome for each metric is highlighted in **bold**). Once more, the proposed GSCEU-Net model demonstrates superiority over all other models on two main segmentation evaluation metrics, with higher scores of 0.0053 and 0.0038 points on *IoU* and *DSC*, respectively, compared to the first runner-up (UNeXt_S). Moreover, based on *accuracy*, GSCEU-Net also surpasses all mainstream models, leaving the second place for U-Net++ which stays behind by 0.0042 points. Even though the proposed GSCEU-Net model ranks only fifth in terms of *sensitivity*, this metric is not prominent, as it indicates that a model considers some non-lesion regions as lesion regions during predictions.

Table 5. Segmentation-performance-comparison results on the private dataset.

Model	Parameters (Million)	GFLOPS	IoU	DSC	Acc	Sen
U-Net	7.770	13.780	0.6254	0.7635	0.9119	0.7868
U-Net++	9.160	34.900	0.6283	0.7664	0.9171	0.7479
Attention-UNet	8.730	16.740	0.6308	0.7695	0.9109	**0.8160**
UNeXt_S	0.300	0.100	0.6397	0.7766	0.9143	0.7934
MALUNet	0.175	0.083	0.6301	0.7698	0.9150	0.7797
GSCEU-Net (*proposed model*)	**0.041**	**0.081**	**0.6450**	**0.7804**	**0.9213**	0.7731

Comparative results of skin lesion segmentation achieved by different models on this dataset are illustrated in Figure 10. It can be observed that GSCEU-Net is capable of predicting the entire lesion more comprehensively and can accurately segment the lesion edges.
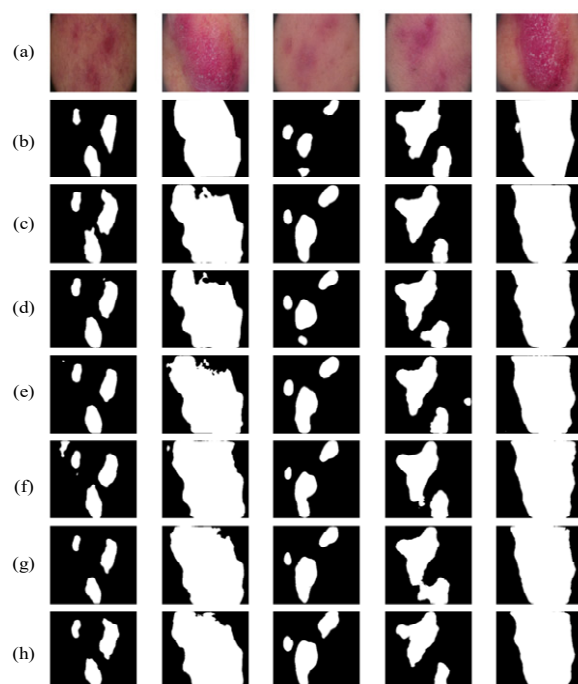

Figure 10. Segmentation results on the private dataset: (**a**) original image; (**b**) ground truth; (**c**) U-Net result; (**d**) U-Net++ result; (**e**) Attention-Unet result; (**f**) UNeXt_S result; (**g**) MALUNet result; (**h**) GSCEU-Net result.

4.5.4. Ablation Study Experiments

To ascertain the efficacy of each individually incorporated module on enhancing the segmentation performance, ablation study experiments were conducted using the U-Net model as a baseline for burn segmentation on the ISIC2018 dataset. The outcomes of these experiments are depicted in Table 6 (the most outstanding outcome for each metric is highlighted in **bold**). Given the complexities inherent in lesions with distinct shapes, colors, and indistinct boundaries within the ISIC2018 dataset, the number of channels was kept the same in U-Net and GSCEU-Net, for fair comparison. Additional modules, suggested in this paper for incorporation into U-Net, were gradually added to it, which reflects changes in parameters, floating-point operations, and evaluation metrics. The model obtained at the final 'U-Net + GSC + ECA' step (i.e., the GSCEU-Net model), demonstrates the best results on all four evaluation metrics used. Although the GSCEU-Net model has three times the number of parameters and floating-point operations compared to the 'U-Net + SConv' combination, it outperforms it by 0.0401 and 0.0267 points on *IoU* and *DSC*, respectively, which are the main metrics used for segmentation performance evaluation. Furthermore, even though the inclusion of the ECA attention module resulted in almost no increase in parameters and floating-point operations, yet it made a significant contribution to the model's segmentation performance.

Table 6. Ablation study results using U-Net as a baseline.

Model	Parameters (Million)	GFLOPS	IoU	DSC	Acc	Sen
U-Net	0.12	0.220	0.7868	0.8758	0.9503	0.8692
U-Net + SConv	**0.01**	**0.024**	0.7744	0.8681	0.9463	0.8653
U-Net + GSC	0.04	0.081	0.7858	0.8763	0.9504	0.8586
U-Net + GSC + ECA (i.e., proposed GSCEU-Net)	0.04	0.081	**0.8145**	**0.8948**	**0.9571**	**0.8743**

5. Discussion

Segmentation of skin lesions requires rapid and accurate prediction, which can provide beneficial assistance to patients. Traditional methods are time-consuming and rely heavily on parameter tuning. On the other hand, complex deep learning models have high computational resource requirements and come with a significant time cost, making it challenging to meet the practical needs of medical skin lesion segmentation. In light of this, a lightweight encoder–decoder model, referred to as GSCEU-Net, has been proposed in this paper by using the U-Net model as a basis with additionally incorporated modules. Firstly, the lightweight idea of PConv [20] has been utilized to design a novel SConv module for incorporation into U-Net, so as to avoid wasting a large portion of input features, utilize 1×1 convolutions to accelerate model prediction, and dynamically compute the scaled output channel numbers at a very low cost to meet the requirements of the U-shaped structure. Secondly, a GSC backbone network has been proposed, which combines upgraded SConv paths with a Ghost module for residual connections, thereby absorbing the features of both module types to enhance the model's robustness and generalization capability. Thirdly, the ECA attention mechanism has been added to the decoding part of the model. After upsampling the output, more prominent features can be extracted for skip connections, introducing minimal additional parameters and negligible computations while achieving significant performance gains. For the loss function, the BCE and DSC losses have been combined to tackle imbalanced samples. The conducted experiments on respected datasets like ISIC2018 and PH2, along with a private dataset, have yielded persuasive results, whereby GSCEU-Net has shown reliability, robustness, and adaptability across tough images by outperforming widely used segmentation models, such as U-Net, U-Net++, and Attention-UNet. Moreover, GSCEU-Net had a lower parameter count and

fewer floating-point operations compared to recently proposed lightweight models, such as UNeXt and MALUNet.

Compared to U-Net, GSCEU-Net was able to improve the *IoU* value by 0.0258, 0.0379, and 0.0196 points on the ISIC2018, PH2, and private datasets, respectively, while also increasing the *DSC* value by 0.0164, 0.0224, and 0.0169 points, respectively. At the same time, the parameter count and floating-point operations were reduced by a factor of 190 and 170, respectively, compared to U-Net.

Compared to the recent skin lesion segmentation models, the proposed model has also demonstrated competitive performance. Importantly, the modules employed by the proposed model (some of which are newly designed ones) have a small memory footprint of only 260 KB after training, compared to 806.5 KB and 1 MB for UNeXt and MALUNet, respectively. This is a significant advantage in practical applications, since the proposed model can be installed on compact mobile devices for standalone execution.

In future research, we plan to explore the following improvements. Firstly, although the model has demonstrated outstanding segmentation results, its training process required a relatively large number of epochs, resulting in slow convergence. If appropriate operations, such as learning rate adjustments or pretraining the model with other datasets, can be incorporated, superior weights can be obtained during the initial training. Secondly, we will explore even more lightweight convolutional modules that can result in state-of-the-art network performance while having an extremely low parameter count and fewer floating-point operations.

6. Conclusions

An improved U-Net network architecture, called GSCEU-Net, has been presented in this paper; it achieves leading performance on various evaluation metrics while ensuring a lightweight design. For this, GSCEU-Net combines newly designed SConv and GSC modules, along with an ECA mechanism, allowing it to achieve *IoU* values of 0.8145, 0.8441, and 0.6450, and *DSC* values of 0.8948, 0.9140, and 0.7804, on the ISIC2018 and PH2 public datasets, and a private dataset, respectively.

Nevertheless, it is essential to acknowledge certain limitations inherent in the proposed model. For instance, the model's training convergence speed is low, and the floating-point operations have not reached the fastest level. We believe that further parameter adjustments to the proposed lightweight model can be made to meet the requirements of real clinical environments.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 7–34. [CrossRef]
2. Madan, V.; Lear, J.T.; Szeimies, R.-M. Non-melanoma skin cancer. *Lancet* **2010**, *375*, 673–685. [CrossRef] [PubMed]
3. Baig, R.; Bibi, M.; Hamid, A.; Kausar, S.; Khalid, S. Deep learning approaches towards skin lesion segmentation and classification from dermoscopic images-a review. *Curr. Med. Imaging* **2020**, *16*, 513–533. [CrossRef] [PubMed]
4. Silveira, M.; Nascimento, J.C.; Marques, J.S.; Marçal, A.R.; Mendonça, T.; Yamauchi, S.; Maeda, J.; Rozeira, J. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 35–45. [CrossRef]

5. Emre Celebi, M.; Wen, Q.; Hwang, S.; Iyatomi, H.; Schaefer, G. Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Res. Technol.* **2013**, *19*, e252–e258. [CrossRef] [PubMed]

6. Garnavi, R.; Aldeen, M.; Celebi, M.E.; Varigos, G.; Finch, S. Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Comput. Med. Imaging Graph.* **2011**, *35*, 105–115. [CrossRef] [PubMed]

7. Guo, Y.; Ashour, A.S.; Smarandache, F. A novel skin lesion detection approach using neutrosophic clustering and adaptive region growing in dermoscopy images. *Symmetry* **2018**, *10*, 119. [CrossRef]

8. Celebi, M.E.; Kingravi, H.A.; Iyatomi, H.; Lee, J.; Aslandogan, Y.A.; Van Stoecker, W.; Moss, R.; Malters, J.M.; Marghoob, A.A. Fast and accurate border detection in dermoscopy images using statistical region merging. In Proceedings of the Medical Imaging 2007: Image Processing, San Diego, CA, USA, 17–22 February 2007; pp. 1297–1306.

9. Celebi, M.E.; Aslandogan, Y.A.; Bergstresser, P.R. Unsupervised border detection of skin lesion images. In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05), Las Vegas, NV, USA, 4–6 April 2005; Volume II, pp. 123–128.

10. 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; pp. 234–241.

11. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef]

12. 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.

13. Zhang, L.; Zhang, J.; Shen, P.; Zhu, G.; Li, P.; Lu, X.; Zhang, H.; Shah, S.A.; Bennamoun, M. Block level skip connections across cascaded V-Net for multi-organ segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 2782–2793. [CrossRef]

14. Alom, M.Z.; Yakopcic, C.; Hasan, M.; Taha, T.M.; Asari, V.K. Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* **2019**, *6*, 014006. [CrossRef]

15. Chen, J.; He, T.; Zhuo, W.; Ma, L.; Ha, S.; Chan, S.-H.G. Tvconv: Efficient translation variant convolution for layout-aware visual processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12548–12558.

16. 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.

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

18. Mendonça, T.; Ferreira, P.M.; Marques, J.S.; Marcal, A.R.; Rozeira, J. PH 2-A dermoscopic image database for research and benchmarking. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 5437–5440.

19. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.

20. Chen, J.; Kao, S.-h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; Chan, S.-H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18 June 2023; pp. 12021–12031.

21. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.

22. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J. Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24261–24272.

23. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical image analysis using convolutional neural networks: A review. *J. Med. Syst.* **2018**, *42*, 226. [CrossRef] [PubMed]

24. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.

25. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [CrossRef]

26. Liu, S.; Song, L.; Liu, X. Neural network for lung cancer diagnosis. In *Computational Intelligence in Cancer Diagnosis*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 89–116.

27. 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, Singapore, 18–22 September 2022; pp. 23–33.

28. Khan, M.A.; Sharif, M.; Akram, T.; Bukhari, S.A.C.; Nayak, R.S. Developed Newton-Raphson based deep features selection framework for skin lesion recognition. *Pattern Recognit. Lett.* **2020**, *129*, 293–303. [CrossRef]

29. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [CrossRef] [PubMed]

30. Manzoor, K.; Majeed, F.; Siddique, A.; Meraj, T.; Rauf, H.T.; El-Meligy, M.A.; Sharaf, M.; Abd Elgawad, A.E.E. A lightweight approach for skin lesion detection through optimal features fusion. *Comput. Mater. Contin.* **2022**, *70*, 1617–1630. [CrossRef]

31. Sebastian, V.B.; Unnikrishnan, A.; Balakrishnan, K. Gray level co-occurrence matrices: Generalisation and some new features. *arXiv* **2012**, arXiv:1205.4831.

32. Ali, A.-R.H.; Li, J.; Yang, G. Automating the ABCD rule for melanoma detection: A survey. *IEEE Access* **2020**, *8*, 83333–83346. [CrossRef]

33. Zafar, M.; Sharif, M.I.; Sharif, M.I.; Kadry, S.; Bukhari, S.A.C.; Rauf, H.T. Skin lesion analysis and cancer detection based on machine/deep learning techniques: A comprehensive survey. *Life* **2023**, *13*, 146. [CrossRef] [PubMed]

34. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]

35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

36. 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), Munich, Germany, 8–14 September 2018; pp. 3–19.

37. Mehta, S.; Rastegari, M.; Caspi, A.; Shapiro, L.; Hajishirzi, H. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 552–568.

38. Wu, B.; Dai, X.; Zhang, P.; Wang, Y.; Sun, F.; Wu, Y.; Tian, Y.; Vajda, P.; Jia, Y.; Keutzer, K. Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10734–10742.

39. Mehta, S.; Rastegari, M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.

40. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

41. Agarap, A.F. Deep learning using rectified linear units (relu). *arXiv* **2018**, arXiv:1803.08375.

42. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.

43. Goyal, P.; Dollár, P.; Girshick, R.; Noordhuis, P.; Wesolowski, L.; Kyrola, A.; Tulloch, A.; Jia, Y.; He, K. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv* **2017**, arXiv:1706.0267.

44. Ruby, U.; Yendapalli, V. Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 5393–5397.

45. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 4th International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.

46. Montazerolghaem, M.; Sun, Y.; Sasso, G.; Haworth, A. U-Net Architecture for Prostate Segmentation: The Impact of Loss Function on System Performance. *Bioengineering* **2023**, *10*, 412. [CrossRef]